Principal component analysis for Hilbertian functional data

Dongwoo Kim[a], Young Kyung Lee[1,b], Byeong U. Park[a]

[a]Department of Statistics, Seoul National University, Korea;
[b]Department of Information Statistics, Kangwon National University, Korea

Abstract

In this paper we extend the functional principal component analysis for real-valued random functions to the case of Hilbert-space-valued functional random objects. For this, we introduce an autocovariance operator acting on the space of real-valued functions. We establish an eigendecomposition of the autocovariance operator and a Karuhnen-Loève expansion. We propose the estimators of the eigenfunctions and the functional principal component scores, and investigate the rates of convergence of the estimators to their targets. We detail the implementation of the methodology for the cases of compositional vectors and density functions, and illustrate the method by analyzing time-varying population composition data. We also discuss an extension of the methodology to multivariate cases and develop the corresponding theory.

Keywords: principal component analysis, functional data, Hilbert space

1. Introduction

Principal component analysis (PCA) is a useful tool for identifying the dominant directions of variation of multivariate data. Functional PCA (FPCA) is its generalization to time-varying data, called functional data or stochastic processes. Functional data may take values in Euclidean spaces, Hilbert spaces, or more generally in metric spaces. Most of the works on functional data have been focused on the Euclidean case (Ramsay and Silverman, 2005; Chiou *et al.*, 2014). These days we often encounter non-Euclidean data. There has been increasing interest in developing suitable statistical methodologies for analyzing non-Euclidean data. The Hilbertian case is central in this development. In this paper, we discuss PCA when the observed data are time-varying and take values in a general Hilbert space.

Recently, Lin and Yao (2019) proposed an approach to FPCA for Riemannian manifolds. Although they are mainly concerned with manifold-valued functional data, their approach is based on an eigenanalysis for Hilbertian functional data, so that one may use the procedure described there for Hilbertian stochastic processes. The eigenanalysis discussed in Lin and Yao (2019) gives eigenfunctions that live in the same space as the data objects. This means that, if one adopts their approach for Hilbertian functional data, then one would have eigenfunctions whose trajectories are time-varying Hilbertian values. In such cases one may face some difficulty in visualizing or interpreting the resulting eigenfunctions.

In this paper we take a different approach. Our approach gives real-valued eigenfunctions regardless of the underlying Hilbert space where the random functions take values. The associated functional

---

[1] Corresponding author: Department of Information Statistics, Kangwon National University, 1 Gangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Korea. E-mail: youngklee@kangwon.ac.kr

principal component (FPC) scores lie in the Hilbert space. For example, consider the case where the Hilbert space under study is a simplex, $\mathcal{S}^{k-1} = \{(x_1, \ldots, x_k) \in (0,1)^k : x_1 + \cdots + x_k = 1\}$. Let $\mathbf{X}$ be a random function defined on a time domain $\mathcal{T} \subset \mathbb{R}$ such that $\mathbf{X}(t)$ for each $t \in \mathcal{T}$ is a random element taking values in $\mathcal{S}^{k-1}$. In this case, our approach gives real-valued eigenfunctions which map $\mathcal{T}$ to $\mathbb{R}$, and FPC scores which live in $\mathcal{S}^{k-1}$. Our procedure specialized to this simplex example may be considered as a generalization of the recent work by Wang *et al.* (2015) who studied the PCA for multivariate compositional data. By introducing an autocovariance operator acting on the space of real-valued functions, we establish an eigendecomposition of the autocovariance operator and a Karuhnen-Loève expansion. We illustrate our proposal through a real data example. We also extend the methodology to multivariate Hilbertian functional data.

2. Methodology

Let $\mathbb{H}$ be a separable Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and the corresponding norm $\| \cdot \|_{\mathbb{H}}$. We denote the zero vector by $\mathbf{0}$, the addition operation by $\oplus$ and the scalar multiplication by $\odot$. There are numerous examples of Hilbert spaces. The most well-known examples are the spaces of square integrable real-valued functions defined on a set $S \subset \mathbb{R}^k$. In this case, $\mathbf{0}$ is a zero function, $\mathbf{a} \oplus \mathbf{b}$ for $\mathbf{a} = a(\cdot)$ and $\mathbf{b} = b(\cdot)$ with $a, b : S \to \mathbb{R}$ is defined by $(a \oplus b)(\mathbf{z}) = a(\mathbf{z}) + b(\mathbf{z})$ and $(c \odot a)(\mathbf{z}) = c \cdot a(\mathbf{z})$ for $c \in \mathbb{R}$. It is well known that these are separable Hilbert spaces. In Section 3 we discuss two other examples. Throughout this paper, we let $L_2^{\mathcal{H}}$, for a Hilbert space $\mathcal{H}$ with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, denote the space of square integrable $\mathcal{H}$-valued functions $f : [0,1] \to \mathcal{H}$, i.e., $L_2^{\mathcal{H}} = \{f : \int_0^1 \|f(t)\|_{\mathcal{H}}^2 \, dt < \infty\}$ where $\|f(t)\|_{\mathcal{H}}^2 = \langle f(t), f(t) \rangle_{\mathcal{H}}$. The associated inner product $\langle \cdot, \cdot \rangle_{L_2^{\mathcal{H}}}$ on $L_2^{\mathcal{H}}$ is defined by $\langle f, g \rangle_{L_2^{\mathcal{H}}} = \int_0^1 \langle f(t), g(t) \rangle_{\mathcal{H}} \, dt$ and its norm $\| \cdot \|_{L_2^{\mathcal{H}}}$ by $\|f\|_{L_2^{\mathcal{H}}}^2 = \langle f, f \rangle_{L_2^{\mathcal{H}}}$.

2.1. Eigenanalysis for Hilbertian functional data

Let $\mathbf{X}$ be a continuous $\mathbb{H}$-valued stochastic process on the unit time interval $[0,1]$. For each $t \in [0,1]$, $\mathbf{X}(t)$ is a random element taking values in $\mathbb{H}$. We assume $E \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^2 \, dt < \infty$. This implies that $E\|\mathbf{X}(t)\|_{\mathbb{H}}^2 < \infty$ a.e. $t \in [0,1]$. It also implies that $\int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^2 \, dt < \infty$ with probability one. For simplicity of presentation, we consider a centered $\mathbf{X}$ such that $E\mathbf{X}(t) = \mathbf{0}$ for all $t \in [0,1]$.

We let $C : [0,1] \times [0,1] \to \mathbb{R}$ be the bivariate function defined by

$$C(s,t) = E\left(\langle \mathbf{X}(s), \mathbf{X}(t) \rangle_{\mathbb{H}}\right). \tag{2.1}$$

In the case where $\mathbb{H} = \mathbb{R}$, the C reduces to the usual 'autocovariance' function $C(s,t) = E(X(s)X(t))$. We consider the integral operator $\mathscr{C} : L_2^{\mathbb{R}} \to L_2^{\mathbb{R}}$ defined by

$$\mathscr{C}(f)(s) = \int_0^1 C(s,t)f(t) \, dt, \quad s \in [0,1], \ f \in L_2^{\mathbb{R}}.$$

Then, $\mathscr{C}$ is nonnegative definite. This follows since

$$\langle \mathscr{C}(f), f \rangle_{L_2^{\mathbb{R}}} = \int_0^1 \mathscr{C}(f)(s) \cdot f(s) \, ds$$

$$= E \left\| \int_0^1 f(t) \odot \mathbf{X}(t) \, dt \right\|_{\mathbb{H}}^2 \geq 0,$$

for all $f \in L_2^{\mathbb{R}}$, where the integral on the right hand side of the second equality is in the Bochner sense. Bochner integration is a notion of integration that generalizes the conventional Lebesgue integrals to Banach-space-valued maps, see Section 2 in Jeon and Park (2020) for details. The kernel $C(\cdot, \cdot)$ of the integral operator $\mathscr{C}$ admits the following eigendecomposition due to Mercer's Theorem:

$$C(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t), \qquad (2.2)$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\{\phi_j\}_{l=1}^{\infty}$ are the associated eigenfunctions of $\mathscr{C}$ that form a complete orthonormal system for the space $L_2^{\mathbb{R}}$. The equality (2.2) is a standard result in functional data analysis, see Theorem 4.6.5 in Hsing and Eubank (2015), for example. The sum of all eigenvalues equals the total variation of $\mathbf{X}$, i.e.,

$$\sum_{j=1}^{\infty} \lambda_j = E \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^2 \, dt.$$

Now, we derive a Karhunen-Loève expansion of $\mathbf{X}$. We note that the standard result such as Theorem 7.2.7 in Hsing and Eubank (2015) do not apply here, since $\phi_j \in L_2^{\mathbb{R}}$ and $\mathbf{X}(\cdot, \omega) \in L_2^{\mathbb{H}}$ belong to different spaces. Let $\{\mathbf{e}_l\}_{l=1}^{L}$ be an orthonormal basis of $\mathbb{H}$, where we allow $L = \infty$. For each $l \geq 1$, define $Z_l(t) = \langle \mathbf{X}(t), \mathbf{e}_l \rangle_{\mathbb{H}}$. They are real-valued stochastic processes. Note that $Z_l(\cdot) \in L_2^{\mathbb{R}}$ with probability one since

$$\int_0^1 |\langle \mathbf{X}(t), \mathbf{e}_l \rangle_{\mathbb{H}}|^2 \, dt \leq \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^2 \, dt \qquad (2.3)$$

and the integral on the right hand side of (2.3) exists with probability one. Since $\{\phi_j\}_{j=1}^{\infty}$ is a complete orthonormal system for $L_2^{\mathbb{R}}$, it holds that

$$Z_l(t) = \sum_{j=1}^{\infty} \left(\int_0^1 \langle \mathbf{X}(s), \mathbf{e}_l \rangle_{\mathbb{H}} \cdot \phi_j(s) \, ds \right) \cdot \phi_j(t). \qquad (2.4)$$

Also, since $\{\mathbf{e}_l\}_{l=1}^{L}$ is an orthonormal basis of $\mathbb{H}$, we get that

$$\begin{aligned}
\mathbf{X}(t) &= \bigoplus_{l=1}^{L} Z_l(t) \odot \mathbf{e}_l \\
&= \bigoplus_{j=1}^{\infty} \bigoplus_{l=1}^{L} \left(\int_0^1 \langle \mathbf{X}(s), \mathbf{e}_l \rangle_{\mathbb{H}} \cdot \phi_j(s) \, ds \cdot \phi_j(t) \right) \odot \mathbf{e}_l \\
&= \bigoplus_{j=1}^{\infty} \left(\bigoplus_{l=1}^{L} \left\langle \int_0^1 \mathbf{X}(s) \odot \phi_j(s) \, ds, \ \mathbf{e}_l \right\rangle_{\mathbb{H}} \odot \mathbf{e}_l \right) \odot \phi_j(t) \\
&= \bigoplus_{j=1}^{\infty} \left(\int_0^1 \mathbf{X}(s) \odot \phi_j(s) \, ds \right) \odot \phi_j(t). \qquad (2.5)
\end{aligned}$$

Here, we have used (2.4) for the second equation. The integrals on the right hand sides of the third and fourth equations in (2.5) are Bochner integrals. Let $\boldsymbol{\xi}_j = \int_0^1 \mathbf{X}(s) \odot \phi_j(s) \, ds$. These are in $\mathbb{H}$ and

plays the role of FPC in the conventional FPCA for real-valued random functions. We call $\boldsymbol{\xi}_j$ the FPC scores of $\mathbf{X}$. The foregoing arguments establish the following theorem. Let $\ominus$ denote the subtraction operation in $\mathbb{H}$ defined by $\mathbf{a} \ominus \mathbf{b} = \mathbf{a} \oplus (-1) \odot \mathbf{b}$.

Theorem 1. *Assume that $\mathbf{X}$ is a continuous $\mathbb{H}$-valued stochastic process on $[0, 1]$ and that $E \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^2 dt < \infty$. Then, it holds that*

$$\lim_{N \to \infty} \sup_{t \in [0,1]} E \left\| \mathbf{X}(t) \ominus \bigoplus_{j=1}^N \boldsymbol{\xi}_j \odot \phi_j(t) \right\|_{\mathbb{H}}^2 = 0.$$

Apparently, $E\boldsymbol{\xi}_j = \mathbf{0}$. Any two FPC scores, $\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_k$ for $j \neq k$, are uncorrelated in the sense that $E(\langle \boldsymbol{\xi}_j, \boldsymbol{\xi}_k \rangle_{\mathbb{H}}) = 0$. Their variances defined by $E(\|\boldsymbol{\xi}_j\|_{\mathbb{H}}^2)$ equal the corresponding eigenvalues λ_j. The latter two properties follow from

$$\begin{aligned}
E\left(\langle \boldsymbol{\xi}_j, \boldsymbol{\xi}_k \rangle_{\mathbb{H}} \right) &= \int_{[0,1]^2} C(s, t) \phi_j(s) \phi_k(t) \, ds \, dt \\
&= \sum_{l \geq 1} \lambda_l \left(\int_0^1 \phi_l(s) \phi_j(s) \, ds \right) \left(\int_0^1 \phi_l(t) \phi_k(t) \, dt \right) \\
&= \lambda_j \cdot I(j = k).
\end{aligned}$$

Remark 1. We are aware of two recent pioneering works on PCA for Riemannian functional data, which are Dai and Müller (2018) and Lin and Yao (2019). Both deal with functional data taking values in Riemannian manifolds. The basic idea of the first is to embed the manifold under consideration in an Euclidean *ambient* space, perform an eigenanalysis in the ambient space and then transform the results back to the manifold via the 'exponential map'. Thus, the actual eigenanalysis is done on a Euclidean space. The latter approach is to transform Riemannian stochastic process to a random element taking values in a tensor Hilbert space and then perform an eigenanalysis in the latter Hilbert space. However, the eigenanalysis in the Hilbert space is different from ours in that it produces eigenfunctions residing in the same space as the transformed stochastic process, while our approach gives eigenfunctions residing in $L_2^{\mathbb{R}}$, not in $L_2^{\mathbb{H}}$, regardless of the nature of $\mathbb{H}$.

2.2. Estimation of eigenfunctions and functional principal component scores

Let $\mathbf{X}_i$, $1 \leq i \leq n$, be independent copies of $\mathbf{X}$ that we actually observe. In this subsection, we describe a method of estimating ϕ_j and $\boldsymbol{\xi}_j$ that we introduced in the previous subsection. Without loss of generality we assume that the sample average $n^{-1} \bigoplus_{i=1}^n \mathbf{X}_i(t) = \mathbf{0}$. First, we estimate $C(s, t)$ by

$$\hat{C}(s, t) = n^{-1} \sum_{i=1}^n \langle \mathbf{X}_i(s), \mathbf{X}_i(t) \rangle_{\mathbb{H}}. \tag{2.6}$$

Define $\hat{\mathscr{C}} : L_2^{\mathbb{R}} \to L_2^{\mathbb{R}}$, the estimator of the integral operator $\mathscr{C}$, by

$$\hat{\mathscr{C}}(f)(s) = \int_0^1 \hat{C}(s, t) f(t) \, dt, \quad s \in [0, 1], \; f \in L_2^{\mathbb{R}}.$$

Then, it is also nonnegative definite since

$$\left\langle \hat{\mathscr{C}}(f), f \right\rangle_{L_2^{\mathbb{R}}} = n^{-1} \sum_{i=1}^n \left\| \int_0^1 f(t) \odot \mathbf{X}_i(t)\, dt \right\|_{\mathbb{H}}^2 \geq 0,$$

for all $f \in L_2^{\mathbb{R}}$. Thus, due to Mercer's Theorem again it holds that there exists $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$ and a complete orthonormal basis $\{\hat{\phi}_j\}_{j=1}^\infty$ of $L_2^{\mathbb{R}}$ such that

$$\hat{C}(s, t) = \sum_{j=1}^\infty \hat{\lambda}_j \hat{\phi}_j(s) \hat{\phi}_j(t).$$

The $\hat{\lambda}_j$ are the estimators of the respective λ_j and $\hat{\phi}_j$ are the estimators of the respective ϕ_j. Let $\boldsymbol{\xi}_{ij} = \int_0^1 \mathbf{X}_i(s) \odot \phi_j(s)\, ds$ be the true FPC scores of $\mathbf{X}_i$. We estimate $\boldsymbol{\xi}_{ij}$ by

$$\hat{\boldsymbol{\xi}}_{ij} = \int_0^1 \mathbf{X}_i(s) \odot \hat{\phi}_j(s)\, ds, \quad j \geq 1. \tag{2.7}$$

The empirical FPC scores $(\hat{\boldsymbol{\xi}}_{ij} : 1 \leq i \leq n)$ and $(\hat{\boldsymbol{\xi}}_{ik} : 1 \leq i \leq n)$ for $j \neq k$ are also uncorrelated in the sense that

$$n^{-1} \sum_{i=1}^n \left\langle \hat{\boldsymbol{\xi}}_{ij}, \hat{\boldsymbol{\xi}}_{ik} \right\rangle_{\mathbb{H}} = 0.$$

Also, we have $\hat{\lambda}_j = n^{-1} \sum_{i=1}^n \|\hat{\boldsymbol{\xi}}_{ij}\|_{\mathbb{H}}^2$ for $1 \leq j \leq d$. Furthermore, using the standard perturbation theory in Bosq (2000) or in Hsing and Eubank (2015), for example, we may show the following theorem.

Theorem 2. *Assume that $E \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^4\, dt < \infty$, and that $\min_{1 \leq l \leq J}(\lambda_l - \lambda_{l+1}) > 0$, where $J \geq 1$ is fixed. Then, for all $1 \leq j \leq J$, it holds that $\int_0^1 |\hat{\phi}_j(t) - \phi_j(t)|^2\, dt = O_p(n^{-1})$. Furthermore, if we assume additionally that $E \int_0^1 \|\mathbf{X}(t)\|_{\mathbb{H}}^{2\beta}\, dt < \infty$ for some $\beta \geq 2$, then $\max_{1 \leq i \leq n} \|\hat{\boldsymbol{\xi}}_{ij} \ominus \boldsymbol{\xi}_{ij}\|_{\mathbb{H}} = O_p(n^{-(\beta-1)/2\beta})$ for all $1 \leq j \leq J$.*

In the above theorem, J is fixed. We may let J depend on the sample size n and grow as $n \to \infty$. In the latter case, the rates of convergence of $\hat{\phi}_j$ and $\hat{\boldsymbol{\xi}}_{ij}$ to their targets depend on how fast $\delta_J := \min_{1 \leq l \leq J}(\lambda_l - \lambda_{l+1})$ tends to zero as $J \to \infty$.

3. Two special cases

Here, we consider two Hilbert spaces and give some more details for the implementation of the methodology presented in the previous section. One is the space of compositional vectors and the other is the space of density functions. The vector operations for these spaces are unconventional. We use isometric isomorphisms that map the respective Hilbert spaces to $\mathbb{R}^k$ and $L_2^{\mathbb{R}}$ where the conventional vector operations apply. We note that, for the space of square integrable real-valued functions on $\mathbb{R}^d$, which is another important class of $\mathbb{H}$, the vector operations are defined by $(f(\cdot) \oplus g(\cdot))(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u})$ and $(c \odot f(\cdot))(\mathbf{u}) = c \cdot f(\mathbf{u})$.

3.1. Space of compositional vectors

Suppose that $\mathbb{H}$ is the standard $(k-1)$-simplex, $\mathcal{S}^{k-1} = \{(v_1, \ldots, v_k) \in (0,1)^k : \sum_{j=1}^{k} v_j = 1\}$. For $\mathbf{v}, \mathbf{w} \in \mathcal{S}^{k-1}$ and $c \in \mathbb{R}$, define vector addition, scalar multiplication and inner product by

$$\mathbf{v} \oplus \mathbf{w} = \left(\frac{v_1 w_1}{v_1 w_1 + \cdots + v_k w_k}, \ldots, \frac{v_k w_k}{v_1 w_1 + \cdots + v_k w_k} \right),$$

$$c \odot \mathbf{v} = \left(\frac{v_1^c}{v_1^c + \cdots + v_k^c}, \ldots, \frac{v_k^c}{v_1^c + \cdots + v_k^c} \right),$$

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{S}^{k-1}} = \frac{1}{2k} \sum_{j=1}^{k} \sum_{l=1}^{k} \log \left(\frac{v_j}{v_l} \right) \log \left(\frac{w_j}{w_l} \right). \tag{3.1}$$

With $\mathbf{0} = (1/k, \ldots, 1/k)$ as the zero vector, $(\mathcal{S}^{k-1}, \oplus, \odot, \langle \cdot, \cdot \rangle_{\mathcal{S}^{k-1}})$ is a separable Hilbert space with the operation structure at (3.1), known as Aitchison geometry (Aitchison, 1986). The map $\mathcal{L} : \mathcal{S}^{k-1} \to \mathbb{R}^k$ defined by $\mathcal{L}(\mathbf{u}) = \mathbf{x}$ with $x_j = \log u_j - k^{-1} \sum_{l=1}^{k} \log u_l$, also known as the 'center log ratio' transform, is an isometry such that $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{S}^{k-1}} = \langle \mathcal{L}(\mathbf{v}), \mathcal{L}(\mathbf{w}) \rangle_E$, where $\langle \cdot, \cdot \rangle_E$ denotes the Euclidean inner product, i.e., $\langle \mathbf{a}, \mathbf{b} \rangle_E = \mathbf{a}^\top \mathbf{b}$. The transformation $\mathcal{L}$ is also an isomorphism. The addition $\mathbf{v} \oplus \mathbf{w}$ and the scalar multiplication $c \odot \mathbf{v}$ in $\mathcal{S}^{k-1}$ may be performed by the Euclidean operation with $\mathcal{L}(\mathbf{v})$ and $\mathcal{L}(\mathbf{w})$ and then transforming the results back to the corresponding compositional vectors. That is,

$$\mathbf{v} \oplus \mathbf{w} = \mathcal{L}^{-1}(\mathcal{L}(\mathbf{v}) + \mathcal{L}(\mathbf{w})), \quad c \odot \mathbf{v} = \mathcal{L}^{-1}(c \cdot \mathcal{L}(\mathbf{v})).$$

As we have seen in the previous section, the FPCA with Hilbertian functional data boils down to the computation of the covariance kernel $\hat{C}$ at (2.6). In the case of compositional functional data $\mathbf{X}_i$ such that $\mathbf{X}_i(t) \in \mathcal{S}^{k-1}$ for all $t \in [0, 1]$, it is nothing else than

$$\hat{C}(s, t) = n^{-1} \sum_{i=1}^{n} \mathcal{L}(\mathbf{X}_i(s))^\top \mathcal{L}(\mathbf{X}_i(t)). \tag{3.2}$$

The eigenfunctions $\hat{\phi}_j$ of the kernel $\hat{C}$ at (3.2) are then readily obtained from the existing method of eigenanalysis for real-valued functional data.

We discuss the computation of the Bochner integrals at (2.7) for $\hat{\boldsymbol{\xi}}_{ij}$. Suppose that we want to compute a Bochner integral $\int_0^1 \mathbf{f}(t) \, dt$ for a Bochner integrable map $\mathbf{f} \in L_2^{\mathcal{S}^{k-1}}$. Then, there exists a sequence of simple maps $\mathbf{f}_n : [0, 1] \to \mathcal{S}^{k-1}$ such that $\int_0^1 \|\mathbf{f}_n(t) \ominus \mathbf{f}(t)\|_{\mathcal{S}^{k-1}} \, dt \to 0$ as $n \to \infty$. Since the center log ratio transform $\mathcal{L}$ is an isometry, $\mathcal{L}(\mathbf{f})$ is Lebesgue integrable. Furthermore, $\int_0 \|\mathcal{L}(\mathbf{f}_n(t)) - \mathcal{L}(\mathbf{f}(t))\|_E \, dt \to 0$, where $\| \cdot \|_E$ denotes the Euclidean norm in $\mathbb{R}^k$. Now, note that the simple maps $\mathbf{f}_n$ take the form $\mathbf{f}_n(t) = \bigoplus_{l=1}^{N} 1_{S_{l,n}}(t) \odot \mathbf{h}_{l,n}$, where $S_{l,n} \subset [0, 1]$ satisfy $S_{l,n} \cap S_{l',n} = \emptyset$ for $l \neq l'$ and $\bigcup_{l=1}^{N} S_{l,n} = [0, 1]$, and $\mathbf{h}_{l,n}$ are constant compositional vectors. For such simple maps, it holds that

$$\int_0^1 \mathbf{f}_n(t) \, dt = \bigoplus_{l=1}^{N} \mu(S_{l,n}) \odot \mathbf{h}_{l,n} = \mathcal{L}^{-1} \mathcal{L} \left(\bigoplus_{l=1}^{N} \mu(S_{l,n}) \odot \mathbf{h}_{l,n} \right)$$

$$= \mathcal{L}^{-1} \left(\sum_{l=1}^{N} \mu(S_{l,n}) \cdot \mathcal{L}(\mathbf{h}_{l,n}) \right)$$

$$= \mathcal{L}^{-1} \left(\int_0^1 \mathcal{L}(\mathbf{f}_n(t)) \, dt \right) \tag{3.3}$$

since $\mathcal{L}$ is injective, where the last integral is in Lebesgue sense. Since $\mathcal{L}$ and its inverse are continuous, we get from (3.3) and the convergence $\int_0^1 \|\mathbf{f}_n(t) \ominus \mathbf{f}(t)\|_{\mathcal{S}^{k-1}} \, dt \to 0$ that

$$\int_0^1 \mathbf{f}(t) \, dt = \lim_{n \to \infty} \mathcal{L}^{-1} \left(\int_0^1 \mathcal{L}(\mathbf{f}_n(t)) \, dt \right) = \mathcal{L}^{-1} \left(\int_0^1 \mathcal{L}(\mathbf{f}(t)) \, dt \right),$$

where the second and third integrals are in Lebesgue sense. This means that we may evaluate $\hat{\boldsymbol{\xi}}_{ij}$ by the Lebesgue integration of $\mathcal{L}(\mathbf{X}_i(t) \odot \hat{\phi}_j(t))$ as follows.

$$\hat{\boldsymbol{\xi}}_{ij} = \mathcal{L}^{-1} \left(\int_0^1 \mathcal{L}\left(\mathbf{X}_i(t) \odot \hat{\phi}_j(t) \right) dt \right), \quad j \geq 1. \tag{3.4}$$

3.2. Space of density functions

For simplicity we consider the space of density functions supported on $[0, 1]$. An extension to the case of a general support on $\mathbb{R}^d$ with finite Lebesgue measure is immediate. This space is actually not a Hilbert space. We consider an equivalence relation such that f and g are equivalent if and only if f is a constant multiple of g. Let $[f]$ denote the class of real-valued functions g such that $g = c \cdot f$ for some constant $c > 0$. Let

$$\mathscr{F} = \left\{ [f] : \int_0^1 (\log f(x))^2 \, dx < \infty \right\}.$$

For this space, we define vector addition, scalar multiplication and inner product as follows.

$$[f] \oplus [g] = [f \cdot g], \quad c \odot [f] = [f^c],$$

$$\langle [f], [g] \rangle_{\mathscr{F}} = \int_{[0,1]^2} \log \left(\frac{f(x)}{f(y)} \right) \log \left(\frac{g(x)}{g(y)} \right) dx \, dy. \tag{3.5}$$

With $\mathbf{0} = [1] = \mathbb{R}$ as the zero vector, $\mathscr{F}$ is a separable Hilbert space with the operation structure at (3.5), see van den Boogaart $et\ al.$ (2014).

We define the map $\mathcal{L} : \mathcal{F} \to L_2^{\mathbb{R}}$ by

$$\mathcal{L}([f])(x) = \log f(x) - \int_0^1 \log f(u) \, du. \tag{3.6}$$

Then, $\mathcal{L}$ is an isometry preserving the metrics on the two spaces:

$$\langle [f], [g] \rangle_{\mathscr{F}} = \langle \mathcal{L}([f]), \mathcal{L}([g]) \rangle_{L_2^{\mathbb{R}}}$$

$$= \int_0^1 \left(\log f(x) - \int_0^1 \log f(u) \, du \right) \left(\log g(x) - \int_0^1 \log g(u) \, du \right) dx.$$

Here we recall that $\langle \cdot, \cdot \rangle_{L_2^{\mathbb{R}}}$ is the usual inner product on $L_2^{\mathbb{R}}$. We note that $\mathcal{L}$ is scale-invariance, so that we may simply write $\mathcal{L}(f)$ instead of $\mathcal{L}([f])$. The transformation $\mathcal{L}$ is also an isomorphism. As in the case of compositional vectors, the addition and scalar multiplication in $\mathcal{F}$ may be performed by doing the corresponding operation on $L_2^{\mathbb{R}}$ and then transforming the results back to the density space. That is,

$$[f] \oplus [g] = \mathcal{L}^{-1}(\mathcal{L}(f) + \mathcal{L}(g)), \quad c \odot [f] = \mathcal{L}^{-1}(c \cdot \mathcal{L}(f)).$$

The covariance kernel $\hat{C}$ at (2.6) in this case is given by

$$\hat{C}(s,t) = n^{-1} \sum_{i=1}^{n} \langle \mathcal{L}(\mathbf{X}_i(s)), \mathcal{L}(\mathbf{X}_i(t)) \rangle_{L_2^{\mathbb{R}}}. \tag{3.7}$$

Here, it is worthwhile to note that $\mathbf{X}_i(s)$ for a given $s \in [0,1]$ is a random density, i.e., if we write it as $X_i(\cdot, s)$, it is a random element such that $X_i(u, s) \geq 0$ and $\int_0^1 X_i(u, s)\,du = 1$ with probability one. For the kernel $\hat{C}$ at (3.7), we may also obtain its eigenfunctions $\hat{\phi}_j$ from the existing method of eigenanalysis for real-valued functional data.

Since $\mathcal{L}$ as defined at (3.6) is an isometry, injective, continuous and has continuous inverse, we may prove (3.4) as well for the space of density functions. In practical implementation, we may discretize densities on a grid of $[0,1]$, which result in functional compositional vectors. Then, we are able to apply the procedure described in Section 3.1 to density functional data. Specifically, we may choose a grid $\{u_1, \ldots, u_D\} \subset [0,1]$ and then apply (3.2) and (3.4) with

$$\mathbf{X}_i(t) = \left(\frac{X_i(u_1, t)}{\sum_{j=1}^{D} X_i(u_j, t)}, \ldots, \frac{X_i(u_D, t)}{\sum_{j=1}^{D} X_i(u_D, t)} \right)$$

as $\mathcal{S}^{D-1}$-valued random elements.

4. Real data example

Here, we illustrate the methodology we presented in Section 2. We analyzed changes in population composition by age as time passes, which is an important element in demographic analysis and a key to understanding the social dynamic. The dataset we analyzed came from the World Population Prospects 2019 (WPP, https://population.un.org/wpp/) offered by the United Nations (UN), available at https://population.un.org/wpp/Download/Files/1_Indicators%20(Standard)/CSV_FILES/WPP. The WPP2019 contains numerous data, among which we took those on population by location, age group and year. The original dataset consists of population estimates from the year 1950 to 2020, and projections from 2020 to 2100. Of these years we took the period 1950–2018. The original age groups were 0–4, 5–$9, \ldots, 95$–99, $100+$. We aggregated them into three age groups, 0–19 (before working age), 20–64 (work force), $65+$ (retired). Locations are given in several types of categories such as country, region, subregion, etc. We chose 'country' as the data unit and the number of countries was 201. Thus, we obtained compositional vectors $\mathbf{x}_i^{\mathrm{o}}(t_j) = (x_{i1}^{\mathrm{o}}(t_j), x_{i2}^{\mathrm{o}}(t_j), x_{i3}^{\mathrm{o}}(t_j))$ for $1 \leq i \leq 201$ and $t_j = 1950 + (j-1)$ for $1 \leq j \leq 69$, where x_{i1}^{o}, x_{i2}^{o}, and x_{i3}^{o} are the population proportions for the age groups, 0–19, 20–64 and $65+$, respectively, for the ith country.

To transform the dataset further into a set of smooth compositional vectors over time, we pre-smoothed $\mathbf{x}_i^{\mathrm{o}}(t_j)$ along the time scale on $[1950, 2018]$. We applied the local linear smoothing technique to $\mathcal{L}(\mathbf{x}_i^{\mathrm{o}}(t_j))$, where $\mathcal{L}$ is the center log transform introduced in Section 3.1, and then transform the results back to compositional functional vectors. Specifically, we computed

$$\mathbf{x}_i(t) = \mathcal{L}^{-1}\left(\underset{\mathbf{a} \in \mathbb{R}^3}{\arg\min} \sum_{j=1}^{69} K_h(t, t_j) \left\| \mathcal{L}(\mathbf{x}_i^{\mathrm{o}}(t_j)) - \mathbf{a} - \mathbf{b}(t_j - t) \right\|_E^2 \right), \quad t \in [1950, 2018],$$

where $K_h(t, t_j)$ with the bandwidth h are kernel weights defined by $K_h(t, t_j) = h^{-1}K((t - t_j)/h)$ for a symmetric nonnegative function K and $\|\cdot\|_E$ denotes the Euclidean norm on $\mathbb{R}^3$. We chose the Epanechnikov kernel $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$ and used $h = 2.0$.
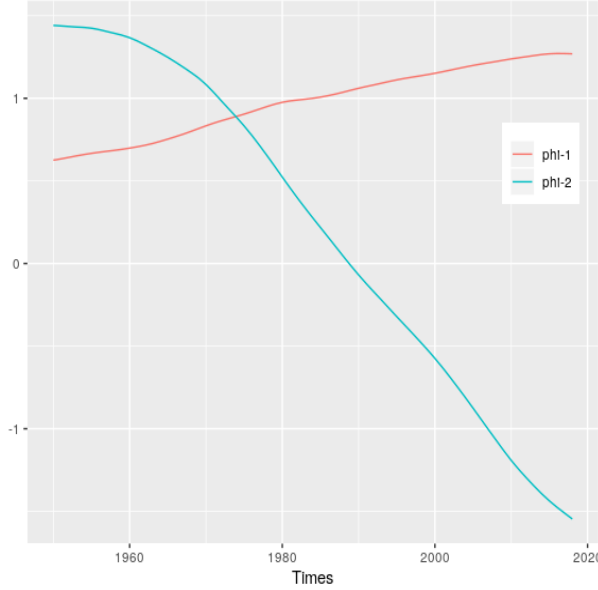
Figure 1: *First two eigenfunctions for time-varying population composition.*

Our primary interest in this example was to see how well a few number of the FPC scores can explain the apparent change in population composition over time. For this we took dominant eigenfunctions based on the criterion called 'fraction of variance explained (FVE)'. The FVE of the first J eigenfunctions is defined by $\text{FVE}_J = \sum_{j=1}^{J} \hat{\lambda}_j / \sum_{j \geq 1} \hat{\lambda}_j$. We chose eigenfunctions which altogether explained more than 90% of the variation of the sample compositional functional vectors $\{\mathbf{x}_i : 1 \leq i \leq 201\}$. It turned out that $J = 2$. In fact, the FVEs of the leading eigenfunctions were $\text{FVE}_1 = 0.899$ and $\text{FVE}_2 = 0.976$. Figure 1 depicts the first two eigenfunctions $\hat{\phi}_1$ and $\hat{\phi}_2$. They represent the principal deviations from the average population composition over time. The first eigenfunction has an increasing trend. This indicates that variability between population composition patterns amongst different countries is mostly explained by increased separation from the mean composition as time passed. This might have something to do with increased variability in recent years among 201 countries in birth rate and human mortality.

To illustrate how well the two leading FPC scores can reproduce the original data, we computed $\hat{\boldsymbol{\xi}}_{ij}$ according to the formula (3.4) and reconstructed $\mathbf{X}_i$ as

$$\hat{\mathbf{X}}_i(t) = \hat{\boldsymbol{\xi}}_{i1} \odot \hat{\phi}_1(t) \oplus \hat{\boldsymbol{\xi}}_{i2} \odot \hat{\phi}_2(t).$$

Figure 2 compares the original $\mathbf{X}_i$ and the reconstructed $\hat{\mathbf{X}}_i$ for Republic of Korea. We find that the reconstruction is quite successful, meaning that the first two FPC scores capture most of time-dynamic features in the population composition of the country. This is also the case with other countries, although we do not present them all here.

We performed a cluster analysis based on the two leading FPC scores. We formed three clusters based on the following distance metric between $\mathbf{x}_i$ and $\mathbf{x}_l$ for $1 \leq i \neq l < 201$.

$$d(\mathbf{x}_i, \mathbf{x}_l) = \sqrt{\left\| \hat{\boldsymbol{\xi}}_{i1} \ominus \hat{\boldsymbol{\xi}}_{l1} \right\|_{\mathcal{S}^2}^2 + \left\| \hat{\boldsymbol{\xi}}_{i2} \ominus \hat{\boldsymbol{\xi}}_{l2} \right\|_{\mathcal{S}^2}^2} = \sqrt{\left\| \mathcal{L}\left(\hat{\boldsymbol{\xi}}_{i1}\right) - \mathcal{L}\left(\hat{\boldsymbol{\xi}}_{l1}\right) \right\|_E^2 + \left\| \mathcal{L}\left(\hat{\boldsymbol{\xi}}_{i2}\right) - \mathcal{L}\left(\hat{\boldsymbol{\xi}}_{l2}\right) \right\|_E^2}.$$

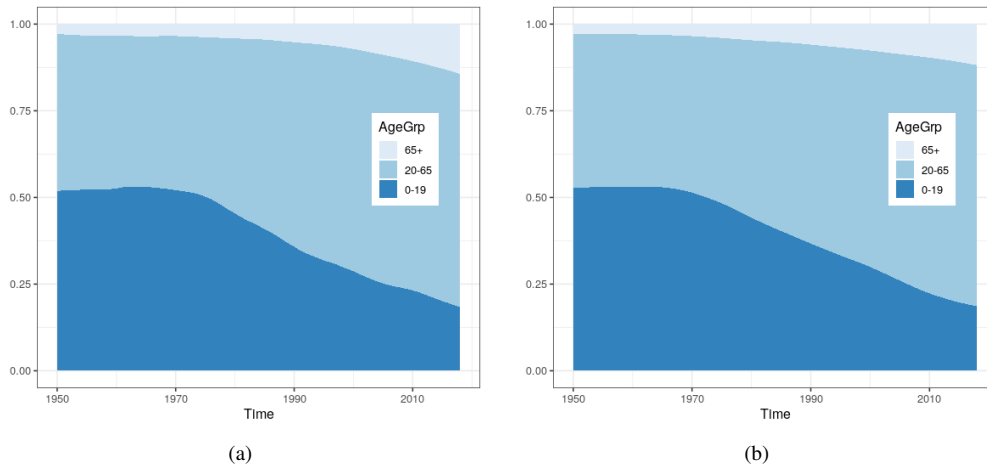(a)   (b)

Figure 2: *Pre-smoothed compositional vector over time (a) and its reconstruction based on two leading eigen-functions (b), for Republic of Korea.*
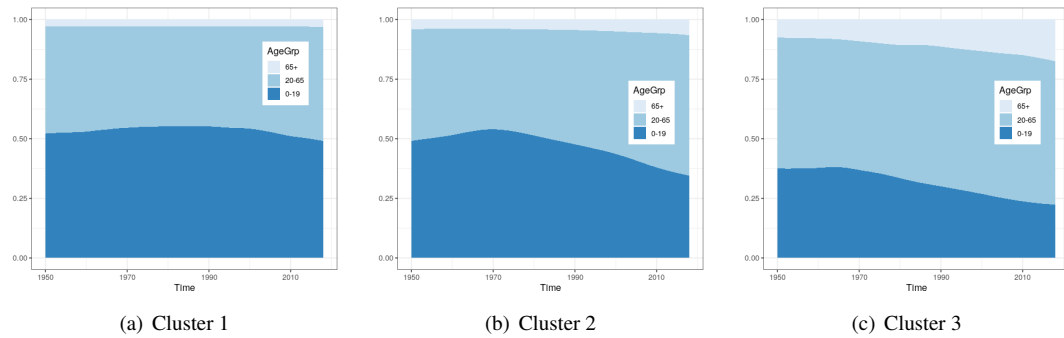

(a) Cluster 1   (b) Cluster 2   (c) Cluster 3

Figure 3: *Mean compositional vectors over time.*

We employed hierarchical clustering based on complete linkage, which takes

$$d(C_j, C_k) = \max_{\mathbf{x} \in C_j, \mathbf{y} \in C_k} d(\mathbf{x}, \mathbf{y}).$$

as the distance between two clusters C_j and C_k. Figure 3 depicts the average compositional vectors over time for each cluster. The figure demonstrates the central features of the clusters. The cluster C_1 consists of countries where population composition does not change much over time. The proportion of the retired ages in the cluster is almost constant over time. Those countries in the clusters C_2 and C_3 experienced some level of decrease in the proportion of the young ages since near 1970, but the proportion in C_2 increased until 1970 while in C_3 they stayed constant until then. Also, both C_2 and C_3 saw some level of increase in the proportion of the retired ages, but C_3 started from a relatively larger elderly proportion and underwent more rapid increase than C_2.

We also compared our clustering result with the UN development groups obtained from the WPP2019. We took the three groups from the WPP2019: countries in more developed regions, least developed countries, and other less developed countries. More developed regions comprise Europe, Northern America, Australia/New Zealand and Japan. Less developed regions comprise all regions of

Table 1: Contingency table for the counts of countries

	Least developed	Other less developed	More developed	Total
Cluster 1	39	17	0	56
Cluster 2	7	80	2	89
Cluster 3	0	13	43	56
Total	46	110	45	201

Africa, Asia (except Japan), Latin America and the Caribbean plus Melanesia, Micronesia and Polynesia. Countries in less developed regions are divided further into two groups, one is the group of 47 least developed countries and the other consists of countries in less developed regions excluding the least developed ones. The criterion for least developed countries can be found at http://unohrlls.org/about-ldcs/criteria-for-ldcs/. Table 1 is a contingency table comparing the two clusterings. We find that the clusters based on the compositional FPC are very close to the UN development groups.

5. Multivariate extension

In this section, we discuss a multivariate extension of the methodology presented in Section 2. Suppose that we now have $\mathbf{X} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(d)})^\top$ where each $\mathbf{X}^{(j)}$ is a $\mathbb{H}$-valued stochastic process with $E \int_0^1 \|\mathbf{X}^{(j)}(t)\|_{\mathbb{H}}^2 \, dt < \infty$. Thus, $\mathbf{X}(t)$ for each $t \in [0, 1]$ takes values in $\mathbb{H}^d$. Without loss of generality we also assume $E\,\mathbf{X}(t) = \mathbf{0}_d$ for all $t \in [0, 1]$, where $\mathbf{0}_d$ is the zero vector in $\mathbb{H}^d$.

To accommodate the dependency between $\mathbf{X}^{(j)}$ and $\mathbf{X}^{(k)}$ for $j \neq k$, we consider the following $d \times d$ matrix of covariance functions:

$$\mathbf{C}(s, t) = \left(C_{jk}(s, t) \right), \quad C_{jk}(s, t) = E\left(\left\langle \mathbf{X}^{(j)}(s), \mathbf{X}^{(k)}(t) \right\rangle_{\mathbb{H}} \right).$$

Define the integral operator $\mathscr{C}^{\otimes} : L_2^{\mathbb{R}^d} \to L_2^{\mathbb{R}^d}$ such that for $\mathbf{f} \in L_2^{\mathbb{R}^d}$

$$\mathscr{C}^{\otimes}(\mathbf{f})(s) = \int_0^1 \mathbf{C}(s, t)\mathbf{f}(t) \, dt, \quad s \in [0, 1].$$

We endow $L_2^{\mathbb{R}^d}$ with the metric $\langle \mathbf{f}, \mathbf{g} \rangle_{L_2^{\mathbb{R}^d}} = \sum_{j=1}^d \langle f_j, g_j \rangle_{L_2^{\mathbb{R}}}$. We may prove that for all $\mathbf{f} \in L_2^{\mathbb{R}^d}$

$$\langle \mathscr{C}^{\otimes}(\mathbf{f}), \mathbf{f} \rangle_{L_2^{\mathbb{R}^d}} = E \left\| \bigoplus_{j=1}^d \int_0^1 f_j(t) \odot \mathbf{X}^{(j)}(t) \, dt \right\|_{\mathbb{H}}^2 \geq 0.$$

Thus, the integral operator $\mathscr{C}^{\otimes}$ is nonnegative definite. Also, its kernel $\mathbf{C}$ is symmetric, i.e., $\mathbf{C}(s, t) = \mathbf{C}(t, s)$ for all $s, t \in [0, 1]$. Thus, due to Mercer's Theorem,

$$\mathbf{C}(s, t) = \sum_{j=1}^{\infty} \lambda_j \boldsymbol{\phi}_j(s) \boldsymbol{\phi}_j(t)^\top,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\{\boldsymbol{\phi}_j\}_{l=1}^{\infty}$ are the associated eigenfunctions of $\mathscr{C}^{\otimes}$ which form a complete orthonormal system for the space $L_2^{\mathbb{R}^d}$.

Let ϕ_{jk} denote the kth entry of $\boldsymbol{\phi}_j$ and $\{\mathbf{e}_l\}_{l=1}^{L}$ be an orthonormal basis of $\mathbb{H}$. Consider the d-variate real-valued random functions $\mathbf{Z}_l(t) = (\langle \mathbf{X}^{(1)}(t), \mathbf{e}_l \rangle_{\mathbb{H}}, \ldots, \langle \mathbf{X}^{(d)}(t), \mathbf{e}_l \rangle_{\mathbb{H}})^\top$ for $1 \leq l \leq L$. Then, we may

write

$$\mathbf{Z}_l(t) = \sum_{j=1}^{\infty} \langle \mathbf{Z}_l, \boldsymbol{\phi}_j \rangle_{L_2^{\mathbb{R}^d}} \cdot \boldsymbol{\phi}_j(t)$$

$$= \sum_{j=1}^{\infty} \left(\sum_{k=1}^{d} \left\langle \int_0^1 \mathbf{X}^{(k)}(s) \odot \phi_{jk}(s)\,ds, \mathbf{e}_l \right\rangle_{\mathbb{H}} \right) \cdot \boldsymbol{\phi}_j(t). \tag{5.1}$$

Write $Z_{lk}(t) = \langle \mathbf{X}^{(k)}(t), \mathbf{e}_l \rangle_{\mathbb{H}}$ for the kth entry of $\mathbf{Z}_l(t)$. Plugging the expression at (5.1) into the right hand side of the equality $\mathbf{X}^{(k)}(t) = \bigoplus_{l=1}^{L} Z_{lk}(t) \odot \mathbf{e}_l$, we get the following Karhunen-Loève expansion of $\mathbf{X}$:

$$\mathbf{X}(t) = \bigoplus_{j=1}^{\infty} \left(\bigoplus_{k=1}^{d} \int_0^1 \mathbf{X}^{(k)}(s) \odot \phi_{jk}(s)\,ds \right) \odot \boldsymbol{\phi}_j(t).$$

The corresponding FPC scores in the above eigenanalysis are

$$\boldsymbol{\xi}_j = \bigoplus_{k=1}^{d} \int_0^1 \mathbf{X}^{(k)}(s) \odot \phi_{jk}(s)\,ds, \quad j \geq 1.$$

The FPC scores reside in $\mathbb{H}$, not in $\mathbb{H}^d$, and the eigenfunctions $\boldsymbol{\phi}_j$ lie in $L_2^{\mathbb{R}^d}$. As in the case of $d = 1$ we treated in Section 2, $\boldsymbol{\xi}_j$ are uncorrelated in the sense that $E\left(\langle \boldsymbol{\xi}_j, \boldsymbol{\xi}_k \rangle_{\mathbb{H}} \right) = 0$ for $j \neq k$ and $E\|\boldsymbol{\xi}_j\|_{\mathbb{H}}^2 = \lambda_j$. Here, we have used the fact that

$$\|\boldsymbol{\phi}_j\|_{L_2^{\mathbb{R}^d}}^2 = \sum_{l=1}^{d} \int_0^1 \phi_{jl}(t)^2\,dt = 1, \quad j \geq 1,$$

$$\langle \boldsymbol{\phi}_j, \boldsymbol{\phi}_k \rangle_{L_2^{\mathbb{R}^d}}^2 = \sum_{l=1}^{d} \int_0^1 \phi_{jl}(t)\phi_{kl}(t)\,dt = 0, \quad j \neq k \geq 1.$$

The following theorem is a multivariate version of Theorem 1. In the theorem, $\mathbf{h} \odot \mathbf{c}$ for $\mathbf{h} \in \mathbb{H}$ and $\mathbf{c} = (c_1, \ldots, c_d)^\top \in \mathbb{R}^d$ means $(c_1 \odot \mathbf{h}, \ldots, c_d \odot \mathbf{h})^\top \in \mathbb{H}^d$.

Theorem 3. *Assume $E \int_0^1 \|\mathbf{X}^{(j)}(t)\|_{\mathbb{H}}^2\,dt < \infty$ for all $1 \leq j \leq d$. Then, it holds that $\mathbf{X}(t) = \bigoplus_{j=1}^{\infty} \boldsymbol{\xi}_j \odot \boldsymbol{\phi}_j(t)$ with probability one.*

We may follow the procedure we described in Section 2.2 to estimate the eigenfunctions and the FPC scores. Let $\mathbf{X}_i$, $1 \leq i \leq n$, be independent copies of $\mathbf{X}$. Without loss of generality we assume that $n^{-1} \bigoplus_{i=1}^{n} \mathbf{X}_i(t) = \mathbf{0}$. Then, we may estimate $\mathbf{C}(s, t)$ by

$$\hat{\mathbf{C}}(s, t) = \left(\hat{C}_{jk}(s, t) \right), \quad \hat{C}_{jk}(s, t) = n^{-1} \sum_{i=1}^{n} \left\langle \mathbf{X}_i^{(j)}(s), \mathbf{X}_i^{(k)}(t) \right\rangle_{\mathbb{H}}. \tag{5.2}$$

This would give the eigendecomposition $\hat{\mathbf{C}}(s, t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\boldsymbol{\phi}}_j(s) \hat{\boldsymbol{\phi}}_j(t)$. We take $\hat{\lambda}_j$ as the estimators of λ_j and $\hat{\boldsymbol{\phi}}_j$ as the estimators of the $\boldsymbol{\phi}_j$. We also estimate the FPC scores of $\mathbf{X}_i$ by

$$\hat{\boldsymbol{\xi}}_{ij} = \bigoplus_{k=1}^{d} \int_0^1 \mathbf{X}_i^{(k)}(s) \odot \hat{\phi}_{jk}(s)\,ds, \quad j \geq 1, \ 1 \leq i \leq n.$$

Then, as in the case of $d = 1$, we have

$$n^{-1} \sum_{i=1}^{n} \left\langle \hat{\boldsymbol{\xi}}_{ij}, \hat{\boldsymbol{\xi}}_{ik} \right\rangle_{\mathbb{H}} = 0, \quad j \neq k \geq 1, \qquad n^{-1} \sum_{i=1}^{n} \left\| \hat{\boldsymbol{\xi}}_{ij} \right\|_{\mathbb{H}}^{2} = \hat{\lambda}_j, \quad j \geq 1.$$

For these estimators of eigenfunctions and FPC scores, we may also obtain an analogue of Theorem 2.

Acknowledgements

References

Aitchison J (1986). *The Statistical Analysis of Compositional Data, Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Bosq D (2000). *Linear Processes in Function Spaces: Theory and Applications*, Springer, New York.

Chiou JM, Chen YT, and Yang YF (2014). Multivariate functional principal component analysis: a normalization approach, *Statistica Sinica*, **24**, 1571–1596.

Dai X and Müller HG (2018). Principal component analysis for functional data on Riemannian manifolds and spheres, *Annals of Statistics*, **46**, 3334–3361.

Hsing T and Eubank R (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley, Singapore.

Jeon JM and Park BU (2020). Additive regression with Hilbertian responses, *Annals of Statistics*, in print.

Lin Z and Yao F (2019). Intrinsic Riemannian functional data analysis, *Annals of Statistics*, **47**, 3533–3577.

Ramsay J and Silverman BW (2005). *Functional Data Analysis*, Springer, New York.

van den Boogaart KG, Egozcue JJ, and Pawlowsky-Glahn V (2014). Bayes Hilbert spaces, *Australian and New Zealand Journal of Statistics*, **56**, 171–194.

Wang H, Shangguan L, Guan R, and Billard L (2015). Principal component analysis for compositional data vectors, *Computational Statistics*, **30**, 1079–1096.