이 학 박 사 학 위 논 문

Additive Regression with Hilbertian Responses

힐버트 반응변수를 갖는 가법 회귀

2018년 8월

서울대학교 대학원

통계학과

전정민

Additive Regression with Hilbertian Responses

By

Jeong Min Jeon

A Thesis
Submitted in Fulfillment of the Requirement
for the Degree of
Doctor of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
August, 2018

ABSTRACT

Additive Regression with Hilbertian Responses

Jeong Min Jeon
The Department of Statistics
The Graduate School
Seoul National University

This paper develops a foundation of methodology and theory for the estimation of structured nonparametric regression models with Hilbertian responses. Our method and theory are focused on the additive model, while the main ideas may be adapted to other structured models. For this, the notion of Bochner integration is introduced for Banach-space-valued maps as a generalization of Lebesgue integration. Several statistical properties of Bochner integrals, relevant for our method and theory, and also of importance in their own right, are presented for the first time. Our theory is complete. The existence of our estimators and the convergence of a practical algorithm that evaluates the estimators are established. These results are non-asymptotic as well as asymptotic. Further-

more, it is proved that the estimators of component maps achieve the univariate error rates in pointwise, L^2 and uniform convergence, and converge jointly in distribution to Gaussian random elements. Our numerical examples include the cases of functional, density-valued and simplex-valued responses, which demonstrate the validity of our approach.
Keywords: Additive model, Smooth backfitting, Bochner integral, Non-Euclidean data, Infinite-dimensional space, Hilbert space, Functional response.
Student Number: 2012 – 20232

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Regression analysis with non-Euclidean data is one of the major challenges in modern statistics. In many cases it is not transparent how one can go beyond traditional Euclidean methods to analyze non-Euclidean objects. The problem we tackle in this paper is particularly the case. We consider the estimation of nonparametric additive models that involve non-Euclidean random objects.

Additivity is a commonly employed structure with which one is able to avoid the curse of dimensionality in nonparametric regression. A powerful kernel-based method for achieving this is the smooth backfitting (SBF) technique originated by Mammen et al. (1999). A full account of the practical issues about the method is given in Nielsen and Sperlich (2005). The idea has been developed for various structured nonparametric models, see Mammen and Park (2006), Yu et al. (2008), Linton et al. (2008), Lee et al. (2010, 2012) and Han and Park (2018+), for example. All of them, however, treated the case of Euclidean response. There have been a few applications of the idea to functional response. Examples include Zhang et al. (2013), Lee et al. (2018) and Park et al. (2018+). But, their techniques and theory are essentially the same as in the case of Euclidean response. They applied the SBF technique to a functional response $Y(\cdot)$ on a domain $\mathcal{T}$ in a *pointwise* manner, i.e., to $Y(t)$ for each $t \in \mathcal{T}$, or to a finite number of its singular components that live in a Euclidean space. These methods have certain drawbacks. The pointwise application does not guarantee a smooth trajectory for $\hat{Y}(\cdot)$ on $\mathcal{T}$ while $Y(\cdot)$ is believed to be smooth. Methods based on singular components require choosing the number of included components in a working model, which is very difficult.

In this paper we develop a unified approach for fitting additive models with a Hilbertian response. Let $\mathbb{H}$ be a separable Hilbert space with a zero vector $\mathbf{0}$, vector addition $\oplus$ and scalar multiplication $\odot$. For a probability space $(\Omega, \mathscr{F}, P)$, we consider a response $\mathbf{Y} : \Omega \to \mathbb{H}$. Let $\mathbf{X} = (X_1, \cdots, X_d)$ be a predictor taking values in a compact subset of $\mathbb{R}^d$, say $[0,1]^d$, and $\boldsymbol{\epsilon}$ be a $\mathbb{H}$-valued error satisfying $\mathrm{E}(\boldsymbol{\epsilon}|\mathbf{X}) = \mathbf{0}$. The additive model we study in this paper is

$$\mathbf{Y} = \mathbf{m}_0 \oplus \bigoplus_{j=1}^{d} \mathbf{m}_j(X_j) \oplus \boldsymbol{\epsilon}, \tag{1.0.1}$$

where $\mathbf{m}_0$ is a constant in $\mathbb{H}$ and $\mathbf{m}_1, \ldots, \mathbf{m}_d : [0,1] \to \mathbb{H}$ are measurable maps. There are numerous examples of Hilbertian variables. In the next section we introduce three examples, which we also treat in our numerical study in Section 5. These are functional variables, density-valued variables and simplex-valued variables. Our approach guarantees that the values of the estimators of $\mathbf{m}_j(x_j)$ at x_j belong to the space $\mathbb{H}$ where the targets live. This is a minimal requirement for a reasonable estimator. For example, in case $\mathbb{H}$ is a space of smooth functions defined on $\mathcal{T}$, as is typically the case with functional data, our approach always produces smooth $\hat{\mathbf{m}}_j(x_j)(\cdot)$. Existing methods where one estimates $\mathbf{m}_j(x_j)(t)$ pointwise in $t \in \mathcal{T}$ do not have this property. Moreover, the computation of our estimators is faster than the pointwise approach as the grid on $\mathcal{T}$ gets denser, since the proposed method estimates $\mathbf{m}_j(x_j)(\cdot)$ on the whole $\mathcal{T}$ all at once.

The SBF technique involves solving a system of integral equations that is based on the integral representations of the conditional expectations of the response. In our case, the traditional Lebesgue integral theory does not apply since we treat random elements taking values in a general Hilbert space. For this, we base our approach on the notion of Bochner integration. The notion, rather new in statistics, is for Banach-space-valued maps. We develop integral formulas for (conditional) expectation, relevant theory for projection operators acting on the spaces of Hilbert-space-valued maps and some topological properties of the space of regression maps under the model (1.0.1). These are essential for investigating the theoretical properties of our estimators. We note that this paper is the first in the statistical application of Bochner integration. We establish the basic building block of structured nonparametric regression for Hilbertian responses. For this we start from the foundation of Bochner integral theory. Some of our results are familiar in Lebesgue integral theory, but their derivation for Bochner integrals requires substantial innovation.

Based on the Bochner integral theory we develop in this paper, we establish the existence of the SBF estimator of the regression map under the model (1.0.1) and the convergence of the SBF algorithm. The results include non-asymptotic versions as well as asymptotic ones. The non-asymptotic results have not been studied before even for the case $\mathbb{H} = \mathbb{R}$. The conventional way of establishing the convergence of the SBF algorithm is to prove that the associated projection operators are compact. We find, however, that this is no longer valid for infinite-dimensional $\mathbb{H}$. Instead, we prove that the space of sums of univariate Hilbert-space-valued maps is closed by a novel use of a result on the equivalence of the 'compatibility' of sum-maps (the condition (c) in Proposition 3.3.1) and the closedness of the sum-space. We also provide a creative way of implementing the proposed algorithm, which reduces the task of iterating abstractly-defined Bochner integration to that of updating real-valued weight functions based on Lebesgue integration, see (2.5.2). Furthermore, we present complete theory for the rates of convergence of the estimators of the component maps $\mathbf{m}_j$ and their asymptotic distributions.

We do not consider the case where the predictor $\mathbf{X}$ in (1.0.1) is of infinite-dimension. The reason is that our approach is based on solving a system of integral equations where each integral is evaluated on the space of $\mathbf{X}$ values. It is well known that there is no nontrivial locally finite translation invariant measure on infinite-dimensional separable Banach spaces, like Lebesgue measure on $\mathbb{R}^k$. Thus, it is not easy to evaluate the integrals in practice when $\mathbf{X}$ takes values in an infinite-dimensional separable Banach space.

There have been a few attempts of dealing with possibly non-Euclidean response. Dabo-Niang and Rhomari (2009), Ferraty et al. (2011) and Ferraty et al. (2012) studied a functional Nadaraya-Watson estimator for Banach- or Hilbert-space-valued response. Lian (2011) and Lian (2012) investigated a functional k-nearest neighbor estimator for Hilbert-space-valued response. But, these are for full-dimensional regression, which would suffer from the curse of dimensionality when the number of predictors increases. Some others for functional response include Chiou et al. (2003), Jiang and Wang (2011), Zhu et al. (2012) and Scheipl et al. (2015). They are differentiated from ours in that their methods or the models under study essentially reduce the problem to the estimation of a regression function for a scalar response. There has been no earlier work on nonparametric regression with density-valued responses, although Petersen and Muller (2016) introduced a transformation approach for density-valued responses and predictors. Recently, Tsagris (2015) considered simplex-valued responses but in a parametric model.

Chapter 2

Bochner Smooth Backfitting

Throughout this paper, we use the symbol $\mathbb{B}$ to denote Banach spaces and $\|\cdot\|$ for their norms. We use the symbol $\mathcal{B}(\mathbb{B})$ to denote the Borel σ-field of $\mathbb{B}$. For a set $S \in \mathcal{B}(\mathbb{B})$, we write $S \cap \mathcal{B}(\mathbb{B})$ for the σ-field $\{S \cap B : B \in \mathcal{B}(\mathbb{B})\}$ on S. We denote Hilbert spaces by $\mathbb{H}$ and their inner products by $\langle \cdot, \cdot \rangle$. We also let Leb_k denote the Lebesgue measure on $\mathbb{R}^k$.

2.1 Examples of Hilbertian response

Here, we introduce three Hilbert spaces. These are the spaces we consider for the response in our numerical study in Section 5.

L^2 and Hilbert-Sobolev spaces. For a subset $S \in \mathcal{B}(\mathbb{R}^k)$, consider $L^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$ and a Hilbert-Sobolev space $W^{l,2}(S)$ for $l \in \mathbb{N}$. It is well known that these are separable Hilbert spaces.

Bayes-Hilbert spaces. Consider a space of density functions on $S \in \mathcal{B}(\mathbb{R}^k)$. Let

$$\mathcal{M} = \{\mu : \mu \text{ is a } \sigma\text{-finite measure on } S \cap \mathcal{B}(\mathbb{R}^k) \text{ such that} \\ \mu \ll \mathrm{Leb}_k \text{ and } \mathrm{Leb}_k \ll \mu\}.$$

For $\mu \in \mathcal{M}$, let $f_\mu = d\mu/d\mathrm{Leb}_k$. For μ, $\nu \in \mathcal{M}$ and $c \in \mathbb{R}$, define $\mu\nu$, $\mu^c : S \cap \mathcal{B}(\mathbb{R}^k) \to [0, \infty]$ by $(\mu\nu)(A) = \int_A f_\mu(\mathbf{s}) f_\nu(\mathbf{s}) d\mathbf{s}$ and $(\mu^c)(A) = \int_A (f_\mu(\mathbf{s}))^c d\mathbf{s}$, respectively. Then, $\mu\nu$, $\mu^c \in \mathcal{M}$. For these measures, $f_{\mu\nu} = f_\mu \cdot f_\nu$ a.e. $[\mathrm{Leb}_k]$ and $f_{\mu^c} = (f_\mu)^c$ a.e. $[\mathrm{Leb}_k]$. Define

$$\mathfrak{B}^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k) = \Big\{[f_\mu] : \mu \in \mathcal{M}, \int_S \big(\log f_\mu(\mathbf{s})\big)^2 d\mathbf{s} < \infty\Big\},$$

where $[f_\mu]$ denotes the class of all measurable functions $g : S \to [0, \infty]$ such that $g = C \cdot f_\mu$ a.e. $[\mathrm{Leb}_k]$ for some constant $C > 0$. Define $\oplus$ and $\odot$ on $\mathfrak{B}^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$ by $[f_\mu] \oplus [f_\nu] = [f_{\mu\nu}] = [f_\mu \cdot f_\nu]$ and $c \odot [f_\mu] = [f_{\mu^c}] = [(f_\mu)^c]$, respectively. Also, define $\langle \cdot, \cdot \rangle$ by

$$\langle [f_\mu], [f_\nu] \rangle = \int_{S^2} \log\left(\frac{f_\mu(\mathbf{s})}{f_\mu(\mathbf{s}')}\right) \log\left(\frac{f_\nu(\mathbf{s})}{f_\nu(\mathbf{s}')}\right) d\mathbf{s} d\mathbf{s}'.$$

Then, $\mathfrak{B}^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$ is a separable Hilbert space with $\mathbf{0} = [f_{\mathrm{Leb}_k}] = [1]$, as proved by van den Boogaart et al. (2014).

Simplices. For $s > 0$, consider the space $\mathcal{S}_s^k = \{(v_1, \cdots, v_k) \in (0, s)^k : \sum_{j=1}^k v_j = s\}$. For $\mathbf{v}, \mathbf{w} \in \mathcal{S}_s^k$ and $c \in \mathbb{R}$, define $\oplus$ and $\odot$, respectively, by $\mathbf{v} \oplus \mathbf{w} = (\frac{s v_1 w_1}{v_1 w_1 + \cdots + v_k w_k}, \ldots, \frac{s v_k w_k}{v_1 w_1 + \cdots + v_k w_k})$ and $c \odot \mathbf{v} = (\frac{s v_1^c}{v_1^c + \cdots + v_k^c}, \ldots, \frac{s v_k^c}{v_1^c + \cdots + v_k^c})$. Define

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^k \sum_{l=1}^k \log(v_j/v_l) \log(w_j/w_l).$$

Then, with $\mathbf{0} = (s/k, \dots, s/k)$, $(\mathcal{S}_s^k, \oplus, \odot, \langle \cdot, \cdot \rangle)$ is a separable Hilbert space.

2.2 Bochner integration

Our method of estimating the additive model (1.0.1) is based on the representation of the conditional means of $\mathbf{m}_k(X_k)$ given X_j for $k \neq j$, in terms of the conditional densities of X_k given X_j. This involves integration of $\mathbf{m}_k(x_k)$ over x_k in the support of the corresponding conditional density. Since each component $\mathbf{m}_k$ is a $\mathbb{H}$-valued map, the conventional Lebesgue integration does not apply to the current problem. In this subsection we study a notion of integration in a more general setting. Specifically, we consider integration of Banach-space-valued maps.

For the integration of $\mathbb{B}$-valued maps, we use a notion of Bochner integral. Let $(\mathcal{Z}, \mathscr{A}, \mu)$ be a measure space. In the classical Bochner integral theory, see Lang (1993) and van Neerven (2008), for example, Bochner integrals are defined for Banach-space-valued μ-measurable maps. Note that a map $\mathbf{f} : \mathcal{Z} \to \mathbb{B}$ is called μ-measurable if it is the μ-almost everywhere limit of a sequence of μ-simple maps. A map $\mathbf{f} : \mathcal{Z} \to \mathbb{B}$ is called μ-simple if $\mathbf{f}(\mathbf{z}) = \bigoplus_{i=1}^n 1_{A_i}(\mathbf{z}) \odot \mathbf{b}_i$ for some $\mathbf{b}_i \in \mathbb{B}$ and disjoint $A_i \in \mathscr{A}$ with $\mu(A_i) < \infty$. However, a μ-measurable map is not necessarily $(\mathscr{A}, \mathcal{B}(\mathbb{B}))$-measurable. Failure of $(\mathscr{A}, \mathcal{B}(\mathbb{B}))$-measurability causes a fundamental problem in statistical applications. To explain why, let $(\Omega, \mathscr{F}, P)$ be a probability space and $\mathbf{Z} : \Omega \to \mathcal{Z}$ be a random element. If $\mathbf{f}$ is not $(\mathscr{A}, \mathcal{B}(\mathbb{B}))$-measurable, then $\mathbf{f}(\mathbf{Z}) : \Omega \to \mathbb{B}$ may not be a random element.

We consider a recently introduced notion of Bochner integration, which has never been studied in statistics, to the best of our knowledge. The new notion is for 'strongly measurable' Banach-space-valued maps. We briefly introduce it here. For more details, see Cohn (2013). For a map $\mathbf{f} : \mathscr{Z} \to \mathbb{B}$, we let $\mathrm{range}(\mathbf{f})$ denote $\{\mathbf{f}(\mathbf{z}) : \mathbf{z} \in \mathscr{Z}\} \subset \mathbb{B}$.

Definition 2.2.1. *A map $\mathbf{f} : (\mathscr{Z}, \mathscr{A}, \mu) \to (\mathbb{B}, \mathcal{B}(\mathbb{B}))$ is called strongly measurable if it is $(\mathscr{A}, \mathcal{B}(\mathbb{B}))$-measurable and $\mathrm{range}(\mathbf{f})$ is separable.*

An immediate example of strongly measurable map is μ-simple map. For a μ-simple map $\mathbf{f}(\mathbf{z}) = \bigoplus_{i=1}^{n} 1_{A_i}(\mathbf{z}) \odot \mathbf{b}_i$, the Bochner integral is defined by $\int \mathbf{f}(\mathbf{z}) d\mu(\mathbf{z}) = \bigoplus_{i=1}^{n} \mu(A_i) \odot \mathbf{b}_i$. It can be shown that, if a map $\mathbf{f}$ is strongly measurable and $\|\mathbf{f}\|$ is Lebesgue integrable with respect to μ, then there exist μ-simple maps $\mathbf{f}_n$ such that $\mathbf{f}(\mathbf{z}) = \lim_{n\to\infty} \mathbf{f}_n(\mathbf{z})$ and $\|\mathbf{f}_n(\mathbf{z})\| \leq \|\mathbf{f}(\mathbf{z})\|$ for all $\mathbf{z}$ and n.

Definition 2.2.2. *A map $\mathbf{f} : (\mathscr{Z}, \mathscr{A}, \mu) \to (\mathbb{B}, \mathcal{B}(\mathbb{B}))$ is called Bochner integrable if it is strongly measurable and $\|\mathbf{f}\|$ is Lebesgue integrable with respect to μ. In this case the Bochner integral of $\mathbf{f}$ is defined by $\int \mathbf{f} d\mu = \lim_{n\to\infty} \int \mathbf{f}_n d\mu$, where $\mathbf{f}_n$ is a sequence of μ-simple maps such that $\mathbf{f}(\mathbf{z}) = \lim_{n\to\infty} \mathbf{f}_n(\mathbf{z})$ and $\|\mathbf{f}_n(\mathbf{z})\| \leq \|\mathbf{f}(\mathbf{z})\|$.*

We present several properties of the Bochner integral that are fundamental in its statistical applications. For $1 \leq p < \infty$, define

$$\mathcal{L}^p((\mathscr{Z}, \mathscr{A}, \mu), \mathbb{B}) = \Big\{ \mathbf{f} : \mathscr{Z} \to \mathbb{B} \,|\, \mathbf{f} \text{ is strongly measurable and } \Big(\int_{\mathscr{Z}} \|\mathbf{f}(\mathbf{z})\|^p d\mu(\mathbf{z}) \Big)^{1/p} < \infty \Big\}.$$

Recall that $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{R})$ can be made into a Banach space by taking its quotient space $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{R})/\mathcal{N}_\mathbb{R}$ with respect to the kernel $\mathcal{N}_\mathbb{R}$ of its norm, $\mathcal{N}_\mathbb{R} = \{f : f = 0 \text{ a.e. } [\mu]\}$. This also holds for $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$. In particular, for $\mathcal{N} = \{\mathbf{f} : \mathbf{f} = \mathbf{0} \text{ a.e. } [\mu]\}$, the quotient space $\mathcal{L}^2((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{H})/\mathcal{N}$ is a Hilbert space with an inner product $\langle \cdot, \cdot \rangle_\mu$ defined by $\langle [\mathbf{f}], [\mathbf{g}] \rangle_\mu = \int_\mathcal{Z} \langle \mathbf{f}(\mathbf{z}), \mathbf{g}(\mathbf{z}) \rangle d\mu(\mathbf{z})$, where $[\mathbf{f}]$ and $[\mathbf{g}]$ denote the equivalence classes of maps $\mathbf{f}$ and $\mathbf{g}$, respectively. We adopt the following convention throughout this paper.

Convention 1. We write $L^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$ for $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})/\mathcal{N}$. We call $L^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$ *Lebesgue-Bochner space*. We will write all elements in $L^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$ using equivalence class notation $[\cdot]$ to distinguish them from the elements in $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$. We say simply 'measurable' for 'strongly measurable' and 'integrable' for 'Bochner integrable'. We say 'μ-integrable' in case we need to specify the underlying measure μ associated with Bochner integration. $\square$

For measure spaces $(\mathcal{Z}, \mathscr{A}, \mu)$ and $(\mathcal{W}, \mathscr{B}, \nu)$, let $\mathscr{A} \otimes \mathscr{B}$ denote the product σ-field and $\mu \otimes \nu$ denote a product measure on $\mathscr{A} \otimes \mathscr{B}$. For a $(\mathscr{A}, \mathscr{B})$-measurable mapping $\mathbf{T} : \mathcal{Z} \to \mathcal{W}$, we let $\mu\mathbf{T}^{-1}$ denote a measure on $(\mathcal{W}, \mathscr{B})$ defined by $\mu\mathbf{T}^{-1}(B) = \mu(\mathbf{T}^{-1}(B)), B \in \mathscr{B}$. For a probability space $(\Omega, \mathscr{F}, P)$ and a random element $\mathbf{Z} : (\Omega, \mathscr{F}, P) \to (\mathcal{Z}, \mathscr{A}, \mu)$ with σ-finite μ, we write $p_\mathbf{Z}$ for its density $dP\mathbf{Z}^{-1}/d\mu$ with respect to μ.

The following two propositions are the basic building blocks of our methodological and theoretical development to be presented later. They are also of interest in their own right. The results are very new in statistics although there are familiar versions in the Lebesgue integral theory. In the propositions and thereafter throughout this paper, $\mathbb{B}$ denotes a *separable* Banach space. Separability is required for the associated maps to be measurable, see Definition 2.2.1.

Proposition 2.2.1. *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space and* $(\mathcal{Z}, \mathscr{A}, \mu)$ *be a σ-finite measure space. Let* $\mathbf{Z} : \Omega \to \mathcal{Z}$ *be a random element such that* $P\mathbf{Z}^{-1} \ll \mu$ *and* $\mathbf{f} : \mathcal{Z} \to \mathbb{B}$ *be a measurable map such that* $\mathrm{E}(\|\mathbf{f}(\mathbf{Z})\|) < \infty$. *Then, it holds that* $\mathrm{E}(\mathbf{f}(\mathbf{Z})) = \int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) \odot p_{\mathbf{Z}}(\mathbf{z}) d\mu(\mathbf{z})$.

Proof. From the condition of the proposition, $\mathbf{f}(\mathbf{Z}) : \Omega \to \mathbb{B}$ is P-integrable so that an application of Lemma 6.1.1 in the Appendix gives that $\mathbf{f}$ is $P\mathbf{Z}^{-1}$-integrable and $\int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) dP\mathbf{Z}^{-1}(\mathbf{z}) = \int_{\Omega} \mathbf{f}(\mathbf{Z}) dP$. According to Lemma 6.1.2 in the Appendix, $\mathbf{f} \odot (dP\mathbf{Z}^{-1}/d\mu)$ is μ-integrable and $\int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) dP\mathbf{Z}^{-1}(\mathbf{z}) = \int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) \odot (dP\mathbf{Z}^{-1}/d\mu)(\mathbf{z}) d\mu(\mathbf{z})$. The proposition now follows. $\square$

Proposition 2.2.2. *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space, and* $(\mathcal{Z}, \mathscr{A}, \mu)$ *and* $(\mathcal{W}, \mathscr{B}, \nu)$ *be σ-finite measure spaces. Let* $\mathbf{Z} : \Omega \to \mathcal{Z}$ *and* $\mathbf{W} : \Omega \to \mathcal{W}$ *be random elements such that* $P(\mathbf{Z}, \mathbf{W})^{-1} \ll \mu \otimes \nu$. *Assume that* $p_{\mathbf{W}}(\mathbf{w}) \in (0, \infty)$ *for all* $\mathbf{w} \in \mathcal{W}$. *Let* $\mathbf{f} : \mathcal{Z} \to \mathbb{B}$ *be a measurable map such that* $\mathrm{E}(\|\mathbf{f}(\mathbf{Z})\|) < \infty$. *Define* $\mathbf{g} : \mathcal{W} \to \mathbb{B}$ *by*

$$\mathbf{g}(\mathbf{w}) = \begin{cases} \int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) \odot \frac{p_{\mathbf{Z},\mathbf{W}}(\mathbf{z},\mathbf{w})}{p_{\mathbf{W}}(\mathbf{w})} d\mu(\mathbf{z}), & \text{if } \mathbf{w} \in D_{\mathcal{W}} \\ \mathbf{g}_0(\mathbf{w}), & \text{otherwise,} \end{cases}$$

where $D_{\mathcal{W}} = \{\mathbf{w} \in \mathcal{W} : \int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\| p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) < \infty\}$ *and* $\mathbf{g}_0 : \mathcal{W} \to \mathbb{B}$ *is any measurable map. Then,* $\mathbf{g}$ *is measurable and* $\mathbf{g}(\mathbf{W})$ *is a version of* $\mathrm{E}(\mathbf{f}(\mathbf{Z})|\mathbf{W})$.

Proof. We first note that the map $(\mathbf{z}, \mathbf{w}) \mapsto \mathbf{f}(\mathbf{z}) \odot p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w})$ is measurable. From Tonelli's theorem, it follows that

$$\begin{aligned} \mathrm{E}(\|\mathbf{f}(\mathbf{Z})\|) &= \int_{\mathcal{Z}} \int_{\mathcal{W}} \|\mathbf{f}(\mathbf{z})\| p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w}) d\nu(\mathbf{w}) d\mu(\mathbf{z}) \\ &= \int_{\mathcal{W}} \int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\| p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}). \end{aligned} \tag{2.2.1}$$

Since $\mathrm{E}(\|\mathbf{f}(\mathbf{Z})\|) < \infty$, (2.2.1) implies that $\int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\| p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \cdot) d\mu(\mathbf{z}) < \infty$ a.e. $[\nu]$. By Lemma 6.1.3 in the Appendix, $\mathbf{h} : \mathcal{W} \to \mathbb{B}$ defined by $\mathbf{h}(\mathbf{w}) = \mathbf{g}(\mathbf{w}) \odot p_{\mathbf{W}}(\mathbf{w})$ is measurable. Thus, $\mathbf{g}$ is measurable and $\mathbf{g}(\mathbf{W})$ is $(\mathbf{W}^{-1}(\mathscr{B}), \mathcal{B}(\mathbb{B}))$-measurable. We also get

$$\int_{\mathcal{W}} \|\mathbf{g}(\mathbf{w})\| dP\mathbf{W}^{-1}(\mathbf{w}) \leq \int_{\mathcal{W}} \int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\| p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}) < \infty.$$

Hence, $\mathbf{g}(\mathbf{W}) \in \mathcal{L}^1((\Omega, \mathbf{W}^{-1}(\mathscr{B}), P), \mathbb{B})$ by Lemma 6.1.1. Now, from Lemmas 6.1.1–6.1.3 and the fact that $\nu(D_{\mathcal{W}}^c) = 0$ it follows that, for all $B \in \mathscr{B}$,

$$\begin{aligned} \int_{\mathbf{W}^{-1}(B)} \mathbf{g}(\mathbf{W}) dP &= \int_{\mathcal{Z} \times \mathcal{W}} \mathbf{f}(\mathbf{z}) \odot [1_B(\mathbf{w}) p_{\mathbf{Z},\mathbf{W}}(\mathbf{z}, \mathbf{w})] d\mu \otimes \nu(\mathbf{z}, \mathbf{w}) \\ &= \int_{\mathbf{W}^{-1}(B)} \mathbf{f}(\mathbf{Z}) dP. \end{aligned}$$

This completes the proof of the proposition. □

2.3 Lebesgue-Bochner spaces of additive maps

We introduce some relevant spaces for the estimation of the additive model (1.0.1). For a probability space $(\Omega, \mathscr{F}, P)$ and a separable Hilbert space $\mathbb{H}$, let $\mathbf{Y} : \Omega \to \mathbb{H}$ be a response with $\mathrm{E}\|\mathbf{Y}\|^2 < \infty$, and $\mathbf{X} : \Omega \to [0, 1]^d$ a d-variate predictor vector. We assume $P\mathbf{X}^{-1} \ll \mathrm{Leb}_d$. For simplicity we write p, instead of $p_{\mathbf{X}}$, to denote its density $dP\mathbf{X}^{-1}/d\mathrm{Leb}_d$. We also write p_j for $dPX_j^{-1}/d\mathrm{Leb}_1$ and p_{jk} for $dP(X_j, X_k)^{-1}/d\mathrm{Leb}_2$.

The conditional means $\mathrm{E}(\mathbf{Y}|X_j)$ and $\mathrm{E}(\mathbf{Y}|\mathbf{X})$, respectively, are $(X_j^{-1}([0,1] \cap \mathcal{B}(\mathbb{R})), \mathcal{B}(\mathbb{H}))$- and $(\mathbf{X}^{-1}([0,1]^d \cap \mathcal{B}(\mathbb{R}^d)), \mathcal{B}(\mathbb{H}))$-measurable maps by definition. In general, for a measurable space $(\mathcal{Z}, \mathscr{A})$, a random element $\mathbf{V} : \Omega \to \mathbb{B}$ and a random element $\mathbf{Z} : \Omega \to \mathcal{Z}$, it holds that $\mathbf{V}$ is $(\mathbf{Z}^{-1}(\mathscr{A}), \mathcal{B}(\mathbb{B}))$-measurable if and only if there exists a measurable map $\mathbf{h} : \mathcal{Z} \to \mathbb{B}$ such that $\mathbf{V} = \mathbf{h}(\mathbf{Z})$, see Lemma 1.13 in Kallenberg (1997), for example. Thus, there exist measurable maps $\mathbf{h}_j : [0,1] \to \mathbb{H}$ and $\mathbf{h} : [0,1]^d \to \mathbb{H}$ such that $\mathrm{E}(\mathbf{Y}|X_j) = \mathbf{h}_j(X_j)$ and $\mathrm{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{h}(\mathbf{X})$. For such measurable maps, we define $\mathrm{E}(\mathbf{Y}|X_j = \cdot) = \mathbf{h}_j$ and $\mathrm{E}(\mathbf{Y}|\mathbf{X} = \cdot) = \mathbf{h}$.

Let $\mathbf{m} : [0,1]^d \to \mathbb{H}$ be defined by $\mathbf{m}(\mathbf{x}) = \mathbf{m}_0 \oplus \bigoplus_{j=1}^{d} \mathbf{m}_j(x_j)$. We note that $\mathbf{m} = \mathrm{E}(\mathbf{Y}|\mathbf{X} = \cdot)$. As the space where $\mathrm{E}(\mathbf{Y}|\mathbf{X} = \cdot)$ belongs, we consider

$$\mathcal{L}_2^{\mathbb{H}}(p) := \mathcal{L}^2(([0,1]^d, [0,1]^d \cap \mathcal{B}(\mathbb{R}^d), P\mathbf{X}^{-1}), \mathbb{H})$$

and endow $L_2^{\mathbb{H}}(p) := L^2(([0,1]^d, [0,1]^d \cap \mathcal{B}(\mathbb{R}^d), P\mathbf{X}^{-1}), \mathbb{H})$ with the norm $\|\cdot\|_2$ defined by

$$\|[\mathbf{f}]\|_2^2 = \int_{[0,1]^d} \|\mathbf{f}(\mathbf{x})\|^2 dP\mathbf{X}^{-1}(\mathbf{x}) = \int_{[0,1]^d} \|\mathbf{f}(\mathbf{x})\|^2 p(\mathbf{x}) d\mathbf{x}.$$

As subspaces of $\mathcal{L}_2^{\mathbb{H}}(p)$, define

$$\mathcal{L}_2^{\mathbb{H}}(p_j) := \left\{ \mathbf{f} \in \mathcal{L}_2^{\mathbb{H}}(p) : \exists \text{ a univariate map } \mathbf{f}_j \text{ such that } \mathbf{f}(\mathbf{x}) = \mathbf{f}_j(x_j) \right\}$$

and define $L_2^{\mathbb{H}}(p_j) := \mathcal{L}_2^{\mathbb{H}}(p_j)/\mathcal{N}$. We note that $\mathcal{L}_2^{\mathbb{H}}(p_j)$ depends on p only through its marginalization p_j since, for $\mathbf{f} \in \mathcal{L}_2^{\mathbb{H}}(p_j)$, it holds that

$$\int_{[0,1]^d} \|\mathbf{f}(\mathbf{x})\|^2 p(\mathbf{x}) d\mathbf{x} = \int_0^1 \|\mathbf{f}_j(x_j)\|^2 p_j(x_j) dx_j,$$

where $\mathbf{f}_j$ is a univariate map such that $\mathbf{f}(\mathbf{x}) = \mathbf{f}_j(x_j)$. Let $S^{\mathbb{H}}(p)$ be the sum-space defined by

$$S^{\mathbb{H}}(p) = \left\{ \bigoplus_{j=1}^{d} [\mathbf{f}_j] : [\mathbf{f}_j] \in L_2^{\mathbb{H}}(p_j),\ 1 \le j \le d \right\} \subset L_2^{\mathbb{H}}(p).$$

To define empirical versions of $\mathcal{L}_2^{\mathbb{H}}(p), L_2^{\mathbb{H}}(p), \mathcal{L}_2^{\mathbb{H}}(p_j), L_2^{\mathbb{H}}(p_j)$ and $S^{\mathbb{H}}(p)$, we let $K : \mathbb{R} \to [0, \infty)$ be a baseline kernel function. Throughout this paper, we assume that K vanishes on $\mathbb{R} \setminus [-1, 1]$ and satisfies $\int_{-1}^1 K(u)du = 1$. For a bandwidth $h > 0$ we write $K_h(u) = K(u/h)/h$. Define a normalized kernel $K_h(u, v)$ by

$$K_h(u, v) = \frac{K_h(u - v)}{\int_0^1 K_h(t - v)dt}$$

whenever $\int_0^1 K_h(t - v)dt > 0$ and we set $K_h(u, v) = 0$ otherwise. This type of kernel function has been used in the smooth backfitting literature, see Mammen et al. (1999), for example. Note that $\int_0^1 K_h(u, v)du = 1$ for all $v \in [0, 1]$ and

$$K_h(u, v) = K_h(u - v) \quad \text{for all } (u, v) \in [2h, 1 - 2h] \times [0, 1]. \tag{2.3.1}$$

We also have

$$\int_{-1}^0 K(u)du \wedge \int_0^1 K(u)du \le \int_0^1 K_h(u - v)du \le \int_{-1}^1 K(u)du$$

for all $v \in [0, 1]$ and $h \le 1/2$. Hence, if $\int_{-1}^0 K(u)du \wedge \int_0^1 K(u)du > 0$, then

$$K_h(u - v) \le K_h(u, v) \le \frac{K_h(u - v)}{\int_{-1}^0 K(u)du \wedge \int_0^1 K(u)du}$$

for all $u, v \in [0,1]$ and $h \leq 1/2$.

Suppose that we observe $(\mathbf{Y}_i, \mathbf{X}_i), 1 \leq i \leq n$, which follow the model (1.0.1). We estimate $p_j(x_j)$ and $p_{jk}(x_j, x_k)$ by

$$\hat{p}_j(x_j) = n^{-1}\sum_{i=1}^{n} K_{h_j}(x_j, X_{ij}),$$
$$\hat{p}_{jk}(x_j, x_k) = n^{-1}\sum_{i=1}^{n} K_{h_j}(x_j, X_{ij}) K_{h_k}(x_k, X_{ik}),$$

respectively, where X_{ij} denotes the jth entry of $\mathbf{X}_i$. Here, we allow the bandwidths h_j to be different for different j. Because of the normalization in defining $K_h(\cdot, \cdot)$, it holds that

$$\int_0^1 \hat{p}_j(x_j) dx_j = 1, \quad \int_0^1 \hat{p}_{jk}(x_j, x_k) dx_k = \hat{p}_j(x_j).$$

Let $\hat{p}$ be the multivariate kernel density estimator of p defined by $\hat{p}(\mathbf{x}) = n^{-1}\sum\limits_{i=1}^{n} \prod_{j=1}^{d} K_{h_j}(x_j, X_{ij})$. The density estimator $\hat{p}$ also have the marginalization properties as p:

$$\int_{[0,1]^{d-1}} \hat{p}(\mathbf{x}) d\mathbf{x}_{-j} = \hat{p}_j(x_j), \quad \int_{[0,1]^{d-2}} \hat{p}(\mathbf{x}) d\mathbf{x}_{-j,k} = \hat{p}_{jk}(x_j, x_k)$$

for $1 \leq j \neq k \leq d$, where $\mathbf{x}_{-j}$ and $\mathbf{x}_{-j,k}$ denote the respective $(d-1)$- and $(d-2)$-vector resulting from omitting x_j and (x_j, x_k) in $\mathbf{x} = (x_1, \ldots, x_d)$.

Now, define a measure $\hat{P}\mathbf{X}^{-1}$ on $[0,1]^d \cap \mathcal{B}(\mathbb{R}^d)$ by $\hat{P}\mathbf{X}^{-1}(B) = \int_B \hat{p}(\mathbf{x}) d\mathbf{x}$. With this measure, we define $\mathcal{L}_2^{\mathbb{H}}(\hat{p}), L_2^{\mathbb{H}}(\hat{p}), \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ and $L_2^{\mathbb{H}}(\hat{p}_j)$ as $\mathcal{L}_2^{\mathbb{H}}(p), L_2^{\mathbb{H}}(p), \mathcal{L}_2^{\mathbb{H}}(p_j)$ and $L_2^{\mathbb{H}}(p_j)$ with $P\mathbf{X}^{-1}$ in the definition of $\mathcal{L}_2^{\mathbb{H}}(p)$ and $L_2^{\mathbb{H}}(p)$ being replaced by $\hat{P}\mathbf{X}^{-1}$. We endow $L_2^{\mathbb{H}}(\hat{p})$ with the norm $\|\cdot\|_{2,n}$ defined by

$$\|[\mathbf{f}]\|_{2,n}^2 = \int_{[0,1]^d} \|\mathbf{f}(\mathbf{x})\|^2 d\hat{P}\mathbf{X}^{-1}(\mathbf{x}) = \int_{[0,1]^d} \|\mathbf{f}(\mathbf{x})\|^2 \hat{p}(\mathbf{x}) d\mathbf{x}.$$

Also, define an analogue of $S^{\mathbb{H}}(p)$ by

$$S^{\mathbb{H}}(\hat{p}) = \left\{ \bigoplus_{j=1}^{d} [\mathbf{f}_j] : [\mathbf{f}_j] \in L_2^{\mathbb{H}}(\hat{p}_j),\ 1 \leq j \leq d \right\} \subset L_2^{\mathbb{H}}(\hat{p}).$$

Convention 2. It is often convenient to treat $\mathbf{f}$ in $\mathcal{L}_2^{\mathbb{H}}(p_j)$ or in $\mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ as a univariate map and write $\mathbf{f}(x_j)$ instead of $\mathbf{f}(\mathbf{x})$. This convention is particularly useful when we express a system of Bochner integral equations in Section 2.4, see (2.4.6) below, for example. Conversely, we may embed a univariate map $\mathbf{f} : [0, 1] \to \mathbb{H}$ into $\mathcal{L}_2^{\mathbb{H}}(p_j)$ or $\mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ by considering its version $\mathbf{f}_j^*$ defined by $\mathbf{f}_j^*(\mathbf{x}) = \mathbf{f}(x_j)$ for $\mathbf{x} \in [0, 1]^d$. We take the above convention throughout this paper. With this convention, we may put $\mathbf{m}_j$ into $\mathcal{L}_2^{\mathbb{H}}(p_j)$ if $\mathrm{E}(\|\mathbf{m}_j(X_j)\|^2) < \infty$. $\square$

2.4 Bochner integral equations and backfitting algorithm

In this section, we describe the estimation of the component maps $\mathbf{m}_j$ in the model (1.0.1) using Bochner integrals. Throughout this paper, we assume that $\mathbf{m}_j \in \mathcal{L}_2^{\mathbb{H}}(p_j)$ for all $1 \leq j \leq d$. Furthermore, we make the following assumptions on the densities p_j and p_{jk}.

Condition (A). *The p_j and p_{jk} for all $1 \leq j \neq k \leq d$ satisfy $p_j(x_j) > 0$ and $\int_0^1 p_{jk}^2(x_j, x_k)/p_k(x_k)dx_k < \infty$ for all $x_j \in [0, 1]$, and*

$$\int_{[0,1]^2} \frac{p_{jk}^2(x_j, x_k)}{p_j(x_j)p_k(x_k)} dx_j dx_k < \infty.$$

We also use the following analogue of the condition (A) for $\hat{p}_j$ and $\hat{p}_{jk}$.

Condition (S). *The $\hat{p}_j$ and $\hat{p}_{jk}$ for all $1 \le j \neq k \le d$ satisfy $\hat{p}_j(x_j) > 0$ and $\int_0^1 \hat{p}_{jk}^2(x_j, x_k)/\hat{p}_k(x_k)dx_k < \infty$ for all $x_j \in [0,1]$, and*

$$\int_{[0,1]^2} \frac{\hat{p}_{jk}^2(x_j, x_k)}{\hat{p}_j(x_j)\hat{p}_k(x_k)} dx_j dx_k < \infty.$$

We note that the condition (S) always holds under weak conditions on the bandwidths and baseline kernel function. Let $X_{(1),j} < \cdots < X_{(n),j}$ denote the order statistics of $(X_{ij} : 1 \le i \le n)$. Suppose that h_j and K satisfy

(S1) $h_j > \max\left\{X_{(1),j}, 1-X_{(n),j}, \max_{1\le i\le n-1}(X_{(i+1),j}-X_{(i),j})/2\right\}$ for all $1 \le j \le d$.

(S2) K is bounded and $\inf_{u\in[-c,c]} K(u) > 0$, where

$$c = \max_{1\le j\le d} h_j^{-1} \max\left\{X_{(1),j}, 1-X_{(n),j}, \max_{1\le i\le n-1}(X_{(i+1),j}-X_{(i),j})/2\right\} < 1.$$

Then, it is easy to see that

$$\inf_{x_j\in[0,1]} \hat{p}_j(x_j) > 0, \qquad \sup_{x_j, x_k\in[0,1]} \hat{p}_{jk}(x_j, x_k) < \infty$$

for all $1 \le j \neq k \le d$. Hence, (S1) and (S2) imply the condition (S).

From the basic properties of conditional expectation and the model (1.0.1), we get

$$\mathrm{E}(\mathbf{Y}|X_j) = \mathbf{m}_0 \oplus \mathbf{m}_j(X_j) \oplus \bigoplus_{k\neq j} \mathrm{E}(\mathbf{m}_k(X_k)|X_j), \quad 1 \le j \le d. \tag{2.4.1}$$

Under the condition (A) we also get that

$$\int_0^1 \|\mathbf{m}_k(x_k)\| p_{jk}(x_j, x_k) dx_k < \infty \tag{2.4.2}$$

for all $x_j \in [0,1]$ and $1 \le j \ne k \le d$. The property (2.4.2) is a simple consequence of an application of Hölder's inequality. Then, by Proposition 2.2.2, we may write (2.4.1) as

$$\mathrm{E}(\mathbf{Y}|X_j) = \mathbf{m}_0 \oplus \mathbf{m}_j(X_j) \oplus \bigoplus_{k \ne j} \int_0^1 \mathbf{m}_k(x_k) \odot \frac{p_{jk}(X_j, x_k)}{p_j(X_j)} dx_k, \quad 1 \le j \le d.$$

By the definition of $\mathrm{E}(\mathbf{Y}|X_j = \cdot)$, we may also write it as

$$\mathrm{E}(\mathbf{Y}|X_j = x_j) = \mathbf{m}_0 \oplus \mathbf{m}_j(x_j) \oplus \bigoplus_{k \ne j} \int_0^1 \mathbf{m}_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} dx_k, \quad 1 \le j \le d. \tag{2.4.3}$$

For the identifiability of $\mathbf{m}_j$ in the model, we put the constraints $\mathrm{E}(\mathbf{m}_j(X_j)) = \mathbf{0}, 1 \le j \le d$. By Proposition 2.2.1, the constraints are equivalent to

$$\int_0^1 \mathbf{m}_j(x_j) \odot p_j(x_j) dx_j = \mathbf{0}, \quad 1 \le j \le d. \tag{2.4.4}$$

The constraints entail $\mathbf{m}_0 = \mathrm{E}(\mathbf{Y})$.

Now we describe the estimation of $\mathbf{m}_j$ based on the Bochner integral equations at (2.4.3). We estimate $\mathrm{E}(\mathbf{Y}|X_j = x_j)$ by the Nadaraya-Watson-type estimator

$$\tilde{\mathbf{m}}_j(x_j) = \left[\hat{p}_j(x_j)^{-1} n^{-1}\right] \odot \bigoplus_{i=1}^n K_{h_j}(x_j, X_{ij}) \odot \mathbf{Y}_i \tag{2.4.5}$$

and $\mathrm{E}(\mathbf{Y})$ by the sample mean $\bar{\mathbf{Y}} = n^{-1} \odot \bigoplus_{i=1}^n \mathbf{Y}^i$. Let $\ominus$ be defined by $\mathbf{b}_1 \ominus \mathbf{b}_2 = \mathbf{b}_1 \oplus (-1 \odot \mathbf{b}_2)$. We solve the estimated system of Bochner integral equations

$$\hat{\mathbf{m}}_j(x_j) = \tilde{\mathbf{m}}_j(x_j) \ominus \bar{\mathbf{Y}} \ominus \bigoplus_{k \neq j} \int_0^1 \hat{\mathbf{m}}_k(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k, \quad 1 \leq j \leq d \tag{2.4.6}$$

for $(\hat{\mathbf{m}}_1, \cdots, \hat{\mathbf{m}}_d)$ in the space of d-tuple maps $\{(\mathbf{f}_1, \ldots, \mathbf{f}_d) : \mathbf{f}_j \in \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j),\ 1 \leq j \leq d\}$, subject to the constraints

$$\int_0^1 \hat{\mathbf{m}}_j(x_j) \odot \hat{p}_j(x_j) dx_j = \mathbf{0}, \quad 1 \leq j \leq d. \tag{2.4.7}$$

We note that the Bochner integrals at (2.4.6) are well-defined for $\hat{\mathbf{m}}_j \in \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ under the condition (S).

In the next section we will show that there exists a solution $(\hat{\mathbf{m}}_j : 1 \leq j \leq d)$ of (2.4.6) satisfying (2.4.7) and that their sum $\bigoplus_{j=1}^d \hat{\mathbf{m}}_j$ is unique, only under the condition (S). The estimator of the regression map $\mathbf{m} := \mathrm{E}(\mathbf{Y}|\mathbf{X} = \cdot) : [0,1]^d \to \mathbb{H}$ is defined by $\hat{\mathbf{m}}$, where $\hat{\mathbf{m}}(\mathbf{x}) = \bar{\mathbf{Y}} \oplus \bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j(x_j)$. For the estimator $\hat{\mathbf{m}}$, we will also prove that the component tuple $(\hat{\mathbf{m}}_j : 1 \leq j \leq d)$ is uniquely determined under some additional assumption. Our estimator of $(\mathbf{m}_1, \ldots, \mathbf{m}_d)$ is then the solution $(\hat{\mathbf{m}}_1, \ldots, \hat{\mathbf{m}}_d)$. We call $\hat{\mathbf{m}}$ and $\hat{\mathbf{m}}_j$ *Bochner smooth backfitting estimators* or *B-SBF estimators* in short, and the system of equations (2.4.6) *Bochner smooth backfitting equation*, or *B-SBF equation* in short.

Our approach guarantees that $\hat{\mathbf{m}}_j(x_j)$ and $\hat{\mathbf{m}}(\mathbf{x})$ belong to $\mathbb{H}$, the space of the true values of the maps $\mathbf{m}_j$ and $\mathbf{m}$ as well as the values of $\mathbf{Y}$. For example, in the case where $\mathbb{H}$ is a space of smooth functions in $L^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$, as is typically assumed in functional data analysis, our approach always produces a smooth trajectory $\hat{\mathbf{m}}_j(x_j)(\cdot) : S \to \mathbb{R}$ for each x_j. Here, one should not confuse the smoothness of $\hat{\mathbf{m}}_j(x_j)(\cdot) : S \to \mathbb{R}$ with that of $\hat{\mathbf{m}}_j$ as maps from $[0,1]$ to $\mathcal{L}^2(S, S \cap \mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$. In the case where $\mathbb{H}$ is a Bayes-Hilbert space $\mathfrak{B}^2(S, S\cap\mathcal{B}(\mathbb{R}^k), \mathrm{Leb}_k)$ or a simplex space $\mathcal{S}_s^k$, our approach gives automatically densities on S or k-dimensional compositional vectors with nonnegative entries, respectively, as the estimators of $\hat{\mathbf{m}}_j(x_j)$ and $\hat{\mathbf{m}}(\mathbf{x})$, that integrate or sum into one.

To solve the B-SBF equation (2.4.6), we take an initial estimator $(\hat{\mathbf{m}}_1^{[0]}, \cdots, \hat{\mathbf{m}}_d^{[0]})$ that satisfies the constraints (2.4.7). We update the estimator $(\hat{\mathbf{m}}_1^{[r]}, \cdots, \hat{\mathbf{m}}_d^{[r]})$ for $r \geq 1$ by

$$\begin{aligned}\hat{\mathbf{m}}_j^{[r]}(x_j) = \tilde{\mathbf{m}}_j(x_j) \ominus \bar{\mathbf{Y}} \ominus \bigoplus_{k<j} \int_0^1 \hat{\mathbf{m}}_k^{[r]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \\ \ominus \bigoplus_{k>j} \int_0^1 \hat{\mathbf{m}}_k^{[r-1]}(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k, \quad 1 \leq j \leq d.\end{aligned} \tag{2.4.8}$$

We let $\hat{\mathbf{m}}^{[r]}(\mathbf{x}) = \bar{\mathbf{Y}} \oplus \bigoplus_{j=1}^d \hat{\mathbf{m}}_j^{[r]}(x_j)$. We call (2.4.8) *Bochner smooth backfitting algorithm* or *B-SBF algorithm* in short. In the next section we will show that the B-SBF algorithm converges always in $\|\cdot\|_{2,n}$ norm under the condition (S). We will also show that it converges in $\|\cdot\|_2$ norm with probability tending to one under weak conditions on p, K and h_j. We note that, if the initial estimator $(\hat{\mathbf{m}}_1^{[0]}, \cdots, \hat{\mathbf{m}}_d^{[0]})$ satisfies the constraints (2.4.7), then all subsequent updates $(\hat{\mathbf{m}}_1^{[r]}, \cdots, \hat{\mathbf{m}}_d^{[r]})$ for $r \geq 1$ also satisfy (2.4.7) due to the normalization property $\int_0^1 K_{h_j}(u, \cdot)du \equiv 1$ on $[0,1]$.

2.5 Practical implementation

Bochner integrals are defined in an abstract way. Thus, one can not evaluate the integrals at (2.4.8) with the usual numerical integration techniques. In this subsection we present an innovative way of implementing the B-SBF algorithm. The key idea is to use the fact that, for any measure space $(\mathcal{Z}, \mathscr{A}, \mu)$,

$$\text{(Bochner)} \int_{\mathcal{Z}} f(\mathbf{z}) \odot \mathbf{b}\, d\mu(\mathbf{z}) = \text{(Lebesgue)} \int_{\mathcal{Z}} f(\mathbf{z}) d\mu(\mathbf{z}) \odot \mathbf{b}, \tag{2.5.1}$$

where f is a real-valued integrable function on $\mathcal{Z}$ and $\mathbf{b}$ is a constant in a Banach space. Suppose that we choose

$$\hat{\mathbf{m}}_j^{[0]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} w_{ij}^{[0]}(x_j) \odot \mathbf{Y}_i$$

as the initial estimators with the weights $w_{ij}^{[0]}(x_j) \in \mathbb{R}$ satisfying $\int_0^1 w_{ij}^{[0]}(x_j)\hat{p}_j(x_j)dx_j = 0$. This is not a restriction since we can take $w_{ij}^{[0]} \equiv 0$ for all $1 \le j \le d$ and $1 \le i \le n$. Then, we may express the Bochner integrals on the right hand side of the equation at (2.4.8) in terms of the corresponding Lebesgue integrals as follows.

$$\begin{aligned} \hat{\mathbf{m}}_j^{[r]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} \Bigg(& \frac{K_{h_j}(x_j, X_j^i)}{\hat{p}_j(x_j)} - 1 - \sum_{k<j} \int_0^1 w_{ik}^{[r]}(x_k) \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \\ & - \sum_{k>j} \int_0^1 w_{ik}^{[r-1]}(x_k) \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \Bigg) \odot \mathbf{Y}_i \\ =: n^{-1} \odot \bigoplus_{i=1}^{n} & w_{ij}^{[r]}(x_j) \odot \mathbf{Y}_i, \quad 1 \le j \le d. \end{aligned} \tag{2.5.2}$$

Thus, it turns out that the algorithm (2.4.8) reduces to a simple iteration scheme that updates the weight functions $w_{ij}^{[r]}$ based on

Lebesgue integrals.

The equation (2.5.2) reveals that $\hat{\mathbf{m}}_j^{[r]}$ for $r \geq 1$ are linear smoothers if the initial $\hat{\mathbf{m}}_j^{[0]}$ are. It also demonstrates explicitly that the values of $\hat{\mathbf{m}}_j^{[r]}(x_j)$ for each x_j belong to $\mathbb{H}$, the space of the values of $\mathbf{Y}_i$ and $\mathbf{m}_j(x_j)$. The idea of using (2.5.1) in the evaluation of Bochner integrals appears to be important in the analysis of more general object-oriented data belonging to a Banach space. One may develop a similar idea for nonparametric structural regression dealing with various types of random objects.

Chapter 3

Existence and Algorithm Convergence

3.1 Projection operators

Our theory for the existence of the B-SBF estimators and the convergence of the B-SBF algorithm relies heavily on the theory of projection operators that map $L_2^{\mathbb{H}}(p)$ to $L_2^{\mathbb{H}}(p_j)$, or $L_2^{\mathbb{H}}(\hat{p})$ to $L_2^{\mathbb{H}}(\hat{p}_j)$. We start with a proposition that characterizes $L_2^{\mathbb{H}}(p_j)$ and $L_2^{\mathbb{H}}(\hat{p}_j)$, respectively, as closed subspaces of $L_2^{\mathbb{H}}(p)$ and $L_2^{\mathbb{H}}(\hat{p})$. These topological properties of $L_2^{\mathbb{H}}(p_j)$ and $L_2^{\mathbb{H}}(\hat{p}_j)$ are essential to defining relevant projection operators. We write $\mathbb{B}_1 \leq \mathbb{B}_2$ if $\mathbb{B}_1$ is a *closed* subspace of a Banach space $\mathbb{B}_2$. The following proposition is immediate from Lemma 6.1.5 in the Appendix and the fact that a complete subspace of a metric space is closed.

Proposition 3.1.1. *$L_2^{\mathbb{H}}(p_j) \leq L_2^{\mathbb{H}}(p)$ and $L_2^{\mathbb{H}}(\hat{p}_j) \leq L_2^{\mathbb{H}}(\hat{p})$.*

We define the operators $\pi_j : L_2^{\mathbb{H}}(p) \to L_2^{\mathbb{H}}(p_j)$ by $\pi_j([\mathbf{f}]) = [\pi_{j\mathbf{f}}]$, where

$$\pi_{j\mathbf{f}}(x_j) = \begin{cases} \int_{[0,1]^{d-1}} \mathbf{f}(\mathbf{x}) \odot \frac{p(\mathbf{x})}{p_j(x_j)} d\mathbf{x}_{-j}, & \text{if } x_j \in D_j(\mathbf{f}) \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

where $D_j(\mathbf{f}) = \{x_j \in [0,1] : \int_{[0,1]^{d-1}} \|\mathbf{f}(\mathbf{x})\| p(\mathbf{x}) d\mathbf{x}_{-j} < \infty\}$. Likewise, we define the operators $\hat{\pi}_j : L_2^{\mathbb{H}}(\hat{p}) \to L_2^{\mathbb{H}}(\hat{p}_j)$ with p and p_j being replaced by $\hat{p}$ and $\hat{p}_j$, respectively. The following proposition demonstrates that both π_j and $\hat{\pi}_j$ are projection operators on the respective spaces.

Proposition 3.1.2. *If $p_j(x_j) > 0$ for all $x_j \in [0,1]$, then, π_j is an orthogonal projection operator. Also, if $\hat{p}_j(x_j) > 0$ for all $x_j \in [0,1]$, then, $\hat{\pi}_j$ is an orthogonal projection operator.*

For Banach spaces $\mathbb{B}_1$ and $\mathbb{B}_2$, we let $\mathcal{L}(\mathbb{B}_1, \mathbb{B}_2)$ denote the space of all bounded linear operators from $\mathbb{B}_1$ to $\mathbb{B}_2$. We write simply $\mathcal{L}(\mathbb{B})$ for $\mathcal{L}(\mathbb{B}, \mathbb{B})$. Let $\pi_j | L_2^{\mathbb{H}}(p_k) : L_2^{\mathbb{H}}(p_k) \to L_2^{\mathbb{H}}(p_j)$ denote the operator π_j restricted to $L_2^{\mathbb{H}}(p_k)$ for $k \neq j$. Under the condition (A), $\pi_j | L_2^{\mathbb{H}}(p_k)$ are integral operators with the kernel $\mathbf{k}_{jk} : [0,1]^d \times [0,1]^d \to \mathcal{L}(\mathbb{H})$ defined by $\mathbf{k}_{jk}(\mathbf{u}, \mathbf{v})(\mathbf{h}) = \mathbf{h} \odot \frac{p_{jk}(u_j, v_k)}{p_j(u_j) p_k(v_k)}$. To see this, we note that the condition (A) implies $\int_{[0,1]^{d-1}} \|\mathbf{f}_k(\mathbf{x})\| p(\mathbf{x}) d\mathbf{x}_{-j} < \infty$ for all $x_j \in [0,1]$ if $\mathbf{f}_k \in \mathcal{L}_2^{\mathbb{H}}(p_k)$, so that $D_j(\mathbf{f}_k) = [0,1]$ for all $\mathbf{f}_k \in \mathcal{L}_2^{\mathbb{H}}(p_k)$. Thus, it holds that

$$\begin{aligned} \pi_{j\mathbf{f}_k}(u_j) &= \int_{[0,1]^d} \mathbf{f}_k(\mathbf{x}) \odot \frac{p_{jk}(u_j, x_k)}{p_j(u_j) p_k(x_k)} dP\mathbf{X}^{-1}(\mathbf{x}) \\ &= \int_{[0,1]^d} \mathbf{k}_{jk}(\mathbf{u}, \mathbf{x})(\mathbf{f}_k(\mathbf{x})) dP\mathbf{X}^{-1}(\mathbf{x}). \end{aligned}$$

Similarly, under the condition (S), $\hat{\pi}_j | L_2^{\mathbb{H}}(\hat{p}_k)$ are integral operators with the kernel $\hat{\mathbf{k}}_{jk} : [0,1]^d \times [0,1]^d \to \mathcal{L}(\mathbb{H})$ defined by $\hat{\mathbf{k}}_{jk}(\mathbf{u}, \mathbf{v})(\mathbf{h}) = \mathbf{h} \odot \frac{\hat{p}_{jk}(u_j, v_k)}{\hat{p}_j(u_j) \hat{p}_k(v_k)}$.

3.2 Compactness of projection operators

In the case where $\mathbb{H} = \mathbb{R}$, a common approach to establishing the existence of the SBF estimators and the convergence of the SBF algorithm is to prove that $\pi_j|L_2^{\mathbb{H}}(p_k)$ or $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ for all $1 \leq j \neq k \leq d$ are *compact operators*, see Mammen et al. (1999) or a more recent Mammen et al. (2014), for example. Indeed, it follows from Proposition A.4.2 in Bickel et al. (1993) that, if $\pi_j|L_2^{\mathbb{H}}(p_k)$ for all $1 \leq j \neq k \leq d$ are compact, then

$$S^{\mathbb{H}}(p) \leq L_2^{\mathbb{H}}(p). \tag{3.2.1}$$

Moreover, according to Corollary 4.3 in Xu and Zikatanov (2002), (3.2.1) implies

$$\|T\|_{\mathcal{L}(S^{\mathbb{H}}(p))} < 1, \tag{3.2.2}$$

where T is an operator in $\mathcal{L}(S^{\mathbb{H}}(p))$ defined by $T = (I - \pi_d) \circ \cdots \circ (I - \pi_1)$, where I is the identity operator. The same properties hold for $S^{\mathbb{H}}(\hat{p})$ and for $\hat{T}$, defined in the same way as T with π_j being replaced by $\hat{\pi}_j$, if $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ are compact. The two properties at (3.2.1) and (3.2.2) are essential to the existence of the B-SBF estimators and the convergence of the B-SBF algorithm.

The compactness of $\pi_j|L_2^{\mathbb{H}}(p_k)$ or $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ has been unknown when $\mathbb{H} \neq \mathbb{R}$. Some sufficient conditions for the compactness of integral operators defined on Lebesgue-Bochner spaces of 'μ-measurable maps' were studied by Busby et al. (1972) and Vath (2000) among others. But, the case for 'strongly measurable maps', which are relevant in statistical applications and on which our theoretical development is based, has never been studied. Below, we present two general theorems in the latter case. The first one gives a sufficient condition for compactness, and the second is about non-compactness for certain integral operators. The two theorems have important implications in our theoretical development, while they are also of interest in their own right.

In the statements of the two theorems, $(\mathcal{Z}, \mathscr{A}, \mu)$ and $(\mathcal{W}, \mathscr{B}, \nu)$ are measure spaces and $\mathbb{B}_1$ and $\mathbb{B}_2$ are separable Banach spaces. We denote by $\|\cdot\|_{\mathcal{L}(\mathbb{B}_1,\mathbb{B}_2)}$ the operator norm of $\mathcal{L}(\mathbb{B}_1, \mathbb{B}_2)$. Let $1 < p, q < \infty$ satisfy $p^{-1} + q^{-1} = 1$. Let $\mathbf{k} : \mathcal{Z} \times \mathcal{W} \to \mathcal{L}(\mathbb{B}_1, \mathbb{B}_2)$ be a *measurable* map such that $\int_{\mathcal{Z}\times\mathcal{W}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\|^q_{\mathcal{L}(\mathbb{B}_1,\mathbb{B}_2)} d\mu\otimes\nu(\mathbf{z}, \mathbf{w}) < \infty$. Define $L : L^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B}_1) \to L^q((\mathcal{W}, \mathscr{B}, \nu), \mathbb{B}_2)$ by $L([\mathbf{f}]) = [L_{\mathbf{f}}]$, where

$$L_{\mathbf{f}}(\mathbf{w}) = \begin{cases} \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w})(\mathbf{f}(\mathbf{z}))d\mu(\mathbf{z}), & \text{if } \mathbf{w} \in D_{\mathcal{W}} \\ L_0(\mathbf{f})(\mathbf{w}), & \text{otherwise,} \end{cases} \tag{3.2.3}$$

where $D_{\mathcal{W}} = \{\mathbf{w} \in \mathcal{W} : \int_{\mathcal{Z}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\|^q_{\mathcal{L}(\mathbb{B}_1,\mathbb{B}_2)} d\mu(\mathbf{z}) < \infty\}$ and L_0 is any linear map from $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B}_1)$ to $\{\mathbf{g} : \mathcal{W} \to \mathbb{B}_2 \,|\, \mathbf{g} \text{ is measurable}\}$. Finally, we let $\mathcal{C}(\mathbb{B}_1, \mathbb{B}_2)$ denote the space of all compact operators from $\mathbb{B}_1$ to $\mathbb{B}_2$.

Theorem 3.2.1. *L is a bounded linear operator. Furthermore, if* $\mathrm{range}(\mathbf{k}) \subset \mathcal{C}(\mathbb{B}_1, \mathbb{B}_2)$*, then L is compact.*

Theorem 3.2.1 tells that, if the kernel of an integral operator takes values in the space of compact operators, then the integral operator is compact. To apply the theorem to $\pi_j|L_2^{\mathbb{H}}(p_k)$ or $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ we need to check the measurability of $\mathbf{k}_{jk}$ and $\hat{\mathbf{k}}_{jk}$. This is not trivial since the Banach space $\mathcal{C}(\mathbb{B}_1, \mathbb{B}_2)$ is not separable in general. In Lemma 6.1.6 in the Appendix we establish that both

$\mathbf{k}_{jk}$ and $\hat{\mathbf{k}}_{jk}$ are measurable. One can also show that $\mathbf{k}_{jk}(\mathbf{u}, \mathbf{v})$ and $\hat{\mathbf{k}}_{jk}(\mathbf{u}, \mathbf{v})$ belong to $\mathcal{C}(\mathbb{H}, \mathbb{H})$ for all $\mathbf{u}, \mathbf{v} \in [0,1]^d$ under the conditions (A) and (S), respectively, if $\mathbb{H}$ is finite-dimensional.

Corollary 3.2.1. *Suppose that $\mathbb{H}$ is finite-dimensional. Then, for all $1 \le j \ne k \le d$, $\pi_j | L_2^{\mathbb{H}}(p_k)$ and $\hat{\pi}_j | L_2^{\mathbb{H}}(\hat{p}_k)$ are compact under the conditions (A) and (S), respectively.*

At the beginning we thought that $\pi_j | L_2^{\mathbb{H}}(p_k)$ and $\hat{\pi}_j | L_2^{\mathbb{H}}(\hat{p}_k)$ might be also compact when $\mathbb{H}$ is infinite-dimensional. However, we find that the conclusion of Corollary 3.2.1 is not valid for infinite-dimensional $\mathbb{H}$, which follows from an application of the following theorem.

Theorem 3.2.2. *Suppose that $\mu(\mathcal{Z}) < \infty$. Let $\kappa : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$ be a measurable function such that $\int_{\mathcal{Z}\times\mathcal{W}} |\kappa(\mathbf{z}, \mathbf{w})|^q d\mu \otimes \nu(\mathbf{z}, \mathbf{w}) < \infty$ and $0 < \int_{\mathcal{W}} \left| \int_{\mathcal{Z}} \kappa(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) \right|^q d\nu(\mathbf{w}) < \infty$. Let $C \in \mathcal{L}(\mathbb{B}_1, \mathbb{B}_2)$ be a non-compact operator. Then, the operator L at (3.2.3) with $\mathbf{k}(\mathbf{z}, \mathbf{w})(\mathbf{h}) = \kappa(\mathbf{z}, \mathbf{w}) \odot C(\mathbf{h})$ is a non-compact bounded linear operator.*

For the application of Theorem 3.2.2 to $\pi_j | L_2^{\mathbb{H}}(p_k)$ and $\hat{\pi}_j | L_2^{\mathbb{H}}(\hat{p}_k)$, we take $\kappa_{jk} : [0,1]^d \times [0,1]^d \to \mathbb{R}$ defined by

$$\kappa_{jk}(\mathbf{u}, \mathbf{v}) = p_{jk}(u_j, v_k)/(p_j(u_j) p_k(v_k))$$

for κ in the theorem, and the identity operator $I_{\mathbb{H}} : \mathbb{H} \to \mathbb{H}$ for C. Note that $I_{\mathbb{H}}$ is non-compact since the unit closed balls in infinite-dimensional Hilbert spaces are not compact. Also, κ_{jk} satisfies the conditions of κ in Theorem 3.2.2 under the condition (A). The same holds for $\hat{\kappa}_{jk}$ defined by $\hat{\kappa}_{jk}(\mathbf{u}, \mathbf{v}) = \hat{p}_{jk}(u_j, v_k)/(\hat{p}_j(u_j) \hat{p}_k(v_k))$

under the condition (S). Therefore, surprisingly we have the following corollary of Theorem 3.2.2.

Corollary 3.2.2. *Suppose that $\mathbb{H}$ is infinite-dimensional. Then, $\pi_j|L_2^{\mathbb{H}}(p_k)$ and $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ for all $1 \leq j \neq k \leq d$ are non-compact under the conditions (A) and (S), respectively.*

3.3 Existence of B-SBF estimators

Non-compactness of $\pi_j|L_2^{\mathbb{H}}(p_k)$ and $\hat{\pi}_j|L_2^{\mathbb{H}}(\hat{p}_k)$ raises a major difficulty in proving (3.2.1) and (3.2.2) since the earlier proofs of them for the case $\mathbb{H} = \mathbb{R}$ use the compactness of the respective projection operators. To tackle the difficulty, we rely on the following equivalence result on (3.2.1) and (3.2.2), which is a direct consequence of applying Lemma 6.1.7 in the Appendix and Proposition 3.1.2. We state the result only for the empirical versions $S^{\mathbb{H}}(\hat{p})$ and $\hat{T}$, but an obvious analogue holds for $S^{\mathbb{H}}(p)$ and T as well. Let $\overline{S^{\mathbb{H}}(\hat{p})}$ denote the closure of $S^{\mathbb{H}}(\hat{p})$.

Proposition 3.3.1. *Assume that $\hat{p}_j(x_j) > 0$ for all $x_j \in [0,1]$ and $1 \leq j \leq d$. Then, the followings are equivalent.*

(a) $S^{\mathbb{H}}(\hat{p}) \leq L_2^{\mathbb{H}}(\hat{p})$.

(b) $\|\hat{T}\|_{\mathcal{L}(\overline{S^{\mathbb{H}}(\hat{p})})} < 1$.

(c) There exists $\hat{c} > 0$ such that, for all $[\mathbf{f}] \in S^{\mathbb{H}}(\hat{p})$, there exist $[\mathbf{f}_1] \in L_2^{\mathbb{H}}(\hat{p}_1), \ldots, [\mathbf{f}_d] \in L_2^{\mathbb{H}}(\hat{p}_d)$ satisfying $\bigoplus_{j=1}^d [\mathbf{f}_j] = [\mathbf{f}]$ and $\sum_{j=1}^d \|[\mathbf{f}_j]\|_{2,n}^2 \leq \hat{c}\,\|[\mathbf{f}]\|_{2,n}^2$.

The most difficulty is that the above proposition does not say that one of (a)–(c) is true, which has never been known. Proving or disproving any of the statements in the proposition is not easy. We find that standard inequalities such as Hölder's and those in Diaz and Metcalf (1966), for example, are not helpful. However, we are able to show that the 'compatibility' condition (c) for sum-maps holds, with an innovative use of Corollary 3.2.1.

Theorem 3.3.1. *Assume that (S) holds. Then, the statements in Proposition 3.3.1 are true.*

Proof. We only need to prove the theorem for infinite-dimensional separable $\mathbb{H}$ since the case of finite-dimensional $\mathbb{H}$ is implied by Corollary 3.2.1 and Proposition 3.3.1. We prove (c) in Proposition 3.3.1. Let $[\mathbf{f}] \in S^{\mathbb{H}}(\hat{p})$ be given and $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal basis of $\mathbb{H}$. Then, $\mathbf{f}(\mathbf{x}) = \bigoplus_{k=1}^{\infty} \langle \mathbf{f}(\mathbf{x}), \mathbf{e}_k \rangle \odot \mathbf{e}_k$ and $\|\mathbf{f}(\mathbf{x})\|^2 = \sum_{k=1}^{\infty} \langle \mathbf{f}(\mathbf{x}), \mathbf{e}_k \rangle^2$ for all $\mathbf{x} \in [0,1]^d$. Thus, we have

$$\|[\mathbf{f}]\|_{2,n}^2 = \int_{[0,1]^d} \sum_{k=1}^{\infty} \langle \mathbf{f}(\mathbf{x}), \mathbf{e}_k \rangle^2 \hat{p}(\mathbf{x}) d\mathbf{x} = \sum_{k=1}^{\infty} \|[\langle \mathbf{f}(\cdot), \mathbf{e}_k \rangle]\|_{2,n}^2,$$

where with slight abuse of the notation for the norm $\|\cdot\|_{2,n}$, we write $\|[g]\|_{2,n}^2$ for real-valued maps $[g] \in L_2^{\mathbb{R}}(\hat{p})$ as well, meaning that $\|[g]\|_{2,n}^2 = \int |g(\mathbf{x})|^2 \hat{p}(\mathbf{x}) d\mathbf{x}$. By applying Corollary 3.2.1 and Proposition 3.3.1 with $\mathbb{H} = \mathbb{R}$, we can argue that there exists $\hat{c} > 0$ such that, for any $[g] \in S^{\mathbb{R}}(\hat{p})$, there exist $[g_j] \in L_2^{\mathbb{R}}(\hat{p}_j)$ for $1 \le j \le d$ satisfying $[g] = \sum_{j=1}^{d} [g_j]$ and $\sum_{j=1}^{d} \|[g_j]\|_{2,n}^2 \le \hat{c} \|[g]\|_{2,n}^2$. For this we have used the condition (S). Since $[\langle \mathbf{f}(\cdot), \mathbf{e}_k \rangle] \in S^{\mathbb{R}}(\hat{p})$ for all $k \ge 1$, this entails that, for each $k \ge 1$, there exist $[f_{kj}] \in L_2^{\mathbb{R}}(\hat{p}_j)$, $1 \le j \le d$, such that $[\langle \mathbf{f}(\cdot), \mathbf{e}_k \rangle] = \sum_{j=1}^{d} [f_{kj}]$ and $\sum_{j=1}^{d} \|[f_{kj}]\|_{2,n}^2 \le \hat{c} \|[\langle \mathbf{f}(\cdot), \mathbf{e}_k \rangle]\|_{2,n}^2$. Thus, it holds that

$$\sum_{j=1}^{d} \sum_{k=1}^{\infty} \|[f_{kj}]\|_{2,n}^2 \le \hat{c} \sum_{k=1}^{\infty} \|[\langle \mathbf{f}(\cdot), \mathbf{e}_k \rangle]\|_{2,n}^2 = \hat{c}\, \|[\mathbf{f}]\|_{2,n}^2 < \infty. \quad (3.3.1)$$

Now, (3.3.1) implies that, for each $1 \le j \le d$, the sequence $\{\bigoplus_{k=1}^{n}[f_{kj}(\cdot)\odot \mathbf{e}_k]\}_{n\ge 1}$ is Cauchy in $L_2^{\mathbb{H}}(\hat{p}_j)$ since

$$\Big\| \bigoplus_{k=m+1}^{n} [f_{kj}(\cdot)\odot \mathbf{e}_k]\Big\|_{2,n}^2 = \int_{[0,1]^d} \sum_{k=m+1}^{n} \|f_{kj}(\mathbf{x})\odot \mathbf{e}_k\|^2 \hat{p}(\mathbf{x})d\mathbf{x}$$
$$= \sum_{k=m+1}^{n} \|[f_{kj}]\|_{2,n}^2 \to 0$$

as $n > m \to \infty$. Denote the limit of the Cauchy sequence in $L_2^{\mathbb{H}}(\hat{p}_j)$ by $[\mathbf{f}_j]$. Then, there exists a subsequence $\{\bigoplus_{k=1}^{n_{jl}}[f_{kj}(\cdot)\odot \mathbf{e}_k]\}_{l\ge 1}$ of $\{\bigoplus_{k=1}^{n}[f_{kj}(\cdot)\odot \mathbf{e}_k]\}_{n\ge 1}$ such that

$$\lim_{l\to\infty} \bigoplus_{k=1}^{n_{jl}} f_{kj}(\mathbf{x})\odot \mathbf{e}_k = \mathbf{f}_j(\mathbf{x}) \text{ a.e. } [\hat{P}\mathbf{X}^{-1}].$$

Then,

$$\begin{aligned}
\bigoplus_{j=1}^{d} \mathbf{f}_j(\mathbf{x}) &= \bigoplus_{k=1}^{\infty} \Big(\sum_{j=1}^{d} \langle \lim_{l\to\infty} \bigoplus_{i=1}^{n_{jl}} f_{ij}(\mathbf{x})\odot \mathbf{e}_i, \mathbf{e}_k\rangle\Big)\odot \mathbf{e}_k \\
&= \bigoplus_{k=1}^{\infty} \Big(\sum_{j=1}^{d} \lim_{l\to\infty} \langle \bigoplus_{i=1}^{n_{jl}} f_{ij}(\mathbf{x})\odot \mathbf{e}_i, \mathbf{e}_k\rangle\Big)\odot \mathbf{e}_k \\
&= \bigoplus_{k=1}^{\infty} \Big(\sum_{j=1}^{d} f_{kj}(\mathbf{x})\Big)\odot \mathbf{e}_k \\
&= \bigoplus_{k=1}^{\infty} \langle \mathbf{f}(\mathbf{x}), \mathbf{e}_k\rangle \odot \mathbf{e}_k = \mathbf{f}(\mathbf{x})
\end{aligned}$$

a.e. $[\hat{P}\mathbf{X}^{-1}]$. Moreover, using the fact that $\mathbf{h}_n \to \mathbf{h} \in \mathbb{H}$ and $\mathbf{h}_n^* \to \mathbf{h}^* \in \mathbb{H}$ imply $\langle \mathbf{h}_n, \mathbf{h}_n^*\rangle \to \langle \mathbf{h}, \mathbf{h}^*\rangle$, we get

$$\sum_{j=1}^{d} \|[\mathbf{f}_j]\|_{2,n}^2 = \sum_{j=1}^{d} \int_{[0,1]^d} \Big(\lim_{l\to\infty} \sum_{k=1}^{n_{jl}} f_{kj}^2(\mathbf{x})\Big)\hat{p}(\mathbf{x})d\mathbf{x} = \sum_{j=1}^{d}\sum_{k=1}^{\infty} \|[f_{kj}]\|_{2,n}^2 \le \hat{c}\,\|[\mathbf{f}]\|_{2,n}^2,$$

where the inequality follows from (3.3.1). □

We are now ready to discuss the existence of the B-SBF estimators. For this we consider an objective functional $\hat{F} : S^{\mathbb{H}}(\hat{p}) \to \mathbb{R}$ defined by

$$\hat{F}([\mathbf{f}]) = \int_{[0,1]^d} n^{-1} \sum_{i=1}^{n} \|\mathbf{Y}_i \ominus \mathbf{f}(\mathbf{x})\|^2 \cdot \prod_{j=1}^{d} K_{h_j}(x_j, X_{ij}) d\mathbf{x}.$$

$\hat{F}$ is well-defined since $\hat{F}([\mathbf{f}]) \leq 2\big(\max_{1\leq i\leq n} \|\mathbf{Y}_i\|^2 + \|[\mathbf{f}]\|_{2,n}^2\big) < \infty$. Now, for $[\mathbf{f}], [\mathbf{g}] \in S^{\mathbb{H}}(\hat{p})$,

$$\begin{aligned}
&\lim_{\epsilon\to 0} \epsilon^{-1} \left[\hat{F}([\mathbf{f}] \oplus \epsilon \odot [\mathbf{g}]) - \hat{F}([\mathbf{f}])\right] \qquad (3.3.2)\\
&= -2 \int_{[0,1]^d} n^{-1}\sum_{i=1}^{n}\langle \mathbf{Y}_i \ominus \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x})\rangle \prod_{j=1}^{d} K_{h_j}(x_j, X_{ij}) d\mathbf{x}\\
&=: D\hat{F}([\mathbf{f}])([\mathbf{g}]).
\end{aligned}$$

Clearly, $D\hat{F}([\mathbf{f}]) : S^{\mathbb{H}}(\hat{p}) \to \mathbb{R}$ is a linear operator. It is also bounded, which we may verify by using Hölder's inequality. Hence, $\hat{F}$ is Gâteaux differentiable.

Theorem 3.3.2. *Assume that the condition (S) holds. Then, there exists a solution* $(\hat{\mathbf{m}}_1, \cdots, \hat{\mathbf{m}}_d) \in \prod_{j=1}^{d} \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ *of (2.4.6) satisfying (2.4.7). Moreover, their sum is unique in the sense that if* $(\hat{\mathbf{m}}_1^*, \cdots, \hat{\mathbf{m}}_d^*)$ *is another solution, then* $\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j(x_j) = \bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j^*(x_j)$ *a.e.* $[\hat{P}\mathbf{X}^{-1}]$*. Furthermore, if* $\hat{p}(\mathbf{x}) > 0$ *for all* $\mathbf{x} \in [0,1]^d$*, then the decomposition of the sum is unique in the sense that* $\hat{\mathbf{m}}_j(x_j) = \hat{\mathbf{m}}_j^*(x_j)$ *a.e.* $[\mathrm{Leb}_1]$ *for all* $1 \leq j \leq d$.

Proof. First, we note that $\hat{F}$ is a convex and continuous functional satisfying $\hat{F}([\mathbf{f}]) \to \infty$ as $\|[\mathbf{f}]\|_{2,n} \to \infty$. These with Theorem 3.3.1 and Lemma 4 in Beltrami (1967) imply that there exists a minimizer of $\hat{F}$ in $S^{\mathbb{H}}(\hat{p})$. Now, $[\hat{\mathbf{f}}]$ being a minimizer of $\hat{F}$ is equivalent to $D\hat{F}([\hat{\mathbf{f}}])([\mathbf{g}]) = 0$ for all $[\mathbf{g}] \in S^{\mathbb{H}}(\hat{p})$, where $D\hat{F}([\mathbf{f}])$ is defined at (3.3.2). With specification of $[\mathbf{g}] \in S^{\mathbb{H}}(\hat{p})$ to $[\mathbf{g}_j] \in L_2^{\mathbb{H}}(\hat{p}_j)$ for each $1 \le j \le d$, we find that this is equivalent to

$$\int_{[0,1]^{d-1}} n^{-1} \odot \left(\bigoplus_{i=1}^{n} (\mathbf{Y}_i \ominus \hat{\mathbf{f}}(\mathbf{x})) \odot \prod_{j=1}^{d} K_{h_j}(x_j, X_{ij}) \right) d\mathbf{x}_{-j} = \mathbf{0} \tag{3.3.3}$$

a.e. $x_j \in [0,1]$ with respect to Leb_1, for all $1 \le j \le d$. Let $\hat{\mathbf{f}} = \mathbf{f}_0 \oplus \bigoplus_{j=1}^{d} \hat{\mathbf{f}}_j$ be a decomposition of $\hat{\mathbf{f}}$ such that $\hat{\mathbf{f}}_j \in \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ and $\int_0^1 \hat{\mathbf{f}}_j(x_j) \odot \hat{p}_j(x_j) dx_j = 0$ for all $1 \le j \le d$. Plugging the decomposition into the left hand side of (3.3.3) and by using $\int_0^1 K_{h_j}(x_j, X_{ij}) dx_j \equiv 1$, we see that $\mathbf{f}_0 = \bar{\mathbf{Y}}$ and $(\hat{\mathbf{f}}_j : 1 \le j \le d)$ satisfies

$$\hat{\mathbf{f}}_j(x_j) = \tilde{\mathbf{m}}_j(x_j) \ominus \bar{\mathbf{Y}} \ominus \bigoplus_{k \ne j} \int_0^1 \hat{\mathbf{f}}_k(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k$$

a.e. $x_j \in [0,1]$ with respect to Leb_1, for all $1 \le j \le d$. Define the right hand side by $\hat{\mathbf{m}}_j(x_j)$ for all $x_j \in [0,1]$. Then, $(\hat{\mathbf{m}}_j : 1 \le j \le d) \in \prod_{j=1}^{d} \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$ and it satisfies (2.4.6) and (2.4.7).

From (2.4.6), we may verify that $[\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j] = \hat{T}([\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j]) \oplus [\tilde{\mathbf{s}}]$ and $[\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j^*] = \hat{T}([\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j^*]) \oplus [\tilde{\mathbf{s}}]$ where

$$[\tilde{\mathbf{s}}] = [\tilde{\mathbf{m}}_d \ominus \bar{\mathbf{Y}}] \oplus (I - \hat{\pi}_d)([\tilde{\mathbf{m}}_{d-1} \ominus \bar{\mathbf{Y}}]) \oplus \cdots \oplus (I - \hat{\pi}_d) \circ \cdots \circ (I - \hat{\pi}_2)([\tilde{\mathbf{m}}_1 \ominus \bar{\mathbf{Y}}]) \in S^{\mathbb{H}}(\hat{p}).$$

Since $\|\hat{T}\|_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))} < 1$ from Theorem 3.3.1, it holds that $[\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j] = [\bigoplus_{j=1}^{d} \hat{\mathbf{m}}_j^*]$. This proves the first part of the theorem.

For the proof of the second part, suppose that $\bigoplus_{j=1}^{d} \hat{\mathbf{g}}_j(x_j) = \mathbf{0}$ a.e. $[\hat{P}\mathbf{X}^{-1}]$ with $\hat{\mathbf{g}}_j$ satisfying (2.4.7). Since $\hat{p} > 0$ on $[0,1]^d$ by the assumption, this implies $\bigoplus_{j=1}^{d} \hat{\mathbf{g}}_j(x_j) = \mathbf{0}$ a.e. on $[0,1]^d$ with respect to Leb_d, so that, for any map $\boldsymbol{\delta}_j \in \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$, we get

$$\left\langle \bigoplus_{k=1}^{d} \hat{\mathbf{g}}_k(x_k) \odot \hat{p}_{\mathbf{X}_{-j}}(\mathbf{x}_{-j}),\, \boldsymbol{\delta}_j(x_j) \right\rangle = 0 \text{ a.e. on } [0,1]^d \text{ w.r.t. } \mathrm{Leb}_d. \tag{3.3.4}$$

Because of the marginalization property of $\hat{p}_{\mathbf{X}_{-j}}$ such that

$$\int_{[0,1]^{d-2}} \hat{p}_{\mathbf{X}_{-j}}(\mathbf{x}_{-j}) d\mathbf{x}_{-j,k} = \hat{p}_k(x_k)$$

and the constraints (2.4.7), the equation (3.3.4) implies that

$$\begin{aligned} 0 &= \sum_{k=1}^{d} \int_{[0,1]^d} \langle \hat{\mathbf{g}}_k(x_k) \odot \hat{p}_{\mathbf{X}_{-j}}(\mathbf{x}_{-j}),\, \boldsymbol{\delta}_j(x_j) \rangle d\mathbf{x} \\ &= \sum_{k \neq j}^{d} \int_0^1 \left\langle \int_0^1 \hat{\mathbf{g}}_k(x_k) \odot \hat{p}_k(x_k) dx_k,\, \boldsymbol{\delta}_j(x_j) \right\rangle dx_j + \int_0^1 \langle \hat{\mathbf{g}}_j(x_j),\, \boldsymbol{\delta}_j(x_j) \rangle dx_j \\ &= \int_0^1 \langle \hat{\mathbf{g}}_j(x_j),\, \boldsymbol{\delta}_j(x_j) \rangle dx_j \end{aligned}$$

for all $\boldsymbol{\delta}_j \in \mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$. This implies $\hat{\mathbf{g}}_j(x_j) = \mathbf{0}$ a.e. on $[0,1]$ with respect to Leb_1. This proves the second part of the theorem. □

3.4 Convergence of B-SBF algorithm

In this subsection we establish the convergence of the B-SBF algorithm (2.4.8). We first consider convergence in the empirical norm, $\|\cdot\|_{2,n}$, for fixed n and given observations $(\mathbf{X}_i, \mathbf{Y}_i)$, $1 \leq i \leq n$. Then, we study convergence in $\|\cdot\|_2$ norm, where we let n diverge to infinity. We note that all works in the smooth backfitting literature treated only the latter asymptotic version for $\mathbb{H} = \mathbb{R}$. Throughout this section we assume that the initial estimators $\hat{\mathbf{m}}_j^{[0]}$ are measurable and satisfy $\max_{1\leq j\leq d} \int_0^1 \|\hat{\mathbf{m}}_j^{[0]}(x_j)\|^2 \hat{p}_j(x_j) dx_j < C$

for an absolute constant $0 < C < \infty$. This is not restrictive since we can take $\hat{\mathbf{m}}_j^{[0]} \equiv \mathbf{0}$ for all $1 \le j \le d$. Under this condition on the initial estimators and the condition (S) one can verify that all the subsequent updates $\hat{\mathbf{m}}_j^{[r]}$ are also measurable and satisfy $\max_{1 \le j \le d} \int_0^1 \|\hat{\mathbf{m}}_j^{[r]}(x_j)\|^2 \hat{p}_j(x_j) dx_j < \infty$. The following theorem is a non-asymptotic version of the convergence of the B-SBF algorithm.

Theorem 3.4.1. *Assume that the condition (S) holds. Then,* $\|\hat{T}\|_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))} < 1$ *and there exists* $\hat{c} > 0$ *such that*

$$\int_{[0,1]^d} \left\|\hat{\mathbf{m}}(\mathbf{x}) \ominus \hat{\mathbf{m}}^{[r]}(\mathbf{x})\right\|^2 \hat{p}(\mathbf{x}) d\mathbf{x} \le \hat{c}\, \|\hat{T}\|^r_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))} \quad \textit{for all } r \ge 0.$$

Proof. We embed $\tilde{\mathbf{m}}_j, \hat{\mathbf{m}}_j$ and $\hat{\mathbf{m}}_j^{[r]}$ into $\mathcal{L}_2^{\mathbb{H}}(\hat{p}_j)$. Then, from (2.4.6) and (2.4.8)

$$\begin{aligned}
[\hat{\mathbf{m}}_j] &= [\tilde{\mathbf{m}}_j] \ominus [\bar{\mathbf{Y}}] \ominus \bigoplus_{k \ne j} \hat{\pi}_j([\hat{\mathbf{m}}_k]), \quad 1 \le j \le d, \\
[\hat{\mathbf{m}}_j^{[r]}] &= [\tilde{\mathbf{m}}_j] \ominus [\bar{\mathbf{Y}}] \ominus \bigoplus_{k < j} \hat{\pi}_j([\hat{\mathbf{m}}_k^{[r]}]) \ominus \bigoplus_{k > j} \hat{\pi}_j([\hat{\mathbf{m}}_k^{[r-1]}]), \quad 1 \le j \le d.
\end{aligned} \tag{3.4.1}$$

Define $\hat{\mathbf{s}} = \bigoplus_{j=1}^d \hat{\mathbf{m}}_j$ and $\hat{\mathbf{s}}^{[r]} = \bigoplus_{j=1}^d \hat{\mathbf{m}}_j^{[r]}$. Then, the two systems of equations at (3.4.1) are expressed as $[\hat{\mathbf{s}}] = \hat{T}([\hat{\mathbf{s}}]) \oplus [\tilde{\mathbf{s}}]$ and $[\hat{\mathbf{s}}^{[r]}] = \hat{T}([\hat{\mathbf{s}}^{[r-1]}]) \oplus [\tilde{\mathbf{s}}]$, respectively, where

$$[\tilde{\mathbf{s}}] = [\tilde{\mathbf{m}}_d \ominus \bar{\mathbf{Y}}] \oplus (I - \hat{\pi}_d)([\tilde{\mathbf{m}}_{d-1} \ominus \bar{\mathbf{Y}}]) \oplus \cdots \oplus (I - \hat{\pi}_d) \circ \cdots \circ (I - \hat{\pi}_2)([\tilde{\mathbf{m}}_1 \ominus \bar{\mathbf{Y}}]) \in S^{\mathbb{H}}(\hat{p}).$$

Since $\|\hat{T}\|_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))} < 1$ from Theorem 3.3.1, it holds that $\bigoplus_{k=0}^\infty \hat{T}^k([\tilde{\mathbf{s}}])$ exists in $S^{\mathbb{H}}(\hat{p})$, $\bigoplus_{k=0}^\infty \hat{T}^k([\tilde{\mathbf{s}}]) = \hat{T}(\bigoplus_{k=0}^\infty \hat{T}^k([\tilde{\mathbf{s}}])) \oplus [\tilde{\mathbf{s}}]$ and thus $\bigoplus_{k=0}^\infty \hat{T}^k([\tilde{\mathbf{s}}]) = [\hat{\mathbf{s}}]$. This entails

$$\|[\hat{\mathbf{s}}^{[r]} \ominus \hat{\mathbf{s}}]\|_{2,n} \le \frac{\|\hat{T}\|^r_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))}}{1 - \|\hat{T}\|_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))}} \left(\|[\hat{\mathbf{s}}^{[0]}]\|_{2,n} + \|[\tilde{\mathbf{s}}]\|_{2,n} \right). \tag{3.4.2}$$

The inequality (3.4.2) gives the theorem with the choice $\hat{c} = \big(\|[\hat{\mathbf{s}}^{[0]}]\|_{2,n} + \|[\tilde{\mathbf{s}}]\|_{2,n}\big)/(1 - \|\hat{T}\|_{\mathcal{L}(S^{\mathbb{H}}(\hat{p}))})$. $\square$

We now turn to the asymptotic version of the convergence of the B-SBF algorithm in $\|\cdot\|_2$ norm. For this we need the following additional conditions.

Condition (B).

(B1) $E(\|\mathbf{Y}\|^{\alpha}) < \infty$ *for some* $\alpha > 2$.

(B2) p *is bounded away from zero and infinity on* $[0,1]^d$, *and* p_{jk} *are continuous on* $[0,1]^2$ *for* $1 \leq j \neq k \leq d$.

(B3) K *is Lipschitz continuous and* $\int_{-1}^{0} K(u)du \wedge \int_{0}^{1} K(u)du > 0$.

(B4) $h_j, \sqrt{\log n/(nh_jh_k)} = o(1)$ *and* $\inf_n n^{c_j} h_j \geq$ *(const.) for some* $c_j < (\alpha-2)/\alpha$ *for* $1 \leq j \neq k \leq d$.

Theorem 3.4.2. *Assume the condition (B). Then, there exist constants* $c > 0$ *and* $\gamma \in (0,1)$ *such that*

$$\lim_{n\to\infty} P\Big(\max_{1\leq j\leq d} \int_0^1 \|\hat{\mathbf{m}}_j(x_j) \ominus \hat{\mathbf{m}}_j^{[r]}(x_j)\|^2 p_j(x_j)dx_j \leq c\,\gamma^r \text{ for all } r \geq 0\Big) = 1.$$

Theorem 3.4.2 is about the L^2-convergence of the B-SBF algorithm, like all other results in the literature on smooth backfitting for $\mathbb{H} = \mathbb{R}$. Here, we add a new convergence result, which is also of interest. We note that the theorem implies $\sum_{r=1}^{\infty} \int_0^1 \|\hat{\mathbf{m}}_j(x_j) \ominus \hat{\mathbf{m}}_j^{[r]}(x_j)\|^2 p_j(x_j)dx_j < \infty$ with probability tending to one. This entails that, with probability tending to one, $\sum_{r=1}^{\infty} \|\hat{\mathbf{m}}_j(x_j) \ominus \hat{\mathbf{m}}_j^{[r]}(x_j)\|^2 p_j(x_j) < \infty$ a.e. $x_j \in [0,1]$ with respect to Leb_1, which gives the following corollary.

Corollary 3.4.1. *Assume that the condition (B) holds. Then, for* $1 \leq j \leq d$,

$$\lim_{n\to\infty} P\left(\hat{\mathbf{m}}_j^{[r]}(x_j) \to \hat{\mathbf{m}}_j(x_j) \text{ as } r \to \infty \text{ a.e. } x_j \text{ with respect to } \mathrm{Leb}_1\right) = 1.$$

Chapter 4

Asymptotic properties

4.1 Rates of convergence

Below we collect the assumptions for our asymptotic theory.

Condition (C).

(C1) $E(\|\mathbf{Y}\|^\alpha) < \infty$ *for some* $\alpha > 5/2$.

(C2) The true maps $\mathbf{m}_j$ *for* $1 \leq j \leq d$ *are twice continuously Fréchet differentiable on* $[0, 1]$.

(C3) The condition (B2) in Section 3.4 holds. In addition, p_{jk} *are* C^1 *on* $[0, 1]^2$ *for* $1 \leq j \neq k \leq d$.

(C4) The condition (B3) in Section 3.4 holds. In addition, $\int_{-1}^{1} uK(u)du = 0$.

(C5) $n^{1/5}h_j \to \alpha_j$ *for some positive constant* α_j, $1 \leq j \leq d$.

The moment condition on $\mathbf{Y}$ and the Fréchet differentiability of the maps $\mathbf{m}_j : [0, 1] \to \mathbb{H}$, respectively, are natural generalizations of the usual moment condition on Euclidean errors and the smoothness assumptions on real-valued functions. In the theory, we need functional calculus for Fréchet derivatives and Bochner integrals. Other assumptions on the baseline kernel K and the density p are typical in the kernel smoothing theory.

Let $I_j = [2h_j, 1-2h_j]$ and I_j^c denote its complement in $[0,1]$. The following theorem demonstrates that our estimators achieve the univariate error rates.

Theorem 4.1.1. *Assume that the condition (C) holds. Then, the followings hold for* $1 \leq j \leq d$.

(i) (Pointwise convergence)

$$\begin{aligned}
\|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\| &= O_p(n^{-2/5}) \quad \textit{for } x_j \in I_j, \\
\|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\| &= O_p(n^{-1/5}) \quad \textit{for } x_j \in I_j^c.
\end{aligned}$$

(ii) (L_2 convergence)

$$\begin{aligned}
\int_{I_j} \|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\|^2 p_j(x_j) dx_j &= O_p(n^{-4/5}), \\
\int_0^1 \|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\|^2 p_j(x_j) dx_j &= O_p(n^{-3/5}).
\end{aligned}$$

(iii) (Uniform convergence)

$$\begin{aligned}
\sup_{x_j \in I_j} \|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\| &= O_p(n^{-2/5}\sqrt{\log n}), \\
\sup_{x_j \in [0,1]} \|\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)\| &= O_p(n^{-1/5}).
\end{aligned}$$

4.2 Asymptotic distribution and asymptotic independence

Recall that, for a mean zero random element $\mathbf{Z} : \Omega \to \mathbb{H}$, its covariance operator $C : \mathbb{H} \to \mathbb{H}$ is characterised by

$$\langle C(\mathbf{h}), \mathbf{g} \rangle = \mathrm{E}\left(\langle \mathbf{Z}, \mathbf{h} \rangle \cdot \langle \mathbf{Z}, \mathbf{g} \rangle \right), \quad \mathbf{h}, \mathbf{g} \in \mathbb{H}.$$

Also, recall that a $\mathbb{H}$-valued random element $\mathbf{Z}$ is called Gaussian if $\langle \mathbf{Z}, \mathbf{h} \rangle$ is normally distributed for any $\mathbf{h} \in \mathbb{H}$. We denote a Gaussian random element with mean zero and covariance operator C, by $\mathbf{G}(\mathbf{0}, C)$.

For brevity, we write

$$w_{ij}(u) = \big(\sum_{i=1}^{n} K_{h_j}(u, X_{ij}) \big)^{-1} \sqrt{n h_j} K_{h_j}(u, X_{ij}).$$

Then, for the marginal estimators $\tilde{\mathbf{m}}_j$ defined at (2.4.5), we may write

$$\sqrt{n h_j} \odot \tilde{\mathbf{m}}_j(x_j) = \bigoplus_{i=1}^{n} w_{ij}(x_j) \odot \mathbf{Y}_i.$$

From the standard kernel smoothing theory and the fact (2.3.1), it follows that $\sum_{i=1}^{n} w_{ij}(x_j)^2$ converges to $p_j(x_j)^{-1} \int_{-1}^{1} K^2(u) du$ in probability for each $x_j \in I_j$ under suitable conditions on p_j, h_j and K. Let $\{\mathbf{e}_l\}_{l=1}^{L}$ be an orthonormal basis of $\mathbb{H}$, where we allow $L = \infty$ for infinite-dimensional $\mathbb{H}$. Define

$$a_{j,kl}(x_j) = p_j(x_j)^{-1} \int_{-1}^{1} K^2(u) du \cdot \mathrm{E}\left(\langle \boldsymbol{\epsilon}, \mathbf{e}_k \rangle \cdot \langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \mid X_j = x_j \right)$$

for the $\mathbb{H}$-valued error $\boldsymbol{\epsilon}$ in the model (1.0.1). Then, as we show in the proof of the following theorem, the conditional covariance operator of $\bigoplus_{i=1}^n w_{ij}(x_j) \odot \boldsymbol{\epsilon}_i$ given $X_j = x_j$ is approximated by the operator $C_{j,x_j} : \mathbb{H} \to \mathbb{H}$ characterised by

$$\langle C_{j,x_j}(\mathbf{h}), \mathbf{e}_k \rangle = \sum_{l=1}^{L} \langle \mathbf{h}, \mathbf{e}_l \rangle \cdot a_{j,lk}(x_j). \tag{4.2.1}$$

The following theorem plays an important role in determining the distributions of the stochastic parts of $\hat{\mathbf{m}}_j(x_j)$.

Theorem 4.2.1. *Assume that the condition (B3) on K holds. Fix $\mathbf{x} \in I_1 \times \cdots \times I_d$ and assume that, for all $1 \le j \le d$ and k, l, (i) $\mathrm{E}(\|\boldsymbol{\epsilon}\|^\alpha) < \infty$ for some $\alpha > 2$ and $E(\langle \boldsymbol{\epsilon}, \mathbf{e}_k \rangle \cdot \langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \,|\, X_j = \cdot)$ are continuous on $[x_j - h_j, x_j + h_j]$, respectively; (ii) p_j are continuous on $[x_j - h_j, x_j + h_j]$, respectively, and $p_j(x_j) > 0$; (iii) $h_j \to 0$ and $nh_j \to \infty$ as $n \to \infty$. Then,*

$$\left(\bigoplus_{i=1}^n w_{i1}(x_1) \odot \boldsymbol{\epsilon}_i, \, \ldots, \, \bigoplus_{i=1}^n w_{id}(x_d) \odot \boldsymbol{\epsilon}_i \right) \xrightarrow{d} \left(\mathbf{G}(\mathbf{0}, C_{1,x_1}), \ldots, \mathbf{G}(\mathbf{0}, C_{d,x_d}) \right),$$

where $\mathbf{G}(\mathbf{0}, C_{1,x_1}), \cdots, \mathbf{G}(\mathbf{0}, C_{d,x_d})$ are independent.

Now, we are ready to present a theorem that demonstrates the asymptotic distribution and independence of our estimators of the component maps $\mathbf{m}_j$. In addition to (C), we need the following condition.

Condition (D). *For all $1 \le j \le d$ and k, l, the followings hold.*

(D1) $E(\langle \boldsymbol{\epsilon}, \mathbf{e}_k \rangle \cdot \langle \boldsymbol{\epsilon}, \mathbf{e}_l \rangle \,|\, X_j = \cdot)$ are continuous on $[0, 1]$.

(D2) $\partial p(\mathbf{x}) / \partial x_j$ exist and are bounded on $[0, 1]^d$.

To state the theorem we need to introduce more terminologies. For a twice Fréchet differentiable $\mathbf{f} : [0, 1] \to \mathbb{H}$, we let $D\mathbf{f} : [0, 1] \to$

$\mathcal{L}(\mathbb{R}, \mathbb{H})$ denote its first Fréchet derivative, and $D^2\mathbf{f} : [0,1] \to \mathcal{L}\big(\mathbb{R}, \mathcal{L}(\mathbb{R}, \mathbb{H})\big)$ its second Fréchet derivative. Let p_j' denote the first derivative of p_j and define

$$\begin{aligned}
\boldsymbol{\delta}_j(x_j) &= \left[\frac{p_j'(x_j)}{p_j(x_j)} \cdot \int_{-1}^{1} u^2 K(u)du\right] \odot D\mathbf{m}_j(x_j)(1), \\
\boldsymbol{\delta}_{jk}(x_j, x_k) &= \left[\frac{\partial p_{jk}(x_j, x_k)/\partial x_k}{p_{jk}(x_j, x_k)} \cdot \int_{-1}^{1} u^2 K(u)dt\right] \odot D\mathbf{m}_k(x_k)(1), \\
\tilde{\boldsymbol{\Delta}}_j(x_j) &= \alpha_j^2 \odot \boldsymbol{\delta}_j(x_j) \oplus \bigoplus_{k \neq j} \int_0^1 \boldsymbol{\delta}_{jk}(x_j, x_k) \odot \left[\alpha_k^2 \frac{p_{jk}(x_j, x_k)}{p_j(x_j)}\right] dx_k.
\end{aligned}$$

Let $(\boldsymbol{\Delta}_1, \cdots, \boldsymbol{\Delta}_d) \in \prod_{j=1}^{d} \mathcal{L}_2^{\mathbb{H}}(p_j)$ be a solution of the system of equations

$$\boldsymbol{\Delta}_j(x_j) = \tilde{\boldsymbol{\Delta}}_j(x_j) \ominus \bigoplus_{k \neq j} \int_0^1 \boldsymbol{\Delta}_k(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} dx_k, \quad 1 \leq j \leq d \tag{4.2.2}$$

satisfying the constraints

$$\int_0^1 \boldsymbol{\Delta}_j(x_j) \odot p_j(x_j) dx_j = \alpha_j^2 \odot \int_0^1 \boldsymbol{\delta}_j(x_j) \odot p_j(x_j) dx_j, \quad 1 \leq j \leq d. \tag{4.2.3}$$

Below in Theorem 4.2.2 we prove that the equation (4.2.2) subject to (4.2.3) has a unique solution. Define $\mathbf{c}_j(x_j) = \frac{1}{2}\int_{-1}^{1} u^2 K(u)du \odot D^2\mathbf{m}_j(x_j)(1)(1)$ and $\boldsymbol{\Theta}_j(x_j) = \alpha_j^2 \odot \mathbf{c}_j(x_j) \oplus \boldsymbol{\Delta}_j(x_j)$. Define $\tilde{C}_{j,x_j} : \mathbb{H} \to \mathbb{H}$ by $\tilde{C}_{j,x_j}(\mathbf{h}) = \alpha_j^{-1} \odot C_{j,x_j}(\mathbf{h})$, where C_{j,x_j} are the covariance operators defined at (4.2.1).

Theorem 4.2.2. *Assume the conditions (C) and (D). Then, there exists a solution of (4.2.2) subject to (4.2.3) and the solution is unique in the sense that if $(\boldsymbol{\Delta}_1^\star, \cdots, \boldsymbol{\Delta}_d^\star)$ is another solution, then*

$\boldsymbol{\Delta}_j(x_j) = \boldsymbol{\Delta}_j^\star(x_j)$ *a.e.* $[\mathrm{Leb}_1]$. *Furthermore, for a.e.* $\mathbf{x} \in \prod_{j=1}^d I_j$ *with respect to* Leb_d*, it holds that*

$$\begin{pmatrix} n^{2/5} \odot (\hat{\mathbf{m}}_1(x_1) \ominus \mathbf{m}_1(x_1)) \\ \vdots \\ n^{2/5} \odot (\hat{\mathbf{m}}_d(x_d) \ominus \mathbf{m}_d(x_d)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \boldsymbol{\Theta}_1(x_1) \oplus \mathbf{G}(\mathbf{0}, \tilde{C}_{1,x_1}) \\ \vdots \\ \boldsymbol{\Theta}_d(x_d) \oplus \mathbf{G}(\mathbf{0}, \tilde{C}_{d,x_d}) \end{pmatrix},$$

where $\boldsymbol{\Theta}_1(x_1) \oplus \mathbf{G}(\mathbf{0}, \tilde{C}_{1,x_1}), \cdots, \boldsymbol{\Theta}_d(x_d) \oplus \mathbf{G}(\mathbf{0}, \tilde{C}_{d,x_d})$ *are independent. Moreover,*

$$n^{2/5} \odot (\hat{\mathbf{m}}(\mathbf{x}) \ominus \mathbf{m}(\mathbf{x})) \xrightarrow{d} \bigoplus_{j=1}^d \boldsymbol{\Theta}_j(x_j) \oplus \mathbf{G}\Big(\mathbf{0}, \sum_{j=1}^d \tilde{C}_{j,x_j}\Big).$$

Let $\hat{\mathbf{m}}_j^{\mathrm{ora}}$ be the oracle estimator of $\mathbf{m}_j$ under the knowledge of all other component maps $\mathbf{m}_k, k \neq j$. Using Theorem 4.2.1, we may prove that for $x_j \in I_j$,

$$n^{2/5}(\hat{\mathbf{m}}_j^{\mathrm{ora}}(x_j) \ominus \mathbf{m}_j(x_j)) \xrightarrow{d} \alpha_j^2 \odot [\boldsymbol{\delta}_j(x_j) \oplus \mathbf{c}_j(x_j)] \oplus \mathbf{G}(\mathbf{0}, \tilde{C}_{j,x_j}).$$

Therefore, $\hat{\mathbf{m}}_j$ and $\hat{\mathbf{m}}_j^{\mathrm{ora}}$ have the same asymptotic covariance operator, but differ in their asymptotic biases. The difference of asymptotic biases is $[\alpha_j^2 \odot \boldsymbol{\delta}_j(x_j)] \ominus \boldsymbol{\Delta}_j(x_j) =: \boldsymbol{\beta}_j(x_j)$ and it holds that $E(\boldsymbol{\beta}_j(X_j)) = \int_0^1 \boldsymbol{\beta}_j(x_j) \odot p_j(x_j) dx_j = \mathbf{0}$ by (4.2.3).

Chapter 5

Numerical Study

In the simulation and real data examples presented here, we took Epanechnikov kernel $K(u) = (3/4)(1-u^2)I(|u| < 1)$. We chose the initial estimators

$$\hat{\mathbf{m}}_j^{[0]}(x_j) = n^{-1} \odot \bigoplus_{i=1}^{n} \Big(\frac{K_{h_j}(x_j, X_j^i)}{\hat{p}_j(x_j)} - 1 \Big) \odot \mathbf{Y}^i =: n^{-1} \odot \bigoplus_{i=1}^{n} w_{ij}^{[0]}(x_j) \odot \mathbf{Y}_i,$$

so that they satisfy $\int_0^1 w_{ij}^{[0]}(x_j)\hat{p}_j(x_j)dx_j = 0$. For the convergence criterion of the B-SBF algorithm we set

$$\max_{1 \le j \le d} \int_0^1 \|\hat{\mathbf{m}}_j^{[r]}(x_j) \ominus \hat{\mathbf{m}}_j^{[r-1]}(x_j)\|^2 \hat{p}_j(x_j)dx_j < 10^{-8}.$$

5.1 Bandwidth selection

Searching for the bandwidths h_j on a full-dimensional grid is not feasible when d is large. One way often adopted in multivariate smoothing is to set $h_1 = \cdots = h_d$ and perform one-dimensional grid search. Obviously, this is not desirable since it ignores different degrees of smoothness for different target functions. Recently, Han et al. (2018) and Han and Park (2018+) used a method called 'bandwidth shrinkage'. The method first selects $\hat{h}_j$ for each j that is good for estimating marginal regression function of X_j and then tunes $c > 0$ for $(c\hat{h}_1, \cdots, c\hat{h}_d)$. The latter method also searches bandwidths on a restricted class of options.

Here, we suggest a new scheme called 'CBS(Coordinate-wise Bandwidth Selection)' based on cross-validation. We used the CBS method, as described below, in our numerical study. Let $\mathrm{CV}(h_1, \ldots, h_d)$ denote a cross-validatory criterion for bandwidths $h_1, \ldots, h_d$.

CBS algorithm. Take a grid $\mathcal{G} = \prod_{j=1}^d \{g_{j1}, \ldots, g_{jL_j}\}$. Choose an initial bandwidth $h_j^{(0)}$ from $\{g_{j1}, \ldots, g_{jL_j}\}$ for $1 \le j \le d$. For $t = 1, 2, \cdots$, find

$$h_j^{(t)} = \underset{g_j \in \{g_{j1}, \cdots, g_{jL_j}\}}{\arg\min} \mathrm{CV}(h_1^{(t)}, \ldots, h_{j-1}^{(t)}, g_j, h_{j+1}^{(t-1)}, \ldots, h_d^{(t-1)}), \quad 1 \le j \le d.$$

Repeat the procedure until $(h_1^{(t)}, \cdots, h_d^{(t)}) = (h_1^{(t-1)}, \cdots, h_d^{(t-1)})$.

$\square$

In our numerical study, we chose $\mathcal{G} = \prod_{j=1}^d \{a_j + 0.005 \times k : k = 0, \cdots, 40\}$ for some small values a_j that satisfy (S1) in Section 2.4 and used a 10-fold cross-validation. The grid actually covered optimal bandwidths. Let

$$T = \min \left\{ t \ge 1 : (h_1^{(t)}, \ldots, h_d^{(t)}) = (h_1^{(t-1)}, \ldots, h_d^{(t-1)}) \right\}. \quad (5.1.1)$$

We note that T is finite since the grid size is finite. In our numerical work, the algorithm converged very fast. In all cases $T \le 4$. We also note that $(h_1^{(T)}, \ldots, h_d^{(T)})$ is a coordinate-wise minimum that satisfies

$$\mathrm{CV}(h_1^{(T)}, \cdots, h_d^{(T)}) = \min_j \min_{g_j} \mathrm{CV}\big(h_1^{(T)}, \ldots, h_{j-1}^{(T)}, g_j, h_{j+1}^{(T)}, \ldots, h_d^{(T)}\big).$$

Although a coordinate-wise minimum does not always match with a global minimum, they coincided in most cases in our numerical study.

5.2 Simulation study with density response

We considered the case where $Y(\cdot)$ is a probability density on a domain $S \in \mathcal{B}(\mathbb{R})$ such that $\mathbf{Y} := [Y(\cdot)] \in \mathfrak{B}^2(S, S \cap \mathcal{B}(\mathbb{R}), \mathrm{Leb}_1)$. In this case, simply writing $w_{i,j,r}(x_j) = n^{-1} w_{ij}^{[r]}(x_j)$ for brevity we get

$$\begin{aligned}
\hat{\mathbf{m}}_j^{[r]}(x_j) &= \left[\left(\int_S \prod_{i=1}^n Y_i(s)^{w_{i,j,r}(x_j)} ds \right)^{-1} \prod_{i=1}^n Y_i(\cdot)^{w_{i,j,r}(x_j)} \right], \\
\bar{\mathbf{Y}} \oplus \bigoplus_{j=1}^d \hat{\mathbf{m}}_j^{[r]}(x_j) &= \left[\left(\int_S \prod_{i=1}^n Y_i(s)^{n^{-1} + \sum_{j=1}^d w_{i,j,r}(x_j)} ds \right)^{-1} \times \right. \\
&\qquad \left. \prod_{i=1}^n Y_i(\cdot)^{n^{-1} + \sum_{j=1}^d w_{i,j,r}(x_j)} \right]
\end{aligned} \tag{5.2.1}$$

whenever the denominators are nonzero and finite. We predicted $Y(\cdot)$ at $\mathbf{X} = \mathbf{x}$ for an out-of-sample $(\mathbf{X}, Y(\cdot))$ by

$$\left(\int_S \prod_{i=1}^n Y_i(s)^{n^{-1} + \sum_{j=1}^d w_{i,j,r}(x_j)} ds \right)^{-1} \times \prod_{i=1}^n Y_i(\cdot)^{n^{-1} + \sum_{j=1}^d w_{i,j,r}(x_j)}$$

We note that the denominators are nonzero and finite for all $w_{ij}^{[r]}(x_j) \in \mathbb{R}$ if $Y_i(\cdot)$'s are essentially bounded away from zero and infinity on S and $\mathrm{Leb}_1(S) < \infty$. In this simulation study, our focus is to demonstrate that (i) the CBS algorithm for bandwidth selection works well, and (ii) the prediction based on the proposed estimators $\hat{\mathbf{m}}_j$ and $\hat{\mathbf{m}}$ is valid for small sample sizes, avoiding the curse of dimensionality.

We generated $Y(\cdot)$ on $S = [-1/2, 1/2]$ according to the following formula.

$$Y(\cdot) = \left(\int_S \prod_{j=1}^{2} f_j(X_j)(s)\epsilon(s)ds \right)^{-1} \cdot \prod_{j=1}^{2} f_j(X_j)(\cdot)\epsilon(\cdot), \quad (5.2.2)$$

where $f_j(x_j)(\cdot) : S \to \mathbb{R}$ are some measurable functions, ϵ is an error process, X_1 and X_2 are uniform $[0, 1]$ random variables. Specifically, we considered $f_1(x_1)(s) = -\exp(x_1)|s| + 2$ and $f_2(x_2)(s) = \cos(s\pi/2)^{x_2+x_2^3}$ and $\epsilon(s) = \exp(-Zs^4)$ with Z being a uniform $[-1, 1]$ random variable. By considering the operations $\oplus$ and $\odot$ for the quotient space $\mathbb{H} = \mathfrak{B}^2(S, S \cap \mathcal{B}(\mathbb{R}), \mathrm{Leb}_1)$ and the equivalence class $[Y(\cdot)]$ as introduced in Section 2, we clearly see that (5.2.2) falls into the additive model (1.0.1) with $d = 2$. We also considered a non-additive model for a sensitivity analysis. For this, we took

$$Y(s) = \frac{\exp(X_1^2 \cos(2\pi s) + X_2^2 \sin(2\pi s) + X_1 X_2 |s|)\epsilon(s)}{\int_{-1/2}^{1/2} \exp(X_1^2 \cos(2\pi s) + X_2^2 \sin(2\pi s) + X_1 X_2 |s|)\epsilon(s)ds} \quad (5.2.3)$$

with $\epsilon(s) = \exp(-Zs^2)$.

We repeatedly generated a training sample of size n and a test sample of size $N = 100$ for $M = 200$ times. As a measure of performance we computed the mean squared prediction error (MSPE) defined by

$$\mathrm{MSPE} = M^{-1} \sum_{m=1}^{M} N^{-1} \sum_{i=1}^{N} \left\| [Y_i^{\mathrm{test}(m)}(\cdot)] \ominus [\hat{Y}_i^{\mathrm{test}(m)}(\cdot)] \right\|^2, \quad (5.2.4)$$

where $Y_i^{\mathrm{test}(m)}(\cdot)$ is the ith response in the mth test sample and $\hat{Y}_i^{\mathrm{test}(m)}(\cdot)$ is the prediction of $Y_i^{\mathrm{test}(m)}(\cdot)$ based on the mth training

Table 5.1: The percentages of the iteration number T defined in (5.1.1) at which the CBS algorithm stops, based on $M = 200$ pseudo samples. Ratio indicates (average computing time for the full-dimensional grid search)/(average computing time for the CBS algorithm).

Scenario	n	$T = 2$	$T = 3$	$T = 4$	Ratio
(5.2.2)	100	48%	49.5%	2.5%	6.09
	400	44%	55.5%	0.5%	8.11
(5.2.3)	100	53%	42.5%	4.5%	6.45
	400	62.5%	36%	1.5%	8.72

sample. We note that

$$\left\| [Y_i^{\text{test}(m)}(\cdot)] \ominus [\hat{Y}_i^{\text{test}(m)}(\cdot)] \right\|^2$$
$$= \int_{[-1/2,1/2]^2} \left[\log \left(\frac{Y_i^{\text{test}(m)}(s)}{Y_i^{\text{test}(m)}(s')} \right) - \log \left(\frac{\hat{Y}_i^{\text{test}(m)}(s)}{\hat{Y}_i^{\text{test}(m)}(s')} \right) \right]^2 ds ds'.$$

Table 5.1 suggests that the CBS algorithm for bandwidth selection converges very fast. Its computation was much faster than the full-dimensional grid search. If the grid $\mathcal{G}$ is denser or d is larger, then the ratios of computing time would increase geometrically. Table 5.2 reveals that the selected bandwidths from the CBS algorithm and the full-dimensional grid search matched in most cases. This may be due to the fact that $\mathrm{CV}(h_1, \cdots, h_d)$ is coordinate-wise convex as is often the case in practice. The results demonstrate that the larger n, the more often the two bandwidth choices coincide. Even in the case where the two were different, the CBS bandwidths gave comparable prediction results to the full-dimensional grid search, as the ratios in the last column of the table shows.

Table 5.2: The percentages of the cases where the CBS algorithm gave the same bandwidth choices as the full-dimensional grid search, based on $M = 200$ pseudo samples. The 'MSPE ratio' means the ratio of the MSPE value with bandwidths from the full-dimensional grid search, to that with CBS bandwidths. In the computation of the MSPE values according to the formula (5.2.4), the cases where CBS=Full are deleted.

Scenario	n	CBS=Full	MSPE ratio for CBS≠Full
(5.2.2)	100	79.5%	0.97
	400	98.5%	1
(5.2.3)	100	88.5%	0.98
	400	98.5%	1.02

In the simulation we also compared the prediction based on our approach with those based on full-dimensional estimators. We considered the functional Nadaraya-Watson estimator proposed by Dabo-Niang and Rhomari (2009), Ferraty et al. (2011) and Ferraty et al. (2012) and the kernel-based functional k-nearest neighbor estimator proposed by Lian (2011) and Lian (2012). For these full-dimensional estimators we used Epanechnikov kernel, and tuned bandwidth and k, respectively, by 10-fold cross-validation on ranges that cover optimal bandwidth and k. Table 5.3 demonstrates that the proposed method outperforms these methods in both additive (5.2.2) and non-additive (5.2.3) scenarios.

To see how our approach performs in higher dimension and in the estimation of the component maps, we tried $d = 4$ and considered the case where $f_j(x_j)(\cdot) = \beta_j(\cdot)^{g_j(x_j)}$ for some real-valued functions β_j and g_j with $\mathrm{E}(g_j(X_j)) = 0$. In this way, we

Table 5.3: The ratios of the MSPE values for the functional Nadaraya-Watson and the kernel-based functional k-NN methods, to that for our proposed method.

Scenario	n	Proposed with CBS	Functional Nadaraya-Watson	Kernel-based functional k-NN
(5.2.2)	100	1	1.99	2.07
	400	1	1.34	1.42
(5.2.3)	100	1	1.47	1.54
	400	1	1.12	1.16

have $\mathrm{E}([f_j(X_j)(\cdot)]) = \mathbf{0}$ since

$$\mathrm{E}([f_j(X_j)(\cdot)]) = \mathrm{E}(g_j(X_j)) \odot [\beta_j(\cdot)] = 0 \odot [\beta_j(\cdot)] = \mathbf{0}, \quad (5.2.5)$$

satisfying the constraints (2.4.4). The first equation at (5.2.5) follows from (2.5.1). Specifically, we generated $Y(\cdot)$ according to

$$Y(\cdot) = \frac{\beta_0(\cdot)\beta_1(\cdot)^{\cos(2\pi X_1)}\beta_2(\cdot)^{\sin(2\pi X_2)}\beta_3(\cdot)^{\cos(\pi X_3)}\beta_4(\cdot)^{2X_4-1}\epsilon(\cdot)}{\int_{-1/2}^{1/2}\beta_0(s)\beta_1(s)^{\cos(2\pi X_1)}\beta_2(s)^{\sin(2\pi X_2)}\beta_3(s)^{\cos(\pi X_3)}\beta_4(s)^{2X_4-1}\epsilon(s)ds}, \quad (5.2.6)$$

where $\epsilon(s) = \alpha(s)^Z$, X_1, X_2, X_3, X_4 are uniform $[0,1]$ random variables and Z is a uniform $[-1,1]$ random variable. We chose for $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$, $\beta_3(\cdot)$, $\beta_4(\cdot)$ and $\alpha(\cdot)$, respectively, the probability density functions of Cauchy$(0, 0.2)$, $N(0, 0.5^2)$, t-distribution with df $= 0.25$, Laplace$(0,1)$, $N(-0.3, 0.2^2)/2 + N(0.3, 0.2^2)/2$ and Logistic$(0,1)$, all truncated on $[-1/2, 1/2]$. With these choices, the simulation model (5.2.6) involves component maps whose values $f_j(x_j)(\cdot)$ take various shapes: light- and heavy-tails, sharp peaks, bimodality etc.

We compared the proposed estimator, denoted by $\hat{m}(1,2,3,4)$, based on the four-dimensional predictor (X_1, X_2, X_3, X_4), with oracle estimators. Let $\hat{m}(1,2)$ denote the oracle estimator that one gets by applying our B-SBF techniques based on the two predictors X_1 and X_2 with the knowledge of f_3 and f_4. Likewise, let $\hat{m}(3,4)$ denote the one based on the knowledge of f_1 and f_2. For $M = 200$, we computed

$$
\begin{aligned}
\mathrm{IMSE}_j &= \int_0^1 M^{-1} \sum_{m=1}^{M} \left\| [f_j(x_j)(\cdot)] \ominus [\hat{f}_j^{(m)}(x_j)(\cdot)] \right\|^2 dx_j = \mathrm{ISB}_j + \mathrm{IV}_j, \\
\mathrm{ISB}_j &= \int_0^1 \left\| [f_j(x_j)(\cdot)] \ominus M^{-1} \odot \bigoplus_{m=1}^{M} [\hat{f}_j^{(m)}(x_j)(\cdot)] \right\|^2 dx_j, \\
\mathrm{IV}_j &= \int_0^1 M^{-1} \sum_{m=1}^{M} \left\| [\hat{f}_j^{(m)}(x_j)(\cdot)] \ominus M^{-1} \odot \bigoplus_{m=1}^{M} [\hat{f}_j^{(m)}(x_j)(\cdot)] \right\|^2 dx_j.
\end{aligned}
$$

The results are contained in Table 5.4 and Figure 5.1. The values in the table reveal that the performance of $\hat{m}(1,2,3,4)$ is comparable with those of the oracle estimators $\hat{m}(1,2)$ and $\hat{m}(3,4)$. This suggests that the proposed method does not suffer from the curse of dimensionality.

5.3 Real data analysis with functional response

We analyzed 'CanadianWeather' data in the R package 'fda'(version 2.4.4), which contains daily temperatures measured on 35 locations, averaged over 35 years from 1960 to 1994. We performed the prediction of temperature curves based on the two-dimensional predictor (latitude, longitude). In this example, $\mathbb{H} = L^2([0,1], [0,1]\cap$

Table 5.4: The values of IMSE, ISB and IV, multiplied by 10^3, of the proposed $\hat{m}(1,2,3,4)$ and of the oracle $\hat{m}(1,2)$ in the estimation of the two component maps f_1 and f_2, and of the oracle $\hat{m}(3,4)$ in the estimation of f_3 and f_4, based on $M = 200$ pseudo samples. All bandwidths were selected by the CBS algorithm.

		First component		Second component	
n	Criterion	$\hat{m}(1,2,3,4)$	$\hat{m}(1,2)$	$\hat{m}(1,2,3,4)$	$\hat{m}(1,2)$
100	IMSE	0.1654	0.1671	0.1346	0.1418
	ISB	0.0089	0.0154	0.0163	0.0231
	IV	0.1565	0.1517	0.1183	0.1187
400	IMSE	0.0350	0.0346	0.0332	0.0330
	ISB	0.0007	0.0011	0.0019	0.0025
	IV	0.0343	0.0335	0.0313	0.0305
		Third component		Fourth component	
n	Criterion	$\hat{m}(1,2,3,4)$	$\hat{m}(3,4)$	$\hat{m}(1,2,3,4)$	$\hat{m}(3,4)$
100	IMSE	0.0372	0.0319	0.0972	0.0885
	ISB	0.0037	0.0030	0.0067	0.0042
	IV	0.0335	0.0289	0.0905	0.0843
400	IMSE	0.0090	0.0087	0.0252	0.0247
	ISB	0.0003	0.0003	0.0009	0.0009
	IV	0.0087	0.0084	0.0243	0.0238

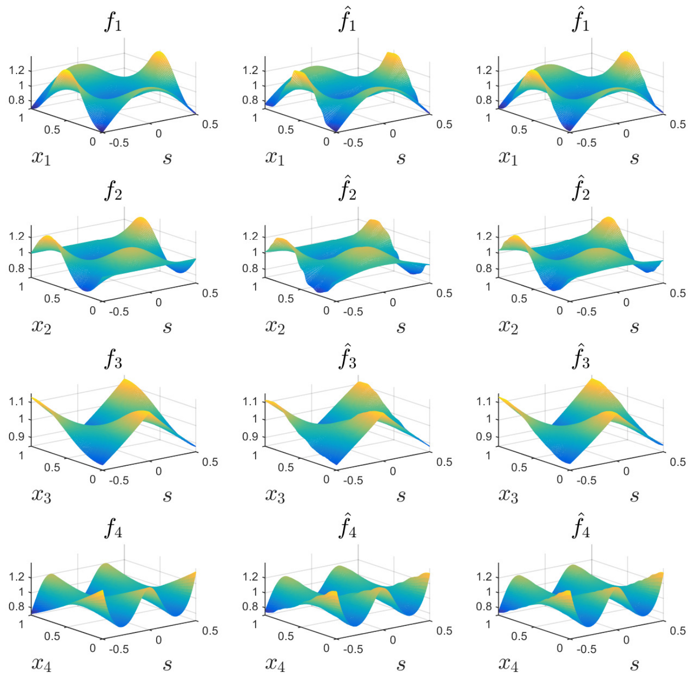

Figure 5.1: True component maps (left) and their estimates for $n = 100$ (middle) and for $n = 400$ (right), based on the median performance sample, i.e., the one for which the value of the total integrated squared error $\sum_{j=1}^{4} \int_0^1 \| [f_j(x_j)(\cdot)] \ominus [\hat{f}_j^{(m)}(x_j)(\cdot)] \|^2 dx_j$ is the median among the 200 values for the whole pseudo samples.

$\mathcal{B}(\mathbb{R}), \mathrm{Leb}_1)$ and $Y_i(\cdot)$ is the pre-smoothed temperature curve for the ith location. We computed the leave-one-curve-out average squared prediction error

$$\mathrm{ASPE} = n^{-1}\sum_{i=1}^{n}\|Y_i(\cdot)\ominus\hat{Y}_i^{(-i)}(\cdot)\|^2 = n^{-1}\sum_{i=1}^{n}\int_0^1 (Y_i(s)-\hat{Y}_i^{(-i)}(s))^2 ds$$

with $n = 35$, where $\hat{Y}_i^{(-i)}(\cdot)$ is the prediction of $Y_i(\cdot)$ based on the sample without the ith observation. We also measured the smoothness of $\hat{Y}_i^{(-i)}(\cdot)$ using *fractal dimension*. Fractal dimension is a measure of smoothness for curves and surfaces. In the case of curves, it takes values in $[1, 2]$ where '1' means that the curve is perfectly smooth and '2' indicates that the curve is extremely wiggly. For the definition of fractal dimension, see Gneiting et al. (2012). We used 'fd.estimate' function in the R package 'fractaldim'(version 0.8-4) and used the madogram estimator suggested by Gneiting et al. (2012). Let $\hat{FD}_i$ denote the estimated fractal dimension of the curve $\hat{Y}_i^{(-i)}(\cdot)$. We computed the average estimated fractal dimension $\mathrm{AEFD} = n^{-1}\sum_{i=1}^{n}\hat{FD}_i$.

For this example, we compared our method with those of Chiou et al. (2003) and Scheipl et al. (2015), and with the functional Nadaraya-Watson and the kernel-based functional k-nearest neighbor estimators. To implement the method of Chiou et al. (2003), we used 'FQR' function in the matlab package 'PACE'(version 2.17) with bandwidth for mean curve being selected by leave-one-curve-out cross-validation and bandwidth for covariance surface being selected by GCV. For the method of Scheipl et al. (2015), we used 'pffr' function in the R package 'refund'(version 0.1-16) with 100 cubic B-spline basis functions and smoothing parameter

Table 5.5: Comparison of ASPE and AEFD for CanadianWeather data.

Method	ASPE	AEFD
B-SBF with CBS	9.10	1
Pointwise SBF with CBS	9.59	1.43
Kernel-based functional k-NN	11.31	1.11
Functional Nadaraya-Watson	14.74	1.13
Chiou et al. (2003)	16.11	1
Scheipl et al. (2015)	19.21	1

selected by GCV. We also computed the pointwise smooth backfitting estimate $\hat{Y}_i^{(-i)}(s)$ for each s using the standard smooth backfitting procedure, as in Mammen et al. (1999), and aggregated them to produce the curve $\hat{Y}_i^{(-i)}(\cdot)$. Table 5.5 and Figure 5.2 contain the results, which suggest that our method outperforms all competitors in terms of prediction performance and smoothness of estimated curves.

5.4 Real data analysis with simplex-valued response

Here, we analyzed 'gemas' data in the R-package 'robCompositions'(version 2.0.5), which contains a simplex-valued response. It is a geochemical dataset about agricultural and grazing land soil in European regions. The dataset has 2,108 observations on 30 variables. Among the variables, we chose the composition of three soil types as the response: (sand, silt, clay) with the sum of the three entries being equal to 1, and (annual mean temperature, an-

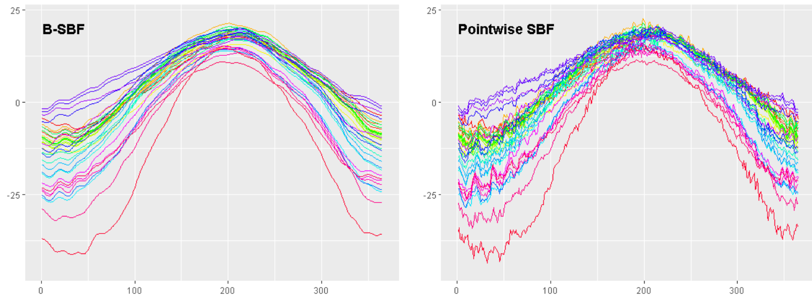

Figure 5.2: Predicted temperature curves for CanadianWeather data based on our B-SBF method(left) and the pointwise SBF method(right). Each of the 35 curves depicts $\hat{Y}_i^{(-i)}(\cdot)$ for the ith location.

nual mean precipitation) as the two-dimensional predictor. In this example, $\mathbb{H} = \mathcal{S}_1^3$. We deleted 26 observations which contain zero proportion in some soil type. We divided the remaining 2,082 observations into 10 partitions $S_k, 1 \le k \le 10$, with each of the first 9 having 208 observations and the last one containing the remainder. We then computed the 10-fold average squared prediction error (ASPE) defined by $10^{-1} \sum_{k=1}^{10} |S_k|^{-1} \sum_{i \in S_k} \|\mathbf{Y}_i \ominus \hat{\mathbf{Y}}_i^{(-S_k)}\|^2$, where $|S_k|$ is the number of observations in S_k and $\hat{\mathbf{Y}}_i^{(-S_k)}$ is the prediction of $\mathbf{Y}_i$ based on the sample without the observations in S_k.

We compared our method with the alpha-transformation method of Tsagris (2015). For the latter, we used 'alfa.reg' function in the R-package 'Compositional' (version 2.5) where 'alpha' was tuned on $\{-1+0.1\times k : 0 \le k \le 20\}$ by 10-fold cross-validation. The proposed method with the CBS algorithm gave ASPE $= 0.98$, while the method of Tsagris (2015) resulted in ASPE $= 1.69$. Figure 5.3

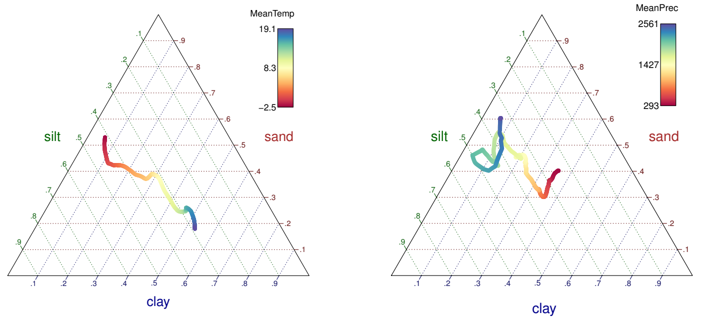

Figure 5.3: The values of the fitted component maps for gemas data based on the B-SBF method, depicted on the simplex $\mathcal{S}_1^3$, for the annual mean temperature(left) and for the annual mean precipitation(right).

depicts the fitted component maps.

Chapter 6

Appendix (Additional Results and Selected Proofs)

6.1 Lemmas and additional propositions

We collect below several lemmas and additional propositions that are used to prove the propositions and theorems in Sections 2–4. We note that $\mathbb{H}$ and $\mathbb{B}$ in Lemmas 6.1.4, 6.1.6 and 6.1.7 do not need to be separable.

Lemma 6.1.1. *Let $(\mathcal{Z}, \mathscr{A}, \mu)$ be a measure space and $(\mathcal{W}, \mathscr{B})$ be a measurable space. Let $\mathbf{T} : \mathcal{Z} \to \mathcal{W}$ be $(\mathscr{A}, \mathscr{B})$-measurable and $\mathbf{g} : \mathcal{W} \to \mathbb{B}$ be measurable. Then, $\mathbf{g} \in \mathcal{L}^1((\mathcal{W}, \mathscr{B}, \mu\mathbf{T}^{-1}), \mathbb{B})$ if and only if $\mathbf{g}(\mathbf{T}) \in \mathcal{L}^1((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$, in which case $\int_{\mathcal{W}} \mathbf{g}(\mathbf{w}) d\mu\mathbf{T}^{-1}(\mathbf{w}) = \int_{\mathcal{Z}} \mathbf{g}(\mathbf{T}(\mathbf{z})) d\mu(\mathbf{z})$.*

Lemma 6.1.2. *Let $(\mathcal{Z}, \mathscr{A})$ be a measurable space and λ and μ be σ-finite measures on $(\mathcal{Z}, \mathscr{A})$ such that $\lambda \ll \mu$. Let $\mathbf{f} : \mathcal{Z} \to \mathbb{B}$ be λ-integrable. Then, $\mathbf{f} \odot (d\lambda/d\mu)$ is μ-integrable and $\int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) d\lambda(\mathbf{z}) = \int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) \odot (d\lambda/d\mu)(\mathbf{z}) d\mu(\mathbf{z})$.*

The next lemma is a general type of Fubini's theorem for $\mathbb{B}$-valued maps. There are versions of Fubini's theorem for σ-finite measure spaces. Lemma 6.1.3 does not require σ-finiteness. In the case $\mathbb{B} = \mathbb{R}$, there are some results that do not require σ-finiteness, see Mukherjea (1972), for example.

Lemma 6.1.3. *Let $(\mathcal{Z}, \mathscr{A}, \mu)$ and $(\mathcal{W}, \mathscr{B}, \nu)$ be measure spaces and $\mathbf{k} : \mathcal{Z} \times \mathcal{W} \to \mathbb{B}$ be measurable. Then, (a) for each $\mathbf{w} \in \mathcal{W}$, the map $\mathbf{k}(\cdot, \mathbf{w}) : \mathcal{Z} \to \mathbb{B}$ is measurable; (b) if $\mathbf{k}(\cdot, \mathbf{w}) \in \mathcal{L}^1((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$ a.e. with respect to ν, then $\mathbf{g} : \mathcal{W} \to \mathbb{B}$ defined by*

$$\mathbf{g}(\mathbf{w}) = \begin{cases} \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}), & \text{if } \mathbf{w} \in D_{\mathcal{W}} \\ \mathbf{g}_0(\mathbf{w}), & \text{otherwise} \end{cases}$$

is measurable, where $D_{\mathcal{W}} = \{\mathbf{w} \in \mathcal{W} : \mathbf{k}(\cdot, \mathbf{w}) \in \mathcal{L}^1((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})\}$ and $\mathbf{g}_0 : \mathcal{W} \to \mathbb{B}$ is any measurable map; (c) if $\mathbf{k} \in \mathcal{L}^1((\mathcal{Z} \times \mathcal{W}, \mathscr{A} \otimes \mathscr{B}, \mu \otimes \nu), \mathbb{B})$, then

$$\int_{\mathcal{Z} \times \mathcal{W}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu \otimes \nu(\mathbf{z}, \mathbf{w}) = \int_{\mathcal{W}} \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}).$$

Lemma 6.1.4. *Let $(\mathcal{Z}, \mathscr{A}, \mu)$ be a measure space and $\mathscr{A}_0$ be a field that generates $\mathscr{A}$. Let $\mathbb{B}$ be a Banach space and $p \in [1, \infty)$ be a constant. Then, $\big\{ \bigoplus_{i=1}^n 1_{A_i} \odot \mathbf{b}_i : n \in \mathbb{N}, A_i \in \mathscr{A}_0, \mu(A_i) < \infty, \mathbf{b}_i \in \mathbb{B} \big\}$ is a dense subset of $\mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$.*

Lemma 6.1.5. *Define a σ-field $\mathscr{B}_j = \{[0,1]^{j-1} \times B_j \times [0,1]^{d-j} : B_j \in [0,1] \cap \mathcal{B}(\mathbb{R})\}$ on $[0,1]^d$. We let $\mathscr{B}_j^*$ denote the smallest σ-field* *such that* $\mathscr{B}_j \subset \mathscr{B}_j^*$ *and* $\{B \in [0,1]^d \cap \mathcal{B}(\mathbb{R}^d) : P\mathbf{X}^{-1}(B) = 1\} \subset \mathscr{B}_j^*$. *Then,* $L_2^{\mathbb{H}}(p_j) = L^2(([0,1]^d, \mathscr{B}_j^*, P\mathbf{X}^{-1}), \mathbb{H})$ *and* $L_2^{\mathbb{H}}(\hat{p}_j) = L^2(([0,1]^d, \mathscr{B}_j^*, \hat{P}\mathbf{X}^{-1}), \mathbb{H})$ *for all* $1 \le j \le d$.

Lemma 6.1.6. *Let* $(\mathcal{Z}, \mathscr{A}, \mu)$ *and* $(\mathcal{W}, \mathscr{B}, \nu)$ *be measure spaces and* $\mathbb{B}$ *be a Banach space. Let* $k : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$ *be a measurable function and* $\mathbf{b} \in \mathbb{B}$ *be a constant. Then,* $\mathbf{k} : \mathcal{Z} \times \mathcal{W} \to \mathbb{B}$ *defined by* $\mathbf{k}(\mathbf{z}, \mathbf{w}) = k(\mathbf{z}, \mathbf{w}) \odot \mathbf{b}$ *is measurable.*

The following lemma follows from Theorem 4.6 in Xu and Zikatanov (2002) and Theorem 2.1 in Blot and Cieutat (2016).

Lemma 6.1.7. *Let* $\mathbb{H}$ *be a Hilbert space and* $\mathbb{H}_1, \ldots, \mathbb{H}_d \le \mathbb{H}$. *Define* $\mathbb{H}_\oplus = \left\{ \bigoplus_{j=1}^d \mathbf{h}_j : \mathbf{h}_j \in \mathbb{H}_j, 1 \le j \le d \right\}$, *and let* $P_j : \mathbb{H} \to \mathbb{H}_j, 1 \le j \le d$, *be orthogonal projections. Then, the followings are equivalent: (a)* $\mathbb{H}_\oplus \le \mathbb{H}$; *(b)* $\|(I - P_d) \circ \cdots \circ (I - P_1)\|_{\mathcal{L}(\overline{\mathbb{H}_\oplus})} < 1$; *(c)* $\exists\, c > 0$ *such that for all* $\mathbf{h} \in \mathbb{H}_\oplus$, *there exists a decomposition* $\mathbf{h} = \bigoplus_{j=1}^d \mathbf{h}_j$ *with* $\mathbf{h}_j \in \mathbb{H}_j$, $1 \le j \le d$, *and* $\sum_{j=1}^d \|\mathbf{h}_j\|^2 \le c\|\mathbf{h}\|^2$.

Lemma 6.1.8. *Assume that there exists a constant* $c > 0$ *such that* $p(\mathbf{x}) \ge c p_j(x_j) p_{\mathbf{X}_{-j}}(\mathbf{x}_{-j})$ *for all* $1 \le j \le d$ *and* $\mathbf{x} \in [0,1]^d$. *Let* $\mathbf{f}_j : [0,1] \to \mathbb{H}, 1 \le j \le d$ *be* $([0,1] \cap \mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{H}))$*-measurable maps. If* $\bigoplus_{j=1}^d \mathbf{f}_j(x_j) = \mathbf{0}$ *for a.e.* $\mathbf{x} \in [0,1]^d$ *with respect to* $P\mathbf{X}^{-1}$, *then* $\mathbf{f}_j(x_j) = \mathbf{c}_j$ *for a.e.* $x_j \in [0,1]$ *with respect to* PX_j^{-1} *for* $1 \le j \le d$, *where* $\mathbf{c}_j \in \mathbb{H}$ *are some constants satisfying* $\bigoplus_{j=1}^d \mathbf{c}_j = \mathbf{0}$.

Proposition 6.1.1. *For* $D \in \mathbb{N}$, *let* $\mathbf{U}_i$ *and* $\mathbf{V}_i$ *be iid copies of a* $[0,1]^D$*-valued random vector* $\mathbf{U}$ *and a* $\mathbb{H}$*-valued random element* $\mathbf{V}$, *respectively. Assume (i)* $\mathrm{E}(\|\mathbf{V}\|^\alpha) < \infty$ *for some* $\alpha > 2$; *(ii)* K *is Lipschitz continuous; (iii)* $\inf_n n^{c_1} \prod_{j=1}^d h_j \ge (const.)$ *for some* $c_1 < (\alpha - 2)/\alpha$ *and* $\inf_n n^{c_2} \min_{1 \le j \le d} h_j \ge (const.)$ *for some*

$c_2 \in \mathbb{R}$. *Then, for* $\mathbf{S}_n(\mathbf{u}) := n^{-1} \odot \bigoplus_{i=1}^n \left(\prod_{j=1}^D K_{h_j}(u_j - U_{ij}) \right) \odot \mathbf{V}_i$, *it holds that*

$$\sup_{\mathbf{u} \in [0,1]^D} \|\mathbf{S}_n(\mathbf{u}) \ominus \mathrm{E}\left(\mathbf{S}_n(\mathbf{u})\right)\| = O_p\left((nh_1 \cdots h_D)^{-1/2}\sqrt{\log n}\right).$$

The following proposition is a Lindeberg-type theorem. It complements Theorem 1.1 in Kundu et al. (2000) that is for infinite-dimensional $\mathbb{H}$.

Proposition 6.1.2. *Let* $\mathbb{H}$ *be a finite-dimensional Hilbert space and* $\{\mathbf{b}_k\}_{k=1}^N$ *be an orthonormal basis of* $\mathbb{H}$. *Let* $\mathbf{V}_{n1}, \cdots, \mathbf{V}_{nn}$ *be independent* $\mathbb{H}$*-valued random elements such that* $\mathrm{E}(\mathbf{V}_{ni}) = \mathbf{0}$ *and* $\mathrm{E}(\|\mathbf{V}_{ni}\|^2) < \infty$ *for* $1 \le i \le n$. *For* $\mathbf{S}_n = \bigoplus_{i=1}^n \mathbf{V}_{ni}$, *assume that*

(i) $a_{kl} := \lim_{n\to\infty} \mathrm{E}(\langle \mathbf{S}_n, \mathbf{b}_k \rangle \langle \mathbf{S}_n, \mathbf{b}_l \rangle)$ *exist for all* $1 \le k, l \le N$;

(ii) $\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}(\langle \mathbf{V}_{ni}, \mathbf{b}_k \rangle^2 I(|\langle \mathbf{V}_{ni}, \mathbf{b}_k \rangle| > \eta)) = 0$ *for all* $1 \le k \le N$ *and* $\eta > 0$.

Then, $\mathbf{S}_n \xrightarrow{d} \mathbf{G}(\mathbf{0}, C)$ *for the covariance operator* $C : \mathbb{H} \to \mathbb{H}$ *characterized by* $\langle C(\mathbf{h}), \mathbf{b}_k \rangle = \sum_{l=1}^N \langle \mathbf{h}, \mathbf{b}_l \rangle a_{kl}$.

6.2 Proof of Theorem 3.2.1

The linearity of L follows from the linearity of $\mathbf{k}(\mathbf{z}, \mathbf{w})$ and L_0. Using Lemma 6.1.3 and the fact

$$\|L(\mathbf{f})(\mathbf{w})\|_{\mathbb{B}_2} \le \left(\int_{\mathcal{Z}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\|^q_{\mathcal{L}(\mathbb{B}_1, \mathbb{B}_2)} d\mu(\mathbf{z}) \right)^{1/q} \left(\int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\|^p_{\mathbb{B}_1} d\mu(\mathbf{z}) \right)^{1/p}, \tag{6.2.1}$$

one may prove that L is bounded.

For the compactness, it suffices to prove that there exists a sequence of compact operators, say L_n, that converges to L. Let

$$(\mathscr{A}\otimes\mathscr{B})_0 = \Big\{ \biguplus_{j=1}^{J} (A_j\times B_j) : A_j \in \mathscr{A}, B_j \in \mathscr{B}, \mu(A_j) < \infty, \nu(B_j) < \infty, J \in \mathbb{N}\Big\}.$$

Due to Proposition 9.1 in Kubrusly (2015), $(\mathscr{A}\otimes\mathscr{B})_0$ is a field. We apply Lemma 6.1.4 with the specifications of $\mathbb{B}$ and $\mathcal{L}^p((\mathcal{Z},\mathscr{A},\mu),\mathbb{B})$ there, respectively, to $\mathcal{C}(\mathbb{B}_1,\mathbb{B}_2)$ and $\mathcal{L}^q((\mathcal{Z}\times\mathcal{W},\mathscr{A}\otimes\mathscr{B},\mu\otimes\nu),\mathcal{C}(\mathbb{B}_1,\mathbb{B}_2))$ here. We get that there exist sequences $I_n \in \mathbb{N}$, $C_{ni} \in \mathcal{C}(\mathbb{B}_1,\mathbb{B}_2)$ and $F_{ni} \in (\mathscr{A}\otimes\mathscr{B})_0$ for $1 \le i \le I_n$ such that $\mathbf{k}_n \in \mathcal{L}^q((\mathcal{Z}\times\mathcal{W},\mathscr{A}\otimes\mathscr{B},\mu\otimes\nu),\mathcal{C}(\mathbb{B}_1,\mathbb{B}_2))$, defined by $\mathbf{k}_n(\mathbf{z},\mathbf{w}) = \sum_{i=1}^{I_n} 1_{F_{ni}}(\mathbf{z},\mathbf{w})C_{ni}$, satisfies

$$\left(\int_{\mathcal{Z}\times\mathcal{W}} \|\mathbf{k}_n(\mathbf{z},\mathbf{w}) - \mathbf{k}(\mathbf{z},\mathbf{w})\|^q_{\mathcal{L}(\mathbb{B}_1,\mathbb{B}_2)} d\mu\otimes\nu(\mathbf{z},\mathbf{w})\right)^{1/q} \le n^{-1}. \tag{6.2.2}$$

We take $L_n : L^p((\mathcal{Z},\mathscr{A},\mu),\mathbb{B}_1) \to L^q((\mathcal{W},\mathscr{B},\nu),\mathbb{B}_2)$ defined by $L_n([\mathbf{f}]) = [L_{n\mathbf{f}}]$, where $L_{n\mathbf{f}}(\mathbf{w}) = \int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z},\mathbf{w})(\mathbf{f}(\mathbf{z}))d\mu(\mathbf{z})$. As in the proof of the first part, we may prove that L_n is a bounded linear operator for each $n \ge 1$. One may also prove that $\|L_n - L\|_{\mathrm{op}} \le n^{-1} \to 0$ as $n \to \infty$, where $\|\cdot\|_{\mathrm{op}}$ is the operator norm.

It remains to prove that L_n is compact for each $n \ge 1$. Fix n and take any sequence $\{[\mathbf{f}_k]\}_{k\ge1}$ in the unit ball of $L^p((\mathcal{Z},\mathscr{A},\mu),\mathbb{B}_1)$. Put $F_{ni} = \biguplus_{j=1}^{J_{ni}}(A_{nij}\times B_{nij})$ with $J_{ni} \in \mathbb{N}$, $A_{nij} \in \mathscr{A}$ and $B_{nij} \in \mathscr{B}$, and define $D_{nij} : \mathcal{L}^p((\mathcal{Z},\mathscr{A},\mu),\mathbb{B}_1) \to \mathbb{B}_1$ by $D_{nij}(\mathbf{f}) = \int_{\mathcal{Z}} 1_{A_{nij}}(\mathbf{z}) \odot \mathbf{f}(\mathbf{z})d\mu(\mathbf{z})$. Then,

$$L_{n\mathbf{f}_k}(\mathbf{w}) = \bigoplus_{i=1}^{I_n}\bigoplus_{j=1}^{J_{ni}} 1_{B_{nij}}(\mathbf{w}) \odot C_{ni}(D_{nij}(\mathbf{f}_k)).$$

Since $\sup_{k\geq 1} \|D_{nij}(\mathbf{f}_k)\|_{\mathbb{B}_1} \leq \mu(A_{nij})^{1/q}$ and C_{ni} are compact, $\{L_n([\mathbf{f}_k])\}_{k\geq 1}$ has a convergent subsequence. This completes the proof of the second part.

6.3 Proof of Theorem 3.2.2

Using Theorem 3.2.1 and Lemma 6.1.6, one may show that L is bounded and linear. We show that L is non-compact. Since C is non-compact, there exists a sequence $\{\mathbf{b}_n\}$ in the unit ball of $\mathbb{B}_1$ such that $\{C(\mathbf{b}_n)\}_{n\geq 1}$ has no Cauchy subsequence. Define $\mathbf{f}_n : \mathcal{Z} \to \mathbb{B}_1$ by $\mathbf{f}_n(\mathbf{z}) \equiv \mathbf{b}_n$. Then, $\mathbf{f}_n$ are measurable and $\sup\limits_n \int_{\mathcal{Z}} \|\mathbf{f}_n(\mathbf{z})\|^p d\mu(\mathbf{z}) \leq \mu(\mathcal{Z})$. It suffices to prove that $\{L([\mathbf{f}_n])\}_{n\geq 1}$ does not have a Cauchy subsequence. By the assumption on $k : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$, we get

$$\begin{aligned}\|L([\mathbf{f}_n]) \ominus L([\mathbf{f}_m])\|_{L^q} &= \left(\int_{\mathcal{W}} \left| \int_{\mathcal{Z}} k(\mathbf{w}, \mathbf{z}) d\mu(\mathbf{z}) \right|^q d\nu(\mathbf{w}) \right)^{1/q} \cdot \|C(\mathbf{b}_n) \ominus C(\mathbf{b}_m)\|_{\mathbb{B}_2} \\ &\geq c \cdot \|C(\mathbf{b}_n) \ominus C(\mathbf{b}_m)\|_{\mathbb{B}_2}\end{aligned}$$

for some constant $c > 0$. This proves the theorem.

6.4 Proof of Theorem 3.4.2

One can prove the theorem by arguing as in the proof of Theorem 3.4.1 and using Lemma 6.1.8 and Proposition 6.1.1.

6.5 Proof of Theorem 4.1.1

We only give an outline of the proof. For brevity we write $q_{ij}(x_j) = \hat{p}_j(x_j)^{-1} n^{-1} K_{h_j}(x_j, X_{ij})$. Define

$$\begin{aligned}
\tilde{\mathbf{m}}_j^A(x_j) &= \bigoplus_{i=1}^n q_{ij}(x_j) \odot \boldsymbol{\epsilon}_i, \\
\tilde{\mathbf{m}}_j^B(x_j) &= \bigoplus_{i=1}^n q_{ij}(x_j) \odot (\mathbf{m}_j(X_{ij}) \ominus \mathbf{m}_j(x_j)), \\
\tilde{\mathbf{m}}_{jk}^C(x_j) &= \bigoplus_{i=1}^n q_{ij}(x_j) \odot \boldsymbol{\eta}_{ik},
\end{aligned}$$

where $\boldsymbol{\eta}_{ik} = \int_0^1 (\mathbf{m}_k(X_{ik}) \ominus \mathbf{m}_k(x_k)) \odot K_{h_k}(x_k, X_{ik}) dx_k$. Then, the B-SBF equation (2.4.6) can be written as

$$\begin{aligned}
\hat{\mathbf{m}}_j(x_j) = \mathbf{m}_j(x_j) \oplus \left[E(\mathbf{Y}) \ominus \bar{\mathbf{Y}}\right] \oplus \tilde{\mathbf{m}}_j^A(x_j) \oplus \tilde{\mathbf{m}}_j^B(x_j) \oplus \tilde{\mathbf{m}}_{jk}^C(x_j) \\
\oplus \bigoplus_{k \neq j} \bigoplus_{i=1}^n q_{ij}(x_j) \odot \int_0^1 (\mathbf{m}_k(x_k) \ominus \hat{\mathbf{m}}_k(x_k)) \odot K_{h_k}(x_k, X_{ik}) dx_k, \\
1 \leq j \leq d.
\end{aligned} \tag{6.5.1}$$

Below, we present a lemma for the approximation of $\tilde{\mathbf{m}}_j^B(x_j)$ and $\tilde{\mathbf{m}}_{jk}^C(x_j)$. Recall the definitions of $\boldsymbol{\delta}_j$, $\boldsymbol{\delta}_{jk}$ and $\mathbf{c}_j$ given immediately before Theorem 4.2.2. Define $\mathbf{a}_k(x_k) = \int_0^1 (\frac{v_k - x_k}{h_k}) K_{h_k}(x_k, v_k) dv_k \odot D\mathbf{m}_k(x_k)(1)$. We introduce generic stochastic maps $\mathbf{r}_j : [0,1] \to \mathbb{H}$ such that

$$\sup_{x_j \in I_j} \|\mathbf{r}_j(x_j)\| = o_p(n^{-2/5}), \qquad \sup_{x_j \in [0,1]} \|\mathbf{r}_j(x_j)\| = O_p(n^{-2/5}). \tag{6.5.2}$$

The notation is used to represent various terms in our asymptotic analysis here and in the proof of Theorem 4.2.2.

Lemma 6.5.1. *Under the condition (C) it holds that*

$$\tilde{\mathbf{m}}_j^B(x_j) = h_j^2 \odot \left(\frac{h_j^{-1}}{\int_0^1 K_{h_j}(x_j, v_j) dv_j} \odot \mathbf{a}_j(x_j) \oplus \boldsymbol{\delta}_j(x_j) \oplus \mathbf{c}_j(x_j) \right) \oplus \mathbf{r}_j(x_j),$$

$$\tilde{\mathbf{m}}_{jk}^C(x_j) = h_k^2 \odot \int_0^1 \left(\frac{h_k^{-1}}{\int_0^1 K_{h_k}(x_k, v_k) dv_k} \odot \mathbf{a}_k(x_k) \oplus \boldsymbol{\delta}_{jk}(x_j, x_k) \oplus \mathbf{c}_k(x_k) \right) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \oplus o_p(n^{-2/5}) \quad \textit{uniformly for } x_j \in [0,1].$$

Now, define $\tilde{\boldsymbol{\Delta}}_j^*(x_j) = h_j^2 \odot \boldsymbol{\delta}_j(x_j) \oplus \bigoplus_{k \neq j}^d \int_0^1 \boldsymbol{\delta}_{jk}(x_j, x_k) \odot \left[h_k^2 \frac{p_{jk}(x_j, x_k)}{p_j(x_j)}\right] dx_k$ and

$$\hat{\boldsymbol{\Delta}}_j(x_j) = \hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j) \ominus \tilde{\mathbf{m}}_j^A(x_j) \ominus \left[\frac{h_j}{\int_0^1 K_{h_j}(x_j, v_j) dv_j} \odot \mathbf{a}_j(x_j) \right] \ominus \left[h_j^2 \odot \mathbf{c}_j(x_j)\right] \oplus \mathbf{r}_j(x_j). \tag{6.5.3}$$

Then, from (6.5.1) and Lemma 6.5.1, we may get uniformly for $x_j \in [0,1]$,

$$\hat{\boldsymbol{\Delta}}_j(x_j) = \tilde{\boldsymbol{\Delta}}_j^*(x_j) \ominus \bigoplus_{k \neq j} \int_0^1 \hat{\boldsymbol{\Delta}}_k(x_k) \odot \frac{\hat{p}_{jk}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k \oplus o_p(n^{-2/5}), \quad 1 \leq j \leq d. \tag{6.5.4}$$

Now, standard theory of kernel smoothing completes the proof of the theorem.

6.6 Proof of Theorem 4.2.1

Let $\mathbb{H}^d$ denote the space of tuples $(\mathbf{h}_j : 1 \leq j \leq d)$ with $\mathbf{h}_j \in \mathbb{H}$. Let $\| \cdot \|_{\mathbb{H}^d}$ and $\langle \cdot, \cdot \rangle_{\mathbb{H}^d}$ denote the norm and inner product on $\mathbb{H}^d$, respectively, defined in the standard way. Let $\mathbf{e}_{jl} \in \mathbb{H}^d$

denote $(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{e}_l, \mathbf{0}, \ldots, \mathbf{0})$ where $\mathbf{e}_l$ is placed at the jth entry. Then, $(\mathbf{e}_{jl} : 1 \leq j \leq d, l \geq 1)$ forms an orthonormal basis of $\mathbb{H}^d$. By applying Theorem 1.1 in Kundu et al. (2000) for infinite-dimensional $\mathbb{H}$ and Proposition 6.1.2 for finite-dimensional $\mathbb{H}$, we may prove

$$\left(\bigoplus_{i=1}^{n} w_{i1}(x_1) \odot \boldsymbol{\epsilon}_i, \, \ldots, \, \bigoplus_{i=1}^{n} w_{id}(x_d) \odot \boldsymbol{\epsilon}_i \right) \xrightarrow{d} \mathbf{G}(\mathbf{0}, C_{\mathbf{x}}), \quad (6.6.1)$$

where $C_{\mathbf{x}} : \mathbb{H}^d \to \mathbb{H}^d$ is a covariance operator such that, for all $\mathbf{h} = (\mathbf{h}_1, \ldots, \mathbf{h}_d) \in \mathbb{H}^d$,

$$\begin{aligned} \langle C_{\mathbf{x}}(\mathbf{h}), \mathbf{e}_{jl} \rangle_{\mathbb{H}^d} &= \sum_{k=1}^{d} \sum_{m} \langle \mathbf{h}, \mathbf{e}_{km} \rangle_{\mathbb{H}^d} \cdot a_{kmjl} = \sum_{k=1}^{d} \sum_{m} \langle \mathbf{h}_k, \mathbf{e}_m \rangle \cdot a_{kmjl} \\ &= \sum_{m} \langle \mathbf{h}_j, \mathbf{e}_m \rangle \cdot a_{j,lm}, \quad l \geq 1, \, 1 \leq j \leq d. \end{aligned} \quad (6.6.2)$$

This completes the first part of the theorem.

For the second part of the theorem, let P_j denote the projection operator that maps $(\mathbf{h}_1, \ldots, \mathbf{h}_d) \in \mathbb{H}^d$ to $\mathbf{h}_j$. Then, its adjoint $P_j^* : \mathbb{H} \to \mathbb{H}^d$ is given by $P_j^*(\mathbf{g}) = (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{g}, \mathbf{0}, \cdots, \mathbf{0})$ where $\mathbf{g}$ is placed at the jth entry. We note that the conclusions of Proposition 4.9–4.10 in van Neerven (2008), for P-measurable Gaussian random elements, also hold for strongly measurable Gaussian random elements. The version of Proposition 4.9 implies $P_j(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})) = \mathbf{G}(\mathbf{0}, P_j \circ C_{\mathbf{x}} \circ P_j^*)$. Now, for $\mathbf{g} \in \mathbb{H}$,

$$\begin{aligned} \langle P_j \circ C_{\mathbf{x}} \circ P_j^*(\mathbf{g}), \mathbf{e}_l \rangle &= \langle C_{\mathbf{x}}(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{g}, \mathbf{0}, \ldots, \mathbf{0}), P_j^*(\mathbf{e}_l) \rangle_{\mathbb{H}^d} \\ &= \langle C_{\mathbf{x}}(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{g}, \mathbf{0}, \ldots, \mathbf{0}), \mathbf{e}_{jl} \rangle_{\mathbb{H}^d} \\ &= \sum_{m} \langle \mathbf{g}, \mathbf{e}_m \rangle \cdot a_{j,lm}, \end{aligned}$$

where the last equality follows from (6.6.2). This proves $P_j \circ C_{\mathbf{x}} \circ P_j^* = C_{j,x_j}$, which coupled with (6.6.1) implies

$$P_j\left(\bigoplus_{i=1}^n w_{i1}(x_1) \odot \boldsymbol{\epsilon}_i, \ \ldots, \ \bigoplus_{i=1}^n w_{id}(x_d) \odot \boldsymbol{\epsilon}_i\right) \xrightarrow{d} P_j(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})) = \mathbf{G}(\mathbf{0}, C_{j,x_j}).$$

It remains to prove that $P_j(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}}))$ for different j are independent. By the version of Proposition 4.10 in van Neerven (2008) for strongly measurable Gaussian random elements, it suffices to show that

$$\mathrm{E}\left(\langle P_j(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})), \mathbf{h}_j\rangle \cdot \langle P_k(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})), \mathbf{g}_k\rangle\right) = 0 \tag{6.6.3}$$

for all $\mathbf{h}_j, \mathbf{g}_k \in \mathbb{H}$ and $1 \le j \ne k \le d$. Fix $1 \le j \ne k \le d$ and take $\mathbf{h} = (\mathbf{0}, \cdots, \mathbf{0}, \mathbf{h}_j, \mathbf{0}, \cdots, \mathbf{0}) \in \mathbb{H}^d$ and $\mathbf{g} = (\mathbf{0}, \cdots, \mathbf{0}, \mathbf{g}_k, \mathbf{0}, \cdots, \mathbf{0}) \in \mathbb{H}^d$ where $\mathbf{h}_j$ and $\mathbf{g}_k$ appear in the jth and kth positions of $\mathbf{h}$ and $\mathbf{g}$, respectively. Then,

$$\langle C_{\mathbf{x}}(\mathbf{h}), \mathbf{g}\rangle_{\mathbb{H}^d} = \mathrm{E}\left(\langle P_j(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})), \mathbf{h}_j\rangle \cdot \langle P_k(\mathbf{G}(\mathbf{0}, C_{\mathbf{x}})), \mathbf{g}_k\rangle\right). \tag{6.6.4}$$

On the other hand, using the fact $\mathbf{g} = \sum_l \langle \mathbf{g}_k, \mathbf{e}_l\rangle \mathbf{e}_{kl}$ and (6.6.2), we have

$$\langle C_{\mathbf{x}}(\mathbf{h}), \mathbf{g}\rangle_{\mathbb{H}^d} = \sum_l \langle \mathbf{g}_k, \mathbf{e}_l\rangle \cdot \langle C_{\mathbf{x}}(\mathbf{h}), \mathbf{e}_{kl}\rangle_{\mathbb{H}^d} = \sum_l \sum_m \langle \mathbf{g}_k, \mathbf{e}_l\rangle \cdot \langle \mathbf{0}, \mathbf{e}_m\rangle \cdot a_{k,lm} = 0.$$

This with (6.6.4) gives (6.6.3).

6.7 Proof of Theorem 4.2.2

We only give a sketch of the proof. Recall the definitions of $\tilde{\boldsymbol{\Delta}}_j^*(x_j)$ and $\hat{\boldsymbol{\Delta}}_j(x_j)$ given in Section 6.5. First, we claim that there exists a solution $(\mathbf{\Delta}_1^*, \cdots, \mathbf{\Delta}_d^*) \in \prod_{j=1}^d \mathcal{L}_2^{\mathbb{H}}(p_j)$ of the system of equations

$$\mathbf{\Delta}_j^*(x_j) = \tilde{\mathbf{\Delta}}_j^*(x_j) \ominus \bigoplus_{k\neq j} \int_0^1 \mathbf{\Delta}_k^*(x_k) \odot \frac{p_{jk}(x_j, x_k)}{p_j(x_j)} dx_k, \quad 1 \leq j \leq d, \tag{6.7.1}$$

satisfying the constraints

$$\int_0^1 \mathbf{\Delta}_j^*(x_j) \odot p_j(x_j) dx_j = h_j^2 \odot \int_0^1 \boldsymbol{\delta}_j(x_j) \odot p_j(x_j) dx_j, \quad 1 \leq j \leq d. \tag{6.7.2}$$

To prove the claim, consider a functional $F : S^{\mathbb{H}}(p) \to \mathbb{R}$ defined by

$$F([\mathbf{f}]) = \int_{[0,1]^d} \left\| \bigoplus_{j=1}^d \left(\left[h_j^2 \int_{-1}^1 u^2 K(u) du \frac{\partial p(\mathbf{x})/\partial x_j}{p(\mathbf{x})} \right] \odot D\mathbf{m}_j(x_j)(1) \right) \ominus \mathbf{f}(\mathbf{x}) \right\|^2 p(\mathbf{x}) d\mathbf{x}.$$

F is a convex, continuous and Gâteaux differentiable functional satisfying $F([\mathbf{f}]) \to \infty$ as $\|[\mathbf{f}]\|_2 \to \infty$. The claim follows by arguing as in the proof of Theorem 3.3.2.

Lemma 6.7.1. *Under the conditions of Theorem 4.2.2, it holds that* $\hat{\mathbf{\Delta}}_j(x_j) \ominus \mathbf{\Delta}_j^*(x_j) = \mathbf{r}_j(x_j)$ *a.e.* x_j *with respect to* $\mathrm{Leb}_1, 1 \leq j \leq d$.

This gives that, for a.e. $x_j \in I_j$ with respect to Leb_1,

$$\begin{aligned} n^{2/5} \odot (\hat{\mathbf{m}}_j(x_j) \ominus \mathbf{m}_j(x_j)) =& n^{2/5} \odot \tilde{\mathbf{m}}_j^A(x_j) \oplus [n^{2/5} h_j^2] \odot \mathbf{c}_j(x_j) \\ & \oplus n^{2/5} \odot \mathbf{\Delta}_j^*(x_j) \oplus o_p(1). \end{aligned}$$

By Theorem 4.2.1,

$$(n^{2/5} \odot \tilde{\mathbf{m}}_1^A(x_1), \cdots, n^{2/5} \odot \tilde{\mathbf{m}}_d^A(x_d)) \xrightarrow{d} (\mathbf{G}(\mathbf{0}, \tilde{C}_{1,x_1}), \cdots, \mathbf{G}(\mathbf{0}, \tilde{C}_{d,x_d})).$$

The fact that $n^{2/5}\sup_{x_j\in[0,1]}\|\mathbf{\Delta}_j^*(x_j)\| = O(1)$ and E.6 in Cohn (2013) entail that $(\lim_{n\to\infty} n^{2/5}\odot \mathbf{\Delta}_1^*,\ldots,\lim_{n\to\infty} n^{2/5}\odot \mathbf{\Delta}_d^*)$ satisfies (4.2.2) and (4.2.3). The uniqueness of sum map follows by arguing as in the proof of Theorem 3.3.2 and the uniqueness of decomposition follows from Lemma 6.1.8. Also, $\lim_{n\to\infty}([n^{2/5}h_j^2]\odot \mathbf{c}_j(x_j)\oplus n^{2/5}\odot \mathbf{\Delta}_j^*(x_j)) = \mathbf{\Theta}_j(x_j)$. This proves the first and the second part of the theorem.

For the third part of the theorem, we note that Proposition 4.8 in van Neerven (2008) also holds for strongly measurable Gaussian random elements. Since $\mathbf{G}(\mathbf{0},\tilde{C}_{1,x_1}),\cdots,\mathbf{G}(\mathbf{0},\tilde{C}_{d,x_d})$ in Theorem 4.2.1 are independent, it follows that $\bigoplus_{j=1}^d \mathbf{G}(\mathbf{0},\tilde{C}_{j,x_j}) = \mathbf{G}\big(\mathbf{0},\sum_{j=1}^d \tilde{C}_{j,x_j}\big)$. This completes the third part of the theorem.

6.8 Proof of Lemma 6.1.1

Using Proposition 2.6.8 in Cohn (2013), one may show that $\mathbf{g}\in \mathcal{L}^1((\mathcal{W},\mathscr{B},\mu\mathbf{T}^{-1}),\mathbb{B})$ if and only if $\mathbf{g}(\mathbf{T})\in\mathcal{L}^1((\mathcal{Z},\mathscr{A},\mu),\mathbb{B})$. In which case, there exist $\mu\mathbf{T}^{-1}$-simple maps $\mathbf{g}_n$ such that $\mathbf{g}_n\to\mathbf{g}$ and $\|\mathbf{g}_n\|\le\|\mathbf{g}\|$ on $\mathcal{W}$ by E.2 in Cohn (2013). Using E.6 in Cohn (2013), one can show that

$$\begin{aligned}\int_{\mathcal{W}}\mathbf{g}(\mathbf{w})d\mu\mathbf{T}^{-1}(\mathbf{w}) &= \lim_{n\to\infty}\int_{\mathcal{W}}\mathbf{g}_n(\mathbf{w})d\mu\mathbf{T}^{-1}(\mathbf{w})\\ &= \lim_{n\to\infty}\int_{\mathcal{Z}}\mathbf{g}_n(\mathbf{T}(\mathbf{z}))d\mu(\mathbf{z})\\ &= \int_{\mathcal{Z}}\mathbf{g}(\mathbf{T}(\mathbf{z}))d\mu(\mathbf{z}).\end{aligned}$$

This completes the proof.

6.9 Proof of Lemma 6.1.2

There exist μ-simple maps $\mathbf{f}_n$ such that $\mathbf{f}_n \to \mathbf{f}$ and $\|\mathbf{f}_n\| \leq \|\mathbf{f}\|$ on $\mathcal{Z}$ by E.2 in Cohn (2013). Then, $\mathbf{f}_n \odot (d\lambda)/(d\mu) \to \mathbf{f} \odot (d\lambda)/(d\mu)$ and $\|\mathbf{f}_n\|(d\lambda)/(d\mu) \leq \|\mathbf{f}\|(d\lambda)/(d\mu)$ on $\mathcal{Z}$. Since each $\mathbf{f}_n \odot (d\lambda)/(d\mu)$ is measurable, $\mathbf{f} \odot (d\lambda)/(d\mu)$ is measurable by E.1 in Cohn (2013). Also, $\mathbf{f} \odot (d\lambda)/(d\mu)$ is μ-integrable since $\int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\|(d\lambda)/(d\mu)(\mathbf{z})d\mu(\mathbf{z}) = \int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z})\|d\lambda(\mathbf{z}) < \infty$. Using E.6 in Cohn (2013), one can show that

$$\begin{aligned}\int_{\mathcal{Z}} \mathbf{f}(\mathbf{z})d\lambda(\mathbf{z}) &= \lim_{n\to\infty} \int_{\mathcal{Z}} \mathbf{f}_n(\mathbf{z})d\lambda(\mathbf{z}) \\ &= \lim_{n\to\infty} \int_{\mathcal{Z}} \mathbf{f}_n(\mathbf{z}) \odot (d\lambda)/(d\mu)(\mathbf{z})d\mu(\mathbf{z}) \\ &= \int_{\mathcal{Z}} \mathbf{f}(\mathbf{z}) \odot (d\lambda)/(d\mu)(\mathbf{z})d\mu(\mathbf{z}).\end{aligned}$$

6.10 Proof of Lemma 6.1.3

(a) follows from Lemma 8.1 in Lang (1993).

Now, we prove (b). Since $\mathbf{k}$ is measurable and $D_{\mathcal{W}} \in \mathscr{B}, \mathbf{k} \odot 1_{D_{\mathcal{W}}}$ is measurable. By E.2 in Cohn (2013), there exist maps $\mathbf{k}_n := \bigoplus_{j=1}^{J_n} 1_{C_{nj}} \odot \mathbf{b}_{nj}$, where $J_n \in \mathbb{N}, C_{nj} \in \mathscr{A} \otimes \mathscr{B}$ and $\mathbf{b}_{nj} \in \mathbb{B}$, such that $\mathbf{k}_n \to \mathbf{k} \odot 1_{D_{\mathcal{W}}}$ and $\|\mathbf{k}_n\| \leq \|\mathbf{k}\|1_{D_{\mathcal{W}}}$ on $\mathcal{Z} \times \mathcal{W}$. Then, the maps $\mathbf{k}_n(\cdot, \mathbf{w}) : \mathcal{Z} \to \mathbb{B}$ are written as $\bigoplus_{j=1}^{J_n} 1_{(C_{nj})_{\mathbf{w}}} \odot \mathbf{b}_{nj}$, where $(C_{nj})_{\mathbf{w}} = \{\mathbf{z} \in \mathcal{Z} | (\mathbf{z}, \mathbf{w}) \in C_{nj}\}$. Then, $\mathbf{k}_n(\cdot, \mathbf{w})$ are measurable since $(C_{nj})_{\mathbf{w}} \in \mathscr{A}$. Moreover, they are μ-integrable since $\mathbf{k}(\cdot, \mathbf{w}) \odot 1_{D_{\mathcal{W}}}(\mathbf{w})$ is μ-integrable. Note that $\int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z}, \mathbf{w})d\mu(\mathbf{z}) = \bigoplus_{j=1}^{J_n} \mu((C_{nj})_{\mathbf{w}}) \odot \mathbf{b}_{nj}$. Since the functions $\mathbf{w} \mapsto \mu((C_{nj})_{\mathbf{w}})$ are measurable, the maps $\mathbf{w} \mapsto \int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z}, \mathbf{w})d\mu(\mathbf{z})$ are also measurable. Since $\mathbf{g}_0$ is also measurable, the maps $\mathbf{g}_n : \mathcal{W} \to \mathbb{B}$ defined by

$\mathbf{g}_n(\mathbf{w}) = \int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z}, \mathbf{w})d\mu(\mathbf{z}) \odot 1_{D_{\mathcal{W}}}(\mathbf{w}) \oplus \mathbf{g}_0(\mathbf{w}) \odot 1_{D^c_{\mathcal{W}}}(\mathbf{w})$ are measurable. Also, $\int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z}, \mathbf{w})d\mu(\mathbf{z}) \to \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w})d\mu(\mathbf{z})$ by E.6 in Cohn (2013). Thus, $\mathbf{g}_n \to \mathbf{g}$ on $\mathcal{W}$. Therfore, $\mathbf{g}$ is measurable by E.1 in Cohn (2013). This proves (b).

For the proof of (c), note that $\mathbf{k} \in \mathcal{L}^1(\mu \otimes \nu, \mathbb{B})$ implies that

$$\int_{\mathcal{W}} \int_{\mathcal{Z}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\| d\mu(\mathbf{z}) d\nu(\mathbf{w}) = \int_{\mathcal{Z}\times\mathcal{W}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\| d\mu \otimes \nu(\mathbf{z}, \mathbf{w}) < \infty.$$

This holds by the Fubini's theorem in Mukherjea (1972). Hence, $D_{\mathcal{W}} \in \mathscr{B}$ and $\nu(D^c_{\mathcal{W}}) = 0$. Define $\mathbf{g}_0, \mathbf{g}_n, \mathbf{g}$ and $\mathbf{k}_n$ as in the proof of (b). A similar argument to the proof of (b) shows that the function $g : \mathcal{W} \to \mathbb{R}$ defined by $g(\mathbf{w}) = \int_{\mathcal{Z}} \|\mathbf{k}(\mathbf{z}, \mathbf{w})\| d\mu(\mathbf{z}) 1_{D_{\mathcal{W}}}(\mathbf{w}) + \|\mathbf{g}_0(\mathbf{w})\| 1_{D^c_{\mathcal{W}}}(\mathbf{w})$ is ν-integrable. Since $\|\mathbf{g}\|$ is dominated by g, E.6 in Cohn (2013) shows that

$$\int_{\mathcal{W}} \mathbf{g}_n(\mathbf{w}) d\nu \to \int_{\mathcal{W}} \mathbf{g}(\mathbf{w}) d\nu = \int_{\mathcal{W}} \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}). \tag{6.10.1}$$

On the other hand, the Fubini's theorem in Mukherjea (1972) shows that

$$\begin{aligned} \int_{\mathcal{W}} \mathbf{g}_n(\mathbf{w}) d\nu(\mathbf{w}) &= \int_{\mathcal{W}} \int_{\mathcal{Z}} \mathbf{k}_n(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}) \\ &= \int_{\mathcal{Z}\times\mathcal{W}} \mathbf{k}_n(\mathbf{z}, \mathbf{w}) d\mu \otimes \nu(\mathbf{z}, \mathbf{w}) \qquad (6.10.2) \\ &\to \int_{\mathcal{Z}\times\mathcal{W}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu \otimes \nu(\mathbf{z}, \mathbf{w}). \end{aligned}$$

By combining (6.10.1) and (6.10.2), we have $\int_{\mathcal{W}} \int_{\mathcal{Z}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu(\mathbf{z}) d\nu(\mathbf{w}) = \int_{\mathcal{Z}\times\mathcal{W}} \mathbf{k}(\mathbf{z}, \mathbf{w}) d\mu \otimes \nu(\mathbf{z}, \mathbf{w})$. This completes the proof of (c).

6.11 Proof of Lemma 6.1.4

E.1 in Cohn (2013) implies that for each $\mathbf{f} \in \mathcal{L}^p((\mathcal{Z}, \mathscr{A}, \mu), \mathbb{B})$, there exist μ-simple maps $\mathbf{f}_n = \bigoplus_{i=1}^{I_n} 1_{A_{ni}} \odot \mathbf{b}_{ni}$, where $I_n \in \mathbb{N}$, $A_{ni} \in \mathscr{A}$ and $\mathbf{b}_{ni} \in \mathbb{B}$, such that $\mathbf{f}_n \to \mathbf{f}$ and $\|\mathbf{f}_n\| \leq \|\mathbf{f}\|$ on $\mathcal{Z}$. Then, the Lebesgue's dominated convergence theorem implies that $(\int_{\mathcal{Z}} \|\mathbf{f}_n(\mathbf{z}) \ominus \mathbf{f}(\mathbf{z})\|^p d\mu(\mathbf{z}))^{1/p} \to 0$. Hence, for a given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $(\int_{\mathcal{Z}} \|\bigoplus_{i=1}^{I_N}(1_{A_{Ni}}(\mathbf{z}) \odot \mathbf{b}_{Ni}) \ominus \mathbf{f}(\mathbf{z})\|^p d\mu(\mathbf{z}))^{1/p} < \epsilon/2$. One can show that for each i, there exists $A_i \in \mathscr{A}_0$ such that $\mu(A_{Ni} \Delta A_i) < (\epsilon/(2I_N\|\mathbf{b}_{Ni}\|))^p$. Note that $\mu(A_i) \leq \mu(A_{Ni} \Delta A_i) + \mu(A_{Ni}) < \infty$, and

$$
\begin{aligned}
(\int_{\mathcal{Z}} \|\bigoplus_{i=1}^{I_N}(1_{A_{Ni}}(\mathbf{z}) \odot \mathbf{b}_{Ni}) \ominus \bigoplus_{i=1}^{I_N}(1_{A_i}(\mathbf{z}) \odot \mathbf{b}_{Ni})\|^p d\mu(\mathbf{z}))^{1/p} \\
\leq \sum_{i=1}^{I_N} (\int_{\mathcal{Z}} 1_{A_{Ni} \Delta A_i}(\mathbf{z}) d\mu(\mathbf{z}))^{1/p} \|\mathbf{b}_{Ni}\| \\
< \epsilon/2.
\end{aligned}
$$

Therefore, $(\int_{\mathcal{Z}} \|\mathbf{f}(\mathbf{z}) \ominus \bigoplus_{i=1}^{I_N}(1_{A_i}(\mathbf{z}) \odot \mathbf{b}_{Ni})\|^p d\mu(\mathbf{z}))^{1/p} < \epsilon$. This completes the proof.

6.12 Proof of Lemma 6.1.6

For the measurability, we need to prove that range($\mathbf{k}$) is separable and $\mathbf{k}$ is $(\mathscr{A} \otimes \mathscr{B}, \mathcal{B}(\mathbb{B}))$-measurable. For the separability, define $\mathbb{R}_{\mathbf{b}} = \{r \odot \mathbf{b} | r \in \mathbb{R}\} \subset \mathbb{B}$ and $\mathbb{Q}_{\mathbf{b}} = \{q \odot \mathbf{b} | q \in \mathbb{Q}\} \subset \mathbb{B}$. Note that for any $\epsilon > 0$ and $r \in \mathbb{R}$, there exists $q \in \mathbb{Q}$ such that $\|r \odot \mathbf{b} \ominus q \odot \mathbf{b}\| = |r - q|\|\mathbf{b}\| < \epsilon$. Hence, $\mathbb{Q}_{\mathbf{b}}$ is a countable dense subset of $\mathbb{R}_{\mathbf{b}}$. Thus, $\mathbb{R}_{\mathbf{b}}$ is separable. Since range($\mathbf{k}$) $\subset \mathbb{R}_{\mathbf{b}}$, and

$\mathbb{R}_{\mathbf{b}}$ is a metric space, range($\mathbf{k}$) is also separable. For the $(\mathscr{A} \otimes \mathscr{B}, \mathcal{B}(\mathbb{B}))$-measurability, note that there exist measurable simple functions $k_n : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$ such that $k_n \to k$ on $\mathcal{Z} \times \mathcal{W}$. Since the maps $\mathbf{k}_n : \mathcal{Z} \times \mathcal{W} \to \mathbb{B}$ defined by $\mathbf{k}_n(\mathbf{z}, \mathbf{w}) = k_n(\mathbf{z}, \mathbf{w}) \odot \mathbf{b}$ are also $(\mathscr{A} \otimes \mathscr{B}, \mathcal{B}(\mathbb{B}))$-measurable, and $\mathbf{k}_n \to \mathbf{k}$ on $\mathcal{Z} \times \mathcal{W}$, $\mathbf{k}$ is $(\mathscr{A} \otimes \mathscr{B}, \mathcal{B}(\mathbb{B}))$-measurable by E.1 in Cohn (2013).

6.13 Proof of Proposition 6.1.2

A similar argument to the proof of Theorem 1.1 in Kundu et al. (2000) gives that $(< \mathbf{S}_n, \mathbf{b}_1 >, \cdots, < \mathbf{S}_n, \mathbf{b}_N >)^\top \xrightarrow{d} \mathbf{N}(0_d, A)$, where $0_d = (0, \cdots, 0)^\top \in \mathbb{R}^d$ and A is a matrix whose (k, l)th entry is a_{kl}. Since each $< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_k >$ is normally distributed, $(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_1 >, \cdots, < \mathbf{G}(\mathbf{0}, C), \mathbf{b}_N >)^\top$ follows a multivariate normal distribution. Since $E(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_k >) = 0$ for all k, and the (k, l)th entry of $E((< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_1 >, \cdots, < \mathbf{G}(\mathbf{0}, C), \mathbf{b}_N >)^\top(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_1 >, \cdots, < \mathbf{G}(\mathbf{0}, C), \mathbf{b}_N >))$ is $E(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_k >< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_l >) = a_{kl}$, we have $(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_1 >, \cdots, < \mathbf{G}(\mathbf{0}, C), \mathbf{b}_N >)^\top \overset{d}{\equiv} \mathbf{N}(0_d, A)$. Consider $\mathbf{T} : \mathbb{R}^N \to \mathbb{H}$ defined by $\mathbf{T}(u_1, \cdots, u_N) = \sum_{k=1}^{N} u_k \odot \mathbf{b}_k$. Then, $\mathbf{T}(< \mathbf{S}_n, \mathbf{b}_1 >, \cdots, < \mathbf{S}_n, \mathbf{b}_N >) = \mathbf{S}_n$ and $\mathbf{T}(< \mathbf{G}(\mathbf{0}, C), \mathbf{b}_1 >, \cdots, < \mathbf{G}(\mathbf{0}, C), \mathbf{b}_N >) = \mathbf{G}(\mathbf{0}, C)$. Since $\mathbf{T}$ is a continuous map, Theorem 2.3 in Bosq (2000) implies that $\mathbf{S}_n \xrightarrow{d} \mathbf{G}(\mathbf{0}, C)$.

Bibliography

[1] Aliprantis, C. D. and Border, K. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag Berlin Heidelberg.

[2] Atkinson, K. and Han, W. (2009). *Theoretical Numerical Analysis*. Springer-Verlag New York.

[3] Beltrami, E. J. (1967). On infinite-dimensional convex programs. *J. Comput. System Sci.* **1** 323-329.

[4] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

[5] Blot, J. and Cieutat, P. (2016). Completeness of sums of subspaces of bounded functions and applications. *Commun. Math. Anal.* **19** 43-61.

[6] Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer-Verlag New York.

[7] Busby, R. C., Schochetman, I. and Smith, H. A. (1972).

Integral operators and the compactness of induced representations. *Trans. Am. Math. Soc.* **164** 461-477.

[8] Chiou, J. M., Müller, H.-G. and Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc. Ser. B* **65** 405-423.

[9] Cohn, D. L. (2013). *Measure Theory.* Birkhäuser Basel.

[10] Dabo-Niang, S. and Rhomari, N. (2009). Kernel regression estimation in a Banach space. *J. Statist. Plan. Infer.* **139** 1421-1434.

[11] Diaz, J. B. and Metcalf, F. T. (1966). A complementary triangle inequality in Hilbert and Banach spaces. *P. Am. Math. Soc.* **17** 88-97.

[12] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2011). Kernel regression with functional response. *Electron. J. Statist.* **5** 159-171.

[13] Ferraty, F., Van Keilegom, I. and Vieu, P. (2012). Regression when both response and predictor are functions. *J. Multivariate Anal.* **109** 10-28.

[14] Gneiting, T., Ševčíková, H. and Percival, D. B. (2012). Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statist. Sci.* **27** 247-277.

[15] Han, K., Müller, H.-G. and Park, B. U. (2018). Smooth backfitting for additive modeling with small errors-in-variables,

with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233-1265.

[16] Han, K. and Park, B. U. (2018+). Smooth backfitting for error-in-variables additive models. To appear in *Ann. Statist.*

[17] Jiang, C. I. and Wang, J.-L. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362-388.

[18] Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag New York.

[19] Kubrusly, C. S. (2015). *Essentials of Measure Theory.* Springer International Publishing.

[20] Kundu, S., Majumdar, S. and Mukherjee, K. (2000). Central limit theorems revisited. *Statist. Prob. Lett.* **47** 265-275.

[21] Lang, S. (1993). *Real and Functional Analysis.* Springer-Verlag New York.

[22] Lee, K., Lee, Y. K., Park, B. U. and Yang, S. J. (2018). Time-dynamic varying coefficient models for longitudinal data. *Comput. Stat. Data Anal.* **123** 50-65.

[23] Lee, Y. K., Mammen, E. and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857-2883.

[24] Lee, Y. K., Mammen, E. and Park, B. U. (2012). Flexible generalized varying coefficient regression models. *Ann. Statist.* **40** 1906-1933.

[25] Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electron. J. Statist.* **5** 31-40.

[26] Lian, H. (2012). Convergence of nonparametric functional regression estimates with functional responses. *Electron. J. Statist.* **6** 1373-1391.

[27] Linton, O., Sperlich, S. and van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Statist.* **36** 686-718.

[28] Mammen, E., Linton, O. B. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443-1490.

[29] Mammen, E. and Park, B. U. (2006). A simple smooth backfitting method for additive models. *Ann. Statist.* **34** 2252-2271.

[30] Mammen, E., Park, B. U. and Shienle, M. (2014). Additive models: Extensions and related models. In The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics (J. S. Racine, L. Su and A. Ullah, eds.) 176-211. Oxford Univ. Press, Oxford.

[31] Mükherjea, A. (1972). A remark on Tonelli's theorem on integral in product spaces. *Pacific J. Math.* **42** 177-185.

[32] Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. Roy. Statist. Soc. Ser. B* **67** 43-61.

[33] Park, B. U., Chen, C.-J., Tao, W. and Müller, H.-G. (2018+). Singular additive models for function to function regression. To appear in *Statist. Sinica*

[34] Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183-218.

[35] Scheipl, F., Staicu, A.-M. and Greven, S. (2015). Functional additive mixed models. *J. Comp. Graph. Statist.* **24** 477-501.

[36] Tsagris, M. (2015). Regression analysis with compositional data containing zero values. *Chil. J. Statist.* **6** 47-57.

[37] van den Boogaart, K. G., Egozcue, J. J. and Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. *Aust. N. Z. J. Statist.* **56** 171-194.

[38] van Neerven, J. (2008). *Stochastic evolution equations.* Lecture Notes of the 11th Internet Seminar, TU Delft OpenCourseWare, http://ocw.tudelft.nl.

[39] Väth, M. (2000). *Volterra and Integral Equations of Vector Functions.* CRC Press.

[40] Xu, J. and Zikatanov, L. (2002). The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Amer. Math. Soc.* **15** 573-597.

[41] Yu, K., Park, B. U. and Mammen, E. (2008). Smooth

backfitting in generalized additive models. *Ann. Statist.* **36** 228-260.

[42] Zhang, X., Park, B. U. and Wang, J.-L. (2013). Time-varying additive models for longitudinal data. *J. Amer. Statist. Assoc.* **108** 983-998.

[43] Zhu, H., Li, R. and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist.* **40** 2634-2666.

국문초록

이 논문에서는 힐버트 반응변수를 가진 구조화된 비모수 회귀 모형에서 추정의 방법론과 이론의 기초를 정립한다. 이를 위해 바나흐 공간 값을 갖는 함수의 적분인 보크너 적분을 도입하고 그 통계적 성질을 처음으로 밝힌다. 또한 제안된 추정량의 존재성과 추정량을 얻기 위한 알고리즘의 수렴성을 점근적 측면과 비점근적 측면에서 모두 증명한다. 그리고 각 성분 함수의 추정량이 각 성분 함수로 최적의 오차로 점근 수렴하고, 성분 함수 추정량들의 쌍이 가우시안 확률 변수들의 쌍으로 분포 수렴하며, 수렴된 가우시안 확률 변수의 성분들이 서로 독립임을 보인다. 시뮬레이션과 실제 자료 분석을 통해 제안된 방법이 제곱 적분 가능한 함수, 확률 밀도 함수, 구성비 벡터 등 여러 힐버트 반응변수에서 잘 작동함을 확인한다.

주요어 : 가법 모형, 평활 역적합, 보크너 적분, 비유클리디안 자료, 무한 차원 공간, 힐버트 공간, 함수적 반응변수.

학번 : 2012-20232