Computational considerations in functional principal component analysis

- Francisco A. Ocaña, Ana M. Aguilera, Manuel Escabias

ORIGINAL PAPER

Computational considerations in functional principal component analysis

Francisco A. Ocaña · **Ana M. Aguilera** ·
Manuel Escabias

Abstract   Computing estimates in functional principal component analysis (FPCA) from discrete data is usually based on the approximation of sample curves in terms of a basis (splines, wavelets, trigonometric functions, etc.) and a geometrical structure in the data space (L^2 spaces, Sobolev spaces, etc.). Until now, the computational efforts have been focused in developing ad hoc algorithms to approximate those estimates by previously selecting an efficient approximating technique and a convenient geometrical structure. The main goal of this paper consists of establishing a procedure to formulate the algorithm for computing estimates of FPCA under general settings. The resulting algorithm is based on the classic multivariate PCA of a certain random vector and can thus be implemented in the majority of statistical packages. In fact, it is derived from the analysis of the effects of modifying the norm in the space of coordinates. Finally, an application on real data will be developed to illustrate the so derived theoretic results.

Keywords   Functional data analysis · Hilbert spaces · Principal components · Covariance estimation · Orthogonal projection

F. A. Ocaña (✉) · A. M. Aguilera · M. Escabias
Department of Statistics and O.R., University of Granada, Granada, Spain
e-mail: focana@ugr.es

A. M. Aguilera
e-mail: aaguiler@ugr.es

M. Escabias
e-mail: escabias@ugr.es

1 Introduction

There are many fields of application where data are functions rather than vectors as in the classical multivariate analysis. Taking into account that the existing multivariate methodology was not entirely satisfactory for analyzing functional data, many papers on functional data analysis (FDA) have been developed in last years. Interesting examples of functional data appear in James et al. (2000) and Rice and Wu (2001). The book of Ramsay and Silverman (2002) considers different applications to illustrate how functional data analysis works out in practice in a diverse range of subject areas as criminology, economics, archaeology, rheumatology, psychology, biomechanics and education, among others.

The majority of FDA techniques are extensions of multivariate methodologies to the case of functional data that can be modeled as sample paths of a continuous-time stochastic process (see Ramsay and Dalzell 1991; Ramsay and Silverman 2005). By the way, the classical (multivariate) principal component analysis (PCA) was firstly adapted to functional data by Deville (1973), where data were assumed to be square integrable functions.

Functional PCA (FPCA) was the first method to be considered in the early literature on FDA with the objective of exploring substantial modes of variation of data. An interesting problem in the estimation of FPCA arises when functional data are only observed at discrete time knots, not necessarily evenly spaced. The usual practice to solve this involves the application of a numerical technique for approximating the underlying functional data from the discrete observed values. In this case, the unknown stochastic process that models the functional data is usually *replaced* by a new one whose sample paths belong to a finite-dimensional functional space generated by a basis of functions (Ramsay and Silverman 2005). If the discrete observed values can be assumed errorless, an appropriate process to convert them to a function is interpolation. Indeed, Besse (1991) and Aguilera et al. (1996a) propose to use natural cubic spline interpolation between the knots. However, if the discrete values are perturbed by some observational error, then their conversion to functions may involve a smoothing procedure such as the orthogonal projection of the sample paths (Aguilera et al. 1995).

As data are functions, it is interesting to incorporate functional properties in the FPCA (Cardot 2000). For example, Ramsay et al. (1995) considers the smoothing behaviour of a biomechanical data set by means of a Sobolev inner product in the data space. Silverman (1996) provides a detailed theoretical discussion of the advantages of smoothing by penalizing the roughness of the principal components weights. In our case, the functional properties of data are incorporated in the FPCA by defining an appropriate geometrical structure in the data space. More recently, Yao et al. (2005) proposed a nonparametric method to perform FPCA for the case of sparse longitudinal data and referred it as principal component analysis through conditional expectation (PACE).

In the available literature on computing FPCA, the efforts have been focused in developing ad hoc algorithms to approximate the estimates in the FPCA by

previously selecting an efficient approximating technique and a convenient geometrical structure. The theoretical results obtained in this paper establish a general procedure which leads to derive the algorithm for computing the estimates of the FPCA. The resulting algorithm is based on the classic multivariate PCA, what has the advantage that is already implemented in the majority of statistical packages.

The present paper is sketched out as follows. In Sect. 2 FPCA is introduced in the general context of a Hilbert space. In order to obtain further results, an equivalence involving FPCA and data transformation will be also introduced under the name of filtering PCA. In Sect. 3 the FPCA on a finite-dimensional functional space is studied to point out the role of functional basis in FDA. The effects of modifying the norm in the space of coordinates will be also analyzed in Sect. 4. Finally, Sect. 5 illustrates the methodology proposed in previous sections.

2 Theoretical framework

We will consider a functional data set given by $X_1(t), X_2(t), \ldots, X_N(t)$, where t is an argument continuously varying on a set $\mathcal{T} \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$. Such data functions are assumed in a separable Hilbert space $\mathcal{H}$, whose inner product $\langle *, * \rangle_{\mathcal{H}}$ induces a norm given by $\|h\|_{\mathcal{H}} = \langle h, h \rangle_{\mathcal{H}}^{1/2}$, $\forall h \in \mathcal{H}$. For instance, $\mathcal{H}$ could be the space $L^2(\mathcal{T})$ of square-integrable functions over a real interval $\mathcal{T}$, whose usual inner product is defined by

$$\langle h_1, h_2 \rangle_{L^2(\mathcal{T})} = \int_{\mathcal{T}} h_1(t)\, h_2(t)\, dt, \quad \forall h_1, h_2 \in L^2(\mathcal{T}).$$

Associated to a functional data set, $X = (X_w(t) : w = 1, \ldots, N; t \in \mathcal{T})$, some summary statistics will now be considered.

- The *mean function* is defined by $\bar{X} = N^{-1} \sum_{w=1}^{N} X_w \in \mathcal{H}$. From now on, $\widetilde{X} = (\widetilde{X}_w : w = 1, \ldots, N)$ stands for the centered functional data set, which is given by $\widetilde{X}_w(t) = X_w(t) - \bar{X}(t)$, $\forall t \in \mathcal{T}$, $w = 1, \ldots, N$.
- The linear space $\mathrm{L}_X = \mathrm{Lin}\{\widetilde{X}_w : w = 1, \ldots, N\}$, which is usually called *the data space*, has a dimension less or equal than $N-1$, because of $\sum_{w=1}^{N} \widetilde{X}_w = 0$. Further, as $\mathrm{L}_X \subseteq \mathcal{H}$, L_X is thus a finite-dimensional Hilbert subspace of $\mathcal{H}$ and then $\mathcal{H} = \mathrm{L}_X \oplus \mathrm{L}_X^{\perp}$, where $\mathrm{L}_X^{\perp}$ is the ortogonal subspace of L_X.
- The operator $\mathrm{U}_X : \mathcal{H} \longmapsto \mathbb{R}^N$ defines column vectors denoted by $\mathrm{U}_X(h)$, for every $h \in \mathcal{H}$, whose components are given by $\mathrm{U}_X[w, h] = \langle h, \widetilde{X}_w \rangle_{\mathcal{H}}$, $\forall w \in \{1, \ldots, N\}$. U_X can thus be interpreted as the constructor of any linear combination from $\widetilde{X}$ through the inner product $\langle *, * \rangle_{\mathcal{H}}$, where the weights for each combination are given by an element of $\mathcal{H}$, i.e., a functional generalization of matrix product. However, we must take into account that the most interesting weights for the aforementioned linear combinations are those contained in L_X, because of $\mathrm{Ker}\, \mathrm{U}_X = \mathrm{L}_X^{\perp}$.

- The *covariance operator*, denoted by $\mathcal{C}_X : \mathcal{H} \longmapsto L_X$, is the positive semi-definite symmetric operator defined by $\mathcal{C}_X = n^{-1} \sum_{w=1}^{N} \widetilde{X}_w \otimes \widetilde{X}_w$. It can then be characterized by

$$\langle \mathcal{C}_X(h_1), h_2 \rangle_{\mathcal{H}} = n^{-1} \langle U_X(h_1), U_X(h_2) \rangle_{\mathbb{R}^N}, \quad \forall h_1, h_2 \in \mathcal{H}, \qquad (1)$$

where $n \in \{N, N-1\}$ and depends on whether our functional data set is considered as a population or a sample. Moreover, some straightforward calculations from Equation (1) lead to establish that

$$\mathcal{C}_X(h)(s) = n^{-1} \sum_{w=1}^{N} \langle \widetilde{X}_w, h \rangle_{\mathcal{H}} \widetilde{X}_w(s) = \left\langle n^{-1} \sum_{w=1}^{N} \widetilde{X}_w(s) \widetilde{X}_w, h \right\rangle_{\mathcal{H}}$$
$$= \langle C_X(s, *), h \rangle_{\mathcal{H}}, \quad \forall s \in \mathcal{T} \text{ and } \forall h \in \mathcal{H},$$

where $C_X : \mathcal{T} \times \mathcal{T} \longmapsto \mathbb{R}$ is the covariance function of X, which is defined by $C_X(s, t) = n^{-1} \sum_{w=1}^{N} \widetilde{X}_w(s) \widetilde{X}_w(t), \forall s, t \in \mathcal{T}$.
Indeed we can only considered the restriction of $\mathcal{C}_X$ to L_X, because of $\mathcal{C}_X(L_X^{\perp}) = \{0\}$. Furthermore, as $\mathcal{C}_X(\mathcal{H}) = \mathcal{C}_X(L_X) = L_X$, the rank of $\mathcal{C}_X$ must be less or equal than $N - 1$.

Finally, notice that the summaries provided by the linear operators U_X and $\mathcal{C}_X$ account for the functional variations of data with respect to their mean, i.e., the variability of X immersed in L_X. This means that the finite-dimensional space L_X will establish the linear framework of the data analysis of X. However, $\mathcal{H}$ will be considered as a shorthand to reference to the geometrical structure induced into L_X by $\mathcal{H}$, i.e., $(L_X, \langle *, * \rangle_{\mathcal{H}})$.

As an illustrative example, let us consider that X is a set of functions of $L^2(\mathcal{T})$. It is thus obtained that $U_X[w, h] = \int_{\mathcal{T}} \widetilde{X}_w(s) h(s) ds, \forall w \in \{1, \ldots, N\}$, and that $\mathcal{C}_X(h)(s) = \int_{\mathcal{T}} C_X(s, t) h(t) dt, \forall s \in \mathcal{T}$ and $\forall h \in L^2(\mathcal{T})$, which generalizes the well-known multivariate PCA framework.

The generalized linear combinations obtained by U_X (vectors in $\mathbb{R}^N$) satisfy the following properties:

- $\overline{U_X(h)} = N^{-1} \sum_{w=1}^{N} U_X[w, h] = 0, \quad \forall h \in \mathcal{H}$;
- $\text{Cov}[U_X(h_1), U_X(h_2)] = \langle \mathcal{C}_X(h_1), h_2 \rangle_{\mathcal{H}}, \quad \forall h_1, h_2 \in \mathcal{H}$;
- $\text{Var}[U_X(h)] = \langle \mathcal{C}_X(h), h \rangle_{\mathcal{H}}, \quad \forall h \in \mathcal{H}$.

Indeed these properties state that the second order moments of the generalized linear combinations $U_X(h)$, for every $h \in \mathcal{H}$, are determined by the covariance operator of X.

2.1 Functional principal component analysis

The FPCA associated to a functional data set X is defined to have the same optimal properties as in the multivariate case (Ramsay and Silverman 2005). In fact, the principal components (p.c.'s) are vectors given by

$$\{\xi_i : \ i \in \mathcal{I} = \{1, 2, \ldots\}\} \subset \mathrm{U}_X(\mathcal{S}_{\mathcal{H}}) \subseteq \mathbb{R}^N,$$

where $\mathcal{S}_{\mathcal{H}} = \{f \in \mathcal{H} : \|f\|_{\mathcal{H}} = 1\}$, and defined by the following stepwise procedure:

- $\xi_1 = \arg\max \{\mathrm{Var}[\eta] : \ \eta \in \mathrm{U}_X(\mathcal{S}_{\mathcal{H}})\}$
- if $i > 1$, then $\xi_i = \arg\max \left\{ \mathrm{Var}[\eta] \ \middle/ \ \begin{matrix} \eta \in \mathrm{U}_X(\mathcal{S}_{\mathcal{H}}) \\ \mathrm{Cov}[\eta, \xi_j] = 0, \ \forall j < i \end{matrix} \right\}.$

The previous optimization problem is equivalent with the spectral decomposition of the covariance operator (Riesz and Sz-Nagy 1990). In fact, the FPCA is obtained from the *principal component curves* or *principal factors* (p.f.'s), which are the solutions of the eigensystem given by

$$\mathcal{C}_X(f_i) = \lambda_i f_i, \quad \forall i \in \mathcal{I}, \tag{2}$$

subject to $\langle f_i, f_j \rangle_{\mathcal{H}} = \delta_{(i,j)}, \forall i, j \in \mathcal{I}$, where $\lambda_i \geq \lambda_{i+1} \geq 0, \forall i \in \mathcal{I}$. Their corresponding p.c.'s are thus obtained from $\xi_i = \mathrm{U}_X(f_i), \forall i \in \mathcal{I}$.

For example, in the case of $\mathcal{H} = L^2(\mathcal{T})$, the p.f.'s are the solutions of the eigenequation $\lambda_i f_i(t) = \int_{\mathcal{T}} C_X(t, s) f_i(s) ds, \forall t \in \mathcal{T}, i \in \mathcal{I}$. The *i-th* p.c. scores are then given by $\xi_i[w] = \int_{\mathcal{T}} f_i(s) \widetilde{X}_w(s) ds, \forall w = 1, \ldots, N$.

Taking into account that $\mathcal{C}_X(\mathcal{H}) = \mathrm{L}_X$, the set $\{f_i : i \in \mathcal{I}\}$ is thus an orthonormal basis of $(\mathrm{L}_X, \langle *, * \rangle_{\mathcal{H}})$. In fact, as the rank of $\mathcal{C}_X$ is not greater than $N - 1$, the maximum number of nonzero eigenvalues is less than N, which also implies that $\mathcal{I} \subseteq \{1, \ldots, N - 1\}$. Therefore, the p.c.'s provide a decomposition of X in terms of its variability components as follows:

$$X_w(t) = \bar{X}(t) + \sum_{i \in \mathcal{I}} \xi_i[w] f_i(t), \quad \forall t \in \mathcal{T} \ \text{and} \ \forall w. \tag{3}$$

The p.c's are uncorrelated vectors whose variances are their associated eigenvalues, such as is derived from the equality $\mathrm{Cov}[\xi_i, \xi_j] = \langle \mathcal{C}_X(f_i), f_j \rangle_{\mathcal{H}} = \lambda_i \delta_{(i,j)}$, $\forall i, j \in \mathcal{I}$. In fact, it follows from the data expansion in Equation (3) that the data variance explained by the ith p.c. can be quantified by $V_i = \lambda_i / V$, where

$$V = n^{-1} \sum_{w=1}^{N} \|\widetilde{X}_w\|_{\mathcal{H}}^2 = \sum_{i \in \mathcal{I}} \mathrm{Var}[\xi_i] = \sum_{i \in \mathcal{I}} \lambda_i$$

accounts for the total variability of data.

In practice, the PCA expansion given in Equation (3) could be truncated in the qth term ($q \leq N - 1$), $X_w^{(q)}(t) = \bar{X}(t) + \sum_{i=1}^{q} \xi_i[w] f_i(t)$, to provide the best q-dimensional linear approximation to X in the least square sense, whose mean squared error is thus given by $n^{-1} \sum_{w=1}^{N} \left\| X_w - X_w^{(q)} \right\|_{\mathcal{H}}^2 = \sum_{j > q} \lambda_j$.

2.2 Filtering PCA

In the framework given by Hilbert-valued random variables, Ocaña et al. (1999) have established the equivalence between the PCA of the original data with

respect to a well suited geometry and the PCA of transformed data with respect to a given geometry, where the considered transform is induced by the geometry initially considered in the original data space. This equivalence included in a framework named as *filtering PCA* provides the theoretic support for the reasonings included in this work. Because of this, the theoretic results here considered are now to be sketched out.

Let $F : \mathcal{H} \longmapsto (\mathcal{G}, \langle *, * \rangle_{\mathcal{G}})$ be an one-to-one linear operator defined from the vector space $\mathcal{H}$ containing the *original data* onto a (separable) Hilbert space $\mathcal{G}$ (*the space of the transformed data*). Such a linear transform F induces in $\mathcal{H}$ a geometrical structure defined by the following inner product:

$$\langle h_1, h_2 \rangle_{\mathcal{H}} = \langle F(h_1), F(h_2) \rangle_{\mathcal{G}}, \quad \forall h_1, h_2 \in \mathcal{H}.$$

Under this setting, F is an isometry between the spaces $(\mathcal{H}, \langle *, * \rangle_{\mathcal{H}})$ and $(F(H), \langle *, * \rangle_{\mathcal{G}})$. From now on $(\mathcal{H}, \langle *, * \rangle_{\mathcal{H}})$ will be assumed a Hilbert space without lack of generality. Otherwise, as F is an isometry, in particular a continuous linear operator, the completion of $\mathcal{H}$ and the continuous extension of F should be then considered to achieve the Hilbert assumption on $\mathcal{H}$.

Let $X = (X_w : w = 1, \dots, N)$ be a functional data set of $\mathcal{H}$. Suppose that each function of X is transformed by F. This way a new functional data set, which is denoted by $F(X) = (FX_w = F(X_w) : w = 1, \dots, N) \subset \mathcal{G}$, is so obtained. The following lemma provides the relationships between the characteristics of X, with respect to $(\mathcal{H}, \langle *, * \rangle_{\mathcal{H}})$, and the corresponding ones of $F(X)$, with respect to $(\mathcal{G}, \langle *, * \rangle_{\mathcal{G}})$.

Proposition 1 *Under the above settings, the following properties hold:*

1. $\overline{F(X)} = F\bar{X}$.
2. $\mathrm{U}_X = \mathrm{U}_{F(X)} \circ F$.
3. $\mathcal{C}_X = F^{-1} \circ \mathcal{C}_{F(X)} \circ F$.
4. $\sigma(\mathcal{C}_X) = \sigma(\mathcal{C}_{F(X)})$, *where* $\sigma(\mathcal{C}_X)$ *and* $\sigma(\mathcal{C}_{F(X)})$ *are the spectrum of the covariance operators* $\mathcal{C}_X$ *and* $\mathcal{C}_{F(X)}$, *respectively.*
5. $F(E_\lambda) = V_\lambda$, *where* $E_\lambda = Ker(\mathcal{C}_X - \lambda I_{\mathcal{H}})$ *and* $V_\lambda = Ker(\mathcal{C}_{F(X)} - \lambda I_{\mathcal{G}})$, *for each eigenvalue* λ.
6. *It can be said that the PCA of X with respect to $(\mathcal{H}, \langle *, * \rangle_{\mathcal{H}})$ is equivalent to the PCA of $F(X)$ with respect to the geometry $(\mathcal{G}, \langle *, * \rangle_{\mathcal{G}})$. In fact, their corresponding PCA expansions are related as follows:*

$$X_w = \bar{X} + \sum_{i \in \mathcal{I}} \xi_i[w] f_i \quad (\text{in } \mathcal{H}) \quad \text{and}$$
$$FX_w = F\bar{X} + \sum_{i \in \mathcal{I}} \xi_i[w] Ff_i \quad (\text{in } \mathcal{G}).$$

In any PCA expansion [see Equation (3)], two kind of structures can be considered. On the one hand, the p.f.'s are those basis functions (directions) in the data space presenting maximum variability for the considered functional

data. On the other hand, each p.c. stands for the weight of each p.f. for any functional datum. Proposition 1 establishes then that, when the functions of X are transformed by F, which is the transform establishing the relationship between the two considered inner products, the PCA's of X and $F(X)$ are essentially the same (they are called equivalent). In such a case the p.c.'s structures are both the same and the p.f.'s are the only structures related through F. An important consequence from this fact is that, if one of these two PCA expansions is derived, then the other is straightforwardly determined.

3 Functional basis based PCA

In this section the equivalence between a PCA of a set of functions contained in a finite-dimensional space and a certain multivariate PCA of its coordinates with respect to a given basis (not necessarily orthonormal) is analyzed.

Let X be a functional data set onto a m-dimensional functional space $\mathbb{E}$, whose inner product is denoted by $\langle *, * \rangle_{\mathbb{E}}$. For a given basis $\mathcal{B} = \{e_1, \ldots, e_m\}$ of $\mathbb{E}$, each function of X can be expanded,

$$X_w = \sum_{k=1}^{m} Y_k[w] \, e_k, \quad \forall w \in \{1, \ldots, N\}, \tag{4}$$

in terms of each row of the $N \times m$-matrix $\mathbf{Y} = \left[Y_1 \vdots \ldots \vdots Y_m \right]$ with $Y_k = (Y_k[1], \ldots, Y_k[N])^{\mathrm{T}}$. We can say that the matrix $\mathbf{Y}$ contains the coordinates of X with respect to the basis $\mathcal{B}$.

Associated to a basis $\mathcal{B}$ of $\mathbb{E}$, let $\Gamma = \Gamma_{\mathcal{B}} : \mathbb{R}^m \longmapsto (\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$ be the linear isomorphism defined by

$$\Gamma(\alpha) = \sum_{k=1}^{m} \alpha_k \, e_k = \underline{e}^{\mathrm{T}} \alpha, \quad \forall \alpha = (\alpha_1, \ldots, \alpha_m)^{\mathrm{T}} \in \mathbb{R}^m, \tag{5}$$

where the functions of $\mathcal{B}$ are contained in $\underline{e} = (e_1, \ldots, e_m)^{\mathrm{T}}$. This way each data function is thus obtained from the coordinate vector
$\mathbf{Y}[w] = (Y_1[w], \ldots, Y_m[w])^{\mathrm{T}}$, i.e., $X_w = \Gamma(\mathbf{Y}[w])$, $\forall w \in \{1, \ldots, N\}$.

Proposition 2 *Under the framework given in this section, the following PCA's are equivalent:*

1. *The PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$, whose PCA expansions are given by*

$$\mathbf{Y}[w] = \bar{Y} + \sum_{i \in \mathcal{I}} \xi_i[w] \, v_i,$$

where $\bar{Y}$ stands for the mean $(m \times 1)$ vector of $\mathbf{Y}$ and $\{v_i\} \subset \mathbb{R}^m$ and $\{\xi_i\} \subset \mathbb{R}^N$ are its p.f.'s and p.c.'s, respectively, with respect to $\langle *, * \rangle_{\mathbf{P}}$. This inner product is defined by

$$\langle \alpha, \beta \rangle_{\mathbf{P}} = \langle \Gamma(\alpha), \Gamma(\beta) \rangle_{\mathbb{E}} = \alpha^{\mathrm{T}} \mathbf{P} \beta, \quad \forall \alpha, \beta \in \mathbb{R}^m, \tag{6}$$

where $\mathbf{P}$ is the $m \times m$ matrix given by

$$\mathbf{P} = \langle \underline{e}\, \underline{e}^{\mathrm{T}} \rangle_{\mathbb{E}} = (\langle e_k, e_l \rangle_{\mathbb{E}} \ : \ k, l = 1, \ldots, m)_{m \times m}. \tag{7}$$

2. *The PCA of X on $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$, whose PCA expansions are given by*

$$X_w = \Gamma(\bar{Y}) + \sum_{i \in \mathcal{I}} \xi_i[w]\, \Gamma(v_i),$$

where $\Gamma(\bar{Y})$ is the mean function of X and ξ_i and $\Gamma(v_i)$ are its corresponding ith p.c. and ith p.f., respectively.

3. *The PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ on $(\mathbb{R}^m, \langle *, * \rangle_u)$, with PCA expansions given by*

$$\mathbf{Q}\,\mathbf{Y}[w] = \mathbf{Q}\,\bar{Y} + \sum_{i \in \mathcal{I}} \xi_i[w]\, \mathbf{Q}\, v_i,$$

where $\mathbf{Q}$ is a $m \times m$ matrix satisfying that $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{P}$ and $\langle *, * \rangle_u$ denotes the usual inner product (in $\mathbb{R}^m$).

Remark 1 • As $\mathbf{P}$ is a symmetric and positive definite matrix, then, among others matrices, $\mathbf{Q}$ could be its well-known square root, $\mathbf{P}^{1/2}$.

• Taking into account that $\mathbf{Y}^{\mathrm{T}} = \left[\mathbf{Y}[1]\vdots \ldots \vdots \mathbf{Y}[N]\right]_{m \times N}$, the transformation of every vector $\mathbf{Y}[w]$ by $\mathbf{Q}$ could be obtained from $\mathbf{Q}(\mathbf{Y}) = \mathbf{Y}\mathbf{Q}^{\mathrm{T}}$.

Proof • *Equivalence between statements* (1) *and* (2).

The geometrical structure in $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$ induces in $\mathbb{R}^m$ a geometry given by the inner product $\langle *, * \rangle_{\mathbf{P}}$ in Equation (6). This is a particular case of the filtering PCA with $F = \Gamma$, $\mathcal{H} = (\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$ and $\mathcal{G} = (\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$.

Let $\mathcal{C}_Y^{(\mathbf{P})}$ be the covariance matrix of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$. As $X = \Gamma(Y)$, Proposition 1 states that $\Gamma(\bar{Y}) = \bar{X}$, $\mathcal{C}_Y^{(\mathbf{P})} = \Gamma^{-1} \circ \mathcal{C}_X \circ \Gamma$ and that the PCA of X on $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$ is equivalent to the PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$. This assures that if $\{v_j\}_j$ is a set of p.f.'s from the PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$, then $\{\Gamma(v_j)\}_j$ are p.f.'s from the PCA of X on $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$, where both PCA's have the same set of p.c.'s.

• *Equivalence between statements* (1) *and* (3).

To find out the matrix $\mathbf{Q}$, the filtering PCA with $F = \mathbf{Q}$, $(\mathcal{H}, \langle *, * \rangle_{\mathcal{H}}) = (\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$ and $(\mathcal{G}, \langle *, * \rangle_{\mathcal{G}}) = (\mathbb{R}^m, \langle *, * \rangle_u)$ should be imposed. In fact, Proposition 1 leads to establish that

$$\langle \alpha, \beta \rangle_{\mathbf{P}} = \alpha^{\mathrm{T}} \mathbf{P} \beta = \langle \mathbf{Q}\alpha, \mathbf{Q}\beta \rangle_u = \alpha^{\mathrm{T}} \mathbf{Q}^{\mathrm{T}} \mathbf{Q} \beta, \quad \forall \alpha, \beta \in \mathbb{R}^m,$$

which means that $\mathbf{P} = \mathbf{Q}^{\mathrm{T}} \mathbf{Q}$.

The characteristics of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$ can be related with those obtained by using the usual geometry in $\mathbb{R}^m$. In fact, it follows from Proposition 1 that $\overline{Q(Y)} = \mathbf{Q}\bar{Y}$ and $\mathcal{C}_Y^{(\mathbf{P})} = \mathbf{Q}^{-1}\mathcal{C}_{Q(Y)}\mathbf{Q} = \mathcal{C}_Y\mathbf{P}$, where $\mathcal{C}_{Q(Y)}$ and $\mathcal{C}_Y$ are the covariance matrices of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ and $\mathbf{Y}$, respectively, with respect to $\langle *, * \rangle_u$. Further, the PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$ is equivalent to the PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ on $(\mathbb{R}^m, \langle *, * \rangle_u)$. In fact, if $\{v_j\}_j$ is a set of p.f.'s from the PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$, then $\{\mathbf{Q}v_j\}_j$ are p.f.'s from the PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ on $\mathbb{R}^m$, where both PCA's have the same set of p.c.'s. $\qquad\square$

An interesting case appears when $\mathcal{B}$ is an orthonormal basis of $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$. In such a case, the matrix inner product $\langle *, * \rangle_{\mathbf{P}}$ becomes the usual inner product in $\mathbb{R}^m$ ($\mathbf{P} = \mathbf{I}_m$). This implies the well-known equivalence between the usual multivariate PCA of $\mathbf{Y}$ and the FPCA of X, for orthonormal basis. Furthermore, Proposition 2 also assures that the FPCA of X can be even obtained from the multivariate PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$, for any orthogonal matrix $\mathbf{Q}$. This equivalence means that the FPCA does not depend on the considered orthogonal basis in $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$. Notice that $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ contains of the coordinate vector of X with respect to the orthonormal basis $\mathbf{Q}\underline{e}$, for any orthogonal matrix $\mathbf{Q}$.

However, FPCA is not necessarily subject to orthonormal basis. In fact, there exist interesting non orthonormal bases in functional spaces (splines, non orthogonal wavelets, etc.) which can be considered in FPCA (Ramsay and Silverman 2005). Proposition 2 establishes how to perform a FPCA with any kind of functional basis.

Moreover, in applications with functional data, the inner product in the data space is usually suggested by the mathematical properties to be assumed for the functional data. However, the implementation of the PCA in the majority of statistical softwares is only given by routines performing the multivariate PCA on $\mathbb{R}^m$ (with the usual geometry). In this context, Proposition 2 provides mathematical results for the algorithmic implementation of the FPCA of X through the available multivariate PCA routines.

Algorithm for computing FPCA

Proposition 2 establishes that the PCA of X on $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$ can be obtained from the classic multivariate PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$. In this sense, the usual PCA of $\mathbf{Y}\mathbf{Q}^{\mathrm{T}}$ is obtained from the diagonalization of its covariance matrix $\mathcal{C}_{Q(Y)}$,

$$\mathcal{C}_{Q(Y)}\, u_j = \mathbf{Q}\, \mathcal{C}_Y\, \mathbf{Q}^{\mathrm{T}}\, u_j = \lambda_j\, u_j, \quad \forall j \in \mathcal{I}, \tag{8}$$

where u_j stands for the jth p.f. and λ_j is the variance of the jth p.c., which is defined by

$$\xi_j[w] = \mathrm{U}_{Q(Y)}[w, u_j] = u_j^{\mathrm{T}}\mathbf{Q}\left(\mathbf{Y}[w] - \bar{Y}\right), \quad \forall w \in \{1, \ldots, N\}. \tag{9}$$

This way the p.c.'s of the FPCA of X and their corresponding variances are already obtained. Further, it follows from Proposition 2 that $\{\mathbf{Q}^{-1}u_j\}_j$ are p.f.'s

of the PCA of $\mathbf{Y}$ on $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{P}})$ and that the p.f.'s of the FPCA of X can thus be obtained by

$$f_j = \Gamma \left(\mathbf{Q}^{-1} u_j \right), \quad \forall j \in \mathcal{I}. \tag{10}$$

Therefore, the FPCA decomposition of X is given by

$$X_w = \Gamma(\bar{Y}) + \sum_{i \in I} \xi_i[w] \, \Gamma \left(\mathbf{Q}^{-1} u_i \right), \quad \forall w \in \{1, \ldots, N\}.$$

Finally, these reasonings for computing the PCA of a functional data set X such as in Equation (4) are sketched in the following algorithm.

    Input: A coordinate matrix of X, $\mathbf{Y}_{N \times m}$, for a basis $\mathcal{B}$ of $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$.
  Output: The FPCA of X on $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$.
  Step 1: Compute $\mathbf{P}$ by using Equation (7) and obtain a matrix $\mathbf{Q}$ such that $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{P}$.
  Step 2: Perform the usual multivariate PCA of $\mathbf{YQ}^{\mathrm{T}}$, obtaining its p.f.'s $\{u_j\}_j \subset \mathbb{R}^m$, its eigenvalues $\{\lambda_j\}$ and its p.c.'s $\{\xi_j\}_j$.
              The p.c.'s of the FPCA of X and its corresponding eigenvalues, $\{\xi_j\}_j$ and $\{\lambda_j\}$, have been already obtained.
  Step 3: The p.f.'s $\{f_j\}_j$ of the FPCA are derived by using Equation (10).

4 Modifying the norm in the space of coordinates

On the one hand, the implementation of FPCA is based on multivariate PCA routines applied to a transform of the coordinates of the functional data, such as is derived in previous section. On the other hand, the multivariate PCA has its own strategies to extract relevant properties from data, which are mainly based on data transformations, as well. For example, the interpretation of a phenomenon described by an initial data set is not usually clear when its PCA is trivial (the variability explained by the first p.c. is closed to 100%). An usual practice in multivariate PCA studies consists of performing PCA to transformed data. Therefore, as such strategies are usually implemented in statistical softwares, we can attempt to adapt them to the FPCA setting. In this section, we will investigate the consequences of such data transform based multivariate techniques on the FPCA.

Let us consider the same settings as those in Sect. 3, where X is a functional data set in $\mathbb{E}$ and $\mathbf{Y}$ is its $N \times m$ matrix of coordinates, such as in Equation (4). For a nonsingular matrix $\mathbf{D} \in M_m(\mathbb{R})$, the multivariate PCA of $\mathbf{YD}^{\mathrm{T}}$ with the usual geometry in $\mathbb{R}^m$ could be considered to extract relevant properties from data. The meaning of such a multivariate PCA will be now derived in the functional data framework.

Proposition 3 *Given a nonsingular matrix $\mathbf{D} \in M_m(\mathbb{R})$, the following PCA's are equivalent:*

1. *The PCA of* $\mathbf{Y}\mathbf{D}^{\mathrm{T}}$ *on* $(\mathbb{R}^m, \langle *, * \rangle_u)$,

$$\mathbf{D}\mathbf{Y}[w] = \mathbf{D}\bar{Y} + \sum_{i \in \mathcal{I}} \pi_i[w] \, g_i.$$

2. *The PCA of* $\mathbf{Y}$ *on* $(\mathbb{R}^m, \langle *, * \rangle_{\mathbf{K}})$,

$$\mathbf{Y}[w] = \bar{Y} + \sum_{i \in \mathcal{I}} \pi_i[w] \, \mathbf{D}^{-1} g_i,$$

where $\mathbf{K} = \mathbf{D}^{\mathrm{T}}\mathbf{D} \in M_m(\mathbb{R})$ and the inner product $\langle *, * \rangle_{\mathbf{K}}$ is defined by

$$\langle \alpha, \beta \rangle_{\mathbf{K}} = \alpha^{\mathrm{T}} \mathbf{K} \beta, \quad \forall \alpha, \beta \in \mathbb{R}^m.$$

3. *The FPCA of* X *on* $(\mathbb{E}, \langle *, * \rangle_K)$,

$$X_w = \Gamma(\bar{Y}) + \sum_{i \in \mathcal{I}} \pi_i[w] \, \Gamma\left(\mathbf{D}^{-1} g_i\right),$$

where the inner product is defined by

$$\langle f, h \rangle_K = \vec{f}^{\,\mathrm{T}} \mathbf{K} \vec{h}, \quad \forall f = \underline{e}^{\mathrm{T}}\vec{f}, \, h = \underline{e}^{\mathrm{T}}\vec{h} \in \mathbb{E}.$$

4. *The FPCA of* $L(X)$ *on* $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$,

$$LX_w = L\Gamma(\bar{Y}) + \sum_{i \in \mathcal{I}} \pi_i[w] \, L\Gamma(\mathbf{D}^{-1} g_i),$$

where $L : \mathbb{E} \longmapsto \mathbb{E}$ is one of the operators defined by

$$L(f) = \underline{e}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{R} \mathbf{D} \vec{f}, \quad \forall f = \underline{e}^{\mathrm{T}}\vec{f} \in \mathbb{E}, \tag{11}$$

for any orthogonal $m \times m$-matrix $\mathbf{R}$, and $\mathbf{Q}$ is the matrix in Proposition 1.

Proof This result is an application of Proposition 1.

- *Equivalence between statements* (1) *and* (2).
  This equivalence is immediate by considering $F = \mathbf{D}$, $\mathcal{H} = \mathbb{R}^m$ and $\mathcal{G} = (\mathbb{R}^m, \langle *, * \rangle_u)$ in Proposition 1, where the induced inner product in $\mathcal{H}$ is then defined by $\langle \alpha, \beta \rangle_{\mathbf{K}} = \langle \mathbf{D}\alpha, \mathbf{D}\beta \rangle_u = \alpha^{\mathrm{T}} \mathbf{D}^{\mathrm{T}} \mathbf{D} \beta = \alpha^{\mathrm{T}} \mathbf{K} \beta$, $\forall \alpha, \beta \in \mathbb{R}^m$.
- *Equivalence between statements* (2) *and* (3).
  Let us consider the operator $\Gamma^{-1} : \mathbb{E} \longmapsto (\mathbb{R}^m, \langle *, * \rangle_{\mathbf{K}})$ [see Equation (5)], whose induced inner product in $\mathbb{E}$ is given by

$$\langle f, h \rangle_K = \left\langle \Gamma^{-1}(f), \Gamma^{-1}(h) \right\rangle_{\mathbf{K}} = \vec{f}^{\,\mathrm{T}} \mathbf{K} \vec{h}, \quad \forall f = \underline{e}^{\mathrm{T}}\vec{f}, \, h = \underline{e}^{\mathrm{T}}\vec{h} \in \mathbb{E}.$$

Proposition 1 establishes the equivalence between the two considered PCA's by considering $F = \Gamma^{-1}$, $\mathcal{H} = (\mathbb{E}, \langle *, * \rangle_K)$ and $\mathcal{G} = (\mathbb{R}^m, \langle *, * \rangle_{\mathbf{K}})$.

- *Equivalence between statements* (3) *and* (4).
  Now the idea consists of finding a linear operator $L : (\mathbb{E}, \langle *, * \rangle_K) \longmapsto (\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$ such that $\langle f, h \rangle_K = \langle L(f), L(h) \rangle_{\mathbb{E}}$, $\forall f, h \in \mathbb{E}$. Once this operator L was found, Proposition 1 assures the equivalence between both considered PCA's.

As $\mathbb{E}$ is a m-dimensional space, there exists a matrix $\mathbf{B} \in M_m(\mathbb{R})$ such that $L(f) = \underline{e}^{\mathrm{T}} \mathbf{B} \vec{f}$, $\forall f = \underline{e}^{\mathrm{T}} \vec{f} \in \mathbb{E}$. Indeed, $\mathbf{B}$ will verify that

$$\langle L(f), L(h) \rangle_{\mathbb{E}} = \vec{f}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \langle \underline{e} \, \underline{e}^{\mathrm{T}} \rangle_{\mathbb{E}} \mathbf{B} \vec{h} = \vec{f}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \mathbf{P} \mathbf{B} \vec{h}$$
$$= \vec{f}^{\mathrm{T}} \mathbf{K} \vec{h}, \quad \forall f = \underline{e}^{\mathrm{T}} \vec{f}, \, h = \underline{e}^{\mathrm{T}} \vec{h} \in \mathbb{E}.$$

It follows from Proposition 2 that $\mathbf{B}$ verifies that

$$\mathbf{K} = \mathbf{B}^{\mathrm{T}} \mathbf{P} \mathbf{B} = (\mathbf{Q} \mathbf{B})^{\mathrm{T}} (\mathbf{Q} \mathbf{B}).$$

As $\mathbf{K} = \mathbf{D}^{\mathrm{T}} \mathbf{D}$, we can consider that $\mathbf{Q} \mathbf{B} = \mathbf{R} \mathbf{D}$, for any orthogonal matrix $\mathbf{R}$. Therefore, it is concluded that $\mathbf{B} = \mathbf{Q}^{-1} \mathbf{R} \mathbf{D}$. □

Among other consequences, Proposition 3 establishes that the operator L is not unique. In fact, an infinite set of operators verifying Statement (4) in Proposition 3 could be considered. However, such operators have a common behaviour such as we will see.

Let $\{v_k\}_{k=1}^m \subset \mathbb{E}$ be the basis of $\mathbb{E}$ defined by $\underline{v}^{\mathrm{T}} = (v_1, \ldots, v_m) = \underline{e}^{\mathrm{T}} \mathbf{Q}^{-1}$. This new basis is orthonormal in $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$, because of

$$\langle \underline{v} \, \underline{v}^{\mathrm{T}} \rangle_{\mathbb{E}} = (\mathbf{Q}^{\mathrm{T}})^{-1} \langle \underline{e} \, \underline{e}^{\mathrm{T}} \rangle_{\mathbb{E}} \mathbf{Q}^{-1} = (\mathbf{Q}^{\mathrm{T}})^{-1} \mathbf{P} \mathbf{Q}^{-1} = \mathbf{I}_m.$$

This means that the term $\underline{e}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{R} = \underline{v}^{\mathrm{T}} \mathbf{R}$ in Equation (11) can be interpreted as the expression of any orthonormal basis of $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$. Therefore, any operator L verifying Statement (4) in Proposition 3 is given by $L(f) = L_{\underline{u}}(f) = \underline{u}^{\mathrm{T}} \mathbf{D} \vec{f}$, $\forall f = \underline{e}^{\mathrm{T}} \vec{f} \in \mathbb{E}$, for any orthonormal basis $\underline{u}$ of $(\mathbb{E}, \langle *, * \rangle_{\mathbb{E}})$.

5 Application with data

To illustrate the theoretic results derived in this paper, they will be applied on the functional data accounting the accumulated number of university professors nominated in Spain during 1992, for a random sample of twenty Spanish universities. The number of new professors was drawn day by day from the BOE (the official Spanish State gazette), for each university. Let $\mathcal{N}_w(t)$ be the number of professors nominated in university w until day t over 1992. Each data function $\mathcal{N}_w$ is thus piecewise constant on random intervals, with limits defined by the days which new professors are nominated in.

Case I: PCA with the usual geometry

The functional PCA of $\mathcal{N}$, with respect to the usual geometry of $L^2(\mathcal{T})$, is given from the solutions of Equation (2). An efficient procedure to solve such an eigenequation is provided by the orthogonal projection method (OPM) which is proposed in Aguilera et al. (1995). As the OPM is suitable when we have information about the nature of the sample paths, as is our case, a functional space with mimic mimic the characteristics of our functional data is the space $\mathbb{E}_m$ ($m = 12$) of piecewise constant functions over subintervals of a given partition Π_m, which has been previously chosen in the observation interval $\mathcal{T} = (0, 366]$. In our study, the year 1992 was divided by natural months, i.e., the knots of Π_m are $a_0 = 0$, $a_1 = 31$, $a_2 = 60$, $a_3 = 91$, $a_4 = 121$, $a_5 = 152$, $a_6 = 182$, $a_7 = 213$, $a_8 = 244$, $a_9 = 274$, $a_{10} = 305$, $a_{11} = 335$, $a_{12} = 366$.

Let us consider the orthonormal basis of $(\mathbb{E}_m, \langle *, * \rangle_{L^2(\mathcal{T})})$ defined as $e_k(t) = \alpha_k^{-1/2} I_k(t)$, $\forall t \in [0, 366]$, where $\alpha_k = a_k - a_{k-1}$ and I_k is the indicator function of the interval $(a_{k-1}, a_k]$, $\forall k = 1, \ldots, m$. The projected functions can then be represented as

$$X_w(t) = \mathcal{P}_m \mathcal{N}_w(t) = \sum_{k=1}^{m} Y_k[w] \, e_k(t) = \sum_{k=1}^{m} M_k[w] \, I_k(t), \quad \forall \, (w, t) \in \Omega \times \mathcal{T},$$

where

$$Y_k[w] = \int_0^{366} \mathcal{N}_w(t) \, e_k(t) \, dt = \sqrt{\alpha_k} \, M_k[w], \quad \forall \, k = 1, \ldots, m,$$

and $M_k[w] = \alpha_k^{-1} \int_{a_{k-1}}^{a_k} \mathcal{N}_w(t) \, dt$, which is interpreted as the mean of the monthly accumulated number of professors for each university w. Let us consider the matrices $\mathbf{Y} = \left[Y_1 \vdots \ldots \vdots Y_m \right]$ and $\mathbf{M} = \left[M_1 \vdots \ldots \vdots M_m \right]$. In this framework, the OPM consists of approximating the PCA of $\mathcal{N}$ by the PCA of $X = \mathcal{P}_m \mathcal{N}$. Further, Proposition 2 establishes that the functional PCA of X can be derived from the multivariate PCA of $\mathbf{Y} = [Y_1 \vdots \ldots \vdots Y_m]$ with respect to the usual inner product in $\mathbb{R}^m$. Indeed their corresponding PCA expansions are given by $X_w = \mu + \sum_{l=1}^{m} \psi_l[w] \, \Gamma(u_l)$ and $\mathbf{Y}[w] = \bar{Y} + \sum_{l=1}^{m} \psi_l[w] \, u_l$, where $\mu = \overline{X}$.

The principal values and percentages of variance explained by the p.c.'s of the aforementioned functional PCA appear in Table 1. The first p.c. explains more than a 96% of the total variability. Aguilera et al. (1996b) have shown that the error committed by reconstructing the data functions with only the first p.c. is not too small. In addition, such an error is drastically reduced by introducing in the model noisy p.c.'s (associated to the smallest variances). Therefore, the

Table 1 Principal values of the projected process and their percentages of explained variance in the following cases: I [usual geometry in $L^2(0, 366)$], II (the geometry induced by standardization of the coordinates), III (the geometry induced by standardization of the projected process)

p.c.	Case I		Case II		Case III	
	Principal values	Expl. var.	Principal values	Expl. var.	Principal values	Expl. var.
1	199408.796	96.142	9.29470	77.4558	283.88157	77.5633
2	5439.012	2.622	2.37300	19.7750	71.93605	19.6547
3	1920.881	0.926	0.20808	1.7340	6.39434	1.7471
4	403.674	0.195	0.08694	0.7245	2.65513	0.7254
5	149.870	0.072	0.01926	0.1605	0.59296	0.1620
6	40.641	0.020	0.01292	0.1077	0.38381	0.1049
7	23.759	0.011	0.00200	0.0167	0.06148	0.0168
8	12.038	0.006	0.00147	0.0123	0.04468	0.0122
9	5.438	0.003	0.00083	0.0070	0.02540	0.0070
10	3.185	0.002	0.00059	0.0049	0.01820	0.0050
11	2.383	0.001	0.00015	0.0013	0.00468	0.0013
12	1.237	0.001	0.00006	0.0005	0.00172	0.0005
	Total variance		Total variance		Total variance	
	207410.913		12		366	

so obtained PCA is considered trivial and thus a new functional PCA decomposition could be searched from the coordinate matrix **Y**.

Case II: PCA with the geometry induced by standardization of its coordinates
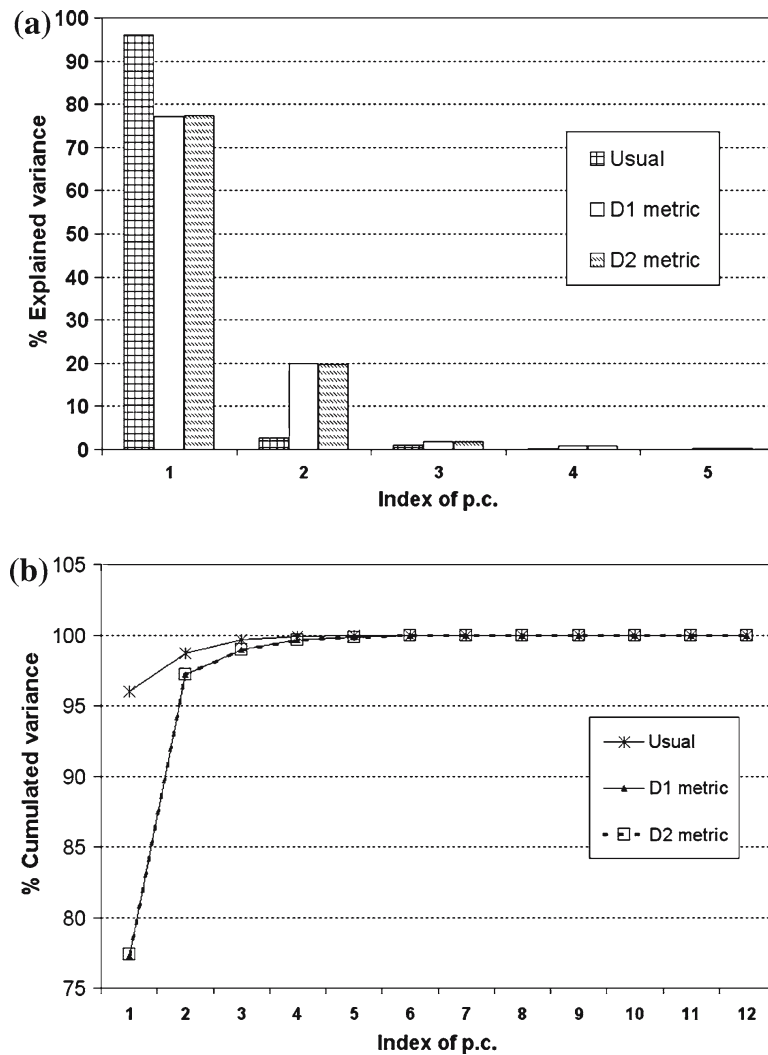
In multivariate analysis it is usual the data standardization to solve the problem of trivial PCA. Taking into account Proposition 2, the PCA of the standardization of coordinates **Y** with respect to the usual geometry in $\mathbb{R}^m$ can be carried out, i.e., the multivariate PCA of $\mathbf{YD_1^T}$ with respect to $(\mathbb{R}^m, \langle *, * \rangle_u)$, where

$$\mathbf{D_1} = \mathrm{diag}\left(\sigma_{Y_1}^{-1}, \ldots, \sigma_{Y_m}^{-1}\right).$$

Proposition 3 establishes that such a PCA is equivalent to the functional PCA of $L_1 X$ with respect to the usual geometry in $L^2(\mathcal{T})$, where the transformed data curves are given by

$$L_1 X_w = \underline{e}^{\mathrm{T}} \mathbf{D_1} \mathbf{Y}[w] = \sum_{k=1}^{m} \frac{Y_k[w]}{\sigma_{Y_k}} e_k = \sum_{k=1}^{m} \alpha_k^{-1/2} \frac{M_k[w]}{\sigma_{M_k}} I_k.$$

The variances explained by the so obtained p.c.'s appear in Table 1. Let us observe the reduction of the variance explained by the first p.c. in comparison with Case I (see Fig. 1).

(a)

(b)

Fig. 1 Distribution of variances explained by the p.c.'s for the three considered PCA's

Case III: PCA with the geometry induced by its own standardization

The standardization of the vector of coordinates of the projected functional data, X, is not necessarily related with direct standardization of such functional data. In fact, the functions of X could not be necessarily contained in $\mathbb{E}_m$. However, in the particular case of piecewise constant functions on fixed intervals, there exists a certain relation that will be now presented.
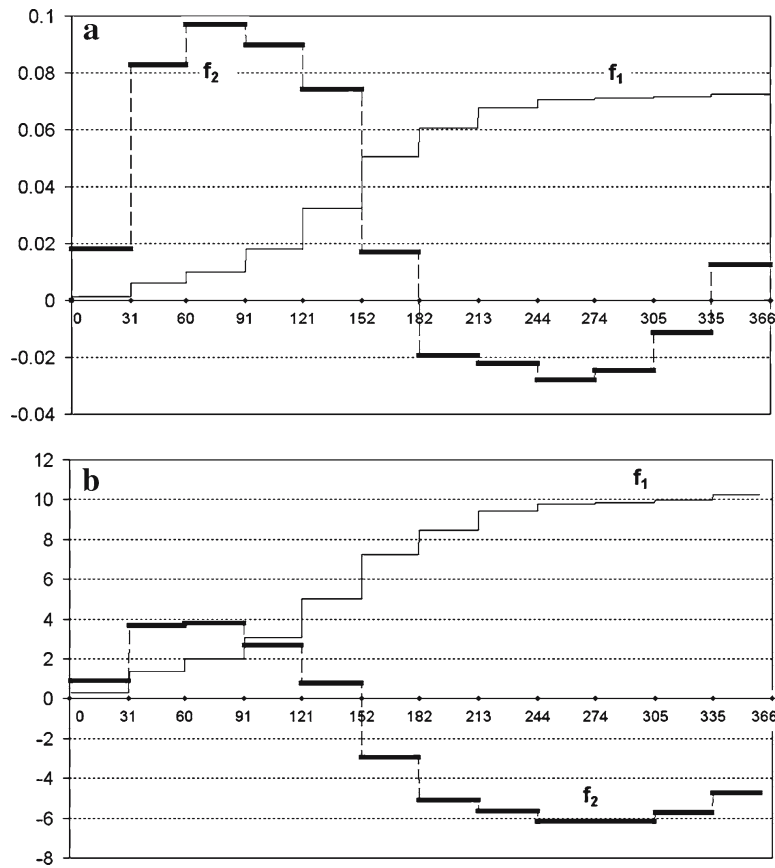
In our example, the standard deviation function of X is given by

$$\sigma_{X(t)} = \sum_{k=1}^{m} \sigma_{Y_k} e_k(t) = \sum_{k=1}^{m} \sigma_{M_k} I_k(t), \quad \forall t \in \mathcal{T}.$$

It follows that the standardized functional data is obtained by

$$\frac{X_w(t)}{\sigma_{X(t)}} = \sum_{k=1}^{m} \frac{M_k[w]}{\sigma_{M_k}} I_k(t) = \sum_{k=1}^{m} \sqrt{\alpha_k} \frac{Y_k[w]}{\sigma_{Y_k}} e_k(t).$$

Fig. 2 First two principal factors obtained with respect to the geometry **a** of $L^2\,(0, 366)$, **b** induced by standardization of the coordinates. For the geometry induced by standardization of the projected process, the principal factors are almost proportional to the ones obtained in **b**

Let us denote by $L_2(X)$ the standardized projection, $L_2(X)(w) = \underline{e}^\mathrm{T}\,\mathbf{D_2}\,\mathbf{Y}[w]$, where

$$\mathbf{D}_2 = \mathrm{diag}\left(\frac{\sqrt{\alpha_1}}{\sigma_{Y_1}}, \ldots, \frac{\sqrt{\alpha_m}}{\sigma_{Y_m}}\right) = \mathrm{diag}\left(\frac{1}{\sigma_{M_1}}, \ldots, \frac{1}{\sigma_{M_m}}\right).$$

Proposition 3 assures that the PCA of $L_2(X)$ with respect to the usual geometry in $L^2\,(\mathcal{T})$ is equivalent to classic multivariate PCA of $\mathbf{Y}\mathbf{D}_2^\mathrm{T}$ with respect to $(\mathbb{R}^m, \langle *, * \rangle_u)$.

The variances explained by the p.c.'s associated with this PCA appear again in Table 1. Notice that they are very similar to the ones obtained in the Case II (see Fig. 1). It is due to the matrices $\mathbf{D_1}$ and $\mathbf{D_2}$ are almost proportional. In fact, the amplitudes of the months have very similar values $\{29, 30, 31\}$. As the two first principal factors for Cases II and III are almost proportional, the ones for Case III have not been considered in Fig. 2.

Roughly speaking, in both Cases II and III the coordinates are transformed by a diagonal matrix $\mathbf{D} = \mathrm{diag}(w_1, \ldots, w_m)$, where w_k stands for the weight of coordinates for the basis function e_k. Proposition 3 specifies the induced inner product in $\mathbb{E}_m$ for $\mathbf{D}$, which is now given by $\langle f, h \rangle_K = \vec{f}^\mathrm{T}\,\mathbf{D}^2\,\vec{h} = \int w^2(s)f(s)\,g(s)ds$, $\forall f = \underline{e}^\mathrm{T}\vec{f}, \ h = \underline{e}^\mathrm{T}\vec{h} \in \mathbb{E}$, where $w(s) = \sum_{k=1}^m w_k I_k(s)$. Moreover, the idea

underlying this new inner product is found in the restriction imposed for each p.f.'s, i.e., $\|f_i\|_K^2 = \sum_{k=1}^m w_k^2 f_{k,i} = 1, \forall i$, where $f_i = \sum_{k=1}^m f_{k,i} e_k$. As we have considered in both cases that $w_k \propto 1/\sigma_{Y_k}$, we can say that the modes of variations given by the p.f.'s are focused on the intervals where σ_{Y_k} is high; in fact, if $\sigma_{Y_k} \approx 0$, then $f_{k,i} \approx 0$.

6 Concluding remarks

In this paper it has been shown how PCA of functional data arising in a finite-dimensional space can take advantage of the multivariate PCA procedures. This result is very useful to simplify computational aspects in practice, because, in almost all applications, the approximation of functional PCA from discrete-time observations implies the approximation of data functions on suitable finite-dimensional functional spaces. By the way, we have also proved that performing multivariate PCA of a transformation of the coordinate vectors is equivalent to a certain functional PCA of the original functional data.

References

Aguilera AM, Gutiérrez R, Ocaña FA, Valderrama MJ (1995) Computational approaches to estimation in the principal component analysis of a stochastic process. Appl Stoch Models Data Anal 11:279–299

Aguilera AM, Gutiérrez R, Valderrama MJ (1996a) Approximation of estimators in the PCA of a stochastic process using B-splines. Commun Stat Simul Comput 25:671–690

Aguilera AM, Ocaña FA, Valderrama MJ (1996b) ACP de un proceso estocástico con funciones muestrales escalonadas. Qüestiiò 20:7–28

Besse P (1991) Approximation spline de l'analyse en composantes principales d'une variable aléatoire hilbertienne. Ann Fac Sci Toulouse 12:329–346

Cardot H (2000) Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. J Nonparametr Stat 12:503–538

Deville JC (1973) Estimation of the eigenvalues and of the eigenvectors of a covariance operator, Technical report. Ann L'INSEE

James GM, Hastie TJ, Sugar CA (2000) Principal component models for sparse functional data. Biometrika 87(3):587–602

Ocaña FA, Aguilera AM, Valderrama MJ (1999) Functional principal component analysis by choice of norm. J Multivariate Anal 71(2):262–276

Ramsay JO, Dalzell JC (1991) Some tools for functional data analysis. J R Stat Soc Ser B 53:539–572

Ramsay JO, Flanagan R, Wang X (1995) The functional data analysis of the pinch force of human fingers. J R Stat Soc Ser C 44:17–30

Ramsay JO, Silverman BW (2002) Applied functional data analysis: methods and case studies. Springer, New York

Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer, New York

Rice JA, Wu CO (2001) Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 57:253–259

Riesz F, Sz-Nagy B (1990) Leçons d'Analyse Fonctionnelle. Gauthier–Villars, Paris

Silverman BW (1996) Smoothed functional principal components analysis by choice of norm. Ann Stat 24(1):1–24

Yao F, Müller HG, Wang JL (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590