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

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

May 21, 2024

Abstract

Dimension reduction is crucial 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 diagonalization of the covariance operator. With the increasing size and complexity of functional datasets, estimating the covariance operator has become more challenging. Therefore, there is a growing need for efficient methodologies to estimate the eigencomponents. Using the duality of the space of observations and the space of functional features, we propose to use the inner-product between the curves to estimate the eigenelements of multivariate and multidimensional functional datasets. The relationship between the eigenelements of the covariance operator and those 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 characterized as functions. These functions could represent measurements taken over time or space, such as temperature readings over a yearly period 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 references on FDA.

Functional principal component analysis (FPCA) is an extension of principal component analysis (PCA, a commonly used tool for dimension reduction in multivariate data) to functional data. FPCA 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 finite-dimensional vectors 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, e.g., regression and classification models. Recently, FPCA has been extended to multivariate functional data, which are data that consist of multiple functions that are observed simultaneously. This extension is referred to as multivariate functional principal component analysis (MFPCA). As for FPCA, a key benefit of MFPCA is that it allows one to identify and visualize the main sources of variation in the multivariate functional data. This can be useful in 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) combine the multivariate curves into one big curve and then perform a usual FPCA via an eigendecomposition of the covariance structure. This methodology can only be run for data that are defined on the same unidimensional domain, that exhibit similar amounts of variability and are measured in the same units. Jacques and Preda (2014) propose to represent each feature of the multivariate function separately using a basis function expansion. This results in a different set of coefficients for each univariate curve. The eigendecomposition is then run on the matrix of 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 of the sampling points. Happ and Greven (2018) extend the estimation of multivariate principal components to functional data defined on different dimensional domains. Their estimation procedure is based on carrying out FPCA on each univariate feature, and then using a weighted combination of the resulting principal components to obtain the multivariate eigencomponents. Finally, Berrendero et al. (2011) develop a different method to estimate the eigencomponents as they perform a principal components analysis for each sampling time point.

The key motivation of this 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 multivariate 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 corresponds 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 or 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 are the same as the ones obtained from a PCA run on the columns of the matrix. The choice of method to use, based on criteria such as computational time or data storage needs, is thus left to the statistician. This concept has been widely studied in 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, while Kneip and Utikal (2001) and Benko et al. (2009) utilized it to compute principal components for dense univariate functional data. Chen et al. (2017) also employ 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 components that are observed on possibly different 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 different data characteristics. The paper concludes with a discussion and an outlook in Section 7.

2 Model

The structure of the data we consider, referred to as *multivariate functional data*, is similar to that presented in Happ and Greven (2018). The data consist of independent trajectories of a vector-valued stochastic process $X = (X^{(1)}, \dots, X^{(P)})^\top$, $P \geq 1$. (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, or feature, $X^{(p)} : \mathcal{T}_p \to \mathbb{R}$ is assumed to belong to $\mathcal{L}^2(\mathcal{T}_p)$, the Hilbert space of square-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 $\mathbf{t} = (t_1, \dots, t_P)$ 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 $\mathcal{H} := \mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_P)$.

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

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

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

$\mathcal{H}$ is a Hilbert space with respect to the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$(Happ and Greven, 2018). 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, $\mu(\mathbf{t}) := \mathbb{E}(X(\mathbf{t}))$, $\mathbf{t} \in \mathcal{T}$. Let C denote the $P \times P$ matrix-valued covariance function which, for $\mathbf{s}, \mathbf{t} \in \mathcal{T}$, is defined as

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

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 features of the process X:

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

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 feature of $\Gamma f(\mathbf{t})$ is given by

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

Let us consider a set of N independent 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. 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 feature. 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 the set $\mathcal{X}$, the inner-product matrix, also called the Gram matrix, $\mathbf{M}$ is defined as a matrix of size $N \times N$ with entries

$$\mathbf{M}_{nn'} = \sqrt{\pi_n \pi_{n'}} \langle\!\langle X_n - \mu, X_{n'} - \mu \rangle\!\rangle, \quad n, n' = 1, \ldots, N. \tag{1}$$

This matrix is symmetric, positive definite, and interpretable as a proximity matrix, each entry being the similarity between the weighted observations.

This setting can be generalized to incorporate a general weighting scheme, as Chiou et al. (2014) or Happ and Greven (2018). We consider the function $\langle\!\langle \cdot, \cdot \rangle\!\rangle_w : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$,

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

Let $\Gamma_w : \mathcal{H} \to \mathcal{H}$ denote the weighted 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 feature of $\Gamma_w f(\mathbf{t})$ is given by

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

Similarly, we define the weigthed inner-product matrix $\mathbf{M}_w$, the matrix of size $N \times N$ with entries

$$\mathbf{M}_{w,nn'} = \sqrt{\pi_n \pi_{n'}} \langle\langle X_n - \mu, X_{n'} - \mu \rangle\rangle_w, \quad n, n' = 1, \ldots, N.$$

Remark 1. *The observation weights π_n and the feature weights w_p are not the same and should not be confused. The observation weights are used to give more (or less) importance to specific observations, while the feature weights control for different levels of variation for the features. A discussion on the feature weights is provided in Section 3.4.*

2.1 Inference

If we observe the set $\mathcal{X}$, the ideal estimators of the mean and covariance function would be

$$\widetilde{\mu}(\mathbf{t}) = \sum_{n=1}^{N} \pi_n X_n(\mathbf{t}), \quad \text{and} \quad \widetilde{C}(\mathbf{s}, \mathbf{t}) = \sum_{n=1}^{N} \pi_n \{X_n(\mathbf{s}) - \widetilde{\mu}(\mathbf{s})\}\{X_n(\mathbf{t}) - \widetilde{\mu}(\mathbf{t})\}^\top, \quad \mathbf{s}, \mathbf{t} \in \mathcal{T}.$$

Similarly, one could estimate the inner-product $\mathbf{M}$ by replacing the mean by their estimators. The resulting matrix $\widetilde{\mathbf{M}}$ has entries

$$\widetilde{\mathbf{M}}_{nn'} = \sqrt{\pi_n \pi_{n'}} \sum_{p=1}^{P} \int_{\mathcal{T}_p} \{X_n^{(p)}(t_p) - \widetilde{\mu}^{(p)}(t_p)\}\{X_{n'}^{(p)}(t_p) - \widetilde{\mu}^{(p)}(t_p)\} \mathrm{d}t_p, \quad n, n' = 1, \ldots, N.$$

In many applications, the elements of the set $\mathcal{X}$ are observed with error and on a finite grid of points in $\mathcal{T}$. For each $1 \leq n \leq N$, and given a vector of positive integers $\mathsf{M}_n = (\mathsf{M}_n^{(1)}, \ldots, \mathsf{M}_n^{(P)})$, let $\mathrm{T}_{n,\mathsf{m}} = (\mathrm{T}_{n,m_1}^{(1)}, \ldots, \mathrm{T}_{n,m_P}^{(P)}), 1 \leq m_p \leq \mathsf{M}_n^{(p)}, 1 \leq p \leq P$, be the random observation times for the curve X_n. These times are obtained as independent realizations of a random variable $\mathbf{T}$ taking values in $\mathcal{T}$. The vectors $\mathsf{M}_1, \ldots, \mathsf{M}_N$ represent an independent sample of an interger-valued random vector M with known expectation. We assume that the realizations of X, M and $\mathbf{T}$ are mutually independent. The observations associated with an observation X_n consist of the pairs $(Y_{n,\mathsf{m}}, \mathrm{T}_{n,\mathsf{m}}) \in \mathbb{R} \times \mathcal{T}$, where $\mathsf{m} = (m_1, \ldots, m_P), 1 \leq m_p \leq \mathsf{M}_n^{(p)}, 1 \leq p \leq P$ and $Y_{n,\mathsf{m}}$ is defined as

$$Y_{n,\mathsf{m}} = X_n(\mathrm{T}_{n,\mathsf{m}}) + \varepsilon_{n,\mathsf{m}}, \quad 1 \leq n \leq N, \tag{2}$$

with the $\varepsilon_{n,\mathsf{m}}$ being independent copies of a centered error random vector $\varepsilon \in \mathbb{R}^P$ with finite variance σ^2.

Let $\widehat{X}_n$ be a suitable estimator of X_n applied to the pairs $(Y_{n,\mathsf{m}}, \mathrm{T}_{n,\mathsf{m}})$, such as P-splines (e.g. Eilers et al. (2015)) or local polynomials (e.g. Fan and Gijbels (1996)). We define the estimator of the mean function as

$$\widehat{\mu}(\mathbf{t}) = \sum_{n=1}^{N} \pi_n \widehat{X}_n(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}.$$

Concerning the estimation of the covariance of the pth feature, we distinguish the diagonal from the non-diagonal points. The estimation of the covariance function of the non-diagonal points is defined as

$$\widehat{C}_{p,p}(s_p, t_p) = \sum_{n=1}^{N} \pi_n \left(\{\widehat{X}_n^{(p)}(s_p) - \widehat{\mu}^{(p)}(s_p)\}\{\widehat{X}_n^{(p)}(t_p) - \widehat{\mu}^{(p)}(t_p)\} \right), \quad s_p \neq t_p, \quad s_p, t_p \in \mathcal{T}_p.$$

The variance function $C_{p,p}(s_p, t_p)$ induces a singularity when estimating the covariance function $C_{p,p}(\cdot, \cdot)$ (see Yao et al. (2005); Zhang and Wang (2016)). We define the estimator of the diagonal of the covariance as

$$\widehat{C}_{p,p}(s_p, s_p) = \sum_{n=1}^{N} \pi_n \{\widehat{X}_n^{(p)}(s_p) - \widehat{\mu}^{(p)}(s_p)\}^2 - \sigma_p^2, \quad s_p \in \mathcal{T}_p.$$

This singularity does not appear in the estimation of the cross-covariance function. We define the estimator of the cross-covariance function between the pth and qth features by

$$\widehat{C}_{p,q}(s_p, t_q) = \sum_{n=1}^{N} \pi_n \left(\{\widehat{X}_n^{(p)}(s_p) - \widehat{\mu}^{(p)}(s_p)\}\{\widehat{X}_n^{(q)}(t_q) - \widehat{\mu}^{(q)}(t_q)\} \right), \quad s_p \in \mathcal{T}_p, t_q \in \mathcal{T}_q.$$

The estimator of the inner-product matrix can be defined by replacing the curves by their estimators. Following Benko et al. (2009); Grith et al. (2018), we define the estimator of the inner-product matrix as

$$\widehat{\mathbf{M}}_{nn'} = \sqrt{\pi_n \pi_{n'}} \sum_{p=1}^{P} \int_{\mathcal{T}_p} \{\widehat{X}_n^{(p)}(t_p) - \widehat{\mu}^{(p)}(t_p)\}\{\widehat{X}_{n'}^{(p)}(t_p) - \widehat{\mu}^{(p)}(t_p)\} \mathrm{d}t_p, \quad n, n' = 1, \ldots, N.$$

As for the covariance estimation, the model implies a bias on the diagonal term of the inner-product matrix. A correction for the diagonal of the Gram matrix is given by

$$\widehat{\mathbf{M}}_{nn} = \pi_n \sum_{p=1}^{P} \left\{ \int_{\mathcal{T}_p} \{\widehat{X}_n^{(p)}(t_p) - \widehat{\mu}^{(p)}(t_p)\}^2 \mathrm{d}t_p - \sigma_p^2 \right\}, \quad n = 1, \ldots, N.$$

For the estimation of the variances σ_p^2, we let the reader refers to Hall and Marron (1990); Hall et al. (1990).

3 On the geometry of multivariate functional data

3.1 Duality diagram

The distinction between the space of rows of a matrix as a sample from a population and the space of columns as the fixed variables on which the observations were measured has been explained in Holmes (2008) and De la Cruz and Holmes (2011) for multivariate data analysis. We propose to define a duality diagram in the context of multivariate functional data. Consider the data matrix defined by the set $\mathcal{X}$. We define an operator $L_X : \mathcal{H} \to \mathbb{R}^N$ by

$$L_X : f \mapsto \begin{pmatrix} \sqrt{\pi_1} \langle\langle X_1 - \mu, f \rangle\rangle \\ \vdots \\ \sqrt{\pi_N} \langle\langle X_N - \mu, f \rangle\rangle \end{pmatrix}.$$

Using the linearity of the inner-product $\langle\langle \cdot, \cdot \rangle\rangle$ and vectors, the operator L_X is linear. Define the operator $L_X^\star : \mathbb{R}^N \to \mathcal{H}$ as

$$L_X^\star : u \mapsto \begin{pmatrix} \sum_{n=1}^N \sqrt{\pi_n} u_n \{X_n^{(1)}(t_1) - \mu^{(1)}(t_1)\} \\ \vdots \\ \sum_{n=1}^N \sqrt{\pi_n} u_n \{X_n^{(P)}(t_P) - \mu^{(P)}(t_P)\} \end{pmatrix}.$$

Lemma 1. *$L_X^\star$ is the adjoint operator of the linear operator L_X.*

The proof of Lemma 1 is given in Appendix A. As an adjoint operator, $L_X^\star$ is a linear operator. The operator Γ and the matrix $\mathbf{M}$ define geometries in $\mathcal{H}$ and $\mathbb{R}^N$, respectively, through

$$\langle\langle f, g \rangle\rangle_\Gamma = \langle\langle f, \Gamma g \rangle\rangle, \quad f, g \in \mathcal{H}, \quad \text{and} \quad (u, v)_{\mathbf{M}} = u^\top \mathbf{M} v, \quad u, v \in \mathbb{R}^N.$$

We denote by $\|\cdot\|_\Gamma$ and $(\!(\cdot)\!)_{\mathbf{M}}$ their associated norms. Using the definition of adjoint operators L_X and $L_X^\star$, we have that

$$(L_X(f), u)_{\mathbf{M}} = \langle\langle f, L_X^\star(u) \rangle\rangle_\Gamma, \quad \text{for all} \quad f \in \mathcal{H}, u \in \mathbb{R}^N.$$

These relationships can be expressed as a duality diagram, see Figure 2. The triplet $(X, \Gamma, \mathbf{M})$ defines a (multivariate) functional data analysis framework. One consequence of this transition between spaces is that the eigencomponents of $\mathcal{X}$ can be estimated equivalently using either the covariance operator Γ or the Gram matrix $\mathbf{M}$. The relationship between the eigencomponents of Γ and the eigencomponents of $\mathbf{M}$ are derived in Section 4.

Remark 2. *In general, in order to avoid confusion, inner-products and norms in function space, $\mathcal{H}$ or $\mathcal{L}^2(\mathcal{T})$, will be refered with angle brackets, $\langle \cdot, \cdot \rangle$, while inner-products and norms in coordinate space, $\mathbb{R}^N$, will be refered with round brackets, $(\cdot, \cdot)$.*

Remark 3. *We present here the duality diagram for the linear integral operator Γ with kernel C. It is however possible to define duality diagrams for more general linear integral operators defined with a continuous symmetric positive definite function as kernel (see González and Muñoz (2010) and Wong and Zhang (2019) for discussions on possible integral operators to represent univariate functional data).*

Figure 2: Duality diagram between the spaces $\mathcal{H}$ and $\mathbb{R}^N$. The operator L_X and its adjoint $L_X^\star$ are linear operators. The covariance operator Γ and the matrix $\mathbf{M}$ define geometries in $\mathcal{H}$ and $\mathbb{R}^N$ respectively. The space $\mathcal{H}^\star$ (resp. $\mathbb{R}^{N\star}$) is the dual space of $\mathcal{H}$ (resp. $\mathbb{R}^N$).

Figure 3: Left: Cloud of observations. Right: Projection of the points onto the elements of $\mathcal{H}$. The observation f (resp. g) is identified by the point M_f (resp. M_g) in the cloud $\mathcal{C}_P$. The point G_μ is the center of gravity of $\mathcal{C}_P$ and the point $\mathrm{O}_\mathcal{H}$ is the origin of the space $\mathcal{H}$.

3.2 Cloud of individuals

Given an element $f \in \mathcal{H}$, let $\{f^{(p)}(t_p),\, t_p \in \mathcal{T}_p,\, p = 1, \ldots, P\}$ be the features set of the element. We identify this set as the point M_f in the space $\mathcal{H}$. The space $\mathcal{H}$ is referred to as the observation space. The cloud of points that represents the set of observations $\mathcal{X}$ in $\mathcal{H}$ is denoted by $\mathcal{C}_P$. Let G_μ be the centre of gravity of the cloud $\mathcal{C}_P$. 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 features are centered, the origin $\mathrm{O}_\mathcal{H}$ of the axes in $\mathcal{H}$ coincides with G_μ.

Let f and g be two elements in $\mathcal{H}$ and denote by M_f and M_g their associated points in $\mathcal{C}_P$ (see Figure 3). The most natural distance between these observations is based on the usual inner product in $\mathcal{H}$, $\langle\!\langle \cdot, \cdot \rangle\!\rangle$, and is defined as

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

This distance measures how different the observations are, and thus gives one characterization of the shape of the cloud $\mathcal{C}_P$. Another description of this shape is to consider the distance between each observation and G_μ, 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_μ, the distance between f and μ is given by

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

Figure 4: Left: Cloud of features. Right: Projection of the points on the elements of $\mathbb{R}^N$. The observation f (resp. g) is identified by the point M_f (resp. M_g) in the cloud $\mathcal{C}_N$. The point G_μ is the center of gravity of $\mathcal{C}_N$ and the point $\mathsf{O}_\mathbb{R}$ is the origin of the space $\mathbb{R}^N$.

Given the set $\mathcal{X}$, the total inertia of $\mathcal{C}_P$, with respect to G_μ and the distance d, is given by

$$\sum_{n=1}^{N} \pi_n d^2(\mathrm{M}_n, \mathrm{G}_\mu) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j d^2(\mathrm{M}_i, \mathrm{M}_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \operatorname{Var} X^{(p)}(t_p)\mathrm{d}t_p. \tag{3}$$

The duality diagram, however, allows us to define another suitable distance to characterize the shape of the cloud $\mathcal{C}_P$. We thus define

$$d_\Gamma^2(\mathrm{M}_f, \mathrm{M}_g) = \|f - g\|_\Gamma^2 .$$

The utilization of the distance measure d_Γ, which accounts for the variability among all the features within the functional data, corresponds to a Mahalanobis-type distance framework for multivariate functional data (see Berrendero et al. (2020) and Martino et al. (2019)). Given the set $\mathcal{X}$, the total inertia of $\mathcal{C}_P$, with respect to G_μ and the distance d_Γ, is given by

$$\sum_{n=1}^{N} \pi_n d_\Gamma^2(\mathrm{M}_n, \mathrm{G}_\mu) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j d_\Gamma^2(\mathrm{M}_i, \mathrm{M}_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \|C_{p\cdot}(t_p, \cdot)\|^2 \,\mathrm{d}t_p. \tag{4}$$

The derivation of these equalities are given in Appendix A.

Remark 4. *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 features are reduced beforehand, such that* $\int_{\mathcal{T}_p} \operatorname{Var} X^{(p)}(t_p)\mathrm{d}t_p = 1$ *for the distance* d *or* $\int_{\mathcal{T}_p} \|C_{p\cdot}(t_p, \cdot)\|^2 \,\mathrm{d}t_p = 1$ *for the distance* d_Γ*, the total inertia of the cloud* $\mathcal{C}_P$ *is equal to the number of components* P*. We are, in general, not interested by the total inertia but how this variance is spread among the features.*

3.3 Cloud of features

Given an element $f \in \mathcal{H}$, let $L_X(f) = \{\sqrt{\pi_n}\langle\langle X_n - \mu, f\rangle\rangle,\ n = 1, \ldots, N\}$ be the set of projections of f onto the centered observations. We identify this set as the point M_f in the space $\mathbb{R}^N$. The space $\mathbb{R}^N$ is referred to as the features' space. The cloud of points that represents the set of observations in $\mathbb{R}^N$ is denoted by $\mathcal{C}_N$. Let G_μ be the centre of gravity of the cloud $\mathcal{C}_N$. In the space $\mathbb{R}^N$, its coordinates are given by $L_X(\mu) = \{\sqrt{\pi_n}\langle\langle X_n - \mu, \mu\rangle\rangle,\ n = 1, \ldots, N\}$. If the data are centered, the origin $\mathsf{O}_\mathbb{R}$ of the axes in $\mathbb{R}^N$ coincides with G_μ.

We consider the usual inner-product in $\mathbb{R}^N$, such that for all $u, v \in \mathbb{R}^N$, $(u, v) = u^\top v$, associated with the norm $(\!(\cdot)\!)$. 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 4). The distance between M_f and M_g is thus defined as

$$\mathsf{d}^2(\mathsf{M}_f, \mathsf{M}_g) = (\!(L_X(f) - L_X(g))\!)^2 = \sum_{n=1}^{N} \pi_n \langle\langle X_n - \mu, f - g\rangle\rangle^2.$$

Similarly to the cloud of individuals, this distance characterizes the shape of the cloud $\mathcal{C}_N$ and we also have access to this characterization through the distance with the center of gravity G_μ. Let f

be an element of $\mathcal{H}$, associated to the point M_f, and μ the element of $\mathcal{H}$ related to G_μ, the distance between M_f and μ is given by

$$\mathsf{d}^2(\mathsf{M}_f, \mathsf{G}_\mu) = (\!(L_X(f) - L_X(\mu))\!)^2 = \sum_{n=1}^{N} \pi_n \langle\!\langle X_n - \mu, f - \mu \rangle\!\rangle^2.$$

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

$$\sum_{n=1}^{N} \pi_n \mathsf{d}^2(\mathsf{M}_n, \mathsf{G}_\mu) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j \mathsf{d}^2(\mathsf{M}_i, \mathsf{M}_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \|\!| C_{p\cdot}(t_p, \cdot) |\!\|^2 \, \mathrm{d}t_p. \tag{5}$$

Using the distances induced by the duality diagram, the total inertia of the cloud $\mathcal{C}_N$ is thus equal to the total inertia of the cloud $\mathcal{C}_P$. This property highlights the duality between the spaces $\mathcal{H}$ and $\mathbb{R}^N$. To further emphasize this duality, the cosine of the angle θ_{fg} formed by the two points M_f and M_g is equal to their correlation coefficient and can be written

$$\cos(\theta_{fg}) = \frac{(L_X(f), L_X(g))}{(\!(L_X(f))\!)\,(\!(L_X(g))\!)} = \frac{\langle\!\langle f, g \rangle\!\rangle_\Gamma}{\|\!| f |\!\|_\Gamma \, \|\!| g |\!\|_\Gamma}.$$

The derivation of these equalities are given in Appendix A.

Remark 5. *Although each axis of the space does not directly represent the features, but rather the projection of an element of $\mathcal{H}$ onto the elements of the set $\mathcal{X}$, we refer to this space as the features' space. We use this terminology to highlight the similarity between multivariate functional data analysis and traditional multivariate data analysis, as well as to emphasize the dimensionality of this space.*

3.4 On centering and reducing

For conducting an MFPCA, the features are usually assumed centred (Happ and Greven, 2018). Prothero et al. (2023) give a complete overview of centering in the context of FDA. Here, we comment on the geometric interpretation of centering in this context and compare with the multivariate scalar case. We focus on the usual centering in FDA, namely $X_n^{(p)}(t_p) - \mu^{(p)}(t_p)$, $t_p \in \mathcal{T}_p$ (refered as *object centering* in Prothero et al. (2023)). The geometric interpretation of object centering is the same if we refer to the observation space $\mathcal{H}$ or the feature space $\mathbb{R}^N$. Within the space $\mathcal{H}$ (resp. $\mathbb{R}^N$), centering is interpreted as translating the centre of gravity of the clouds, G_μ (resp. G_μ), to the origin point $\mathrm{O}_\mathcal{H}$ (resp. $\mathsf{O}_\mathbb{R}$) of the space $\mathcal{H}$ (resp. $\mathbb{R}^N$). This transformation, being a translation, does not change the shape of the cloud $\mathcal{C}_P$ (resp. $\mathcal{C}_N$). The interpretation is the same as for the centering in the multivariate scalar data context within their observation space.

Concerning the standardization of the data, there are two main proposals in the literature. Happ and Greven (2018) propose to consider the weights

$$w_p = \left(\int_{\mathcal{T}_p} \operatorname{Var} X^{(p)}(t_p) \mathrm{d}t_p \right)^{-1}.$$

This standardization is coherent with the derivation of the total inertia of the observation space using the usual distance in $\mathcal{H}$. Chiou et al. (2014) and Jacques and Preda (2014) propose the weighting function

$$w_p(t_p) = \left(\operatorname{Var} X^{(p)}(t_p) \right)^{-1}, \quad t_p \in \mathcal{T}_p.$$

This corresponds to a standardization of the curves by the standard deviation of the component at each sampling point. The standard deviation curve is estimated as the square root of the diagonal of the covariance function estimates, obtained using a local linear smoother of the pooled data. For each functional feature p, this standardization mimics the standardization used for principal components analysis if the number of (scalar) features is infinite. Considering the duality diagram and the total inertia of the clouds with respect to the distance d_Γ, we propose to use the weights

$$w_p = \left(\int_{\mathcal{T}_p} \|\!| C_{p\cdot}(t_p, \cdot) |\!\|^2 \, \mathrm{d}t_p \right)^{-1}.$$

The total inertia of $\mathcal{C}_P$ (and $\mathcal{C}_N$) will thus be equal to the number of components. In every case, we may consider the rescaled elements $\widetilde{X}^{(p)} = w_p^{1/2}\{X^{(p)} - \mu^{(p)}\}$ in place of the elements $X^{(p)}$.

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 (2018), 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

$$\Gamma\phi_k = \lambda_k\phi_k, \quad \text{and} \quad \lambda_k \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty. \tag{6}$$

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, 2018), we obtain the decomposition

$$X(\mathbf{t}) = \mu(\mathbf{t}) + \sum_{k=1}^{\infty} \mathfrak{c}_k\phi_k(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T} \tag{7}$$

where $\mathfrak{c}_k = \langle\langle X - \mu, \phi_k\rangle\rangle$ are the projections 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 (7) as the eigenvalues λ_k, and hence the contribution of $\mathfrak{c}_k$ to (7), becomes negligible as k goes to infinity. Let

$$X_{\lceil K\rceil}(\mathbf{t}) = \mu(\mathbf{t}) + \sum_{k=1}^{K} \mathfrak{c}_k\phi_k(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}, \quad K \geq 1, \tag{8}$$

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

$$X^{(p)}_{\lceil K_p\rceil}(t_p) = \mu^{(p)}(t_p) + \sum_{k=1}^{K_p} \mathfrak{c}_k^{(p)}\varphi_k^{(p)}(t_p), \quad t_p \in \mathcal{T}_p, \quad K_p \geq 1, \quad 1 \leq p \leq P, \tag{9}$$

be the truncated Karhunen-Loève expansion of the pth feature of the process X. For each p, the function $\mu^{(p)}$ is the pth feature of the multivariate function μ and the set $\{\varphi_k^{(p)}\}_{1\leq k\leq K_p}$ is a basis of univariate functions in $\mathcal{L}^2(\mathcal{T}_p)$, whose elements are not the components of the multivariate functions ϕ_k. In (9), the coefficients $\mathfrak{c}_k^{(p)}$ are the projection of the centered curve $X^{(p)}$ onto the eigenfunctions $\varphi_k^{(p)}$ and are not (directly) related to the coefficients $\mathfrak{c}_k$ in (8).

4.1 Diagonalization of the covariance operator

The estimation of the eigencomponents of the covariance Γ by its diagonalization is derived in Happ and Greven (2018) for a general class of multivariate functional data defined on different dimensional domains. They give a direct relationship between the truncated representation (9) of the single elements $X^{(p)}$ and the truncated representation (8) of the multivariate functional data X.

We recall here how to estimate the eigencomponents. Following Happ and Greven (2018, Prop. 5), the multivariate components for X are estimated by a weighted combination of the univariate components computed from each $X^{(p)}$. First, we perform a univariate FPCA on each of the features of X separately. For a feature $X^{(p)}$, the eigenfunctions and eigenvectors are computed using a matrix decomposition of the estimated covariance $C_{p,p}$ from Equation (**??**). This results in a set of eigenfunctions $\{\varphi_k^{(p)}\}_{1\leq k\leq K_p}$ associated with a set of eigenvalues $\{\lambda_k^{(p)}\}_{1\leq k\leq K_p}$ for a given truncation integer K_p. Then, the univariate scores for a realization $X_n^{(p)}$ of $X^{(p)}$ are given by $\mathbf{c}_{nk}^{(p)} = \left\langle X_n^{(p)}, \varphi_k^{(p)}\right\rangle$, $1 \leq k \leq K_p$. These scores might be estimated by numerical integration for example. Considering $K_+ := \sum_{p=1}^{P} K_p$, 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 features of the nth observation: $(\mathbf{c}_{n1}^{(1)}, \ldots, \mathbf{c}_{nK_1}^{(1)}, \ldots, \mathbf{c}_{n1}^{(P)}, \ldots, \mathbf{c}_{nK_P}^{(P)})$. 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 carried out to estimate the eigenvectors $\boldsymbol{v}_k$ and eigenvalues λ_k. Finally, the multivariate eigenfunctions are estimated as a linear combination of the univariate eigenfunctions using

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

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 as

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

where $\mathcal{Z}_{n,\cdot}$ is the nth row of the matrix $\mathcal{Z}$. We refer the reader to Happ and Greven (2018) for the derivation of the eigencomponents of the covariance operator if the curves are expanded in a general basis of functions.

4.2 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 $\mathbf{M}$, with entries defined in (1) and assuming that all observations are equally weighted, i.e., for all $n = 1, \ldots, N$, $\pi_n = 1/N$. Let $\{l_k\}_{1 \leq k \leq N}$ such that $l_1 \geq \cdots \geq l_N \geq 0$ be the set of eigenvalues and $\{\boldsymbol{u}_k\}_{1 \leq k \leq N}$ be the set of eigenvectors of the matrix $\mathbf{M}$. The relationship between all nonzero eigenvalues of the covariance operator Γ and the eigenvalues of $\mathbf{M}$ is given by

$$\lambda_k = l_k, \quad k = 1, 2, \ldots, N, \tag{10}$$

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

$$\phi_k(\mathbf{t}) = \frac{1}{\sqrt{Nl_k}} \sum_{n=1}^{N} [\boldsymbol{u}_k]_n \left\{X_n(\mathbf{t}) - \mu(\mathbf{t})\right\}, \quad \mathbf{t} \in \mathcal{T}, \quad k = 1, 2, \ldots, N, \tag{11}$$

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

$$\mathfrak{c}_{nk} = \sqrt{Nl_k}[\boldsymbol{u}_k]_n, \quad n = 1, 2, \ldots, N, \quad k = 1, 2, \ldots, N. \tag{12}$$

The derivation of these equalities are given in Appendix B in a slighty more general framework where the observation weights π_n are not equal. These results can be extended in a natural way if all the curves are expanded in a general basis of functions, see Section A in the Supplementary Material.

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 features, we assume that all observations of the feature 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. Here, we also assume that the curves are perfectly observed, and thus no smoothing step is included in the expression of the time complexity. Note that the smoothing step will often have the same impact on complexity between the approaches.

To estimate the time complexity of an algorithm, we count the number of elementary operations performed, considering a fixed execution time for each. Worst-case time complexity is considered. We first give the time complexity for the estimation of the eigencomponents using the covariance operator by explaining the time complexity of each individual step (see Happ and Greven (2018) and Section 4.1). For each feature 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 $\mathcal{O}(NM_pK_p)$. Therefore, the total time complexity is the sum over the p univariate time complexities. 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

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

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 $\mathcal{O}(M^1)$. Since 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 compared to the number of curves N or to the total number of sampling points M^1. Both time complexities 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 compared to the total number of sampling points, it thus seems preferable to use the covariance operator to estimate the eigencomponents, while if the total 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 features P does not have much impact on the computational complexity. These results are confirmed in the simulation (see Section 5.2).

Remark 6. *We can use singular values decomposition (SVD) in both cases to make the algorithm faster as it allows to compute only the first K eigenfunctions. In practice, this might be important as the maximum number of non-zero eigenvalues is the minimum between the number of observations and the number of sampling points.*

5 Empirical analysis

Using simulated data, we compare the estimation of the eigencomponents and the reconstruction of the curves using the diagonalization of the covariance operator, based on univariate FPCA as well as P-splines basis expansion, and the diagonalization of the Gram matrix. For the diagonalization of the covariance operator, we consider the methodology of Happ and Greven (2018). In the case of MFPCA based on the expansion of each univariate feature into univariate principal components, the eigencomposition of the 1-dimensional curves is performed using univariate FPCA and the eigendecomposition of the 2-dimensional curves is calculated using the Functional Canonical Polyadic-Tensor Power Algorithm (FCP-TPA) for regularized tensor decomposition (Allen, 2013). In the case of MFPCA based on P-splines basis expansion, data are expanded in B-splines basis with suitable smoothness penalty (Eilers and Marx, 1996).

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

$$\mathrm{ISE}(\phi_k, \widehat{\phi}_k) = \left\|\phi_k - \widehat{\phi}_k\right\|^2 = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \{\phi_k^{(p)}(t_p) - \widehat{\phi}_k^{(p)}(t_p)\}^2 \mathrm{d}t_p, \quad k = 1, \ldots, K. \tag{13}$$

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

$$\mathrm{RSE}(\lambda_k, \widehat{\lambda}_k) = \left(\lambda_k - \widehat{\lambda}_k\right)^2 / \lambda_k^2, \quad k = 1, \ldots, K. \tag{14}$$

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

$$\mathrm{MRSE}(\mathcal{X}, \widehat{\mathcal{X}}) = \frac{1}{N}\sum_{n=1}^{N} \left\|X_n - \widehat{X}_n\right\|^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\{X_n^{(p)}(t_p) - \widehat{X}_n^{(p)}(t_p)\right\}^2 \mathrm{d}t_p. \tag{15}$$

Each integral is approximated by the trapezoidal rule with an equidistant grid.

5.1 Simulation experiments

The simulation setting is based on the *Setting 3* of the simulation in Happ and Greven (2018). The data generating process is based on a truncated version of the Karhunen-Loève decomposition. We simulate multivariate functional data with $P = 2$ components. For the first component, we generate an orthonormal basis $\{\phi_k^{(1)}\}_{1\leq k\leq K}$ of $\mathcal{L}^2(\mathcal{T}_1)$ on an interval $\mathcal{T}_1 = [0,1]\times[0,0.5]$ as the tensor product of the first Fourier basis functions:

$$\phi_k^{(1)}(s,t) = \psi_l(s)\otimes\psi_m(t), \quad s\in[0,1] \text{ and } t\in[0,0.5], \quad k=1,\dots,K,$$

where ψ_l and ψ_m are elements of the Fourier basis. For the second component, we generate an orthonormal Legendre basis $\{\phi_k^{(2)}\}_{1\leq k\leq K}$ of $\mathcal{L}^2(\mathcal{T}_2)$ on an interval $\mathcal{T}_2 = [-1,1]$. To ensure orthonormality, the basis $\phi_k^{(1)}$ and $\phi_k^{(2)}$ are weighted by random factors $\alpha^{1/2}$ and $(1-\alpha)^{1/2}$, respectively, where $\alpha\sim\mathcal{U}(0.2,0.8)$. Each curve is then simulated using the truncated multivariate Karhunen-Loève expansion (8):

$$X(\mathbf{t}) = \sum_{k=1}^{K}\mathfrak{c}_k\phi_k(\mathbf{t}), \quad \mathbf{t}\in\mathcal{T}:=\mathcal{T}_1\times\mathcal{T}_2,$$

where $\phi_k = (\phi_k^{(1)},\phi_k^{(2)})^\top$ and 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, $N = 50, 100$ and 250 observations. The first component is sampled on a regular grid of $M^{(1)} = 11\times 11, 26\times 26$ and 101×51 sampling points. The second component is sampled on a regular grid of $M^{(2)} = 21, 51$ and 201 sampling points. We set $K = 25$. We also consider data with measurement errors. In that case, we observe data from (2), where $\sigma^2 = 0.25$. Finally, we study sparse data as in Happ and Greven (2018), with medium ($50\% - 70\%$ missings) and high ($90\% - 95\%$ missings) sparsity.

MFPCA using the diagonalization of the covariance operator is based on the univariate estimation of $K^{(1)} = 20$ eigenimages and $K^{(2)} = 15$ eigenfunctions. The estimation of the eigenimages is performed with the FCP-TPA where the smoothing parameters are chosen via cross-validation in $[10^{-5}, 10^5]$ for dense, sparse and noisy data. In the case of sparse data, we also linearly interpolate the observation to get data that are regularly sampled. The estimation of the eigenfunctions is done using the PACE algorithm (Yao et al., 2005) with P-splines to smooth the mean and covariance functions. Penalties are estimated using cross-validation (Eilers and Marx, 1996).

Concerning MFPCA using P-splines basis expansion, $X^{(1)}$ is expanded in tensor products of $K^{(1)} = 13\times 13$ B-splines and $X^{(2)}$ is expanded in $K^{(2)} = 13$ B-splines. For dense data, no penalty is involved, while for sparse and noisy data, penalties are estimated using cross-validation.

Finally, MFPCA using the Gram matrix is based on the observed data points in the dense case, on the linear interpolation of the data in the sparse case (Benko et al., 2009) and on the P-splines smoothing of the data in the noisy case.

5.2 Simulation results

We compared MFPCA results of the different methods in terms of their CT, estimation of eigenvalues (RSE), estimation of eigenfunctions (ISE), and reconstruction of curves (MRSE). We fix the number of retained components to be 12 for each simulation scenario. Each experiment is repeated 200 times. The results are presented below.

Computational time. To compare the computational time of the different methods, we measured the time it took for each method to complete the MFPCA for each simulated dataset in the dense and noiseless case. Figure 5 shows the kernel density estimates of the ratio of CT for each method across all sample sizes and number of sampling points. Except when we have many observations with few sampling points, the diagonalization of the Gram matrix is faster than the other two methods. This result is coherent with the computational complexity derived in Section 4.3.

The shorter computational time of the diagonalization of the Gram matrix makes it a more efficient option for analyzing two and higher-dimensional functional datasets as the number of sampling points tend to grow faster in these case. It is however worth noting that the computational time can still vary depending on the specific implementation of each method, the computational resources available, the complexity of the dataset (number of observations, number of sampling points, etc.) and whether a smoothing need to be included.

Eigenvalues estimation. To compare the estimation of the eigenvalues between the diagonalization of the covariance operator and the diagonalization of the Gram matrix, we calculated

Figure 5: Ratio of computation time in the dense case between the (Tensor) PCA and 2D/1D B-Splines methods and the Gram matrix method. N is the number of observations, M is the number of sampling points per curve (the first two numbers are for the images and the last one is for the curves).

Figure 6: RSE for the estimated eigenvalues for each method in the dense case. N is the number of observations, M is the number of sampling points per curve (the first two numbers are for the images and the last one is for the curves).

the ratio of the log $-$AE (14) between the estimated eigenvalues and the true eigenvalues for each simulated dataset and for the first five eigenvalues. Figure 6 shows the boxplots of the log $-$AE for each method across all sample sizes, number of sampling points and number of features for Scenario 1. We found that the two methods behave similarly for all considered settings.

Eigenfunctions estimation. To compare the estimation of the eigenfunctions between the diagonalization of the covariance operator and the diagonalization of the Gram matrix, we calculated the ratio of the ISE (13) between the estimated eigenfunctions and the true eigenfunctions for each simulated dataset and for the first five eigenfunctions. Figure 7 shows the boxplots of the ISE for each method across all sample sizes, number of sampling points and number of features for Scenario 1. We found that the two methods behave similarly for all considered settings. For $P = 10, 20$ and 50, the results are identical.

Curves reconstruction. To compare the quality of the reconstruction of the curves between the diagonalization of the covariance operator and the diagaonlisation of the Gram matrix, we calculated the ratio of the MISE (15) between the reconstruction of the curves and the true curves for each simulated dataset. Figure 8 shows the boxplots of the ISE for each method across all sample sizes, number of sampling points and number of features for Scenario 1. We found that the diagonalization of the Gram matrix gives slightly better results for all considered settings.

Figure 7: ISE for the estimated eigenfunctions for each method in the dense case. N is the number of observations, M is the number of sampling points per curve (the first two numbers are for the images and the last one is for the curves).

Figure 8: MISE for the reconstructed curves for each method in the dense case. N is the number of observations, M is the number of sampling points per curve (the first two numbers are for the images and the last one is for the curves).

(a) Stephen Curry (b) Duncan Robinson

Figure 9: Make/Miss chart for Stephen Curry and Duncan Robinson.

6 Application

We implement our methodology utilizing data from the National Basketball Association (NBA). To access comprehensive game data, we utilize the Python package `nba_api`[1], which provides access to the APIs of `nba.com`. The dataset encompasses shot location data from all NBA games spanning the seasons between $2018-2019$ and $2022-2023$. Filtering the dataset, we focus solely on players who have made more than 1000 shots during this five-season period, resulting in a cohort of 131 players. These players accounted for a total of 493723 shots attempted, of which 234941 were successful. Subsequently, we exclude shots deemed impossible (e.g., out-of-bounds), leaving us with a dataset comprising 492621 shots (see Figure 9 for the shots chart of Stephen Curry and Kevin Durant). To analyze shooting behavior, we employ 2-dimensional kernel density estimation, utilizing Silverman's rule (Silverman, 1986) for bandwidth estimation. The density estimation is conducted on a regularly spaced grid consisting of 51×51 points.

We estimate the principal components using the decomposition of the Gram matrix and the eigenanalysis of the covariance operator using the FCP-TPA for the univariate components. Remark that the functional principal components, as a representation of the deviation from the mean surface function, may take negative values, while density can only take positive values. For presentation purpose, the functional principal components have been normalized between -1 and 1. Blue corresponds to negative values of the functional components, and thus contribute negatively to the density, while Red corresponds to positive values of the functional components, and thus contribute positively to the density. The bluer the area in the image of the functional components, the smaller/negative the scores, whereas the redder areas correspond to positive values of the scores. The white areas correspond to scores that are closed to 0, and basically have not impact on the shooting density. The obtained functional components can be explained as different shooting styles. For the Gram matrix, the first two principal components and the fourth are related to the shoots under the basket, excluding the other areas. This account for around 95% of the variance. The third and fifth components are related to the three-pointers. These results exhibits the shooting patterns of Stephen Curry. He mostly scores under the basket or behind the three-points line, but almost never in-between. The decomposition of the density of made shots is similar to the ones of attempted shots. We may advance an explanation for that: Stephen Curry does not have a "weak" spot, that is somewhere on the court where he shots a lot but does not score.

7 Discussion and conclusion

MFPCA is a fundamental statistical tool for the analysis of multivariate functional data, which enables us to capture the variability in observations defined by multiple curves. In this paper, we have

[1] https://github.com/swar/nba_api/

(a) Curry density (b) Robinson density

(c) Curry Gram (d) Robinson Gram

(e) Curry covariance (f) Robinson covariance

Figure 10: Reconstruction.

described the duality between rows and columns of a data matrix within the context of multivariate functional data. We have proposed to use this duality to estimate the eigencomponents of the covariance operator in multivariate functional datasets. By comparing the results of the two methods, we provide the researcher with guidelines for determining the most appropriate method within a range of functional data frameworks. Overall, our simulations showed that the diagonalization of the covariance operator or of the Gram matrix gave similar results in terms of the estimation of the eigenvalues, the eigenfunctions and reconstruction of the curves for one-dimensional multivariate functional data, while the performance of the diagonalization of the Gram matrix outperformed the FCP-TPA for higher-dimensional functional datasets. Regarding the computation time, the use of the covariance operator is faster in most cases for multivariate functional data defined on unidimensional domains. The only situation where the use of the Gram matrix is quicker is when $M \gg N$. For data defined on higher-dimensional domains, diagonalization of the Gram matrix is faster. In conclusion, we recommend to use the covariance operator for multivariate functional data with all features defined on a one-dimensional domeins (curves) and for a number of observations larger or comparable to the number of sampling points regardless of the number of features. Regarding the reconstruction error, if the data are defined on multi-dimensional domains (images) or the number of sampling points is much higher than the number of observations, we advise using the Gram matrix.

Utilizing the Gram matrix enables the estimation of the number of components retained via the percentage of variance explained by the multivariate functional data, whereas the decomposition of the covariance operator necessitates the specification of the percentage of variance accounted for by each individual univariate feature. Specifying the percentage of variance explained for each feature does not guarantee that we recover the nominal percentage of variance explained for the multivariate data (Golovkine et al., 2023). Although we have not investigated the extent to which this might be important, the duality relation derived in this work provides a direct solution to the problem. Future work could investigate the settings in which univariate variance-explained cutoffs fail to retain the correct percentage of variance explained in multivariate functional data, and hence where the Gram matrix approach may be preferred.

In practice, observations of (multivariate) functional data are often subject to noise. As we recommend the use of the Gram matrix solely for densely sampled functional datasets, individual curve smoothing should suffice to approximate the Gram matrix in such cases. The estimation

of the Gram matrix in the context of sparsely sampled functional data is however deemed irrelevant, given our findings that the utilization of the covariance operator for the estimation of the eigencomponents yields comparable results, while typically requiring less computational time.

The open-source implementation can be accessed at https://github.com/StevenGolovkine/FDApy, while scripts to reproduce the simulations are at https://github.com/FAST-ULxNUIG/geom_mfpca.

A Derivation of the equalities

A.1 Proof of Lemma 1

Using the definition of adjoint operators, we must prove that

$$(L_X(f), u)_{\mathbf{M}} = \langle\!\langle f, L_X^\star(u)\rangle\!\rangle_\Gamma, \quad \text{for all} \quad f \in \mathcal{H}, \quad u \in \mathbb{R}^N. \tag{16}$$

Proof. For all $f \in \mathcal{H}, u \in \mathbb{R}^N$, we have that

$$\begin{aligned}
(L_X(f), u)_{\mathbf{M}} &= L_X(f)^\top \mathbf{M} u \\
&= \sum_{i=1}^N \sum_{j=1}^N \pi_i \sqrt{\pi_j} u_j \langle\!\langle X_i - \mu, X_j - \mu\rangle\!\rangle \langle\!\langle X_i - \mu, f\rangle\!\rangle, \\
\langle\!\langle f, L_X^\star(u)\rangle\!\rangle_\Gamma &= \langle\!\langle \Gamma f, L_X^\star(u)\rangle\!\rangle \\
&= \sum_{p=1}^P \int_{\mathcal{T}_p} (\Gamma f)^{(p)}(t_p) \left\{ \sum_{n=1}^N \sqrt{\pi_n} u_n \left(X_n^{(p)}(t_p) - \mu^{(p)}(t_p) \right) \right\} \mathrm{d}t_p \\
&= \sum_{p=1}^P \int_{\mathcal{T}_p} \left\{ \sum_{q=1}^P \int_{\mathcal{T}_q} C_{p,q}(t_p, s_q) f^{(q)}(s_q) \mathrm{d}s_q \right\} \left\{ \sum_{j=1}^N \sqrt{\pi_j} u_j \left(X_j^{(p)}(t_p) - \mu^{(p)}(t_p) \right) \right\} \mathrm{d}t_p \\
&= \sum_{i=1}^N \sum_{j=1}^N \pi_i \sqrt{\pi_j} u_j \left\{ \sum_{p=1}^P \int_{\mathcal{T}_p} \left(X_i^{(p)}(t_p) - \mu^{(p)}(t_p) \right) \left(X_j^{(p)}(t_p) - \mu^{(p)}(t_p) \right) \mathrm{d}t_p \right\} \times \\
&\qquad \left\{ \sum_{q=1}^P \int_{\mathcal{T}_q} \left(X_i^{(q)}(s_q) - \mu^{(q)}(s_q) \right) f^{(q)}(s_q) \mathrm{d}s_q \right\} \\
&= \sum_{i=1}^N \sum_{j=1}^N \pi_i \sqrt{\pi_j} u_j \langle\!\langle X_i - \mu, X_j - \mu\rangle\!\rangle \langle\!\langle X_i - \mu, f\rangle\!\rangle.
\end{aligned}$$

So, the equality (16) is proved and we conclude that $L_X^\star$ is the adjoint operator of L_X. □

A.2 Derivation of the inertia of the clouds

Recall that

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

We first show that the total inertia of the cloud of individuals $\mathcal{C}_P$ using the distance d is given by equalities in (3):

$$\sum_{n=1}^N \pi_n d^2(\mathrm{M}_n, \mathrm{G}_\mu) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \pi_i \pi_j d^2(\mathrm{M}_i, \mathrm{M}_j) = \sum_{p=1}^P \int_{\mathcal{T}_p} \mathrm{Var}\, X^{(p)}(t_p) \mathrm{d}t_p.$$

Proof.

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

The equalities in Equation (3) are shown. □

Next, we derive the total inertia of the cloud of individuals $\mathcal{C}_P$ using the distance d_Γ, given by equalities in (4):

$$\sum_{n=1}^{N} \pi_n d_\Gamma^2(\mathrm{M}_n, \mathrm{G}_\mu) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j d_\Gamma^2(\mathrm{M}_i, \mathrm{M}_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\| C_{p\cdot}(t_p, \cdot) \right\|^2 \mathrm{d}t_p.$$

Proof.

$$\begin{aligned}
\sum_{n=1}^{N} \pi_n d_\Gamma^2(\mathrm{M}_n, \mathrm{G}_\mu) &= \sum_{n=1}^{N} \pi_n \langle\!\langle X_n - \mu, X_n - \mu \rangle\!\rangle_\Gamma \\
&= \sum_{n=1}^{N} \pi_n \left\| X_n \right\|_\Gamma^2 - \left\| \mu \right\|_\Gamma^2 \\
&= \sum_{p=1}^{P} \left(\sum_{n=1}^{N} \pi_n \left\| X_n^{(p)} \right\|_\Gamma^2 - \left\| \mu^{(p)} \right\|_\Gamma^2 \right) \\
&= \sum_{p=1}^{P} \sum_{q=1}^{P} \int_{\mathcal{T}_p} \int_{\mathcal{T}_q} C_{p,q}(t_p, s_q) C_{p,q}(t_p, s_q) \mathrm{d}s_q \mathrm{d}t_p \\
&= \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\| C_{p\cdot}(t_p, \cdot) \right\|^2 \mathrm{d}t_p, \\
\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j d_\Gamma^2(\mathrm{M}_i, \mathrm{M}_j) &= \sum_{i=1}^{N} \pi_i \left\| X_i \right\|_\Gamma^2 - \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j \langle\!\langle X_i, X_j \rangle\!\rangle_\Gamma \\
&= \sum_{i=1}^{N} \pi_i \left\| X_i \right\|_\Gamma^2 - \left\| \mu \right\|_\Gamma^2 \\
&= \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\| C_{p\cdot}(t_p, \cdot) \right\|^2 \mathrm{d}t_p.
\end{aligned}$$

The equalities in Equation (4) are shown. □

Finally, we derive the inertia of the cloud of features $\mathcal{C}_N$ using the distance d, given by equalities in (5):

$$\sum_{n=1}^{N} \pi_n \mathsf{d}^2(\mathsf{M}_n, \mathsf{G}_\mu) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_i \pi_j \mathsf{d}^2(\mathsf{M}_i, \mathsf{M}_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \left\| C_{p\cdot}(t_p, \cdot) \right\|^2 \mathrm{d}t_p.$$

Proof.

$$\begin{aligned}
\sum_{n=1}^{N} \pi_n \mathsf{d}^2(\mathsf{M}_n, \mathsf{G}_\mu) &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle \\
&= \sum_{p=1}^{P}\sum_{q=1}^{P} \int_{\mathcal{T}_p}\int_{\mathcal{T}_q} C_{p,q}(t_p, s_q) C_{p,q}(t_p, s_q) \mathrm{d}s_q \mathrm{d}t_p \\
&= \sum_{p=1}^{P} \int_{\mathcal{T}_p} \|C_{p\cdot}(t_p, \cdot)\|^2 \, \mathrm{d}t_p, \\
\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j \mathsf{d}^2(\mathsf{M}_i, \mathsf{M}_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{N} \pi_i \pi_j \pi_n \langle\!\langle X_n - \mu, X_i - X_j \rangle\!\rangle \langle\!\langle X_n - \mu, X_i - X_j \rangle\!\rangle \\
&= 2\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \pi_j \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle \\
&= 2\sum_{p=1}^{P} \int_{\mathcal{T}_p} \|C_{p\cdot}(t_p, \cdot)\|^2 \, \mathrm{d}t_p.
\end{aligned}$$

The equalities in Equation (5) are shown. □

B Derivation of the eigencomponents

B.1 General framework

In this section, we calculate the relationships between the eigenelements of the covariance operator Γ and the ones of the Gram matrix $\mathbf{M}$ of a functional dataset. We then prove the equalities (10), (11) and (12).

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

$$\mathbf{M}\boldsymbol{u}_k = l_k \boldsymbol{u}_k. \tag{17}$$

Let $X = (X_1 - \mu, \ldots, X_N - \mu)^\top$ and denote $\widetilde{X} = \mathrm{diag}\{\sqrt{\pi_1}, \ldots, \sqrt{\pi_N}\}X$, the matrix of weighted observations. Recall that, in the case of P-dimensional process, the realisations of the process $X_n,\ n = 1, \cdots, N$ and μ are vectors of functions of length P, and thus X (and $\widetilde{X}$) is a matrix of functions of size $N \times P$. By left multiplying Equation (17) by $\widetilde{X}^\top$, we obtain

$$\widetilde{X}^\top \mathbf{M}\boldsymbol{u}_k = l_k \widetilde{X}^\top \boldsymbol{u}_k. \tag{18}$$

Expanding Equation (18), for each component $p = 1, \ldots, P$, we have,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \sqrt{\pi_j} [\boldsymbol{u}_k]_j \Big\{ X_i^{(p)}(\cdot) - \mu^{(p)}(\cdot) \Big\} \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle = l_k \sum_{n=1}^{N} \sqrt{\pi_n} [\boldsymbol{u}_k]_n \Big\{ X_n^{(p)}(\cdot) - \mu^{(p)}(\cdot) \Big\}. \tag{19}$$

Here and in the following, we note $[a]_p$ the pth entry of the vector a. Starting from the left side of

Equation (19), we get

$$
\begin{aligned}
[\widetilde{X}^\top \mathbf{M}\boldsymbol{u}_k]_p &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i \sqrt{\pi_j}[\boldsymbol{u}_k]_j \left\{X_i^{(p)}(\cdot) - \mu^{(p)}(\cdot)\right\} \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle \\
&= \sum_{q=1}^{P} \int_{\mathcal{T}_q} \sum_{i=1}^{N} \pi_i \left\{X_i^{(p)}(\cdot) - \mu^{(p)}(\cdot)\right\} \left\{X_i^{(q)}(s_q) - \mu^{(q)}(s_q)\right\} \\
&\qquad \times \sum_{j=1}^{N} \sqrt{\pi_j}[\boldsymbol{u}_k]_j \left\{X_j^{(q)}(s_q) - \mu^{(q)}(s_q)\right\} \mathrm{d}s_q \\
&= \sum_{q=1}^{P} \int_{\mathcal{T}_q} C_{p,q}(\cdot, s_q) \sum_{j=1}^{N} \sqrt{\pi_j}[\boldsymbol{u}_k]_j \left\{X_j^{(q)}(s_q) - \mu^{(q)}(s_q)\right\} \mathrm{d}s_q \\
&= \sum_{j=1}^{N} \langle\!\langle C_{p\cdot}(\cdot,\cdot), \sqrt{\pi_j}[\boldsymbol{u}_k]_j \{X_j - \mu\} \rangle\!\rangle \\
&= \Gamma\left(\sum_{j=1}^{N} \sqrt{\pi_j}[\boldsymbol{u}_k]_j \{X_j - \mu\}\right)^{(p)}(\cdot),
\end{aligned}
\tag{20}
$$

and, starting from the right side of Equation (19),

$$
[l_k \widetilde{X}^\top \boldsymbol{u}_k]_p = l_k \sum_{n=1}^{N} \sqrt{\pi_n}[\boldsymbol{u}_k]_n \left\{X_n^{(p)}(\cdot) - \mu^{(p)}(\cdot)\right\}. \tag{21}
$$

From Equation (20) and Equation (21), we obtain

$$
\Gamma\left(\sum_{j=1}^{N} \sqrt{\pi_j}[\boldsymbol{u}_k]_j \{X_j - \mu\}\right)^{(p)}(\cdot) = l_k \sum_{n=1}^{N} \sqrt{\pi_n}[\boldsymbol{u}_k]_n \left\{X_n^{(p)}(\cdot) - \mu^{(p)}(\cdot)\right\}, \quad \text{for all } p = 1, \ldots, P.
$$

By identification in Equation (6), we find that, for each components p,

$$
\lambda_k = l_k \quad \text{and} \quad \phi_k^{(p)}(\cdot) = \sum_{n=1}^{N} \sqrt{\pi_n}[\boldsymbol{u}_k]_n \left\{X_n^{(p)}(\cdot) - \mu^{(p)}(\cdot)\right\}, \quad k \geq 1. \tag{22}
$$

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

$$
\begin{aligned}
\|\!|\phi_k|\!\|^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} \sqrt{\pi_i \pi_j}[\boldsymbol{u}_k]_i [\boldsymbol{u}_k]_j \langle\!\langle X_i - \mu, X_j - \mu \rangle\!\rangle = \sum_{i=1}^{N} [\boldsymbol{u}_k]_i \sum_{j=1}^{N} \mathbf{M}_{ij}[\boldsymbol{u}_k]_j \\
&= \sum_{i=1}^{N} [\boldsymbol{u}_k]_i l_k [\boldsymbol{u}_k]_i = l_k \left(\!(\boldsymbol{u}_k)\!\right)^2 = l_k.
\end{aligned}
$$

Therefore, in order to have an orthonormal basis of eigenfunctions, we normalise the eigenfunctions ϕ_k from Equation (22) by $1/\sqrt{l_k}$. Concerning the estimation of the scores, for $n = 1, \ldots, N$, for $k \geq 1$, we have

$$
\begin{aligned}
\mathfrak{c}_{nk} = \langle\!\langle X_n - \mu, \phi_k \rangle\!\rangle &= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} \sqrt{\pi_j}[\boldsymbol{u}_k]_j \langle\!\langle X_n - \mu, X_j - \mu \rangle\!\rangle \\
&= \frac{1}{\sqrt{l_k \pi_n}} \sum_{j=1}^{N} [\boldsymbol{u}_k]_j \mathbf{M}_{nj} = \sqrt{\frac{l_k}{\pi_n}}[\boldsymbol{u}_k]_n.
\end{aligned}
$$

If we assume that the observations are equally weighted, i.e., $\pi_n = 1/N, n = 1, \ldots, N$, we get the equalities (10), (11) and (12).

C More results

C.1 Simulation

C.2 Application

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.

Berrendero, J. R., Bueno-Larraz, B., and Cuevas, A. (2020). On Mahalanobis Distance in Functional Settings. *Journal of Machine Learning Research*, 21(9):1–33.

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.

De la Cruz, O. and Holmes, S. (2011). The Duality Diagram in Data Analysis: Examples of Modern Applications. *The annals of applied statistics*, 5(4):2266–2277.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Eilers, P. H. C., Marx, B. D., and Durbán, M. (2015). Twenty years of P-splines. *SORT-Statistics and Operations Research Transactions*, 39(2):149–186.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. CRC Press.

Golovkine, S., Gunning, E., Simpkin, A. J., and Bargary, N. (2023). On the estimation of the number of components in multivariate functional principal component analysis.

González, J. and Muñoz, A. (2010). Representing functional data in reproducing Kernel Hilbert Spaces with applications to clustering and classification. *DES - Working Papers. Statistics and Econometrics. WS*, (ws102713).

Grith, M., Wagner, H., Härdle, W. K., and Kneip, A. (2018). Functional Principal Component Analysis for Derivatives of Multivariate Curves. *Statistica Sinica*, 28(4):2469–2496.

Hall, P., Kay, J. W., and Titterington, D. M. (1990). Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression. *Biometrika*, 77(3):521–528.

Hall, P. and Marron, J. S. (1990). On Variance Estimation in Nonparametric Regression. *Biometrika*, 77(2):415–419.

Happ, C. and Greven, S. (2018). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, 113(522):649–659.

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

Holmes, S. (2008). Multivariate data analysis: The French way. *Probability and Statistics: Essays in Honor of David A. Freedman*, 2:219–234.

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.

Martino, A., Ghiglietti, A., Ieva, F., and Paganoni, A. M. (2019). A k-means procedure based on a Mahalanobis type distance for clustering multivariate functional data. *Statistical Methods & Applications*, 28(2):301–322.

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

Prothero, J., Hannig, J., and Marron, J. S. (2023). New Perspectives on Centering. *The New England Journal of Statistics in Data Science*, pages 1–21.

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.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC Press.

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.

Wong, R. K. W. and Zhang, X. (2019). Nonparametric operator-regularized covariance function estimation for functional data. *Computational Statistics & Data Analysis*, 131:131–144.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321.