On the usage of the Gram matrix for the multivariate functional principal components analysis

Steven Golovkine* Edward Gunning† Andrew J. Simpkin‡
Norma Bargary§

March 28, 2023

Abstract

Dimension reduction has a major importance in functional data analysis (FDA). The key tool to reduce the dimension of the data is functional principal component analysis. Existing approaches for functional principal component analysis usually involve the diagonalisation of the covariance operator. Using the duality of the space of observations and the space of functional components, we propose to use the inner-product between the curves to estimate the eigenelements of, eventually multivariate and multidimensional, functional datasets. The relationship between the eigenelements of the covariance operator and the ones of the inner-product matrix is established. We explore the application of these methodologies in several FDA settings and provide general guidance on their usability.

Keywords— Dimension Reduction; Functional Data Analysis; Functional Principal Components; Multivariate Functional Data

1 Introduction

Functional data analysis (FDA) is a statistical methodology for analyzing data that can be represented as functions. These functions could represent measurements taken over time or space, such as temperature readings over a period of time or spatial patterns of disease occurrence. The goal of FDA is to extract meaningful information from these functions and to model their behavior. See, e.g., Ramsay and Silverman (2005); Horvàth and Kokoszka (2012); Wang et al. (2016); Kokoszka et al. (2017) for some recent references on functional data analysis.

Functional principal component analysis (FPCA) is an extension of principal component analysis (PCA), which is a commonly used tool for dimension reduction in multivariate data, to functional data. This tool was introduced by Karhunen (1947) and Loève (1945) and developed by Dauxois et al. (1982). Since then, FPCA has become a prevalent tool in FDA due to its ability to convert infinite-dimensional functional data into a finite-dimensional vector of random scores. These scores are a countable sequence of uncorrelated random variables that can be truncated to a finite vector in practical applications. By applying multivariate data analysis tools to these random scores, FPCA can achieve the goal of dimension reduction while assuming mild assumptions about the underlying stochastic process. FPCA is usually used as a preprocessing step to feed regression and classification models. Recently, FPCA has been extended to the analysis of multivariate functional data, which are data that consist of multiple functions that are observed simultaneously. This extension is referred to as multivariate functional data analysis (MFPCA). As for FPCA, a key benefit of MFPCA is that it allows to identify and visualize the main sources of variation in the multivariate functional data. This can be useful in many different applications, such as identifying patterns of movements in sport biomechanics (Warmenhoven et al., 2019), analyzing changes in brain activity in neuroscience (Song and Kim, 2022), or comparing countries' competitiveness in economics (Krzyśko et al., 2022).

In MFPCA, we seek to decompose the covariance structure of the multivariate functional data into a set of orthogonal basis functions, named the principal components, which capture the main

*MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland steven.golovkine@ul.ie
†MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland edward.gunning@ul.ie
‡School of Mathematical and Statistical Sciences, University of Galway, Ireland andrew.simpkin@nuigalway.ie
§MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland norma.bargary@ul.ie

sources of variation in the data. There are multiple approaches to estimate the principal components of a multivariate functional dataset. Ramsay and Silverman (2005) stack the multivariate curves into one big curve and then perform an usual FPCA by doing an eigendecomposition of the covariance structure. This methodology can only be run for data that are defined on the same unidimensional domain and that exhibit similar variations. Jacques and Preda (2014) propose to expand each univariate component into a basis of functions. It results in a set of coefficients for each univariate curve. The eigendecomposition is then run on the matrix of the stacked coefficients. To consider the normalization issue of Ramsay and Silverman (2005), Jacques and Preda (2014) and Chiou et al. (2014) propose to normalize the data by the standard deviation of the curves at each sampling points. Happ and Greven (2015) extend the estimation to functional data defined on different dimensional domains. The estimation is based on FPCA for each univariate component and use a weighted combinaison of them to obtain the multivariate eigencomponents. Finally, Berrendero et al. (2011) develop a very different method to estimate the eigencomponents as they perform a principal components analysis for each sampling time points.

The key motivation of the paper is to investigate the duality between rows and columns of a data matrix to estimate the eigencomponents of a multivariate functional dataset. The duality between rows and columns of a data matrix is a fundamental concept in classical statistics (Escofier, 1979; Saporta, 1990). A data matrix typically represents a set of observations of multiple features, each row corresponds to an individual observation and each column correspond to an individual feature. The duality between rows and columns refers to the fact that many statistical methodologies can be conducted either on the rows of the columns of the data matrix, and the results will be related to each other. For example, the principal components obtained from a PCA run on the rows of the data matrix can be used to compute PCA on the columns of the matrix. The choice of method to use, based on criteria such as computational time or data storage, is thus left to the statistician. This concept has been widely studied for multivariate statistics (see, e.g., Pagès (2014); Härdle and Simar (2019)). In the context of functional data, this principle has received limited attention despite being mentioned in the seminal paper of FDA (Ramsay, 1982). Ramsay and Silverman (2005) briefly commented on it in a concluding remark of Chapter 8. Kneip and Utikal (2001); Benko et al. (2009) utilized it to compute principal components for dense univariate functional data. Chen et al. (2017) also mention it to gain computational advantage when univariate functional data are sampled on a very dense grid. To the best of our knowledge, however, there is no available literature on its application to multivariate functional data that are observed on different dimensional domains. Our aim is therefore to investigate this duality for multivariate functional data observed on different dimensional domains and provide guidelines to statisticians on which method to use in different cases.

The remainder of the paper is organized as follows. In Section 2, we define multivariate functional data with the coordinates possibly having different definition domains. In Section 3, we develop the duality between the observations' space and the functional components space. The relationship between the eigencomponents of the covariance operator of the multivariate functional datasets and the eigencomponents of the inner-product matrix between the observations is derived in Section 4. Extensive simulations are given in Section 5. We also provide guidelines on which method to use with respect to data characteristics. The paper concludes with a discussion and an outlook in Section 6.

2 Model

The structure of the data we consider, referred to as *multivariate functional data*, is very similar to that presented in Happ and Greven (2015). The data consist of independent trajectories of a vector-valued stochastic process $X = (X^{(1)}, \ldots, X^{(P)})^\top$, $P \geq 1$. (Here and in the following, for any matrix A, $A^\top$ denotes its transpose.) For each $1 \leq p \leq P$, let $\mathcal{T}_p$ be a rectangle in some Euclidean space $\mathbb{R}^{d_p}$ with $d_p \geq 1$, *e.g.*, $\mathcal{T}_p = [0,1]^{d_p}$. Each coordinate $X^{(p)} : \mathcal{T}_p \to \mathbb{R}$ is assumed to belong to $\mathcal{L}^2(\mathcal{T}_p)$, the Hilbert space of squared-integrable real-valued functions defined on $\mathcal{T}_p$, having the usual inner product that we denote by $\langle \cdot, \cdot \rangle$, and $\|\cdot\|$ the associated norm. Thus X is a stochastic process indexed by $\mathbf{t} = (t_1, \ldots, t_P)$ belonging to the $P-$fold Cartesian product $\mathcal{T} := \mathcal{T}_1 \times \cdots \times \mathcal{T}_P$ and taking values in the $P-$fold Cartesian product space $\mathcal{H} := \mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_P)$.

We consider the function $\langle\!\langle \cdot, \cdot \rangle\!\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$,

$$\langle\!\langle f, g \rangle\!\rangle := \sum_{p=1}^{P} \langle f^{(p)}, g^{(p)} \rangle = \sum_{p=1}^{P} \int_{\mathcal{T}_p} f^{(p)}(t_p) g^{(p)}(t_p) \mathrm{d}t_p, \quad f, g \in \mathcal{H}.$$

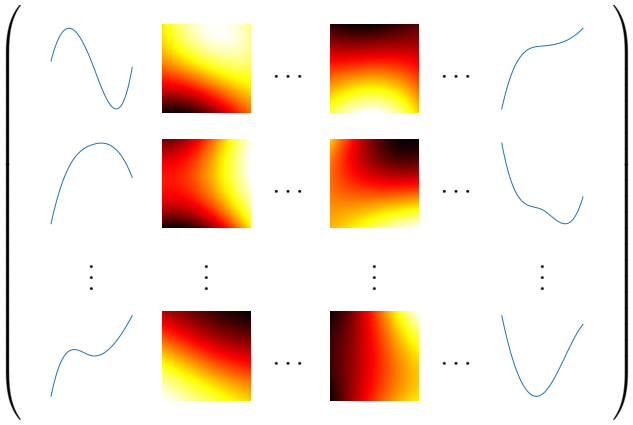

Figure 1: Functional data matrix, adapted from Berrendero et al. (2011).

$\mathcal{H}$ is a Hilbert space with respect to the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$(Happ and Greven, 2015). We denote by $|\!|\!| \cdot |\!|\!|$, the norm induced by $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Let $\mu : \mathcal{T} \to \mathcal{H}$ denote the mean function of the process X, $\mu(\mathbf{t}) \coloneqq \mathbb{E}(X(\mathbf{t})), \mathbf{t} \in \mathcal{T}$. Let C denote the $P \times P$ matrix-valued covariance function which, for $\mathbf{s}, \mathbf{t} \in \mathcal{T}$, is defined as

$$C(\mathbf{s}, \mathbf{t}) \coloneqq \mathbb{E}\left(\{X(\mathbf{s}) - \mu(\mathbf{s})\}\{X(\mathbf{t}) - \mu(\mathbf{t})\}^\top\right), \quad \mathbf{s}, \mathbf{t} \in \mathcal{T}.$$

More precisely, for $1 \leq p, q \leq P$, the (p, q)th entry of the matrix $C(\mathbf{s}, \mathbf{t})$ is the covariance function between the pth and the qth components of the process X:

$$C_{p,q}(s_p, t_q) \coloneqq \mathbb{E}\left(\{X^{(p)}(s_p) - \mu^{(p)}(s_p)\}\{X^{(q)}(t_q) - \mu^{(q)}(t_q)\}\right), \quad s_p \in \mathcal{T}_p, t_q \in \mathcal{T}_q.$$

Let $\Gamma : \mathcal{H} \to \mathcal{H}$ denote the covariance operator of X, defined as an integral operator with kernel C. That is, for $f \in \mathcal{H}$ and $\mathbf{t} \in \mathcal{T}$, the pth component of $\Gamma f(\mathbf{t})$ is given by

$$(\Gamma f)^{(p)}(t_p) \coloneqq \langle\!\langle C_{p,\cdot}(t_p, \cdot), f(\cdot) \rangle\!\rangle = \langle\!\langle C_{\cdot,p}(\cdot, t_p), f(\cdot) \rangle\!\rangle, \quad t_p \in \mathcal{T}_p.$$

Let us consider a set of N curves $\mathcal{X} = \{X_1, \ldots, X_n, \ldots, X_N\}$ generated as a random sample of the P-dimensional stochastic process X with continuous trajectories. Unless otherwise stated, the data are assumed to be observed without error. The data can be viewed as a table with N rows and P columns where each entry is a curve, potentially on a multidimensional domain (see Figure 1). Each row of this matrix represents an observation; while each column represents a functional variable. At the intersection of row n and column p, we thus have $X_n^{(p)}$ which is the curve that concerns the (functional) feature p for the individual n.

For $n \in \{1, \ldots, N\}$, each observation n is attributed the weight π_n such that $\sum_n \pi_n = 1$, e.g., $\pi_n = 1/N$. For a given $p \in \{1, \ldots, P\}$, the mean curve of the pth component along the N observations is denoted by $\mu^{(p)}$. This quantity can be computed using

$$\mu^{(p)}(t_p) = \sum_{n=1}^{N} \pi_n X_n^{(p)}(t_p), \quad t_p \in \mathcal{T}_p, \quad p \in \{1, \ldots, P\}.$$

The covariance function of the pth component along the N observations can be computed using

$$C_{p,p}(s_p, t_p) = \sum_{n=1}^{N} \pi_n X_n^{(p)}(s_p) X_n^{(p)}(t_p) - \mu^{(p)}(s_p)\mu^{(p)}(t_p), \quad s_p, t_p \in \mathcal{T}_p, \quad p \in \{1, \ldots, P\}. \tag{1}$$

2.1 Basis decomposition

In many practical situations, data are noisy and only observed at specific time points. To extract the underlying functional features of the data, smoothing and interpolation techniques are commonly employed. These techniques involve approximating the true curve of the data by finite-dimensional set of basis functions. Assume that for each component $p = 1, \ldots, P$, there exists a set of basis of functions $\Psi^{(p)} = \{\psi_k^{(p)}\}_{1 \leq k \leq K_p}$ such that each component of each curve $n = 1, \ldots, N$ can be expanded into this basis:

$$X_n^{(p)}(t_p) = \sum_{k=1}^{K_p} c_{nk}^{(p)} \psi_k^{(p)}(t_p), \quad t_p \in \mathcal{T}_p,$$

where $\{c_{nk}^{(p)}\}_{1\leq k\leq K_p}$ is a set of coefficients for the component p of the curve n. We denote by $\overline{c}_k^{(p)} = \sum_{n=1}^{N} \pi_n c_{nk}^{(p)}$ the mean coefficient of the component p corresponding to the kth basis function along the N curves. The pth component of the mean function can be then expanded in the same basis as:

$$\mu^{(p)}(t_p) = \sum_{k=1}^{K_p} \overline{c}_k^{(p)} \psi_k^{(p)}(t_p), \quad t_p \in \mathcal{T}_p.$$

In a similar way, the covariance function of the pth component is given by:

$$C_{p,p}(s_p, t_p) = \sum_{k=1}^{K_p}\sum_{l=1}^{K_p} \left(\sum_{n=1}^{N} \pi_n c_{nk}^{(p)} c_{nl}^{(p)} - \overline{c}_k^{(p)} \overline{c}_l^{(p)} \right) \psi_k^{(p)}(s_p)\psi_l^{(p)}(t_p), \quad s_p, t_p \in \mathcal{T}_p.$$

These formula can be written with a matrix formulation. For $\mathbf{t} \in \mathcal{T}$, we have that $X(\mathbf{t}) = C\Psi(\mathbf{t})$ where $X(\mathbf{t})$ is a matrix with N rows and P columns with entries $X_n^{(p)}(t_p),\ t_p \in \mathcal{T}_p,\ 1 \leq p \leq P,\ 1 \leq n \leq N$,

$$C = \begin{pmatrix} C^{(1)} & \cdots & C^{(P)} \end{pmatrix}, \quad \text{and} \quad \Psi(\mathbf{t}) = \operatorname{diag}\{\Psi^{(1)}(t_1), \ldots, \Psi^{(P)}(t_P)\},$$

where

$$C^{(p)} = \begin{pmatrix} c_{11}^{(p)} & \cdots & c_{1K_p}^{(p)} \\ \vdots & \ddots & \vdots \\ c_{N1}^{(p)} & \cdots & c_{NK_p}^{(p)} \end{pmatrix} \quad \text{and} \quad \Psi^{(p)}(t_p) = \begin{pmatrix} \psi_1^{(p)}(t_p) \\ \vdots \\ \psi_{K_p}^{(p)}(t_p) \end{pmatrix}.$$

Using the basis expansion and denoting $\Pi^\top = (\pi_1, \ldots, \pi_N)$, the mean and covariance functions are given by

$$\mu(\mathbf{t}) = \Psi(\mathbf{t})^\top C^\top \Pi \quad \text{and} \quad C(\mathbf{s}, \mathbf{t}) = \Psi(\mathbf{s})^\top C^\top \left(\operatorname{diag}\{\pi_1, \ldots, \pi_N\} - \Pi\Pi^\top\right) C\Psi(\mathbf{t}).$$

Finally, we denote by W the matrix of inner products of the functions in the basis Ψ. The matrix W is a block-diagonal matrix such that $W = \operatorname{diag}\{W^{(1)}, \ldots, W^{(P)}\}$ where each entry is given by

$$W_{k,l}^{(p)} = \langle \psi_k^{(p)}, \psi_l^{(p)} \rangle, \quad 1 \leq k, l \leq K_p, \quad 1 \leq p \leq P.$$

We remark that, if the basis Ψ is an orthonormal basis, the matrix W is equal to the identity matrix of size $\sum_{p=1}^{P} K_p$.

3 On the geometry of multivariate functional data

3.1 Cloud of individuals

Given $n \in \{1, \ldots, N\}$, let $\{X_n^{(p)}(t_p),\ t_p \in \mathcal{T}_p,\ p = 1, \ldots, P\}$ be the features set for a particular observation n. We identify this set as the point M_n in the space $\mathcal{H}$. The space $\mathcal{H}$ is refered as the observations' space. The cloud of points that represents the set of observations is denoted by $\mathcal{C}_N$. Let G_N be the centre of gravity of the cloud $\mathcal{C}_N$. In the space $\mathcal{H}$, its coordinates are given by $\{\mu^{(p)}(t_p),\ t_p \in \mathcal{T}_p,\ p = 1, \ldots, P\}$. If the variables are centered, the origin $\mathrm{O}_{\mathcal{H}}$ of the axes in $\mathcal{H}$ coincides with G_N.

Let f and g be two elements in $\mathcal{H}$ and denote by M_f and M_g their associated points in $\mathcal{C}_N$ (see Figure 2). The distance between these observations is defined as

$$d^2(f, g) = \|p - g\|^2 = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\{ f^{(p)}(t_p) - g^{(p)}(t_p) \right\}^2 \mathrm{d}t_p.$$

This distance measures how different the observations are, and thus characterized the shape of the cloud $\mathcal{C}_N$. Another description of this shape is to consider the distance between each observation and G_N, the center of the cloud. Let f be an element of $\mathcal{H}$, associated to the point M_f, and μ the element of $\mathcal{H}$ related to G_N, the distance between f and μ is given by

$$d^2(f, \mu) = \|f - \mu\|^2 = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\{ f^{(p)}(t_p) - \mu^{(p)}(t_p) \right\}^2 \mathrm{d}t_p.$$

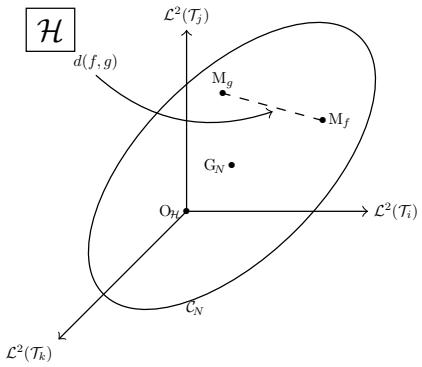

Figure 2: Cloud of observations. The observation f (resp. g) is identified by the point M_f (resp. M_g) in the cloud $\mathcal{C}_N$. The point G_N is the center of gravity of $\mathcal{C}_N$ and the point $\mathrm{O}_\mathcal{H}$ is the origin of the space $\mathcal{H}$.

Given the set $\mathcal{X}$, the total inertia of $\mathcal{C}_N$, with respect to G_N, is given by

$$\sum_{n=1}^{N} \pi_n d^2(X_n, \mu) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j d^2(X_i, X_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \operatorname{Var} X^{(p)}(t_p) \mathrm{d}t_p.$$

The derivation of these equalities are given in Appendix A.

Remark 1. *These results have the same interpretation as for multivariate scalar data. This is also the multivariate analogue of the relation between variance and sum of squared differences known for univariate functional data. If the variables are reduced beforehand, the total inertia of the cloud $\mathcal{C}_N$ is equal to the number of components P. We are, in general, not interested by the total inertia but mostly how this variance is spread among the variables.*

3.2 Cloud of variables

For now, we need that all the variables are defined on the same space $\mathcal{T}_0$.

Let $\{X_n^{(p)}(t), n = 1, \ldots, N\}$ be the observations set for a particular feature p. We identify this set as the point M_p in the space $\mathcal{G} \coloneqq \mathcal{L}^2(\mathcal{T}_0)^N$. The set $\mathcal{G}$ is refered as the features' space, or variables' space. The cloud of points that represented the set of variables is denoted by $\mathcal{C}_P$. Let $\mathcal{O}_\mathcal{G}$ be the centre of this space. Its coordinates are given by a vector of functions of length N where each entry is $f(t) = 0$ for all $t \in \mathcal{T}_0$.

We assume that the observations are centered. Consider M_h a point in $\mathcal{C}_P$ and h the element of $\mathcal{G}$ representing by M_h. Let o be the element of $\mathcal{O}_\mathcal{G}$ representing by $\mathcal{O}_\mathcal{G}$. The distance between M_h and $\mathcal{O}_\mathcal{G}$ is defined as

$$d^2(h, o) = \sum_{n=1}^{N} p_n \left\| h_n - \mu_h \right\|^2 = \int_{\mathcal{T}_0} \operatorname{Var} h^{(p)}(t) dt.$$

For now, the cloud of variables is not clear!

3.3 On centering and reducing

Take a look at Prothero et al. (2021)

In MFPCA, the components are usually assumed centred Happ and Greven (2015). If curves are not centered, we replace $X_n^{(p)}(t)$ by $X_n^{(p)}(t) - \mu^{(p)}(t)$. The geometric interpretation of the centering differs if we refer to $\mathcal{H}$ or $\mathcal{G}$. Within the space $\mathcal{H}$, centering is interpreted as translating the centre of gravity of the curves G_N to the the origin point $\mathrm{O}_\mathcal{H}$ of $\mathcal{H}$. This transformation, being a translation, does not change the shape of the cloud $\mathcal{C}_N$. Within the space $\mathcal{G}$, the centering is harder to interpret and has not the same meaning as in the multivariate case (projection on the subspace orthogonal to the constant vector).

Remark 2. *What happened if we project the multivariate curves onto the vector of constant functions? In the space $\mathcal{G}$, the inner product is given by*

$$\langle\!\langle f, g \rangle\!\rangle = \sum_{i=1}^{N} \int_{\mathcal{T}_k} f_i^{(k)}(t) g_i^{(k)}(t) dt, \quad f, g \in \mathcal{G}.$$

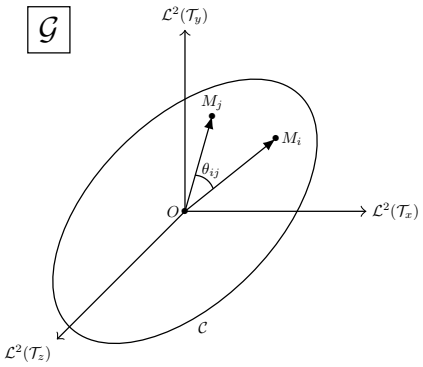

Figure 3: Cloud of features.

Let $\mathbf{1}$ *be the vector of constant function in* $\mathcal{G}$ *and* f *an element of* $\mathcal{G}$*. Then, the projection of* f *onto* $\mathbf{1}$ *is given by*

$$P_{\mathbf{1}} f = \frac{\langle\langle f, \mathbf{1} \rangle\rangle}{\|\mathbf{1}\|} \mathbf{1} = \frac{1}{N|\mathcal{T}_k|} \sum_{i=1}^{N} \int_{\mathcal{T}_k} f_i^{(k)}(t) dt \mathbf{1}$$

In practice, this is equivalent to compute the mean value of the mean curve for each component.

Concerning standardisation, there are different proposal in the literature. Happ and Greven (2015) propose to use $w_k = (\int_{\mathcal{T}_k} \operatorname{Var} X^{(k)}(t) dt)^{-1/2}$, while Chiou et al. (2014) consider the function $w_k(t) = (\operatorname{Var} X^{(k)}(t))^{-1/2}$.

4 Multivariate functional principal components analysis

Assuming that the covariance operator Γ is a compact positive operator on $\mathcal{H}$ and using the results in Happ and Greven (2015), and the theory of Hilbert-Schmidt operators, e.g., Reed and Simon (1980), there exists a complete orthonormal basis $\Phi = \{\phi_k\}_{k \geq 1} \subset \mathcal{H}$ associated to a set of real numbers $\{\lambda_k\}_{k \geq 1}$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ that satisfy

$$\Gamma \phi_k = \lambda_k \phi_k, \quad \text{and} \quad \lambda_k \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty. \tag{2}$$

The set $\{\lambda_k\}_{k \geq 1}$ contains the eigenvalues of the covariance operator Γ and Φ contains the associated eigenfunctions. Using the multivariate Karhunen-Loève theorem (Happ and Greven, 2015), we obtain the decomposition

$$X(\mathbf{t}) = \mu(\mathbf{t}) + \sum_{k=1}^{\infty} \mathfrak{c}_k \phi_k(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T} \tag{3}$$

where $\mathfrak{c}_k = \langle\langle X - \mu, \phi_k \rangle\rangle$ are the projection of the centered curve onto the eigenfunctions. We have that $\mathbb{E}(\mathfrak{c}_k) = 0$, $\mathbb{E}(\mathfrak{c}_k^2) = \lambda_k$ and $\mathbb{E}(\mathfrak{c}_k \mathfrak{c}_{k'}) = 0$ for $k \neq k'$. Note that the coefficients $\mathfrak{c}_k$ are scalar random variables while the multivariate functions ϕ_k are vectors of functions. Let us call Φ the multivariate functional principal component analysis basis. In practice, we use a truncated version of the Karhunen-Loève expansion (3) as the eigenvalues λ_k, , and hence the contribution of $\mathfrak{c}_k$ to Equation (3), become negligible as k goes to infinity. Let

$$X_{\lceil K \rceil}(\mathbf{t}) = \mu(\mathbf{t}) + \sum_{k=1}^{K} \mathfrak{c}_k \phi_k(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}, \quad K \geq 1, \tag{4}$$

be the truncated Karhunen-Loève expansion of the process X and

$$X^{(p)}_{\lceil K_p \rceil}(t_p) = \mu^{(p)}(t_p) + \sum_{k=1}^{K_p} \mathfrak{c}_k^{(p)} \phi_k^{(p)}(t_p), \quad t_p \in \mathcal{T}_p, \quad K_p \geq 1, \quad 1 \leq p \leq P, \tag{5}$$

be the truncated Karhunen-Loève expansion of the components of the process X.

4.1 By diagonalization of the covariance operator

The estimation of the eigencomponents of the covariance Γ by its diagonalization is derived in Happ and Greven (2015) for a general class of multivariate functional data defined on different dimensional domains. They give a direct relationship between the truncated representation (5) of the single elements $X^{(p)}$ and the truncated representation (4) of the multivariate functional data X (see Happ and Greven, 2015, Prop. 5).

We recall here, how to actually estimate the eigencomponents. Following Happ and Greven (2015, Prop. 5), the multivariate components for X are estimated by plugging the univariate components computed from each $X^{(p)}$. These estimations are done as follows. First, we perform an univariate FPCA on each of the components of X separately. For a component $X^{(p)}$, the eigenfunctions and eigenvectors are computed as a matrix analysis of the estimated covariance $C_{p,p}$ from (1). This results in a set of eigenfunctions $(\varphi_1^{(p)}, \dots, \varphi_{K_p}^{(p)})$ associated with a set of eigenvalues $(\lambda_1^{(p)}, \dots, \lambda_{K_p}^{(p)})$ for a given truncation integer K_p. Then, the univariate scores for a realization $X_n^{(p)}$ of $X^{(p)}$ are given by $\mathbf{c}_{nk}^{(p)} = \langle X_n^{(p)}, \varphi_k^{(p)} \rangle$, $1 \leq k \leq K_p$. These scores might be estimated by numerical integration for example. Considering $K_+ \coloneqq \sum_{p=1}^{P} K_p$, we then define the matrix $\mathcal{Z} \in \mathbb{R}^{N \times K_+}$, where on each row we concatenate the scores obtained for the P components of the nth observation: $(\mathbf{c}_{n1}^{(1)}, \dots, \mathbf{c}_{nK_1}^{(1)}, \dots, \mathbf{c}_{n1}^{(P)}, \dots, \mathbf{c}_{nK_P}^{(P)})$. An estimation of the covariance of the matrix $\mathcal{Z}$ is given by $\mathbf{Z} = (N-1)^{-1} \mathcal{Z}^\top \mathcal{Z}$. An eigenanalysis of the matrix $\mathbf{Z}$ is done to estimate the eigenvectors $\boldsymbol{v}_k$ and eigenvalues λ_k. And finally, the multivariate eigenfunctions are estimated with

$$\phi_k^{(p)}(t_p) = \sum_{l=1}^{K_p} [\boldsymbol{v}_k]_l^{(p)} \varphi_l^{(p)}(t_p), \quad t_p \in \mathcal{T}_p, \quad 1 \leq k \leq K_+, \quad 1 \leq p \leq P,$$

where $[\boldsymbol{v}_k]_l^{(p)}$ denotes the lth entry of the pth block of the vector $\boldsymbol{v}_k$. The multivariate scores are estimated with

$$\mathfrak{c}_{nk} = \mathcal{Z}_{n,\cdot} \boldsymbol{v}_k, \quad 1 \leq n \leq N, \quad 1 \leq k \leq K_+.$$

Add computation using basis expansion.

4.2 By diagonalization of the inner product matrix

We can use the duality relation between row and column spaces of a data matrix to estimate the eigencomponents of the covariance operator. Consider the inner-product matrix, also called the Gram matrix, M with entries

$$M_{nn'} = \langle\!\langle X_n - \mu, X_{n'} - \mu \rangle\!\rangle, \quad n, n' = 1, \dots, N.$$

Let $\{l_k\}_{1 \leq k \leq N}$ such that $l_1 \geq \dots \geq l_N \geq 0$ be the set of eigenvalues and $\{v_k\}_{1 \leq k \leq N}$ be the set of eigenvectors of the matrix M. The relationship between all nonzero eigenvalues of the covariance operator Γ and the eigenvalues of M is given by

$$\lambda_k = \frac{l_k}{N}, \quad k = 1, 2, \dots, N,$$

while the relationship between the multivariate eigenfunctions of the covariance operator Γ and the orthonormal eigenvectors of M is given by

$$\phi_k(\mathbf{t}) = \sum_{n=1}^{N} [v_k]_n \left\{ X_n(\mathbf{t}) - \mu(\mathbf{t}) \right\}, \quad \mathbf{t} \in \mathcal{T}, \quad k = 1, 2, \dots, N,$$

where $[v_k]_n$ is the nth entry of the vector v_k. The scores are then computed as the inner-product between the multivariate curves and the multivariate eigenfunctions and are given by

$$\mathfrak{c}_{nk} = \sqrt{l_k} [v_k]_n, \quad n = 1, 2, \dots, N, \quad k = 1, 2, \dots, N.$$

The derivation of these equalities are given in Appendix B.

Add computation using basis expansion.

4.3 Computational complexity

We describe the time complexity for the computation of the MFPCA algorithm using the covariance operator and the inner product matrix. Considering the observation of N curves with P components, we assume that all observations of the component p are sampled on a common grid of M_p points. For $a \in \mathbb{N}$, let $M^a = \sum_p M_p^a$. Let K be the number of multivariate eigenfunctions to estimate. For the estimation of the eigencomponents using the covariance operator, we have $K \leq K_+$. While K has the same interpretation for both the eigendecomposition of the covariance operator and the eigendecomposition of the inner product matrix, in the latter case, it is not computed as the summation over the univariate elements, but rather as the number of components needed to achieve a certain amount of variance explained (see Sections 4.4 and 5.3). Here, we also assume that the curves are perfectly observed, and thus no smoothing step is included in the expression of the time complexity.

To estimate the time complexity of an algorithm, we count the number of elementary operations realized, considering a fixed execution time for each elementary operation. Worst-case time complexity is usually considered; and this is what we do in the following. We first give the time complexity for the estimation of the eigencomponents using the covariance operator by expliciting the time complexity of each individual step (see Happ and Greven (2015) and Section 4.1). For each component p, the time complexity of the estimation of the covariance matrix is $\mathcal{O}(NM_p^2)$, of the eigendecomposition of the matrix is $\mathcal{O}(M_p^3)$ and of the univariate score is $\mathcal{O}(NM_pK_p)$. Considering all the components, the time complexity is the sum over p the univariate time complexity. The covariance matrix $\mathbf{Z}$ of the stacked univariate scores $\mathcal{Z}$ is then computed with a time complexity of $\mathcal{O}(NK_+^2)$, because the dimension of the matrix $\mathcal{Z}$ is $N \times K_+$. The eigendecomposition of the matrix $\mathbf{Z}$ has a time complexity of $\mathcal{O}(K_+^3)$. The final step is to compute the multivariate eigenfunctions and scores. For the estimation of the $K \leq K_+$ multivariate eigenfunctions, the time complexity is $\mathcal{O}(K\sum_p M_pK_p)$ and for the estimation of the scores, the time complexity is $\mathcal{O}(NK^2)$. Gathering all the results, the final complexity of the estimation of the eigencomponents using the eigendecomposition of the covariance operator is

$$\mathcal{O}\left(NM^2 + M^3 + N\sum_{p=1}^{P} M_pK_p + NK_+^2 + K_+^3 + K\sum_{p=1}^{P} M_pK_p + NK^2\right).$$

We now consider the time complexity of the estimation of the eigencomponents using the eigendecomposition of the inner product matrix (see Section 4.2). The inner product between two curves can be estimated in $\mathcal{O}(M^1)$. There are N^2 terms in the matrix. The time complexity for the computation of the inner product matrix is then $\mathcal{O}(N^2M^1)$. The eigendecomposition of this matrix has a time complexity of $\mathcal{O}(N^3)$. For the multivariate eigenfunctions, the time complexity is $\mathcal{O}(KNP)$ and is $\mathcal{O}(KN)$ for the multivariate scores. Gathering all the results, the final complexity of the estimation of eigencomponents using the eigendecomposition of the inner product matrix is

$$\mathcal{O}\left(N^2M^1 + N^3 + KNP + KN\right).$$

The number of components K to estimate is usually small compare to the number of curves N or to the number of sampling points M^1. Both time complexity can then be reduced to $\mathcal{O}(NM^2+M^3)$ for the diagonalization of the covariance operator and to $\mathcal{O}(N^2M^1 + N^3)$ using the Gram matrix. If the number of observations is large compare to the number of sampling points, it thus seems preferable to use the covariance operator to estimate the eigencomponents, while if the number of sampling points is large compare to the number of observations, the use of the Gram matrix seems better. Note that the number of components P does not have much impact on the computational complexity. These results are confirmed in the simulation (see Section 5.2).

Remark 3. *We can use the SVD in both cases to make the algorithm faster as it allows to compute only the first K eigenfunctions and not all of them.*

4.4 Percentage of variance explained

We argue that the percentage of variance explained in Happ and Greven (2015) is not the good one as they consider the variance explained by each of the components separately and not as a all. Using the inner product matrix however gives the right number of eigenfunctions for a given amount of variance explained.

See attached, a very brief simulation. It seems to show that choosing a univariate cut-off within each dimension (e.g., 95%), tends to overestimate the final amount of variance – the sum of the

final eigenvalues is larger than the sum of the true eigenvalues. I did some reading – Happ and Greven considered the effect of pve only on eigenfunction estimation. I've attached the relevant pages from their supplementary material. I'm not sure they considered eigenvalues or the total variation after. Moreover, I think (but not sure) they assume in their simulation that M is known where they say $\min\{M_1 + M_2, M\}$; whereas in practice, it is obviously unknown. But maybe they're just referring to M as the total number of sample eigenfunctions with non-zero eigenvalues (rather than the true number of multivariate eiegenfunctions).

Note that in Happ and Greven (2015), the percentage of variance explained in not specified in the MFPCA algorithm. The number of components is specified and is chosen to be equal to be the minimum between $\sum_{p=1}^{P} K_p$ and K.

5 Empirical analysis

Using simulated data, we compare, in this Section, the behavior of the estimation of the eigencomponents using the diagonalization of the covariance operator and the Gram matrix. The diagonalization of the covariance operator is performed using the methodology of Happ and Greven (2015). As this methodology is based on the expansion of each univariate components, we used univariate FPCA if the curves are unidimensional and the FCP-TPA algorithm for regularized tensor decomposition (Allen, 2013), if the curves are two-dimensional. We choose to use the FCP-TPA algorithm as it is the one used by Happ and Greven (2015) in their algorithm and implemented in their software (Happ-Kurz, 2020). Note that we could also use a two-dimensional basis expansion such as penalized tensor splines or discrete cosine transform, but we do not investigate this expansion here, as we do not want to prespecify a basis of functions.

The results of the simulation are compared using computation times (CT), the integrated squared error (ISE) risk for the multivariate eigenfunctions, the log-absolute error (lAE) risk for the eigenvalues and the mean integreated squared error (MISE) risk for the reconstructed data. Let ϕ_k be the true eigenfunction and $\widehat{\phi}_k$ the estimated eigenfunction defined on $\mathcal{T}$. We then defined the ISE as

$$\text{ISE}(\phi_k, \widehat{\phi}_k) = \left\| \phi_k - \widehat{\phi}_k \right\|^2 = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \{\phi_k^{(p)}(t) - \widehat{\phi}_k^{(p)}(t)\}^2 \mathrm{d}t, \quad k = 1, \ldots, K.$$

Let $\lambda = (\lambda_1, \ldots, \lambda_K)$ be the vector of true eigenvalues and $\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K)$ be the vector of estimated eigenvalues. We then define the MSE as

$$\text{lAE}(\lambda_k, \widehat{\lambda}_k) = \log(|\lambda_k - \widehat{\lambda}_k|), \quad k = 1, \ldots, K.$$

Let $\mathcal{X}$ be the set of true data and $\widehat{\mathcal{X}}$ be the set of reconstructed data. We defined the MISE of the reconstructed data as

$$\text{MISE}(\mathcal{X}, \widehat{\mathcal{X}}) = \frac{1}{N} \sum_{n=1}^{N} \left\| X_n - \widehat{X}_n \right\|^2 = \frac{1}{N} \sum_{n=1}^{N} \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\{ X_n^{(p)} - \widehat{X}_n^{(p)} \right\}^2 \mathrm{d}t.$$

Each integral is approximated by the trapezoidal rule with an equidistant grid. Note $\widehat{\phi}$, $\widehat{\lambda}$ and $\widehat{\mathcal{X}}$ the different estimators using the Gram matrix and $\widetilde{\phi}$, $\widetilde{\lambda}$ and $\widetilde{\mathcal{X}}$ the different estimators using the covariance operator. For each configuration (N, M, P), and each of the 500 samples, we compute the ratios

$$\frac{\text{ISE}(\phi_k, \widehat{\phi}_k)}{\text{ISE}(\phi_k, \widetilde{\phi}_k)}, \quad \frac{\text{lAE}(\lambda_k, \widehat{\lambda}_k)}{\text{lAE}(\lambda_k, \widetilde{\lambda}_k)}, \quad k = 1, \ldots, K, \quad \text{and} \quad \frac{\text{MISE}(\mathcal{X}, \widehat{\mathcal{X}})}{\text{MISE}(\mathcal{X}, \widetilde{\mathcal{X}})},$$

and compare them to 1.

5.1 Simulation experiments

We consider two simulation scenarios. One consists of multivariate functional data with univariate components defined as unidimensional curves and the other consists of univariate functional data of two-dimensional surfaces. Each experiment is repeated 500 times.

Scenario 1. The simulation setting is based on the simulation in Happ and Greven (2015). The data generating process is based on a truncated version of the Karhunen-Loève decomposition. First, we generate a large orthonormal basis $\{\psi_k\}_{1 \leq k \leq K}$ of $\mathcal{L}^2(\mathcal{T})$ on an interval $\mathcal{T} = [0, T] \subset \mathbb{R}$.

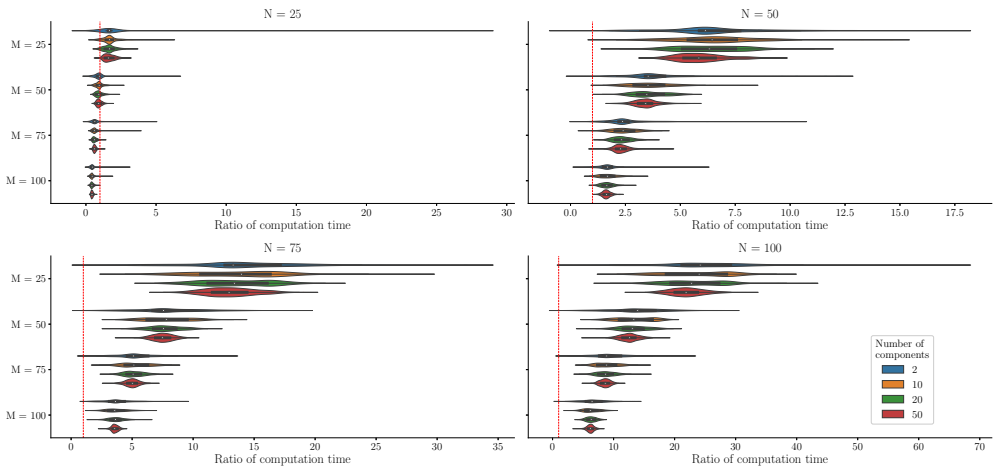

Figure 4: Computation time for multivariate functional data. Each univariate component is defined on a one-dimensional domain. to be modify

We fix $T_1 = 0$ and $T_{P+1} = T$ and we generate $P-1$ cutting points $T_2, \ldots, T_P$ uniformly in $\mathcal{T}$ such that $0 = T_1 < \cdots < T_P < T_{P+1} = T$. Let $s_1, \ldots, s_P \in \{-1, 1\}$ be coefficients that randomly flip the eigenfunctions. The univariate components of the eigenfunctions are then defined as

$$\phi_k^{(p)}(t_p) = s_p \psi_k\big|_{[T_p, T_{p+1}]}\left(\frac{t_p - T_p}{T_{p+1} - T_p}\right), \quad k = 1, \ldots, K, \quad p = 1, \ldots, P.$$

The notation $\phi_k\big|_{[T_p, T_{p+1}]}$ is the restriction of the function ϕ_k to the set $[T_p, T_{p+1}]$. The set of multivariate functions $\{\psi_k\}_{1 \leq k \leq K}$ is an orthonormal system in $\mathcal{H} := \mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_P)$ with $\mathcal{T}_p = [0, 1]$. Each curve is then simulated using the truncated multivariate Karhunen-Loève expansion (4):

$$X(\mathbf{t}) = \sum_{k=1}^{K} \mathfrak{c}_k \phi_k(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T},$$

where the scores $\mathfrak{c}_k$ are sampled as random normal variables with mean 0 and variance λ_k. The eigenvalues λ_k are defined with an exponential decrease, $\lambda_k = \exp(-(k+1)/2)$. We simulate, for each replication of the simulation, $N = 25, 50, 75$ and 100 observations. Similarly, each component is sampled on a regular grid of $M = 25, 50, 75$ and 100 sampling points. We compare the methods for $P = 2, 10, 20$ and 50 components and we set $K = 10$.

Scenario 2. The data generating process is based on a truncated version of the Karhunen-Loève decomposition. First, we generate an orthonormal basis $\{\phi_k\}_{1 \leq k \leq K}$ of $\mathcal{L}^2(\mathcal{T})$ on an interval $\mathcal{T} = [0, 1] \times [0, 1]$ as the tensor product of the first Fourier basis functions:

$$\phi_k(s, t) = \psi_l(s) \otimes \psi_m(t), \quad s, t \in [0, 1], \quad k = 1, \ldots, K,$$

where ψ_l and ψ_m are elements of the Fourier basis functions. Each curve is then simulated using the truncated multivariate Karhunen-Loève expansion (4):

$$X(s, t) = \sum_{k=1}^{K} \mathfrak{c}_k \phi_k(s, t), \quad s, t \in [0, 1],$$

where the scores $\mathfrak{c}_k$ are defined as for the Scenario 1. We simulate, for each replication of the simulations, $N = 25, 50, 75$ and 100 observations. Similarly, each component is sampled on a regular grid of $M = 25 \times 25, 50 \times 50, 75 \times 75$ and 100×100 sampling points. We set $K = 25$.

5.2 Simulation results

Computation time

Eigenvalues estimation

Eigenfunctions estimation

Curves reconstruction

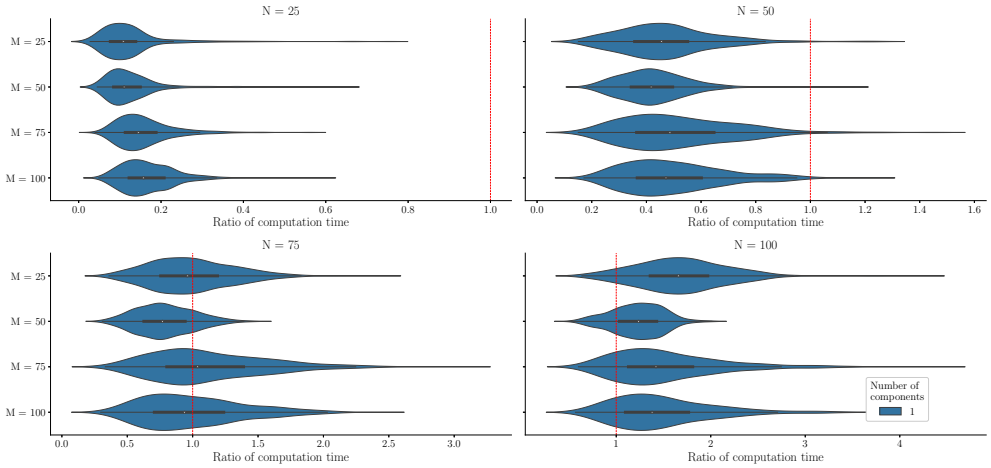

Figure 5: Computation time for univariate functional data of images data to be modify

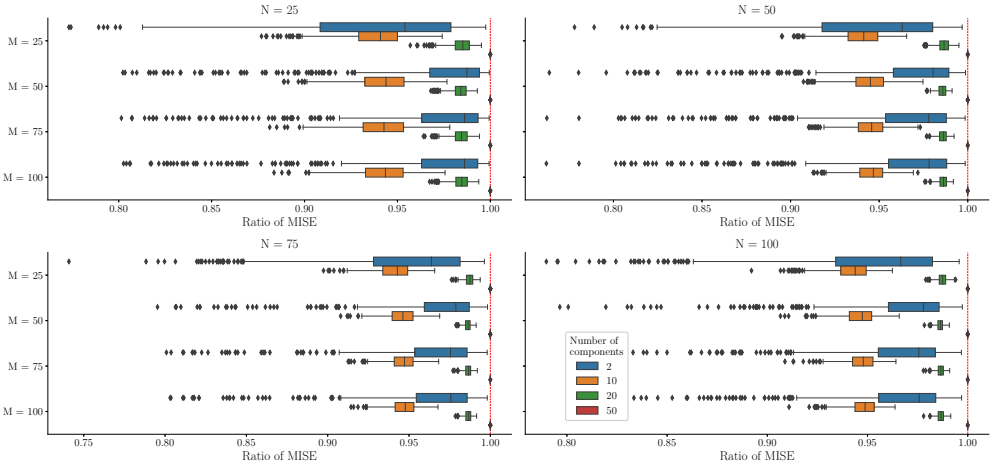

Figure 6: MISE for multivariate functional data. Each univariate component is defined on a one-dimensional domain. to be modify

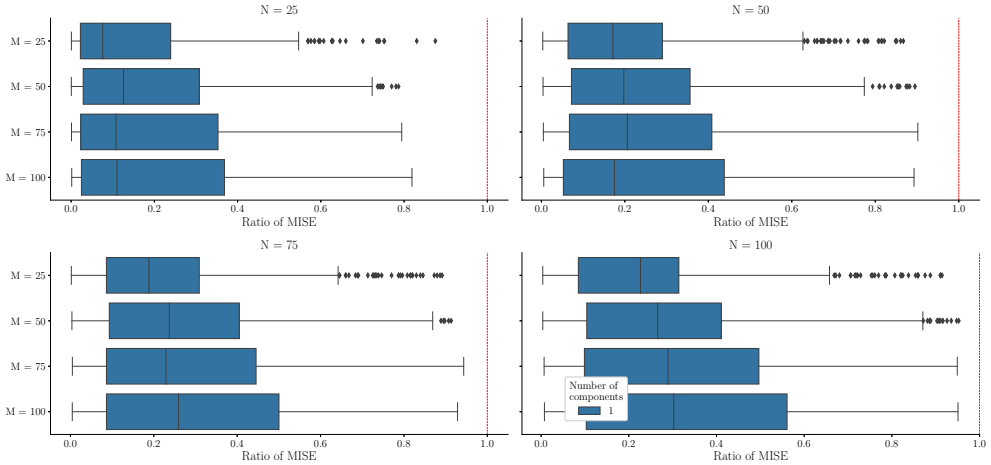

Figure 7: MISE for 2D data to be modify

5.3 Percentage of variance explained

We also consider the estimation of the number of components to retain to reach a prespecify percentage of variance explained by the data based on the Scenario 1 with $P = 2$, $K = 20$ and we use linear and exponential decreasing of the eigenvalues. The true percentage of variance explained by the kth component is given by $\lambda_k / \sum_{k'=1}^{20} \lambda_{k'}$ and the true percentage of variance explained by the first $K(\leq 20)$ components is given by $\sum_{k=1}^{K} \lambda_k / \sum_{k'=1}^{20} \lambda_{k'}$.

Let fix a certain percentage of variance explained $\alpha \in [0, 1]$. We define $\widetilde{K}$ as the minimum number of components to estimate to reach $100\alpha\%$ of the variance explained,

$$\widetilde{K} = \arg\min_{K} \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k'=1}^{20} \lambda_{k'}} \geq \alpha.$$

Using the covariance operator, the implementation of Happ and Greven (2015) does not allow to directly estimated $\widetilde{K}$ from the multivariate functional data. They propose however to estimate the number of univariate eigenfunctions K_1 and K_2 based on the percentage of variance explained α for both elements. The number of multivariate eigenfunctions is then set to be $\min\{K_1 + K_2, K\}$ where K is a given scalar. In the simulation, we first run FPCA on each univariate components to estimate the number of components needed to explain $\alpha\%$ of the variance for each components. Then, we run an MFPCA with $K = K_1 + K_2$. Using the Gram matrix, we directly estimated the number of components needed to explain a certain percentage of the variance from the multivariate functional data.

6 Discussion

Discussion about the methods and possible extension. What about the sparse/noisy data?

A Derivation of the inertia of the clouds

In this Section, we derive the inertia of the cloud $\mathcal{C}_N$. Recall that

$$\mathrm{Var}\{X^{(p)}(t_p)\} = \sum_{n=1}^{N} \pi_n \{X_n^{(p)}(t_p)\}^2 - \{\mu^{(p)}(t_p)\}^2 \quad \text{where} \quad \mu^{(p)}(t_p) = \sum_{n=1}^{N} \pi_n X_n^{(p)}(t_p), \quad t_p \in \mathcal{T}_p.$$

$$\begin{aligned}
\sum_{n=1}^{N} \pi_n d^2(X_n, \mu) &= \sum_{n=1}^{N} \pi_n \sum_{p=1}^{P} \left\| X_n^{(p)} - \mu^{(p)} \right\|^2 \\
&= \sum_{p=1}^{P} \left(\sum_{n=1}^{N} \pi_n \left\| X_n^{(p)} \right\|^2 - \left\| \mu^{(p)} \right\|^2 \right) \\
&= \sum_{p=1}^{P} \int_{\mathcal{T}_p} \mathrm{Var}\, X^{(p)}(t_p) \mathrm{d}t_p \\
\sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j d^2(X_i, X_j) &= \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j \sum_{p=1}^{P} \left\| X_i^{(p)} - X_j^{(p)} \right\|^2 \\
&= \sum_{p=1}^{P} \left(2 \sum_{i=1}^{N} \pi_i \left\| X_i^{(p)} \right\|^2 - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} \langle X_i^{(p)}, X_j^{(p)} \rangle \right) \\
&= \sum_{p=1}^{P} \left(2 \sum_{i=1}^{N} \pi_i \left\| X_i^{(p)} \right\|^2 - 2 \left\| \mu^{(p)} \right\|^2 - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} \langle X_i^{(p)}, X_j^{(p)} \rangle + 2 \left\| \mu^{(p)} \right\|^2 \right) \\
&= 2 \sum_{p=1}^{P} \int_{\mathcal{T}_p} \mathrm{Var}\, X^{(p)}(t_p) \mathrm{d}t_p
\end{aligned}$$

B Derivation of the eigencomponents

Using the Hilbert-Schmidt theorem, there exists a complete orthonormal basis of eigenvectors $\{v_k\}_{1\leq k\leq N}$ of the inner-product matrix M such that

$$Mv_k = l_k v_k. \tag{6}$$

Let $X = (X_1 - \mu, \ldots, X_N - \mu)^\top$. Recall that, in the case of P-dimensional process, the realisations of the process $X_i,\ i = 1, \cdots, N$ and μ are vectors of functions of length P, and thus X is a matrix of functions of size $N \times P$. By left multiplying Equation (6) by $X^\top$, we obtain

$$X^\top M v_k = l_k X^\top v_k. \tag{7}$$

Expanding Equation (7), for each component $p = 1, \ldots, P$, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{N} v_{jk}\left\{X_i^{(p)} - \mu^{(p)}\right\}\langle\langle X_i - \mu, X_j - \mu\rangle\rangle = l_k \sum_{i=1}^{N} v_{ik}\left\{X_i^{(p)} - \mu^{(p)}\right\}. \tag{8}$$

In the following, we note $[a]_p$ the pth entry of the vector a. Starting from the left side of Equation (8), we get

$$\begin{aligned}
[X^\top M v_k]_p &= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{jk}\left\{X_i^{(p)} - \mu^{(p)}\right\}\langle\langle X_i - \mu, X_j - \mu\rangle\rangle \qquad (9)\\
&= \sum_{q=1}^{P}\int_{\mathcal{T}_q}\sum_{i=1}^{N}\left\{X_i^{(p)} - \mu^{(p)}\right\}\left\{X_i^{(q)}(s) - \mu^{(q)}(s)\right\}\sum_{j=1}^{N} v_{jk}\left\{X_j^{(q)}(s) - \mu^{(q)}(s)\right\}\mathrm{d}s\\
&= \sum_{q=1}^{P}\int_{\mathcal{T}_q} N C_{pq}(\cdot, s)\sum_{j=1}^{N} v_{jk}\left\{X_j^{(q)}(s) - \mu^{(q)}(s)\right\}\mathrm{d}s\\
&= N\sum_{j=1}^{N}\langle\langle C_{p\cdot}(\cdot,\cdot), v_{jk}\{X_j - \mu\}\rangle\rangle\\
&= N\Gamma\left(\sum_{j=1}^{N} v_{jk}\{X_j - \mu\}\right)^{(p)}(\cdot)
\end{aligned}$$

and, starting from the right side of Equation (8),

$$[l_k X^\top v_k]_p = l_k \sum_{i=1}^{N} v_{ik}\left\{X_i^{(p)} - \mu^{(p)}\right\}. \tag{10}$$

From Equation (9) and Equation (10), we obtain

$$\Gamma\left(\sum_{j=1}^{N} v_{jk}\{X_j - \mu\}\right)^{(p)}(\cdot) = \frac{l_k}{N}\sum_{i=1}^{N} v_{ik}\left\{X_i^{(p)} - \mu^{(p)}\right\}, \quad \text{for all } p = 1, \ldots, P.$$

By identification in Equation (2), we find that, for each components p,

$$\lambda_k = \frac{l_k}{N} \quad \text{and} \quad \phi_k^{(p)} = \sum_{i=1}^{N} v_{ik}\left\{X_i^{(p)} - \mu^{(p)}\right\}, \quad k \geq 1. \tag{11}$$

For $k \geq 1$, the norm of the eigenfunction is computed as the following:

$$\begin{aligned}
\|\phi_k\|^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik} v_{jk}\langle\langle X_i - \mu, X_j - \mu\rangle\rangle = \sum_{i=1}^{N} v_{ik}\sum_{j=1}^{N} M_{ij} v_{jk}\\
&= \sum_{i=1}^{N} v_{ik} l_k v_{ik} = l_k \|v_k\|^2 = l_k.
\end{aligned}$$

Therefore, in order to have an orthonormal basis of eigenfunctions, we normalise the eigenfunctions ϕ_k from Equation (11) by $1/\sqrt{l_k}$. Concerning the estimation of the scores, for $i = 1, \ldots, N$, for $k \geq 1$, we have

$$\begin{aligned}\widehat{\mathfrak{c}}_{ik} = \langle\langle X_i - \mu, \phi_k \rangle\rangle &= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} \langle\langle X_i - \mu, X_j - \mu \rangle\rangle \\ &= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} M_{ij} = \sqrt{l_k} v_{ik}.\end{aligned}$$

****

Concerning the expansion of the data into the basis of function Ψ, we write

$$M = \left(CW^{1/2}\right)\left(CW^{1/2}\right)^{\top}.$$

We also assume that $\widehat{\phi}_1, \widehat{\phi}_2, \ldots$ the eigenfunctions of $\widehat{\Gamma}_N$ have a decomposition into the basis Ψ

$$\widehat{\phi}_k(t) = \begin{pmatrix} \widehat{\phi}_k^{(1)}(t) \\ \vdots \\ \widehat{\phi}_k^{(P)}(t) \end{pmatrix} = \begin{pmatrix} \psi^{(1)\top}(t) b_{1k} \\ \vdots \\ \psi^{(P)\top}(t) b_{Pk} \end{pmatrix}, \quad \text{where} \quad b_{pk} = \left(b_{pk1}, \ldots, b_{pkK_p}\right)^{\top}.$$

We have, for all $t \in \mathcal{T}_0$, $p = 1, \ldots, P$,

$$\begin{aligned}\left(\widehat{\Gamma}_N \phi_k\right)^{(p)}(t) &= \sum_{q=1}^{P} \int_{\mathcal{T}_0} \widehat{C}_{pq}(t,s) \phi_k^{(q)}(s) ds \\ &= \frac{1}{N} \sum_{q=1}^{P} \int_{\mathcal{T}_0} \Psi(t)^{(p)\top} C^{(p)\top} C^{(q)} \Psi^{(q)}(s) \Psi^{(q)}(s)^{\top} b_{qk} ds \\ &= \frac{1}{N} \Psi(t)^{(p)\top} C^{(p)\top} \sum_{q=1}^{P} C^{(q)} \int_{\mathcal{T}_0} \Psi^{(q)}(s) \Psi(s)^{(q)\top} ds b_{qk} \\ &= \frac{1}{N} \Psi(t)^{(p)\top} C^{(p)\top} \sum_{q=1}^{P} C^{(q)} W^{(q)} b_{qk}.\end{aligned}$$

This equation is true for all $p = 1, \cdots, P$, this can be rewritten with matrices as

$$\widehat{\Gamma}_N \phi_k(t) = \frac{1}{N} \Psi^{\top}(t) C^{\top} C W b_k.$$

From the eigenequation, we have that

$$\widehat{\Gamma}_N \phi_k(t) = \lambda_k \phi_k(t) \Longleftrightarrow \frac{1}{N} \Psi(t)^{\top} C^{\top} C W b_k = \lambda_k \Psi(t)^{\top} b_k, \quad t \in \mathcal{T}_0.$$

Since this equation must be true for all $t \in \mathcal{T}_0$, this imply the equation

$$C^{\top} C W b_k = N \lambda_k b_k. \tag{12}$$

As the eigenfunctions are assumed to be normalized, $\|\phi_k\|^2 = 1$. And so, $b_k^{\top} W b_k = 1$. Let $u_k = W^{1/2} b_k$. Then, from Equation (12), we obtain

$$W^{1/2} C^{\top} C W^{1/2} u_k = N \lambda_k u_k \Longleftrightarrow \left(CW^{1/2}\right)^{\top} \left(CW^{1/2}\right) u_k = N \lambda_k u_k. \tag{13}$$

From the eigendecomposition of the matrix M, we get

$$M v_k = l_k v_k \Longleftrightarrow \left(CW^{1/2}\right)\left(CW^{1/2}\right)^{T} v_k = l_k v_k. \tag{14}$$

The equations (13) and (14) are eigenequations in the classical PCA case, with the duality $X^\top X$ and $XX^\top$. Following Pagès (2014), we find that, for $1 \leq k \leq K$,

$$\lambda_k = \frac{l_k}{N}, \quad v_k = \frac{1}{\sqrt{l_k}} C W^{1/2} u_k \quad \text{and} \quad u_k = \frac{1}{\sqrt{l_k}} W^{1/2} C^\top v_k.$$

And finally, to get the coefficient of the eigenfunctions, for $1 \leq k \leq K$,

$$b_k = W^{-1/2} u_k = \frac{1}{\sqrt{l_k}} C^\top v_k.$$

Acknowledgment

S. Golovkine, A. J. Simpkin and N. Bargary are partially supported by Science Foundation Ireland under Grant No. 19/FFP/7002 and co-funded under the European Regional Development Fund. E. Gunning is supported in part Science Foundation Ireland (Grant No. 18/CRT/6049) and co-funded under the European Regional Development Fund.

References

Allen, G. I. (2013). Multi-way functional principal components analysis. In *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 220–223.

Benko, M., Härdle, W., and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, 37(1):1–34.

Berrendero, J., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55:2619–2634.

Chen, K., Zhang, X., Petersen, A., and Müller, H.-G. (2017). Quantifying Infinite-Dimensional Data: Functional Data Analysis in Action. *Statistics in Biosciences*, 9(2):582–604.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate Functional Principal Component Analysis: A Normalization Approach. *Statistica Sinica*, 24(4):1571–1596.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154.

Escofier, B. (1979). Traitement simultané de variables qualitatives et quantitatives en analyse factorielle. *Cahiers de l'analyse des données*, 4(2):137–146.

Happ, C. and Greven, S. (2015). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, 113.

Happ-Kurz, C. (2020). Object-Oriented Software for Functional Data. *Journal of Statistical Software*, 93(1):1–38.

Härdle, W. K. and Simar, L. (2019). *Applied Multivariate Statistical Analysis*. Springer Nature.

Horvàth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics.

Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, 71:92–106.

Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*. PhD thesis, (Sana), Helsinki.

Kneip, A. and Utikal, K. J. (2001). Inference for Density Families Using Functional Principal Component Analysis. *Journal of the American Statistical Association*, 96(454):519–532.

Kokoszka, P., Oja, H., Park, B., and Sangalli, L. (2017). Special issue on functional data analysis. *Econometrics and Statistics*, 1:99–100.

Krzyśko, M., Nijkamp, P., Ratajczak, W., and Wołyński, W. (2022). Multidimensional economic indicators and multivariate functional principal component analysis (MFPCA) in a comparative study of countries' competitiveness. *Journal of Geographical Systems*, 24(1):49–65.

Loève, M. (1945). Sur les fonctions aléatoires stationnaires de second ordre. *La Revue Scientifique, 5. Série*, 83:297–303.

Pagès, J. (2014). *Multiple Factor Analysis by Example Using R*. CRC Press.

Prothero, J. B., Hannig, J., and Marron, J. S. (2021). New Perspectives on Centering.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Science & Business Media.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics: Functional Analysis*. Academic Press.

Saporta, G. (1990). Simultaneous Analysis of Qualitative and Quantitative Data. *Atti 35° Riunione Scientifica della Societa Italiana di Statistica*, pages pp. 63–72.

Song, J. and Kim, K. (2022). Sparse multivariate functional principal component analysis. *Stat*, 11(1):e435.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295.

Warmenhoven, J., Cobley, S., Draper, C., Harrison, A., Bargary, N., and Smith, R. (2019). Bivariate functional principal components analysis: Considerations for use with multivariate movement signatures in sports biomechanics. *Sports Biomechanics*, 18(1):10–27.