A geometric interpretation of the multivariate functional principal components analysis

Steven Golovkine

November 15, 2022

Abstract

Your abstract.

1 Introduction

We aim to provide a geometric understanding of the multivariate functional principal components analysis (MPFCA) and develop visualisation, as well as interpretation, for this method.

This analysis of biomechanical data shows how the geometric interpretation could be useful in some practical case. These data contains multiple functional variables. Data are recorded on diverse part of the body (hip, knee and ankle) and within several angle (flexion, abduction and rotation) for several runners.

We might also add quantitative variables, such as sex, or qualitative variables, such as age, as supplementary variables in the analysis.

Factorial analysis have been initially introduced to analysed observations described by both quantitative and qualitative variables [4, 10, 7] as an extension of principal components analysis for quantitative variables and multiple correspondence analysis for qualitative variables.

The remainder of the paper is organized as follows. ...

2 Ideas

- Make the connection between [6] and [2], [12]?
- This is related to the fact that the eigenvalues decomposition of $X^\top X$ and related to the eigenvalues decomposition of $XX^\top$. They have the same eigenvalues and there exists transition formulas. Gram matrix for this type of data. Let X be a $N \times P$ matrix, N number of observations and P number of features. Let $V = N^{-1}X^\top X$ be the covariance matrix of the data and $W = N^{-1}XX^\top$ the associated Gram matrix. The eigendecomposition of V and W are given by

$$\begin{aligned} Vu &= \lambda u \quad \text{s.t.} \quad \|u\| = 1, \\ Wv &= \rho v \quad \text{s.t.} \quad \|v\| = 1. \end{aligned}$$

  By replacing, V and W by their expressions using X, we found out that $\lambda = \rho$, $v = (N\lambda)^{-1/2}Xu$ and $u = (N\lambda)^{-1/2}X^\top v$.
- Concerning the estimation of functional principal components from the sample covariance matrix see [9, Chap. 8.4] for univariate fPCA and [6] for multivariate fPCA.
- Multiple correspondence analysis is equivalent to principal composant analysis of a transformed complete disjunctive table, see [8].
- How is this related to kernel PCA?
- Variance of a functional observation:

$$Var(X) = \mathbb{E}\{(X - \mathbb{E}(X))^2\} = \frac{1}{N}\sum_{n=1}^{N}(X_i(t) - \bar{X}(t))^2$$

- Represent the components in a correlation circle where two components are close from one another if they exhibits the same variation in the space of principal components.
- Geometric of the different principal components.
- Comments on the geometric interpretation after the expansion in a common basis of functions such as B-splines or Fourier.
- "[...] in multivariate analysis a substantial interpretation of principal components is often difficult and has to be based on vague arguments concerning the correlation of principal components with original variables. Such a problem does not at all exists in the functional context, where $\gamma_1(t), \gamma_2(t), \ldots$ are functions representing the major modes of variation of $X_i(t)$ over t." [1]. It might however not be the case for multivariate functional data as the setting is more similar to the multivariate data.
- "Our approach for estimating principal components is motivated by the well-known duality relation between row and column spaces of a data matrix [...]" [1]. They provide the relationships to go from one basis to another, it might be extend to multivariate functional data. [1] use both the Gram matrix and the estimation of the covariance operator by numerical integration to estimate the eigencomponents.

3 Model

The structure of our data, referred to as *multivariate functional data*, is very similar to that presented in [6]. 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 $\langle \cdot, \cdot \rangle_2$, 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 $\langle\!\langle \cdot, \cdot \rangle\!\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$,

$$\langle\!\langle f, g \rangle\!\rangle := \sum_{p=1}^{P} \left\langle f^{(p)}, g^{(p)} \right\rangle_2 = \sum_{p=1}^{P} \int_{\mathcal{T}_p} f^{(p)}(t_p) g^{(p)}(t_p) \mathrm{d}t_p, \quad f, g \in \mathcal{H}.$$

$\mathcal{H}$ is a Hilbert space with respect to the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ [6]. 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, $\mu(\mathbf{t}) := \mathbb{E}(X(\mathbf{t})), \mathbf{t} \in \mathcal{T}$.

Following [6], we also consider a generalization to a weighted scalar product on $\mathcal{H}$:

$$\langle\!\langle f, g \rangle\!\rangle_w := \sum_{p=1}^{P} w_p \left\langle f^{(p)}, g^{(p)} \right\rangle_2, \quad f, g \in \mathcal{H},$$

for some positive weights $w_1, \ldots, w_p$.

Let us consider N curves $X_1, \ldots, X_N$ generated as a random sample of the P-dimensional stochastic process X with continuous trajectories. For now, we assume that the curves are observed without error. From a practical point of view, the data can be viewed as a $N \times P$-table where each entry is a curve, eventually multidimensional. So, at each intersection of row n and column p, we have $X_n^{(p)}$ which is the curve of the feature p for the individual n. Each row represents an observation. Each column represents a functional variable. For a given $p \in \{1, \ldots, P\}$, we denote by $\mu_p(t_p)$ the mean curve of the component along the N observations. This quantity can be estimated by

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

4 A geometric point of view of MFPCA

4.1 Cloud of individuals

Let $\{X_n^{(p)}(t), p = 1, \ldots, P\}$ be the features set for a particular observation n. We identify this set as the point M_n in the space $\mathcal{H} := \mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_P)$. The set $\mathcal{H}$ is refered as the observations'

space. The cloud of points that represented the set of observations is denoted by $\mathcal{C}_N$. We can add weight p_n to each of the observation, such that $\sum_n p_n = 1$, *e.g* $p_n = 1/N$. Let G_N be the centre of gravity of the cloud $\mathcal{C}_N$. In the space $\mathcal{H}$, its coordinates are given by $\{\mu^{(p)}(t), p = 1, \dots, P\}$. If the variables are centered, the origin $\mathcal{O}_{\mathcal{H}}$ of the axes in $\mathcal{H}$ coincide with G_N.

Consider M_f and M_g, two points in $\mathcal{C}_N$ and denote by f and g the two elements of $\mathcal{H}$ representing by M_f and M_g. The distance between these observations is defined as

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

This distance measures how different the observations are, and thus charaterized the shape of the cloud $\mathcal{C}_N$. Another description of this shape is to consider the distance between each point M_n and the center of the cloud G_N. Denoting f_n the element of $\mathcal{H}$ related to M_n and f the element of $\mathcal{H}$ related to G_N, the distance between M_n and G_N is given by

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

The total inertia of $\mathcal{C}_N$, with respect to G_N, is given by

$$\sum_{n=1}^{N} p_n d^2(f_n,f) = \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j d^2(f_i,f_j) = \sum_{p=1}^{P} \int_{\mathcal{T}_p} \operatorname{Var} f^{(p)}(t) dt.$$

****

Here, we have $\operatorname{Var} f^{(p)}(t) = \mathbb{E}(f^{(p)}(t)^2) - \mathbb{E}(f^{(p)}(t))^2$. In the case of regularly sampled smoothed functional data, it can be estimated using

$$\operatorname{Var} f^{(p)}(t) = \frac{1}{N} \sum_{n=1}^{N} f_n^{(p)}(t)^2 - \mu^{(p)}(t)^2 \quad \text{where} \quad \mu^{(p)}(t) = \frac{1}{N} \sum_{n=1}^{N} f_n^{(p)}(t)$$

$$\begin{aligned}
\sum_{n=1}^{N} p_n d^2(f_n,f) &= \sum_{n=1}^{N} p_n \sum_{p=1}^{P} \left(\left\| f_n^{(p)} - \mu^{(p)} \right\| \right)^2 \\
&= \sum_{p=1}^{P} \left(\sum_{n=1}^{N} p_n \left\| f_n^{(p)} \right\|^2 - \left\| \mu^{(p)} \right\|^2 \right) \\
&= \sum_{p=1}^{P} \int_{\mathcal{T}_p} \operatorname{Var} f^{(p)}(t) dt
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j d^2(f_i,f_j) &= \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \sum_{p=1}^{P} \left\| f_i^{(p)} - f_j^{(p)} \right\|^2 \\
&= \sum_{p=1}^{P} \left(2 \sum_{i=1}^{N} p_i \left\| f_i^{(p)} \right\|^2 - 2 \sum_{i=1}^{N}\sum_{j=1}^{N} \left\langle f_i^{(p)}, f_j^{(p)} \right\rangle_2 \right) \\
&= \sum_{p=1}^{P} \left(2 \sum_{i=1}^{N} p_i \left\| f_i^{(p)} \right\|^2 - 2 \left\| \mu^{(p)} \right\|^2 - 2 \sum_{i=1}^{N}\sum_{j=1}^{N} \left\langle f_i^{(p)}, f_j^{(p)} \right\rangle_2 + 2 \left\| \mu^{(p)} \right\|^2 \right) \\
&= 2 \sum_{p=1}^{P} \int_{\mathcal{T}_p} \operatorname{Var} f^{(p)}(t) dt
\end{aligned}$$

****

Remark 1 *This results have the same interpretation than in $\mathbb{R}^n$. 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 thus not interested by the total inertia but mostly how this variance is spread among the variables.*

4.2 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

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

For now, the cloud of variables is not clear!

4.3 On centering and reducing

In MFPCA, the components are usually assumed centred [6]. 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 $\mathcal{O}_\mathcal{H}$ 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*

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

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 litterature. [6] propose to use $w_k = (\int_{\mathcal{T}_k} \operatorname{Var} X^{(k)}(t)dt)^{-1/2}$, while [3] consider the function $w_k(t) = (\operatorname{Var} X^{(k)}(t))^{-1/2}$.

5 Functional principal components analysis

5.1 Univariate case

For now, we will consider the univariate case ($P = 1$), and estimate the principal components. Assume the existence of a continuous covariance function

$$C(s, t) = \mathbb{E}(\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}), \quad s, t \in \mathcal{T}_0.$$

The covariance operator of the process X is given by

$$\Gamma f(\cdot) = \int C(s, \cdot) f(s)ds, \quad f \in \mathcal{L}^2(\mathcal{T}_0).$$

Let $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\phi_1, \phi_2, \ldots$ be the eigenvalues and eigenfunctions of the covariance operator Γ. The set $\phi = \{\phi_k\}_{k \geq 1}$ forms a complete orthonormal basis of $\mathcal{L}^2(\mathcal{T}_0)$. Note that ϕ contains an infinite number of elements. Using the Karhunen-Loève decomposition, we get

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \mathfrak{c}_k \phi_k(t), \quad t \in \mathcal{T}_0,$$

where $\mathfrak{c}_k = \langle X - \mu, \phi_k \rangle_2$ 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 $\mathbb{E}(\mathfrak{c}_k \mathfrak{c}_r) = 0$ for $k \neq r$.

Assume we observe N realizations $X_1, \ldots, X_N$ of the process X. Estimators of the mean and covariance function are given by

$$\widehat{\mu}(t) = \frac{1}{N} \sum_{n=1}^{N} X_i(t) \quad \text{and} \quad \widehat{C}(s,t) = \frac{1}{N} \sum_{i=1}^{N} \left(X_i(t) - \widehat{\mu}(t)\right) \left(X_i(s) - \widehat{\mu}(s)\right).$$

And thus, the estimator of the covariance operator follows as

$$\widehat{\Gamma}_N f(\cdot) = \int \widehat{C}(s,\cdot) f(s) ds, \quad f \in \mathcal{L}^2(\mathcal{T}_0).$$

****

Assume that there exists a basis of functions $\{\psi_k\}_{1 \leq k \leq K}$ such that the curves can be extended into this basis

$$X(t) = \sum_{k=1}^{K} c_k \psi_k(t), \quad t \in \mathcal{T}_0.$$

The mean function is given by

$$\mu(t) = \sum_{k=1}^{K} \mathbb{E}(c_k) \psi_k(t), \quad t \in \mathcal{T}_0.$$

The covariance function is given by

$$C(s,t) = \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\mathbb{E}(c_k c_l) - \mathbb{E}(c_k)\mathbb{E}(c_l)\right) \psi_k(s) \psi_l(t), \quad s,t \in \mathcal{T}_0.$$

Using N realizations of the process, we have $X(t) = C\Psi(t)$ where

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_N(t) \end{pmatrix}, \quad C = \begin{pmatrix} c_{11} & \cdots & c_{1K} \\ \vdots & \ddots & \vdots \\ c_{N1} & \cdots & c_{NK} \end{pmatrix} \quad \text{and} \quad \Psi(t) = \begin{pmatrix} \psi_1(t) \\ \vdots \\ \psi_K(t) \end{pmatrix}.$$

The estimation of the mean and covariance functions are given by

$$\widehat{\mu}(t) = \frac{1}{N} \mathbb{1}_N^\top C \Psi(t) \quad \text{and} \quad \widehat{C}(s,t) = \frac{1}{N} \Psi(s)^\top C^\top C \Psi(t).$$

We denote by W the matrix of inner products of the functions of the basis Ψ. The entries of W are given by

$$W_{kl} = \langle \psi_k, \psi_l \rangle_2, \quad 1 \leq k, l \leq K.$$

5.1.1 By diagonalization of the covariance operator

From [9]. The estimation of the eigenvalues and eigenfunctions of the covariance operator is usually performed by estimating the covariance surface on a fine grid and diagonalize it. Let $H = (t_1, \ldots, t_m)$ a grid. In that case, the empirical covariance matrix $\widehat{C}$ is of size $m \times m$. It results to sets of eigenvalues $(\rho_1, \ldots, \rho_m)$ and eigenvectors $(u_1, \ldots, u_m)$ such that

$$\widehat{C} u = \rho u. \tag{1}$$

Given ϕ an eigenfunction of the covariance operator Γ, let $\widetilde{\phi}$ be the vector of length m with entries $\phi(t)$ for $t \in H$. Then, for each $t \in H$,

$$\widehat{\Gamma}\phi(s) = \int \widehat{C}(s,t)\phi(t)dt \approx \frac{|\mathcal{T}_0|}{|H|} \sum_{k=1}^{m} \widehat{C}(s,t_k)\widetilde{\phi}_k.$$

The equation $\Gamma\phi = \lambda\phi$ has the approximate discrete form

$$\frac{|\mathcal{T}_0|}{|H|} \widehat{C}\widetilde{\phi} = \lambda\widetilde{\phi}. \tag{2}$$

The solutions of the equations (1) and (2) are the same with eigenvalues $\lambda = \frac{|\mathcal{T}_0|}{|H|}\rho$. To approximate the eigenfunction ϕ from $\widetilde{\phi}$, we use interpolation techniques.

****

Using the expansion of the realizations of the process into the basis of functions Ψ. The eigenfunctions of the estimation of the covariance operator can also be expanded in Ψ

$$\widehat{\phi}_k(t) = \sum_{l=1}^{K} b_{kl}\psi_l(t) = \Psi(t)^\top b_k, \quad \text{where} \quad b_k = (b_1, \ldots, b_K)^\top.$$

Then,

$$\begin{aligned}
\widehat{\Gamma}_N\widehat{\phi}(t) &= \int \widehat{C}(t,s)\widehat{\phi}(s)ds \\
&= \int \frac{1}{N}\Psi(t)^\top C^\top C\Psi(s)\Psi(s)^\top b ds \\
&= \frac{1}{N}\Psi(t)^\top C^\top CWb \\
&= \lambda\Psi(t)^\top b.
\end{aligned}$$

As this relationship should be true for all t, we get

$$\frac{1}{N}C^\top CWb = \lambda b.$$

5.1.2 By diagonalization of the inner product matrix

In this section, we use the duality relation between row and column spaces of a data matrix (from [1]). Consider the matrix M with entries

$$M_{ij} = \langle X_i - \widehat{\mu}, X_j - \widehat{\mu}\rangle_2, \quad i,j = 1,\ldots,N.$$

The relationship between all nonzero eigenvalues $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots$ of the empirical covariance operator $\widehat{\Gamma}_N$ and the eigenvalues $l_1 \geq l_2 \geq \cdots$ of the matrix M is given by

$$\widehat{\lambda}_k = \frac{l_k}{N}, \quad k = 1,2,\ldots.$$

Let $\widehat{\phi}_1, \widehat{\phi}_2, \ldots$ be the eigenfunctions of $\widehat{\Gamma}_N$ and $v_1, v_2, \ldots$ be the orthonormal eigenvectors of M. These quantities are linked by

$$\widehat{\phi}_k(\cdot) = \frac{1}{\sqrt{l_k}}\sum_{i=1}^{N} v_{ik}(X_i(\cdot) - \widehat{\mu}(\cdot)), \quad k = 1,2,\ldots$$

The empirical scores are given by

$$\widehat{\mathfrak{c}}_{ik} = \sqrt{l_k}v_{ik}, \quad k = 1,2,\ldots$$

****

For the eigenvalues and the eigenfunctions. For $k = 1,2,\ldots$, we have

$$Mv_k = l_k v_k. \tag{3}$$

Let $X = (X_1(\cdot) - \widehat{\mu}(\cdot) \cdots X_N(\cdot) - \widehat{\mu}(\cdot))^\top$. By left multiplying (15) by $X^\top$, we obtain

$$X^\top Mv_k = l_k X^\top v_k. \tag{4}$$

Then, we have

$$\begin{aligned}
X^\top Mv_k &= \sum_{i=1}^{N}(X_i(\cdot) - \widehat{\mu}(\cdot))\sum_{j=1}^{N}\langle X_i(\cdot) - \widehat{\mu}(\cdot), X_j(\cdot) - \widehat{\mu}(\cdot)\rangle_2 v_{jk} \\
&= \int_{\mathcal{T}_0}\sum_{i=1}^{N}(X_i(\cdot) - \widehat{\mu}(\cdot))(X_i(s) - \widehat{\mu}(s))\sum_{j=1}^{N}(X_j(s) - \widehat{\mu}(s))v_{jk}ds \\
&= \int_{\mathcal{T}_0} N\widehat{C}(\cdot,s)\sum_{j=1}^{N}(X_j(s) - \widehat{\mu}(s))v_{jk}ds \\
&= N\widehat{\Gamma}_n\left(\sum_{j=1}^{N}(X_j(\cdot) - \widehat{\mu}(\cdot))v_{jk}\right)(\cdot)
\end{aligned}$$

and

$$l_k X^\top v_k = l_k \sum_{i=1}^{N} \left(X_i(\cdot) - \widehat{\mu}(\cdot)\right) v_{ik}.$$

From (9), we get

$$\widehat{\Gamma}_n \left(\sum_{j=1}^{N} \left(X_j(\cdot) - \widehat{\mu}(\cdot)\right) v_{jk} \right)(t) = \frac{l_k}{N} \sum_{i=1}^{N} \left(X_i(t) - \widehat{\mu}(t)\right) v_{ik}, \quad \text{for all } t \in \mathcal{T}_0.$$

By identification, we find that

$$\widehat{\lambda}_k = \frac{l_k}{N} \quad \text{and} \quad \widehat{\phi}_k(\cdot) = \sum_{i=1}^{N} v_{ik}(X_i(\cdot) - \widehat{\mu}(\cdot)), \quad k = 1, 2, \ldots$$

For $k = 1, 2, \ldots$, we have

$$\begin{aligned}
\left\|\widehat{\phi}_k\right\|^2 &= \left\langle \sum_{i=1}^{N} v_{ik}(X_i(\cdot) - \widehat{\mu}(\cdot)), \sum_{j=1}^{N} v_{jk}(X_j(\cdot) - \widehat{\mu}(\cdot)) \right\rangle_2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik}v_{jk} \left\langle (X_i(\cdot) - \widehat{\mu}(\cdot)), (X_j(\cdot) - \widehat{\mu}(\cdot)) \right\rangle_2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik}v_{jk} \int_{\mathcal{T}_0} (X_i(s) - \widehat{\mu}(s))(X_j(s) - \widehat{\mu}(s)) ds \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik}v_{jk} M_{ij} \\
&= \sum_{i=1}^{N} v_{ik} \sum_{j=1}^{N} M_{ij} v_{jk} \\
&= \sum_{i=1}^{N} v_{ik} l_k v_{ik} \\
&= l_k \left\|v_k\right\|^2 \\
&= l_k
\end{aligned}$$

We normalise $\widehat{\phi}_k$ by $1/\sqrt{l_k}$ to have an orthonormal basis.

For the scores. For $k = 1, 2, \ldots$, we have

$$\begin{aligned}
\widehat{\mathfrak{c}}_{ik} &= \left\langle X_i(\cdot) - \widehat{\mu}(\cdot), \widehat{\phi}_k(\cdot) \right\rangle_2 \\
&= \left\langle X_i(\cdot) - \widehat{\mu}(\cdot), \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk}(X_j(\cdot) - \widehat{\mu}(\cdot)) \right\rangle_2 \\
&= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} \left\langle X_i(\cdot) - \widehat{\mu}(\cdot), X_j(\cdot) - \widehat{\mu}(\cdot) \right\rangle_2 \\
&= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} M_{ij} \\
&= \frac{1}{\sqrt{l_k}} (M v_k)_i \\
&= \frac{1}{\sqrt{l_k}} l_k v_{ik} \\
&= \sqrt{l_k} v_{ik}
\end{aligned}$$

****

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

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

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

$$\widehat{\phi}_k(t) = \sum_{l=1}^{K} b_{kl}\psi_l(t) = \Psi(t)^\top b_k, \quad \text{where} \quad b_k = (b_{k1}, \ldots, b_{kK})^\top .$$

We have, for all $t \in \mathcal{T}_0$,

$$\begin{aligned}
\widehat{\Gamma}_N\phi_k(t) &= \int_{\mathcal{T}_0} \widehat{C}(t,s)\phi_k(s)ds \\
&= \frac{1}{N}\int_{\mathcal{T}_0} \Psi(t)^\top C^\top C\Psi(s)\Psi(s)^\top b_k ds \\
&= \frac{1}{N}\Psi(t)^\top C^\top C \int_{\mathcal{T}_0} \Psi(s)\Psi(s)^\top ds b_k \\
&= \frac{1}{N}\Psi(t)^\top C^\top C W b_k.
\end{aligned}$$

From the eigenequation, we have that

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

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

$$C^\top C W b_k = N\lambda_k b_k \tag{5}$$

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 (13), we obtain

$$W^{1/2}C^\top C W^{1/2}u_k = N\lambda_k u_k \iff \left(CW^{1/2}\right)^\top \left(CW^{1/2}\right) u_k = N\lambda_k u_k. \tag{6}$$

From the eigendecomposition of the matrix M, we get

$$Mv_k = l_k v_k \iff \left(CW^{1/2}\right)\left(CW^{1/2}\right)^T v_k = l_k v_k. \tag{7}$$

The equations (14) and (15) are eigenequations in the classical PCA case, with the duality $X^\top X$ and $XX^\top$. Following [8], we find that, for $1 \le k \le K$,

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

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

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

5.2 Multivariate case

Here, we will consider the multivariate case. Assume that X is a p-dimensional process. We aim to estimate the principal components. Assume the existence of a continuous covariance function

$$C(s,t) = \mathbb{E}(\{X(s) - \mu(s)\} \otimes \{X(t) - \mu(t)\}), \quad s,t \in \mathcal{T},$$

where each entry is given by, for all $p, q = 1, \ldots, P$,

$$C_{pq}(s,t) = \mathbb{E}(\{X^{(p)}(s) - \mu^{(p)}(s)\}\{X^{(q)}(t) - \mu^{(q)}(t)\}), \quad s,t \in \mathcal{T}_0.$$

The covariance operator of the process X is denoted by Γ where each entry is given by

$$(\Gamma f)^{(p)}(\cdot) = \sum_{q=1}^{P} \int_{\mathcal{T}_0} C_{pq}(s,\cdot) f^{(q)}(s)ds, \quad f \in \mathcal{L}^2(\mathcal{T}_0), \quad p = 1, \ldots, P.$$

Let $\lambda_1 \geq \lambda_2 \geq \dots$ and $\phi_1, \phi_2, \dots$ be the eigenvalues and eigenfunctions of the covariance operator Γ. The set $\phi = \{\phi_k\}_{k\geq 1}$ forms a complete orthogonale basis of $\mathcal{H}$. Note that ϕ contains an infinite number of elements. Using the multivariate Karhunen-Loève theorem [6], we get

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

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 $\mathbb{E}(\mathfrak{c}_k \mathfrak{c}_r) = 0$ for $k \neq r$.

Assume we observe N realizations $X_1, \dots, X_N$ of the process X. Estimators of the mean and covariance function are given by

$$\widehat{\mu}^{(p)}(t) = \frac{1}{N} \sum_{n=1}^{N} X_i^{(p)}(t) \quad \text{and} \quad \widehat{C}_{pq}(s,t) = \frac{1}{N} \sum_{i=1}^{N} \left(X_i^{(p)}(t) - \widehat{\mu}^{(p)}(t) \right) \left(X_i^{(q)}(s) - \widehat{\mu}^{(q)}(s) \right), \quad p, q = 1, \dots, P.$$

And thus, the estimator of the covariance operator follows as

$$(\widehat{\Gamma}_N f)^{(p)}(\cdot) = \sum_{q=1}^{P} \int_{\mathcal{T}_0} \widehat{C}_{pq}(s, \cdot) f^{(q)}(s) ds, \quad f \in \mathcal{H}.$$

****

Assume that for all component $p = 1, \dots, P$, there exists a basis of functions $\{\psi_k^{(p)}\}_{1 \leq k \leq K_p}$ such that each component of the curves can be extended into this basis

$$X^{(p)}(t) = \sum_{k=1}^{K_p} c_{pk} \psi_k^{(p)}(t), \quad t \in \mathcal{T}_0.$$

The pth component of the mean function is given by

$$\mu^{(p)}(t) = \sum_{k=1}^{K_p} \mathbb{E}(c_{pk}) \psi_k^{(p)}(t), \quad t \in \mathcal{T}_0.$$

The (p,q)th entry of the covariance function is given by

$$C_{pq}(s,t) = \sum_{k=1}^{K_p} \sum_{l=1}^{K_q} \left(\mathbb{E}(c_{pk} c_{ql}) - \mathbb{E}(c_{pk}) \mathbb{E}(c_{ql}) \right) \psi_k^{(p)}(s) \psi_l^{(q)}(t), \quad s, t \in \mathcal{T}_0.$$

Using N realizations of the process, we have $X(t) = C\Psi(t)$ where

$$X(t) = \begin{pmatrix} X_1^{(1)}(t) & \cdots & X_1^{(P)}(t) \\ \vdots & \ddots & \vdots \\ X_N^{(1)}(t) & \cdots & X_N^{(P)}(t) \end{pmatrix},$$

$$C = \begin{pmatrix} C^{(1)} & \cdots & C^{(P)} \end{pmatrix}, \quad \text{with} \quad C^{(p)} = \begin{pmatrix} c_{1p1} & \cdots & c_{1pK_p} \\ \vdots & \ddots & \vdots \\ c_{Np1} & \cdots & c_{NpK_p} \end{pmatrix}$$

$$\Psi(t) = \begin{pmatrix} \psi^{(1)}(t) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \psi^{(P)}(t) \end{pmatrix}, \quad \text{with} \quad \Psi^{(p)}(t) = \begin{pmatrix} \psi_1^{(p)}(t) \\ \vdots \\ \psi_{K_p}^{(p)}(t) \end{pmatrix}.$$

The estimation of the mean and covariance functions are given by

$$\widehat{\mu}(t) = \frac{1}{N} \mathbb{1}_N^\top C \Psi(t) \quad \text{and} \quad \widehat{C}(s,t) = \frac{1}{N} \Psi(s)^\top C^\top C \Psi(t).$$

We denote by W the matrix of inner products of the basis Ψ. The entries of W are given by

$$W_{k,l}^{(p)} = \left\langle \psi_k^{(p)}, \psi_l^{(p)} \right\rangle_2, \quad 1 \leq k, l \leq K_p, \quad 1 \leq p \leq P.$$

5.2.1 By diagonalization of the covariance operator

See [6].

5.2.2 By diagonalization of the inner product matrix

In this section, we use the duality relation between row and column spaces of a data matrix to perform the MFPCA. Consider the matrix M with entries

$$M_{ij} = \langle\langle X_i - \widehat{\mu}, X_j - \widehat{\mu} \rangle\rangle, \quad i, j = 1, \ldots, N.$$

The relationship between all nonzero eigenvalues $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots$ of the empirical covariance operator $\widehat{\Gamma}_N$ and the eigenvalues $l_1 \geq l_2 \geq \cdots$ of the matrix M is given by

$$\widehat{\lambda}_k = \frac{l_k}{N}, \quad k = 1, 2, \ldots.$$

Let $\widehat{\phi}_1, \widehat{\phi}_2, \ldots$ be the eigenfunctions of $\widehat{\Gamma}_N$ and $v_1, v_2, \ldots$ be the orthonormal eigenvectors of M. These quantities are linked by

$$\widehat{\phi}_k(\cdot) = \sum_{i=1}^{N} v_{ik}(X_i(\cdot) - \widehat{\mu}(\cdot)), \quad k = 1, 2, \ldots$$

The empirical scores are given by

$$\widehat{\mathfrak{c}}_{ik} = \sqrt{l_k} v_{ik}, \quad k = 1, 2, \ldots$$

****

For the eigenvalues and the eigenfunctions. For $k = 1, 2, \ldots$, we have

$$M v_k = l_k v_k. \tag{8}$$

Let $X = (X_1(\cdot) - \widehat{\mu}(\cdot) \cdots X_N(\cdot) - \widehat{\mu}(\cdot))^\top$. Recall that, in the case of P-dimensional process, the realisations $X_i, i = 1, \cdots, N$ and $\widehat{\mu}$ are vectors of function of size P, and thus X is a matrix of size $N \times P$. By left multiplying (8) by $X^\top$, we obtain

$$X^\top M v_k = l_k X^\top v_k. \tag{9}$$

Then, we have

$$X^\top M v_k = \begin{pmatrix} \sum_{i=1}^N \sum_{j=1}^N (X_i^{(1)} - \widehat{\mu}^{(1)}) \langle\langle X_i - \widehat{\mu}, X_j - \widehat{\mu} \rangle\rangle v_{jk} \\ \vdots \\ \sum_{i=1}^N \sum_{j=1}^N (X_i^{(P)} - \widehat{\mu}^{(P)}) \langle\langle X_i - \widehat{\mu}, X_j - \widehat{\mu} \rangle\rangle v_{jk} \end{pmatrix}$$

and

$$l_k X^\top v_k = \begin{pmatrix} l_k \sum_{i=1}^N (X_i^{(1)} - \widehat{\mu}^{(1)}) v_{ik} \\ \cdots \\ l_k \sum_{i=1}^N (X_i^{(P)} - \widehat{\mu}^{(P)}) v_{ik} \end{pmatrix}$$

Thus, for all $p = 1, \cdots, P$, we have

$$\sum_{i=1}^N \sum_{j=1}^N (X_i^{(p)} - \widehat{\mu}^{(p)}) \langle\langle X_i - \widehat{\mu}, X_j - \widehat{\mu} \rangle\rangle v_{jk} = l_k \sum_{i=1}^N (X_i^{(p)} - \widehat{\mu}^{(p)}) v_{ik} \tag{10}$$

Starting from the left side of Equation (10), we get

$$
\begin{aligned}
[X^\top M v_k]_p &= \sum_{i=1}^{N}\sum_{j=1}^{N}(X_i^{(p)} - \widehat{\mu}^{(p)})\langle\!\langle X_i - \widehat{\mu}, X_j - \widehat{\mu}\rangle\!\rangle v_{jk} \qquad (11)\\
&= \sum_{k=1}^{P}\int_{\mathcal{T}_0}\sum_{i=1}^{N}\left(X_i^{(p)}(\cdot) - \widehat{\mu}^{(p)}(\cdot)\right)\left(X_i^{(k)}(s) - \widehat{\mu}^{(k)}(s)\right)\sum_{j=1}^{N}\left(X_j^{(k)}(s) - \widehat{\mu}^{(k)}(s)\right)v_{jk}ds\\
&= \sum_{k=1}^{P}\int_{\mathcal{T}_0} NC_{pk}(\cdot,s)\sum_{j=1}^{N}\left(X_j^{(k)}(s) - \widehat{\mu}^{(k)}(s)\right)v_{jk}ds\\
&= N\langle\!\langle C_{p\cdot}(\cdot,\cdot), \sum_{j=1}^{N}\left(X_j(s) - \widehat{\mu}(s)\right)v_{jk}\rangle\!\rangle\\
&= N\widehat{\Gamma}_N\left(\sum_{j=1}^{N}\left(X_j(\cdot) - \widehat{\mu}(\cdot)\right)v_{jk}\right)^{(p)}(\cdot)
\end{aligned}
$$

and

$$
[l_k X^\top v_k]_p = l_k\sum_{i=1}^{N}\left(X_i^{(p)}(\cdot) - \widehat{\mu}^{(p)}(\cdot)\right)v_{ik}. \qquad (12)
$$

From (11) and (12), we get

$$
\widehat{\Gamma}_N\left(\sum_{j=1}^{N}\left(X_j(\cdot) - \widehat{\mu}(\cdot)\right)v_{jk}\right)^{(p)}(\cdot) = \frac{l_k}{N}\sum_{i=1}^{N}\left(X_i^{(p)}(\cdot) - \widehat{\mu}^{(p)}(\cdot)\right)v_{ik}, \quad \text{for all } t\in\mathcal{T}_0, \quad p = 1,\ldots,P.
$$

By identification, we find that

$$
\widehat{\lambda}_k = \frac{l_k}{N} \quad \text{and} \quad \widehat{\phi}_k^{(p)}(\cdot) = \sum_{i=1}^{N} v_{ik}(X_i^{(p)}(\cdot) - \widehat{\mu}^{(p)}(\cdot)), \quad k = 1,2,\ldots
$$

For $k = 1,2,\ldots$, we have

$$
\begin{aligned}
\left\|\widehat{\phi}_k\right\|^2 &= \langle\!\langle\sum_{i=1}^{N} v_{ik}(X_i(\cdot) - \widehat{\mu}(\cdot)), \sum_{j=1}^{N} v_{jk}(X_j(\cdot) - \widehat{\mu}(\cdot))\rangle\!\rangle\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik}v_{jk}\langle\!\langle X_i(\cdot) - \widehat{\mu}(\cdot), X_j(\cdot) - \widehat{\mu}(\cdot)\rangle\!\rangle\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ik}v_{jk}M_{ij}\\
&= \sum_{i=1}^{N} v_{ik}\sum_{j=1}^{N} M_{ij}v_{jk}\\
&= \sum_{i=1}^{N} v_{ik}l_k v_{ik}\\
&= l_k\left\|v_k\right\|^2\\
&= l_k
\end{aligned}
$$

We normalise $\widehat{\phi}_k$ by $1/\sqrt{l_k}$ to have an orthonormal basis.

For the scores. For $k = 1, 2, \ldots$, we have

$$\begin{aligned}
\widehat{\mathfrak{c}}_{ik} &= \langle\!\langle X_i(\cdot) - \widehat{\mu}(\cdot), \widehat{\phi}_k(\cdot) \rangle\!\rangle \\
&= \langle\!\langle X_i(\cdot) - \widehat{\mu}(\cdot), \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk}(X_j(\cdot) - \widehat{\mu}(\cdot)) \rangle\!\rangle \\
&= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} \langle\!\langle X_i(\cdot) - \widehat{\mu}(\cdot), X_j(\cdot) - \widehat{\mu}(\cdot) \rangle\!\rangle \\
&= \frac{1}{\sqrt{l_k}} \sum_{j=1}^{N} v_{jk} M_{ij} \\
&= \frac{1}{\sqrt{l_k}} (M v_k)_i \\
&= \frac{1}{\sqrt{l_k}} l_k v_{ik} \\
&= \sqrt{l_k} v_{ik}
\end{aligned}$$

****

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

$$C^{\top} C W b_k = N \lambda_k b_k. \tag{13}$$

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 (13), we obtain

$$W^{1/2} C^{\top} C W^{1/2} u_k = N \lambda_k u_k \Longleftrightarrow \left(CW^{1/2}\right)^{\top} \left(CW^{1/2}\right) u_k = N \lambda_k u_k. \tag{14}$$

From the eigendecomposition of the matrix M, we get

$$Mv_k = l_k v_k \iff \left(CW^{1/2}\right)\left(CW^{1/2}\right)^T v_k = l_k v_k. \tag{15}$$

The equations (14) and (15) are eigenequations in the classical PCA case, with the duality $X^\top X$ and $XX^\top$. Following [8], we find that, for $1 \leq k \leq K$,

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

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

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

5.3 On the smoothing

Where should we do the smoothing in the case we estimate the eigencomponents using the inner product matrix?

Using the covariance operator, [9] propose to smooth the eigenfunctions by penalizing their roughness by its integrated squared second derivative.

Using the inner product matrix, we may think about multiple way to do it. First, we smooth all the curves beforehand using kernel regression or local polynomial (see [5] for estimating the optimal bandwidth). Under the twice differentiable curve assumption, we may use the optimal bandwidth as defined by [11]. Second, we could smooth the eigenfunctions directly, maybe using a smoothness penalty as in [9] by cross-validation.

5.4 Sparse data

What you should we do for sparse and irregularly sampled functional data? This question can be rewritten as how to compute the inner product between two curves while they are not sampled on the same grid and contains only few observations? One idea would be to consider the PACE algorithm [13]. We might also only need to smooth each curve in "a good way" to estimate the curves on a common grid and apply the usual algorithm [5].

5.5 Computational complexity

The time complexity is the computational complexity that describes the amount of computer time it takes to run an algorithm. Time complexity is commonly estimated by counting the number of elementary operations performed by the algorithm, supposing each elementary operation takes a fixed amount of time to perform. We usually consider the worst-case time complexity.

Here, we explain the time and space complexity of the computation of MFPCA. First, assume that we observe N curves with P components, and each components p is sampled on a common grid of M_p points. Let $M^a = \sum_{p=1}^P M_p^a$. For the pth components, we consider the estimation of K_p eigenfunctions. Let $K = \sum_{p=1}^P K_p$ be the number of multivariate eigenfunctions that can be estimated. While it 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 element, but rather as the number of components needed for achieving a certain amount of the variance explained (see Section 5.6). Note that the smoothing part is not taken into account here in the computation of the time complexity.

Using the diagonalization of the covariance operator [6]. The algorithm is split into mutiple steps:

- Get the covariance matrix of each components. For each component, the complexity is $\mathcal{O}(NM_p^2)$. So, considering all the components the complexity is $\mathcal{O}(NM^2)$.
- Compute the eigenvalue decomposition of each covariance matrices. For each component, the complexity is $\mathcal{O}(M_p^3)$. So, considering all the components the complexity is $\mathcal{O}(M^3)$.
- Compute the univariate scores. Let K_p be the number of eigencomponents kept for the pth component. For each component, the complexity is $\mathcal{O}(NM_pK_p)$. So, considering all the components the complexity is $\mathcal{O}(N\sum_{p=1}^P M_pK_p)$.

- Compute the covariance matrix of the stacked univariate scores. The matrix of the stacked univariate scores is $N \times K$. The complexity to compute the covariance matrix of this matrix is $\mathcal{O}(NK^2)$.
- Compute the eigenvalue decomposition of the previous matrix. The complexity is $\mathcal{O}(K^3)$.
- Compute the multivariate eigenfunctions and scores. For the estimation of one multivariate eigenfunction, the complexity is $\mathcal{O}(PK_pM_p)$. For the estimation of the K eigenfunctions, the complexity is $\mathcal{O}(KPK_pM_p)$. For the estimation of the scores, the complexity is $\mathcal{O}(NK^2)$.

Gathering all the results, the final complexity of the estimation of MFPCA using the eigendecomposition of the covariance operator is

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

There is something strange in this results. The dependencies on K_p and M_p should not appear like that.

****

Using the inner product matrix. The algorithm is split into multiple steps:

- Estimate the inner product matrix. The inner product between two curves can be estimated in $\mathcal{O}(\sum_{p=1}^{P} M_p)$. There are N^2 terms in the matrix. So, the time complexity is $\mathcal{O}(N^2 \sum_{p=1}^{P} M_p)$.
- Compute the eigenvalue decomposition of this matrix. The complexity is $\mathcal{O}(N^3)$.
- Compute the multivariate eigenfunctions and scores. For the multivariate eigenfunctions the time complexity is $\mathcal{O}(NPK)$ and for the scores $\mathcal{O}(KN)$.

Gathering all the results, the final complexity of the estimation of MFPCA using the eigendecomposition of the inner product matrix is

$$\mathcal{O}\left(N^2\sum_{p=1}^{P} M_p + N^3 + NPK + KN\right).$$

We can use the SVD in both case to make the algortihm faster as it allows to compute only the first k eigenfunctions.

5.6 Percentage of variance explained

We argue that the percentage of variance explained in [6] 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.

5.7 Simulations

We run simulations to check the goodness of fit of the MFPCA using both methods and compare the time complexity of the two ways.

Some ideas:

- Varies number of curves in the sample $N = 50, 100, 200$.
- Varies number of points per curve $M = 20, 50, 100$.
- Varies number of components $P = 1, 5, 10, \ldots$.
- Common grid / Irregularly sampled grid
- Denoising

References

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

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

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

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

[5] S. Golovkine, N. Klutchnikoff, and V. Patilea. Learning the smoothness of noisy curves with application to online curve estimation. *Electronic Journal of Statistics*, 16, Jan. 2022.

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

[7] J. Pagès. Analyse factorielle de données mixtes. *Revue de Statistique Appliquée*, 52(4):93–111, 2004.

[8] J. Pagès. *Multiple Factor Analysis by Example Using R.* CRC Press, Nov. 2014.

[9] J. Ramsay, J. Ramsay, and B. W. Silverman. *Functional Data Analysis.* Springer Science & Business Media, June 2005.

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

[11] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer New York, New York, NY, 2009.

[12] W. Yang, H.-G. Müller, and U. Stadtmüller. Functional singular component analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(3):303–324, 2011.

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