Principal points and elliptical distributions from the multivariate setting to the functional case*

Juan Lucas Bali
Universidad de Buenos Aires and FONCYT, Argentina
and
Graciela Boente
Universidad de Buenos Aires and CONICET, Argentina

Abstract

The k principal points of a random vector $\mathbf{X}$ are defined as a set of points which minimize the expected squared distance between $\mathbf{X}$ and the nearest point in the set. They are thoroughly studied in Flury ([2], [3]), Tarpey [13] and Tarpey, Li and Flury [20]. For their treatment, the examination is usually restricted to the family of elliptical distributions. In this paper, we present an extension of the previous results to the functional case, i.e., when dealing with random elements over a separable Hilbert space $\mathcal{H}$. Principal points for gaussian processes were defined in Tarpey and Kinateder [19]. In this paper, we generalize the concepts of principal points, self-consistent points and elliptical distributions so as to fit them in this functional framework. Results linking self-consistency and the eigenvectors of the covariance operator are re-obtained in this new setting as well as an explicit formula for the $k = 2$ case so as to include elliptically distributed random elements in $\mathcal{H}$.

AMS Subject Classification 2000: Primary 62G99, Secondary 62H25, 62H30.
Key words and phrases:

Corresponding Author
Graciela Boente
Moldes 1855, 3^o A
Buenos Aires, C1428CRA, Argentina
email: gboente@dm.uba.ar

*This research was partially supported by Grants X-018 from the Universidad de Buenos Aires, PID 5505 from CONICET and PAV 120 and PICT 21407 from ANPCYT, Argentina. The research of Lucas Bali was supported by a scholarship of the Agencia Nacional de Promoción Científica y Tecnológica.

1 Introduction

1.1 Motivation

Inside statistics there exists lots of situations where the collected data may not be represented with classic schemes like numbers or numeric vectors and so, sometimes a functional representation is more appropriate. For example, consider results of an electrocardiogram (EGC) or the study of the temperature in a weather station, which lend themselves to this new framework (see, for instance, Ramsay and Silverman [10] for more examples). A classical discretization of the data as a sequence of numbers may loose some functional characteristics like smoothness and continuity. For this reason, in the last decades different methods appeared to handle this new kind of data. In an informal way, we may say that a functional datum is a random variable (element would be a better word) that takes its values in a functional space, instead of a finite dimensional one. In this paper, we will study some fundamental concepts of this development, which in a way will result in a mixture between statistics and functional analysis over Hilbert spaces. The main idea is to mix together, in a very general family of distributions, some notions of principal components and principal points. In a multivariate setting, those developments were mainly done by Flury and Tarpey ([2] to [6], [12] to [17], [18] and [20]) at the beginning of the 90's. The idea here is to adapt the results obtained therein to the functional case.

Maybe the final conclusion of this work is not only the theoretical result obtained. Rather, as it was done previously, our results show about the possibility of doing, with some technical difficulties but not critical ones, an interesting generalization of the classical results from multivariate analysis to a more general framework, so as to gain a better comprehension of the phenomenon, as it often tends to happen when abstraction or generalization of a mathematical concept is made.

In section 2, we will define the notion of elliptical families. We will first remind their definition in the multivariate case and later we will extend this definition to the functional case. The definition of self–consistent points and principal points as well as some of their properties are stated in section 3 where we extend the results given in Flury ([2], [3]), Tarpey [13] and Tarpey and Flury [18] to the case of random elements lying in a separable Hilbert space. We also provide a characterization that, under an hypothesis of ellipticity, allows us to make an important link between principal components and self–consistent points. We conclude with some results that allow to compute principal points in a somewhat specific case.

2 Elliptical families

2.1 Review on some finite–dimensional results

For the sake of completeness and to fix our notation we will remind some results regarding elliptical families before extending them to the functional setting. They can be found in Muirhead [9], Seber [11] and also in Frahm [7].

Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector. We will say that $\mathbf{X}$ has an elliptical distribution, and we will denote it as $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, a positive semidefinite matrix

$\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ and a function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that the characteristic function of $\mathbf{X} - \boldsymbol{\mu}$ is given by $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{t})$, for all $\mathbf{t} \in \mathbb{R}^d$. In some situations, for the sake of simplicity, we will omit the symbol ϕ and will denote $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \mathbf{\Sigma})$.

As it is well known, if $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi)$ and $E(\mathbf{X})$ exists, then $E(\mathbf{X}) = \boldsymbol{\mu}$. Moreover, if the second order moments exist $\mathbf{\Sigma}$ is up to a constant the covariance matrix of $\mathbf{X}$, i.e., $Var(\mathbf{X}) = \alpha\mathbf{\Sigma}$. Even more, it is easy to see that the constant α equals to $-2\phi'(0)$, where ϕ' stands for the derivative of ϕ.

The following theorem is a well known result and will also be extended in the sequel to adapt for functional random elements.

Theorem 2.1. *Let $\mathbf{X} \sim \varepsilon(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi)$ with $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ a semidefinite positive matrix with rank$(\mathbf{\Sigma}) = r$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\mathrm{T}}, \boldsymbol{\mu}_2^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{X}_1$ the vector of the first k coordinates of $\mathbf{X}$ ($k < r$) such that $\mathbf{\Sigma}_{11}$ is not singular. Denote by*

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} \in \mathbb{R}^{d\times d}$$

with submatrixes $\mathbf{\Sigma}_{11} \in \mathbb{R}^{k\times k}$, $\mathbf{\Sigma}_{21} \in \mathbb{R}^{(d-k)\times k}$, $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21}^{\mathrm{T}} \in \mathbb{R}^{k\times(d-k)}$ and $\mathbf{\Sigma}_{22} \in \mathbb{R}^{(d-k)\times(r-k)}$. Then,

a) $\mathbf{X} \overset{\mathcal{D}}{\sim} \boldsymbol{\mu} + \mathcal{R}\mathbf{\Lambda}\mathbf{U}^{(r)}$, where $\mathbf{Y} \overset{\mathcal{D}}{\sim} \mathbf{Z}$ means that the two random vectors $\mathbf{Y}$ and $\mathbf{Z}$ have the same distribution, $\mathbf{U}^{(r)}$ is uniformly distributed over $\mathcal{S}^{r-1} = \{\mathbf{y} \in \mathbb{R}^r \,:\, \|\mathbf{y}\| = 1\}$ and $\mathcal{R}$ and $\mathbf{U}^{(r)}$ are independent.

b) Assume that the conditional random vector $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ exists, then $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ has an elliptical distribution $\varepsilon_{d-k}(\boldsymbol{\mu}^*, \mathbf{\Sigma}^*, \phi^*)$ where

$$\begin{aligned} \boldsymbol{\mu}^* &= \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \\ \mathbf{\Sigma}^* &= \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}, \end{aligned}$$

and ϕ^* corresponds to the characteristic generator of $\mathcal{R}^*\mathbf{U}^{(r-k)}$ with

$$\mathcal{R}^* \overset{\mathcal{D}}{\sim} \mathcal{R}\sqrt{1-\beta}\,\Big|(\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} = \mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)).$$

Here $\mathbf{C}_{11}$ stands for the Cholesky square root of $\mathbf{\Sigma}_{11}$, $\mathbf{U}^{(k)}$ is uniformly distributed in $\mathcal{S}^{k-1}$, $\beta \sim Beta(\frac{k}{2}, \frac{r-k}{2})$ and $\mathcal{R}$, β, $\mathbf{U}^{(k)}$ and $\mathbf{U}^{(r-k)}$ are mutually independent.

2.2 Functional case

In this section, we will extend the definition of elliptical distributions to the case of random elements on a separable Hilbert space. The definition will be based on the one given for the multivariate case.

Definition 2.1. Let V be a random element in a separable Hilbert space $\mathcal{H}$. We will say that V has an elliptical distribution of parameters $\mu \in \mathcal{H}$ and $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$, with $\mathbf{\Gamma}$ a self–adjoint, positive semidefinite and compact operator, and we will denote $\mathbf{V} \sim \mathcal{E}(\mu, \mathbf{\Gamma})$, if for any lineal and bounded operator $A : \mathcal{H} \to I\!R^d$ (that is, such that $\sup_{\|x\|=1} \|Ax\| < \infty$) we have that AV has a multivariate elliptical distribution of parameters $A\mu$ and $\mathbf{\Sigma} = A\mathbf{\Gamma}A^*$, i.e., $AV \sim \varepsilon_d(A\mu, \mathbf{\Sigma})$ where $A^* : \mathbb{R}^p \to \mathcal{H}$ stands for the adjoint operator of A.

The following result shows that elliptical families in Hilbert spaces are closed through linear and bounded transformations.

Lemma 2.1. *Let $V \sim \varepsilon(\mu, \mathbf{\Gamma})$ an elliptical random element in $\mathcal{H}_1$ of parameters μ and $\mathbf{\Gamma}$ and $A : \mathcal{H}_1 \to \mathcal{H}_2$ linear and bounded. Then AV is an elliptical random element $\mathcal{H}_2$ of parameters $A\mu$ and $A\mathbf{\Gamma}A^*$.*

Lemma 2.2 shows that both parameters, μ y $\mathbf{\Gamma}$, that characterizes the element V are respectively the expectation and the covariance operator, provide they exist. Its proof can be found in the Appendix.

Lemma 2.2. *Let V be a random element in a separable Hilbert space $\mathcal{H}$ such that $\mathbf{V} \sim \mathcal{E}(\mu, \mathbf{\Gamma})$.*

a) If $E(V)$ exists, then, $E(V) = \mu$.

b) If the covariance operator, $\mathbf{\Gamma}_V$, exists then, $\mathbf{\Gamma}_V = \alpha\, \mathbf{\Gamma}$, for some $\alpha \in \mathbb{R}$.

Based on the finite dimensional results, one way of obtaining random elliptical elements is through the following transformation. Let V_1 be a gaussian element in $\mathcal{H}$ with zero mean and covariance operator $\mathbf{\Gamma}_{V_1}$, and let Z be a random variable with distribution G independent of V_1. Given $\mu \in \mathcal{H}$, define $V = \mu + Z\, V_1$. Then, V has an elliptical distribution and if $E(Z^2)$ exists $\mathbf{\Gamma}_V = E(Z^2)\, \mathbf{\Gamma}_{V_1}$.

We are interested in obtaining some properties concerning the conditional distribution of elliptical families similar to those existing in the multivariate setting. Let V be a random element belonging to an elliptical family of parameter μ and $\mathbf{\Gamma}$ and let us consider in $\mathcal{H}$ the orthonormal basis, $\{\phi_n\}_{n\in\mathcal{I}}$ ($\mathcal{I}$ countable or finite) constructed using the eigenfunctions of the operator $\mathbf{\Gamma}$ related to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Given $d \in \mathcal{I}$ fixed, define the closed subspaces (and so Hilbert spaces)

$$\mathcal{H}_1 = < \phi_1, \ldots, \phi_d > \qquad \mathcal{H}_2 = < \phi_1, \ldots, \phi_d >^{\perp}$$

Define over these spaces the truncating projections, that is, $P_d = P_{\mathcal{H}_1} : \mathcal{H} \to \mathcal{H}_1$ and $P_{\mathcal{H}_2} : \mathcal{H} \to \mathcal{H}_2$ such that

$$P_{\mathcal{H}_1}(\phi_i) = \begin{cases} \phi_i & i = 1, 2, \ldots, d \\ 0 & i > d \end{cases} \qquad P_{\mathcal{H}_2}(\phi_i) = \begin{cases} \phi_i & i > d \\ 0 & i = 1, 2, \ldots, d \end{cases} . \tag{1}$$

We will make a composition of P_d with the natural operator that identifies $\mathcal{H}_1$ with $\mathbb{R}^d$. That is, we will consider the operator $T_d : \mathcal{H} \to \mathbb{R}^d$ defined as

$$T_d(\phi_i) = \left\{ \begin{array}{ll} \mathbf{e}_i & i = 1, 2, \ldots, d \\ 0 & i > n \end{array} \right. , \tag{2}$$

with $\mathbf{e}_1, \ldots, \mathbf{e}_d$ the vectors of the canonical base of $\mathbb{R}^d$. Then, for any $x \in \mathcal{H}$ we have that $T_d(x) = \sum_{j=1}^{d} < x, \phi_j > \mathbf{e}_j$. We will use T_d instead of P_d as a projector in many situations, because its image is $\mathbb{R}^d$ and we will call each of them truncating projectors.

Based on these projections we can construct $V_1 = P_{\mathcal{H}_1} V \in \mathcal{H}_1$, $W_1 = T_d V \in \mathbb{R}^d$ and $V_2 = P_{\mathcal{H}_2} V \in \mathcal{H}_2$ random elements, both of them elliptical by Lemma 2.1.

We have essentially split the random element in two parts, one of them being finite dimensional which will allow us to define a conditional distribution $V_2|V_1$ following the guidelines previously established.

Theorem 2.2. *Let $\mathcal{H}$ be a separable Hilbert space. Let V be a random element in $\mathcal{H}$ with distribution $\mathcal{E}(\mu, \mathbf{\Gamma})$ with finite second moments. Without loss of generality, we can assume that $\mathbf{\Gamma}$ is the covariance operator. Assume that $\mathbf{\Gamma}$ is Hilbert–Schmidt so that $\sum_{i=1}^{\infty} \lambda_i < \infty$. Let $d \in \mathcal{I}$ fixed and consider $V_1 = P_{\mathcal{H}_1} V$, $W_1 = T_d V$ y $V_2 = P_{\mathcal{H}_2} V$ with $P_{\mathcal{H}_1}$ y $P_{\mathcal{H}_2}$ defined in (1) and T_d defined in (2). Let $\lambda_1 \geq \lambda_2 \geq \ldots$ be the eigenvalues of $\mathbf{\Gamma}$ and assume that $\lambda_d > 0$. Then,*

a) the covariance matrix of W_1 given by $\mathbf{\Sigma}_{W_1} = T_d \mathbf{\Gamma} T_d^* = diag\,(\lambda_1, \ldots, \lambda_d)$ is non–singular

b) $E(V_2|W_1) = \mu_2 + \mathbf{\Gamma}_{V_2,W_1} \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1)$,

where $\mathbf{\Gamma}_{V_2,W_1}$ is the covariance operator between V_2 and W_1, $\boldsymbol{\mu}_1 = E(W_1)$ and $\mu_2 = E(V_2)$.

3 Self–consistent points and principal points

As mentioned in the Introduction self–consistent and principal points were studied by Flury ([2], [3]), Tarpey [13] and Tarpey and Flury [18] in the multivariate setting. Later on, Tarpey and Kinateder [19] extended their definition and properties for gaussian processes while Tarpey *et al.* [21] applied principal points to estimate a set of representative longitudinal response curves from a clinical trial. The aim of this section is to extend some of the properties previously obtained to include elliptical families.

For the sake of completeness, we remind the definition of self–consistency and principal points.

Definition 3.1. Let $\mathcal{W} = \{y_1, \ldots, y_k\}$ with $y_i \in \mathcal{H}$, $1 \leq i \leq k$ we define the minimun distance of V to the set $\mathcal{W}$ as $d(V, \{y_1, \ldots, y_k\}) = \min_{1 \leq j \leq k} \|V - y_j\|$.

The set $\mathcal{W}$ induce a partition of the space $\mathcal{H}$ determined by the domains of attraction.

Definition 3.2. Given $\mathcal{W} = \{y_1, \ldots, y_k\}$, the domain of attraction $\mathcal{D}_j$ of y_j consists in all the elements of $\mathcal{H}$ that have y_j as the closest point of $\mathcal{W}$, that is, $\mathcal{D}_j = \{x \in \mathcal{H} : \|x - y_j\| < \|x - y_\ell\| \,, \quad \ell \neq j\}$.
For points $x \in \mathcal{H}$ with equal distance to two or several y_j, we assign them arbitrarily to the set with lower index j.

Definition 3.3. Let V be a random element in $\mathcal{H}$ with expectation $E(V)$. A set $\mathcal{W} = \{y_1, \ldots, y_k\}$ is said to be self-consistent for V if $E(V|V \in \mathcal{D}_j) = y_j$.
A random element W is called self-consistent for V if $E(V|W) = W$.

Definition 3.4. Let V be a random element in $\mathcal{H}$ with finite second moment. The elements $\xi_1, \ldots, \xi_k$ are called principal points of V if

$$D_V(k) = E(d^2(V, \{\xi_1, \ldots, \xi_k\})) = \min_{y_j \in \mathcal{H}} E(d^2(V, \{y_1, \ldots, y_k\}))$$

Lemma 3 in Tarpey and Kinateder [19] establishes for $L^2(\mathcal{I})$ functions, with $\mathcal{I}$ a real bounded interval, the well–known result in multivariate analysis that the mean of a distribution lies in the convex hull of any set of self–consistent points. Moreover, Flury [3] established that principal points of a random vector in $\mathbb{R}^p$ are self–consistent points. This result was generalized to random funcions in $L^2(\mathcal{I})$ by Tarpey and Kinateder [19]. The same arguments allow to establish these results for any separable Hilber space $\mathcal{H}$, we state them without proof.

Lemma 3.1. *Let V be a random element of a separable Hilbert space $\mathcal{H}$ such that $E(V)$ exists. Then,*

a) if $\{y_1, \ldots, y_k\}$ is a self-consistent set, then $E(V)$ is a convex combination of $y_1, \ldots, y_k$.

b) Moreover, if V has finite second moments and the set $\mathcal{W} = \{\xi_1, \ldots, \xi_k\}$ is a set of principal points for V, then it is self-consistent.

As a consequence of Lemma 3.1, if $k = 1$ and V is a random element with self-consistent set $\{y_1\}$, then $y_1 = E(V)$. Moreover, we will have self-consistent points whenever we have principal points.

The following result will allows us to assume, in the sequel, that the random element V has expectation 0. It also generalizes Lemma 2.2 in Tarpey *et al.* [20], to the infinite–dimensional setting. Its proof is given in the Appendix.

Lemma 3.2. *Let V be a random element of a separable Hilbert space $\mathcal{H}$ and define $V_2 = \nu + \rho\, UV$ with $\nu \in \mathcal{H}$, ρ a scalar and $U : \mathcal{H} \to \mathcal{H}$ a unitary operator, i.e., surjective and isometric. Then, we have that* *a) If $\mathcal{W} = \{y_1, \ldots, y_k\}$ is a set of k self-consistent points of V, then $\mathcal{W}_2 = \{\nu + \rho\, U y_1, \ldots, \nu + \rho\, U y_k\}$ is a set of k self-consistent points of V_2.*

b) If $\mathcal{W} = \{y_1, \ldots, y_k\}$ is a set of k principal points of V, then $\mathcal{W}_2 = \{\nu + \rho\, U y_1, \ldots, \nu + \rho\, U y_k\}$ is a set of k principal points of V_2 and $E(d^2(V_2, \mathcal{W}_2)) = \rho^2 E(d^2(V, \mathcal{W}))$.

Lemma 3.3 is analogous to Lemma 2.3 in Tarpey *et al.* [20].

Lemma 3.3. *Let V be a random element with expectation 0. Let $\{y_1, \ldots, y_k\}$ be a set of k self–consistent points of V spanning a subspace $\mathcal{M}$ of dimension q, with an orthonormal basis $\{e_1, \ldots, e_q\}$. Then, the random vector of $\mathbb{R}^q$ defined by $\mathbf{X} = (X_1, \ldots, X_q)^{\mathrm{T}}$ with $X_i =< e_i, V >$ will have $\mathcal{W} = \{\mathbf{w}_j\}_{1 \leq j \leq k}$ with $\mathbf{w}_j = (w_{1j}, \ldots, w_{qj})^{\mathrm{T}}$ and $w_{ij} =< e_i, y_j >$ as self–consistent set.*

The notion of best $k-$point approximation has been considered by Tarpey *et al.* [20] for finite–dimensional random elements. It extends immediately to elements on a Hilbert space.

Definition 3.5. Let $W \in \mathcal{H}$ be a discrete random element, jointly distributed with the random element $V \in \mathcal{H}$ and denote $\mathcal{S}(W)$ the support of W. The random element W is a best $k-$point approximation to V if $\mathcal{S}(W)$ contains exactly k different elements $y_1, \ldots, y_k$ and $E(\|V - W\|^2) \leq E(\|V - Z\|^2)$ for any random element $Z \in \mathcal{H}$ whose support has at most k points, i.e., $\#\mathcal{S}(Z) \leq k$.

The following result is the infinite–dimensional counterpart of Lemma 2.4 in Tarpey *et al.* [20].

Lemma 3.4. *Let W be a best $k-$point approximation to V and denote by $y_1, \ldots, y_k$ a set of k different elements in $\mathcal{S}(W)$ and by $\mathcal{D}_j$ the domain of attraction of y_j. Then,*

a) If $V \in \mathcal{D}_j$ then W equals y_j with probability 1. That is, $W = \sum_{i=1}^{k} y_i \mathbb{I}_{\mathcal{D}_j}(V)$.

b) $\|V - W\| \leq \|V - y_j\|$ a.s. for all $y_j \in \mathcal{S}(W)$.

c) $E(V|W) = W$ a.s., i.e., W is self-consistent for V.

It is worth noticing that given a self-consistent set $\{y_1, \ldots, y_k\}$ of V, as in the finite–dimensional case, we can define in a natural way a random variable $Y = \sum_{i=1}^{k} y_i \mathbb{I}_{V \in \mathcal{D}_i}$ with support $\{y_1, \ldots, y_k\}$ and so, $P(Y = y_j) = P(V \in \mathcal{D}_j)$. Since $\{y_1, \ldots, y_k\}$ is a self-consistent set, we will have that $E(V|Y) = Y$, with probability 1. As in the finite–dimensional setting, Y will not be necessarily a best approximation, unless the set $\{y_1, \ldots, y_k\}$ is a set of k principal points.

As mentioned above, if V is a random element with a self-consistent set of $k = 1$ elements $\{y_1\}$ then $E(V) = y_1$ and so that if we assume $E(V) = 0$, we have $y_1 = 0$. The forthcoming results try to characterize the subspace spanned by the self–consistent points when $k > 1$. They generalize the results obtained in the finite–dimensional case by Tarpey *et al.* [20] and extended to gaussian processes by Tarpey and Kinateder [19]. They also justify the use of the $k-$means algorithm not only for gaussian processes but also for elliptical processes with finite second moments.

Theorem 3.1. *Let V be a random element in a separable Hilbert space $\mathcal{H}$, with finite second moment and assume that $E(V) = 0$. Let $W = \{y_1, \ldots, y_k\}$ be a set of k self–consistent points for V. Then, $y_j \in Ker(\mathbf{\Gamma}_V)^{\perp}$, for all j, where $\mathbf{\Gamma}_V$ denotes the covariance operator of V.*

In particular, if $\mathcal{W}$ denotes the linear space spanned by the k self–consistent points, we have get easily that $Ker(\mathbf{\Gamma}_V) \cap \mathcal{W} = \{0\}$. Moreover, it will also hold that $\mathbf{\Gamma}_V(\mathcal{W}) \cap \mathcal{W}^{\perp} = \{0\}$. This last fact will follow from the properties of semidefinite and diagonalizable operators.

Corollary 3.1. *Let V be a random element in a separable Hilbert space $\mathcal{H}$, with finite second moment and compact covariance operator $\mathbf{\Gamma}_V$, such that $E(V) = 0$. Let $W = \{y_1, \ldots, y_k\}$ be a set with k self–consistent points for V and denote $\mathcal{W}$ the subspace spanned by them. Then,*

a) $Ker(\mathbf{\Gamma}_V) \cap \mathcal{W} = \{0\}$.

b) Denote by $\mathcal{W}$ be the subspace spanned by the set $\{y_1, \ldots, y_k\}$ of $k > 1$ self-consistent points. Then, $\mathbf{\Gamma}_V \mathcal{W} \cap \mathcal{W}^{\perp} = \{0\}$.

The following Theorems provide the desired result relating, for elliptical elements, self–consistency and principal components.

Theorem 3.2. *Let V be a random elliptical element with $E(V) = 0$ and compact covariance operator $\mathbf{\Gamma}_V$. Let $\mathcal{W}$ the subspace spanned by the set $\{y_1, \ldots, y_k\}$ of $k > 1$ self-consistent points. Then, $\mathcal{W}$ is spanned by a set of eigenfunctions of $\mathbf{\Gamma}_V$.*

Theorem 3.3. *Let V be a random elliptical element with $E(V) = 0$ and compact covariance operator $\mathbf{\Gamma}_V$. If k principal points of V generate a subspace $\mathcal{W}$ of dimension q, then this subspace will also be spanned by the q eigenfunctions of $\mathbf{\Gamma}_V$ related to the q largest eigenvalues.*

3.1 Properties of principal points and resolution for the case $k = 2$

As mentioned above, when $k = 1$ the principal point equals the mean of the distribution. The goal of this section is to obtain, as in the finite–dimensional setting, an explicit expression for the principal points when $k = 2$. As is well known, even when dealing with finite–dimensional data, no general result is known for any value of k. The following theorem will be very useful in the sequel and it generalizes a result given, for the finite–dimensional case, by Flury [2]. It is worth noticing that Theorem 3.4 do not require to the random element to have an elliptical distribution.

Theorem 3.4. *Let $\mathcal{H}$ be a separable Hilbert space and $V : \Omega \to \mathcal{H}$ a random element with mean μ and with k principal points $\xi_1, \ldots, \xi_k \in \mathcal{H}$. Then, the dimension of the linear space spanned by $\xi_1 - \mu, \ldots, \xi_k - \mu$ is strictly lower than k.*

In particular, when $k = 1$ we get that the mean is a $1-$principal point. We will now focuss our attention of the case $k = 2$ and the results will be derived for elliptical distributions.

Theorem 3.5 generalizes Theorem 2 in Flury [2] which states an analogous property for the finite dimensional vectors. As in euclidean spaces, the result assumes the existence of self-principal points for real variables, conditions under which this holds are given in Theorem 1 of Flury [2].

Theorem 3.5. *Let V be an elliptical random element of a separable Hilbert space $\mathcal{H}$ with mean μ and covariance operator $\mathbf{\Gamma}$ with finite trace. Denote by $\phi_1 \in \mathcal{H}$ an eigenfunction of $\mathbf{\Gamma}$ with norm 1, related its largest eigenvalue λ_1. Assume that the real random variable $Y =< v, V - \mu >$ has two principal points for any $v \in \mathcal{H}$ and let γ_1, γ_2 the two principal points of the real random variable $< \phi_1, V - \mu >$. Then, V has two principal points $y_1 = \mu + \gamma_1 \phi_1$ and $y_2 = \mu + \gamma_2 \phi_1$.*

A Appendix

Proof of Lemma 2.1 Let $B : \mathcal{H}_2 \to \mathbb{R}^d$ linear and bounded, let us show that BAV is an elliptical multivariate random vectorof mean $BA\mu$ and covariance matrix $BA\mathbf{\Gamma}A^*B^*$.

Let $B \circ A : \mathcal{H}_1 \to \mathbb{R}^d$ the composition. Then, $B \circ A$ is linear and bounded, therefore $BAV = (B \circ A)(V)$ is elliptical with parameters $B \circ A(\mu) = BA\mu$ and $(B \circ A)\mathbf{\Gamma}(B \circ A)^* = BA\mathbf{\Gamma}A^*B^*$, finishing the proof. □

Proof of Lemma 2.2. a) Denote by $\mathcal{H}^*$ the dual space of $\mathcal{H}$, i.e., $\mathcal{H}^*$ is the set of all linear and continuous functions $f : \mathcal{H} \to \mathbb{R}$. Let $f \in \mathcal{H}^*$, then $E(|f(V)|) < \infty$ and since $f : \mathcal{H} \to \mathbb{R}$ is linear and continuous it is linear and bounded. Then, $f(V)$ has an elliptical distribution with parameters $f(\mu)$ and $f\mathbf{\Gamma}f^*$. The existence of $E(V)$ entails that $E(f(V))$ exists and that $E(f(V)) = f(E(V))$. Since $E(f(V)) = f(\mu)$, by uniqueness we get that $E(V) = \mu$.

The proof of b) will follow from the properties of the covariance operator and Lemma 2.1 using the uniqueness of the covariance.

For that purpose, it will be convenient to have defined a series of special operators. Since $\mathcal{H}$ is separable, it admits an orthonormal countable base, that is, there exists $\{\phi_n\}_{n \in I\!N}$ (eventually finite if the space is of finite dimension) orthonormal generating $\mathcal{H}$. We will choose as basis of $\mathcal{H}$ the basis of eigenfunctions of $\mathbf{\Gamma}$ related to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Without loss of generality we can assume that $\lambda_1 > 0$, otherwise $P(V = E(V)) = 1$ and the conclusion would be trivial.

Define $P_n = P_{<\phi_1, \ldots, \phi_n>}$, the orthogonal projection onto the subspace $\mathcal{H}_1$ spanned by $\phi_1, \ldots, \phi_n$ and T_n as in (2).

We want to show that $\mathbf{\Gamma}_V = \alpha \mathbf{\Gamma}$, i.e., that

$$< \alpha \mathbf{\Gamma} u, v >_{\mathcal{H}} = a_V(u, v) = Cov(< v, V >_{\mathcal{H}}, < v, V >_{\mathcal{H}})$$

for any $u, v \in \mathcal{H}$, where we have explicitly written the space where the internal product is taken for clarity.

Let $d \in I\!N$ fixed. Using that V has an elliptical distribution, we get that $T_d V \sim \varepsilon_d(T_d \mu, T_d \mathbf{\Gamma} T_d^*)$. On the other hand, since V has finite second moment, the same holds for $T_d V$ which implies that $E(T_d V) = T_d \mu$ and the covariance matrix of $T_d V$, denoted $\mathbf{\Sigma}_d$, is proportional to $T_d \mathbf{\Gamma} T_d^*$. Therefore, it exists $\alpha_d \in \mathbb{R}$ such that $\mathbf{\Sigma}_d = \alpha_d T_d \mathbf{\Gamma} T_d^*$.

We begin by showing thet α_d does not depend on d. It is easy to see that

$$T_d \mathbf{\Gamma} u = \sum_{i=1}^{d} \lambda_i < \phi_i, u >_{\mathcal{H}} \ \mathbf{e}_i \ .$$

Therefore, using that $< \phi_i, T_d^* \mathbf{x} >_{\mathcal{H}} = < T_d \phi_i, \mathbf{x} >_{\mathbb{R}^d} = < \mathbf{e}_i, \mathbf{x} >_{\mathbb{R}^d} = x_i$ for all $\mathbf{x} \in \mathbb{R}^d$, we obtain that

$$T_d \mathbf{\Gamma} T_d^* = \text{diag}\,(\lambda_1, \ldots, \lambda_d) \ . \tag{A.1}$$

Let $k \leq d$ and $\pi_k : \mathbb{R}^d \to \mathbb{R}^k$ be the usual projection $\pi_k(\mathbf{x}) = \pi_k\left((x_1, \ldots, x_p)^{\mathrm{T}}\right) = (x_1, \ldots, x_k)^{\mathrm{T}} = \mathbf{A}_k \mathbf{x}$, where $\mathbf{A}_k = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. The fact that $\pi_k T_d V = T_k V$ implies that the covariance matrix of $T_k V$ is given by $\mathbf{\Sigma}_k = \mathbf{A}_k \mathbf{\Sigma}_d \mathbf{A}_k^{\mathrm{T}}$ and so, $\alpha_k T_k \mathbf{\Gamma} T_k^* = \alpha_d \mathbf{A}_k (T_d \mathbf{\Gamma} T_d^*) \mathbf{A}_k^{\mathrm{T}}$ which together with (A.1), implies that $\alpha_k = \alpha_d$.

Hence, there exists $\alpha \in \mathbb{R}$ such that for all $d \in I\!N$ the covariance matrix $\mathbf{\Sigma}_d$ of $T_d V$ is equal to $\alpha T_d \mathbf{\Gamma} T_d^*$, implying that

$$< \alpha T_d \mathbf{\Gamma} T_d^* \mathbf{x}, \mathbf{y} >_{\mathbb{R}^d} = Cov(< \mathbf{x}, T_d V >_{\mathbb{R}^d}, < \mathbf{y}, T_d V >_{\mathbb{R}^d}).$$

Using the definition of adjoint of T_d, we have that $< T_d \mathbf{\Gamma} T_d^* \mathbf{x}, \mathbf{y} >_{\mathbb{R}^d} = < \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}}$, meanwhile the right member of the equality can be written as

$$Cov(< \mathbf{x}, T_d V >_{\mathbb{R}^d}, < \mathbf{y}, T_d V >_{\mathbb{R}^d}) = Cov(< T_d^* \mathbf{x}, V >_{\mathcal{H}}, < T_d^* \mathbf{y}, V >_{\mathcal{H}}) \ .$$

Then, we have that for all $d \in I\!N$, $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^d$,

$$< \alpha \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}} = Cov(< T_d^* \mathbf{x}, V >_{\mathcal{H}}, < T_d^* \mathbf{y}, V >_{\mathcal{H}}) = a_V(T_d^* \mathbf{x}, T_d^* \mathbf{y}).$$

Given $u, v \in \mathcal{H}$, define $u_d = P_d u$, $v_d = P_d v$, $\mathbf{x} = T_d u = T_d u_d$ and $\mathbf{y} = T_d v = T_d v_d$. We have that $\lim_{d \to \infty} \|u - u_d\| = 0$ and $\lim_{d \to \infty} \|v - v_d\| = 0$. Then, using that $u_d = T_d^* \mathbf{x}$, $v_d = T_d^* \mathbf{y}$ we get

$$< \alpha \mathbf{\Gamma} u_d, v_d >_{\mathcal{H}} = < \alpha \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}} = a_V(T_d^* \mathbf{x}, T_d^* \mathbf{y}) = a_V(u_d, v_d).$$

The continuity of a_V entails that $\lim_{d \to \infty} a_V(u_d, v_d) = a_V(u, v)$. On the other hand, using that $\mathbf{\Gamma}$ is a self–adjoint, compact operator, we obtain that $\lim_{d \to \infty} < \mathbf{\Gamma} u_d, v_d >_{\mathcal{H}} = < \mathbf{\Gamma} u, v >_{\mathcal{H}}$, which concludes the proof. □

Proof of Theorem 2.2. The proof of a) follows inmediately since $\lambda_d > 0$.
b) It is enough to show that $E(|f(V_2|W_1)|) < \infty$ and $E(f(V_2)|W_1) = f(\mu_2 + \mathbf{\Gamma}_{V_2, W_1} \mathbf{\Sigma}_{W_1}^{-1}(W_1 - \mu_1))$, for any $f \in \mathcal{H}^*$. Let $\mathbf{W} = (W_1, f(V_2)) = TV$. Using that with T is a bounded and linear operator, we get that $\mathbf{W}$ is elliptical of parameters $(\boldsymbol{\mu}_1, f(\mu_2))^{\mathrm{T}}$ and

$$T \mathbf{\Gamma} T^* = \begin{pmatrix} \mathbf{\Sigma}_{W_1} & Cov(W_1, f(V_2)) \\ Cov(W_1, f(V_2))^{\mathrm{T}} & f \mathbf{\Gamma} f^* \end{pmatrix} \ .$$

Using Theorem 2.1, we get that $f(V_2)|W_1$ has also an elliptical distribution with expectation given by $\boldsymbol{\mu}_f = f(\mu_2) + Cov(f(V_2), W_1) \mathbf{\Sigma}_{W_1}^{-1}(W_1 - \boldsymbol{\mu}_1)$. On the other hand, $Cov(f(V_2), W_1) = f \mathbf{\Gamma}_{V_2, W_1}$ which implies that

$$\begin{aligned} E(f(V_2)|W_1) &= f(\mu_2) + Cov(f(V_2), W_1) \mathbf{\Sigma}_{W_1}^{-1}(W_1 - \boldsymbol{\mu}_1) \\ &= f(\mu_2) + f \mathbf{\Gamma}_{V_2, W_1} \mathbf{\Sigma}_{W_1}^{-1}(W_1 - \boldsymbol{\mu}_1) = f(\mu_2 + \mathbf{\Gamma}_{V_2, W_1} \mathbf{\Sigma}_{W_1}^{-1}(W_1 - \boldsymbol{\mu}_1)) \end{aligned}$$

and so, we conclude the proof. $\square$

Proof of Lemma 3.2. a) Using that $\mathcal{W}$ is self-consistent for V, we get that $E(V|V \in \mathcal{D}_j) = y_j$. Let us notice that, since U is a unitary operator, $V \in \mathcal{D}_j$ if and only if $V_2 = \nu + \rho\, UV \in \widetilde{\mathcal{D}_j}$. Therefore, $\widetilde{\mathcal{D}_j}$ is the domain of attraction of $\nu + \rho\, Uy_j$ which implies that $\nu + \rho\, U\mathcal{D}_j \subset \widetilde{\mathcal{D}_j}$. Hence,

$$\begin{aligned} E(V_2|V_2 \in \widetilde{\mathcal{D}_j}) &= E(\nu + \rho UV|V_2 \in \widetilde{D_j}) = \nu + \rho U E(V|\nu + \rho UV \in \nu + \rho U\mathcal{D}_j) \\ &= \nu + \rho U E(V|V \in \mathcal{D}_j) = \nu + \rho U y_j. \end{aligned}$$

b) Let $\xi_1, \ldots, \xi_k$ be any set of points in $\mathcal{H}$ and denote by $\mathcal{A}_1, \ldots, \mathcal{A}_k$ their respective domains of attraction. We have to prove that $E(d^2(V_2, \mathcal{W}_2)) \le E(d^2(V_2, \{\xi_1, \ldots, \xi_k\}))$.

Let z_j such that $\xi_j = \nu + \rho\, Uz_j$, then

$$\begin{aligned} E(d^2(V_2, \{\xi_1, \ldots, \xi_k\})) &= E(\min_{1\le j\le k} \|V_2 - \xi_j\|^2) = E(\min_{1\le j\le k} \|\nu + \rho\, UV - \xi_j\|^2) \\ &= E(\min_{1\le j\le k} \|\nu + \rho\, UV - \nu - \rho\, Uz_j\|^2) = E(\min_{1\le j\le k} \|\rho\, UV - \rho\, Uz_j\|^2) \\ &= \rho^2 E(\min_{1\le j\le k} \|UV - Uz_j\|^2) = \rho^2 E(\min_{1\le j\le k} \|V - z_j\|^2) \end{aligned}$$

where the last inequality holds from the fact that U is an isometry. On the other hand, using that $\mathcal{W}$ is a set of principal points of V, we get that $\{y_1, \ldots, y_k\} = \operatorname{argmin}_{z_1,\ldots,z_k} E(\min_{1\le j\le k} \|V - z_j\|^2)$. Threfore, we have that

$$E(d^2(V_2, \{\xi_1, \ldots, \xi_k\})) = \rho^2 E(\min_{1\le j\le k} \|V - z_j\|^2) \ge \rho^2 E(\min_{1\le j\le k} \|V - y_j\|^2) = E(\min_{1\le j\le k} \|V_2 - \xi_{0,j}\|^2)$$

where $\xi_{0,j} = \nu + \rho U y_j$, which means $\mathcal{W}_2$ are principal points of V_2. Besides, we also obtain that $E(d^2(V_2, \mathcal{W}_2)) = \rho^2 E(d^2(V, \mathcal{W}))$. $\square$

Proof of Lemma 3.3. Let us define $A : \mathcal{H} \to \mathbb{R}^q$ as $A(v) = \mathbf{v} = (v_1, \ldots, v_q)^{\mathrm{T}}$ with $v_i =< e_i, v >$. We want to show that $\mathbf{X} = AV \in \mathbb{R}^q$ has as a self-consistent set $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} = \{Ay_1, \ldots, Ay_k\}$. Denote by $\widetilde{\mathcal{D}_j}$ the domain of attraction of Ay_j and by $\mathcal{D}_j$ that of y_j. Then, by extending the orthonormal basis $\{e_1, \ldots, e_q\}$ of $\mathcal{M}$ to an orthonormal basis $\{e_\ell, \ell \ge 1\}$ of $\mathcal{H}$, and using that $< y_j, e_\ell >= 0$ for $\ell > q$, it is easy to see that $V \in \mathcal{D}_j$ if and only if $AV \in \widetilde{\mathcal{D}_j}$, which implies that

$$E(AV|AV \in \widetilde{\mathcal{D}_j}) = AE(V|AV \in \widetilde{\mathcal{D}_j}) = AE(V|V \in \mathcal{D}_j) = Ay_j = \mathbf{w}_j\ ,$$

concluding the proof. $\square$

Proof of Lemma 3.4. a) We will always suppose that the probability of V being found in the frontier of two domains of attractions $(\mathcal{D}_i \cap \mathcal{D}_j)$ is 0.

Let us suppose that the result is false and denote by Ω the common probability space. That is, let us assume that the set $A = \{\omega : V(\omega) \in \mathcal{D}_j\} \cap \{W(\omega) = y_r\}$ has probability strictly positive, with $r \neq j$ and define a new element $Z(\omega)$ that is equal to $W(\omega)$ if $\omega \notin A$ and is equal to y_i if $\omega \in A$. We will show that Z is a better approximation of V than W.

$$
\begin{aligned}
E(\|V - Z\|^2) &= E(\|V - Z\|^2 \mathbb{I}_{A^c}) + E(\|V - Z\|^2 \mathbb{I}_A) \\
&= E(\|V - W\|^2 \mathbb{I}_{A^c}) + E(\|V - y_j\|^2 \mathbb{I}_A)
\end{aligned}
$$

However, since for any $\omega \in A$, we have that $V(\omega) \in \mathcal{D}_j$, we get that $\|V - y_j\|^2 \mathbb{I}_A < \|V - y_r\|^2 \mathbb{I}_A = \|V - W\|^2 \mathbb{I}_A$. Therefore, using that A is a set with positive probability, we obtain that $E(\|V - y_j\|^2 \mathbb{I}_A) < E(\|V - W\|^2 \mathbb{I}_A)$ and so

$$
E(\|V - W\|^2 \mathbb{I}_{A^c}) + E(\|V - y_j\|^2 \mathbb{I}_A) < E(\|V - W\|^2 \mathbb{I}_{A^c}) + E(\|V - W\|^2 \mathbb{I}_A) = E(\|V - W\|^2) \,,
$$

which entails that $E(\|V - Z\|^2) < E(\|V - W\|^2)$ implying that Z is a better $k-$point approximation than W, concluding the proof of a).
b) If $V(\omega) \in \mathcal{D}_j$ then, using a) we get that, except for a zero probability set, $W(\omega) = y_j$. Then, $\|V - W\|(\omega) = \|V(\omega) - y_j\| \leq \|V(\omega) - y_i\|$ for any y_i since $V(\omega) \in \mathcal{D}_j$.
c) $E(V|W)$ is a measurable function $g(W)$ that minimizes the expected squared distance between V and any measurable function $h(W)$. Using a) we have that, any function $h(W)$ has a support containing at most k points. Then, by the definition of best approximation, we have that W is a better approximation than $h(W)$ for any measurable function h, with the expected squared distance criteria. Then, the function $g(W) = E(V|W)$ equals W. $\square$

In order to prove Theorem 3.2 we will need some technical Lemmas. In particular, these lemmas will allow to derive that the matrix $\mathbf{\Gamma}_{W_1,W_1}$ defined therein is not singular, which is a fundamental step in order to get the desired conclusion.

Lemma A.1. *Let V be a random element in a separable Hilbert space $\mathcal{H}$, with finite second moment and assume that $E(V) = 0$.*

a) If $h \in Ker(\mathbf{\Gamma}_V)$ then $< h, V > \equiv 0 \in \mathbb{R}$ a.s.

b) Denote by $\mathcal{H}_1 = Ker(\mathbf{\Gamma}_V)$, the kernel of the covariance operator and $\mathcal{H}_2 = Ker(\mathbf{\Gamma}_V)^\perp$ its orthogonal. Then, $P\left(P_{\mathcal{H}_1} V = 0\right) = 1$ and $P\left(P_{\mathcal{H}_2} V = V\right) = 1$, i.e, $P\left(V \in Ker(\mathbf{\Gamma}_V)^\perp\right) = 1$.

Proof. The proof of a) follows easily noticing that $E(< h, V >) =< h, E(V) >= 0$ and $Var(< h, V >) = Cov(< h, V >, < h, V >) =< \mathbf{\Gamma}_V h, h >= 0$, since $h \in Ker(\mathbf{\Gamma}_V)$.

b) Note that the separability of $\mathcal{H}$ entails that $V = P_{\mathcal{H}_1} V + P_{\mathcal{H}_2} V$, since $Ker(\mathbf{\Gamma}_V)$ is a closed subspace. Thus, it will be enough to show that $P_{\mathcal{H}_1} V = 0$ with probability 1. For the sake of simplicity, we will assume that both $\mathcal{H}_1$ and $\mathcal{H}_2$ are infinite dimensional spaces. Otherwise, the same calculations hold but using a finite index set as the only significant change. Denote by

$\{e_1, e_2, \ldots\}$ an orthonormal basis of $\mathcal{H}_1$ and extend it to a basis of $\mathcal{H}$ so that, $\{f_1, f_2, \ldots\}$ will denote the orthonormal basis of $\mathcal{H}_2$. Then, $Var(P_{\mathcal{H}_1}V) = E\left(\|P_{\mathcal{H}_1}V - E\left(P_{\mathcal{H}_1}V\right)\|^2\right) = \sum_{i=1}^{\infty} < \mathbf{\Gamma}_{P_{\mathcal{H}_1}V}\ e_i, e_i >= \sum_{i=1}^{\infty} < \mathbf{\Gamma}_V\ e_i, e_i >$ and each summand equals zero since $e_i \in Ker(\mathbf{\Gamma}_V)$. On the other hand, $E(P_{\mathcal{H}_1}V) = P_{\mathcal{H}_1}E(V) = 0$, therefore $E\left(\|P_{\mathcal{H}_1}V\|^2\right) = 0$, which implies that $P_{\mathcal{H}_1}V = 0$ *a.s.*, concluding the proof. □

Corollary A.1. *If* $A \cap Ker(\mathbf{\Gamma}_V)^{\perp} = \emptyset$ *then* $P(V \in A) = 0$.

The proof is inmediately since $P(V \in Ker(\mathbf{\Gamma}_V)^{\perp} = 1$.

Proof of Theorem 3.1. Note that $Ker(\mathbf{\Gamma}_V)^{\perp}$ is a closed subspace and therefore a convex set. For each point y_j, its domain of attraction $\mathcal{D}_j$ is also a convex set, therefore, $Ker(\mathbf{\Gamma}_V)^{\perp} \cap \mathcal{D}_j$ is also convex. By Lemma A.1, the support of the random element V is included in $Ker(\mathbf{\Gamma}_V)^{\perp}$, thus

$$y_j = E(V|V \in \mathcal{D}_j) = E(V|V \in \mathcal{D}_j, V \in Ker(\mathbf{\Gamma}_V)^{\perp}) = E(V|V \in \mathcal{D}_j \cap Ker(\mathbf{\Gamma}_V)^{\perp}).$$

Now the proof follows easily by noticing that the expectation of a random element taking values in a convex set $\mathcal{C}$ will also be in $\mathcal{C}$, i.e., $y_j \in \mathcal{D}_j \cap Ker(\mathbf{\Gamma}_V)^{\perp} \subset Ker(\mathbf{\Gamma}_V)^{\perp}$. □

Proof of Corollary 3.1. a) follows immediately from Theorem 3.1. We have only to prove b). Let $z \in \mathbf{\Gamma}_V(\mathcal{W}) \cap \mathcal{W}^{\perp}$, then $z = \mathbf{\Gamma}_V w$ with $w \in \mathcal{W}$. We want to show that $z = 0$. Let $\{\phi_j\}_{j \in \mathbb{N}}$ be an orthonormal base of eigenfunctions of $\mathbf{\Gamma}_V$ related to the eigenvalues $\mu_1 \geq \ldots \geq \mu_j \geq \ldots$ Then, $w = \sum_{j=1}^{\infty} < w, \phi_j > \phi_j$ which entails that $z = \sum_{j=1}^{\infty} < w, \phi_j > \mu_j \phi_j$. Using that $z \in \mathcal{W}^{\perp}$, we get that $< z, w >= 0$ and so $0 =< z, w >= \sum_{j=1}^{\infty} \mu_j < w, \phi_j >^2$. The fact that μ_j are non-negative, implies that $< w, \phi_j >= 0$ if $\mu_j > 0$ and so $\mathbf{\Gamma}_V w = 0$, which entails that $w \in Ker(\mathbf{\Gamma}_V)$ and $z = 0$ concluding the proof. □

Proof of Theorem 3.2. For the sake of simplicity, we will avoid the index V in $\mathbf{\Gamma}_V$ and will denote $\mathbf{\Gamma}$ the covariance operator.

Let q be the dimension of $\mathcal{W}$ and $\{v_1, \ldots, v_q\}$ an orthonormal basis of $\mathcal{W}$. Let us denote by $\{v_1, \ldots, v_q, v_{q+1}, \ldots,\}$ the extension to an orthonormal basis of $\mathcal{H}$. Define $A_1 : \mathcal{H} \to \mathbb{R}^q$ as $A_1(h) = \sum_{i=1}^{q} < v_i, h > \mathbf{e}_i$, with $\mathbf{e}_i$ the canonical basis of $\mathbb{R}^q$. Then, $A_1^* : \mathbb{R}^q \to \mathcal{H}$, equals $A_1^* \mathbf{x} = \sum_{i=1}^{q} x_i v_i$ with $\mathbf{x} = (x_1, \ldots, x_q)^{\mathrm{T}}$, that is, the image of A_1^* is $\mathcal{W}$. We also define $A_2 : \mathcal{H} \to \mathcal{H}_\infty \subset \mathbb{R}^{\mathbb{N}}$ as

$$A_2(h) = \begin{pmatrix} < v_{q+1}, h > \\ < v_{q+2}, h > \\ \vdots \end{pmatrix}.$$

To ensure the continuity of the second operator, we consider as norm in $\mathcal{H}_\infty$ the norm given by the square root of the sum of squares of the elements of the sequence and as inner product, the one generating this norm. Using Parseval's identity we have that for any $h \in \mathcal{H}$, $\|h\|^2 = \sum_{k=1}^{\infty} <$

$h, v_k >^2$ which implies that A_2 is continuous and with norm equal to 1 since

$$\|A_2(h)\|^2 = \sum_{k=q+1}^{\infty} < h, v_k >^2 \leq \sum_{k=1}^{\infty} < h, v_k >^2 = \|h\|^2 \ ,$$

with equality for any $h \in \mathcal{W}^{\perp}$. Moreover, define $A : \mathcal{H} \to \mathcal{H}_{\infty}$, as $A(h) = \left(A_1(h)^{\mathrm{T}}, A_2(h)^{\mathrm{T}}\right)^{\mathrm{T}}$ and $W = A(V) = \left(A_1(V)^{\mathrm{T}}, A_2(V)^{\mathrm{T}}\right)^{\mathrm{T}} = \left(W_1^{\mathrm{T}}, W_2^{\mathrm{T}}\right)^{\mathrm{T}}$ with W_1 finite–dimensional. Notice that W is an elliptical element in $\mathcal{H}_{\infty}$ with null expectation and covariance operator $\mathbf{\Gamma}_{AV} = A\mathbf{\Gamma}_V A^*$. Thus, if $\mathbf{\Gamma}_{W_1,W_1}$ is non–singular, using Theorem 2.2, we get that

$$E(A_2(V)|A_1(V)) = E(W_2|W_1) = \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1} A_1(V) \tag{A.2}$$

where $\mathbf{\Gamma}_{W_2,W_1} = A_2\mathbf{\Gamma}A_1^*$ and $\mathbf{\Gamma}_{W_1,W_1} = A_1\mathbf{\Gamma}A_1^*$.

Using that $Ker(\mathbf{\Gamma}_V) \cap \mathcal{W} = \{0\}$ and that $\mathbf{\Gamma}_V\mathcal{W} \cap \mathcal{W}^{\perp} = \{0\}$, we will show that $\mathbf{\Gamma}_{W_1,W_1}$ is non–singular. Since $\mathbf{\Gamma}_{W_1,W_1}$ is an endomorphism between finite dimensional vector spaces, we only have to prove that it is a monomorphism and we will automatically have that it is an isomorphism. Let us see the injectivity of this operator. Let us assume that for some $\mathbf{h} \in \mathbb{R}^q$, we have that $A_1\mathbf{\Gamma}A_1^*\mathbf{h} = \mathbf{0}$. We want to show that $\mathbf{h} = \mathbf{0}$. Using that $A_1\mathbf{\Gamma}A_1^*\mathbf{h} = \mathbf{0}$, we get that $\mathbf{\Gamma}A_1^*\mathbf{h} \in Ker(A_1) = \mathcal{W}^{\perp}$. On the other hand, $A_1^*\mathbf{h} \in \mathcal{W}$ and so, $\mathbf{\Gamma}A_1^*\mathbf{h} \in \ \mathbf{\Gamma}(\mathcal{W}) \cap \mathcal{W}^{\perp}$ which implies $\mathbf{\Gamma}A_1^*\mathbf{h} = 0$ by Corollary 3.1. Hence, $A_1^*\mathbf{h} \in Ker(\mathbf{\Gamma}) \cap \mathcal{W} = \{0\}$ by Corollary 3.1 and so, $A_1^*\mathbf{h} = 0$. The fact that A_1^* is injective, lead to $\mathbf{h} = \mathbf{0}$ and therefore, $\mathbf{\Gamma}_{W_1,W_1}$ is non–singular and (A.2) holds.

Define now a random element Y such that $P(Y = y_j) = P(V \in \mathcal{D}_j)$, with $\mathcal{D}_j$ the domain of attraction of y_j, that is, $Y = \sum_{i=1}^{k} y_i \mathbb{I}_{V\in\mathcal{D}_i}$. Using that A_2 is linear and continuous, and that y_j is an element on the self–consistent set, we get that

$$E(A_2V|Y = y_j) = E(A_2V|V \in \mathcal{D}_j) = A_2E(V|V \in \mathcal{D}_j) = A_2y_j = 0$$

where the last equality holds since $y_j \in \mathcal{W}$, that is the kernel of A_2. Therefore, $P(E(A_2V|Y) = 0) = 1$ and so, $0 = E(A_2V|Y) = E(E(A_2V|A_1V)|Y)$ with probability 1. Then, using (A.2), we get

$$\begin{aligned} 0 = E(A_2V|Y) &= E(\mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1}A_1(V)|Y) = \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1}A_1 \ E(V|Y) \\ &= A_2\mathbf{\Gamma}A_1^*(A_1\mathbf{\Gamma}A_1^*)^{-1}A_1 \ Y \quad a.s. \end{aligned}$$

where the last equality follows using Lemma 3.4.

Using that the support of Y spans $\mathcal{W}$, we get that the support of A_1Y spans $\mathbb{R}^q$. Then, using that $(A_1\mathbf{\Gamma}A_1^*)^{-1}$ is non–singular, the fact that $P\left(A_2\mathbf{\Gamma}A_1^*(A_1\mathbf{\Gamma}A_1^*)^{-1}A_1Y = 0\right) = 1$ implies that $A_2\mathbf{\Gamma}A_1^*(A_1\mathbf{\Gamma}A_1^*)^{-1}A_1 \ y_j = 0$, for $1 \leq j \leq k$, and so, $\mathbf{\Gamma}\mathcal{W} \cap \mathcal{W}^{\perp} = \{0\}$. Therefore, $\forall \mathbf{x} \in \mathbb{R}^q$, $A_2\mathbf{\Gamma}A_1^*\mathbf{x} = 0$, or equivalently, $A_2\mathbf{\Gamma}A_1^* : \mathbb{R}^q \to \mathcal{H}_{\infty}$ is the null operator and the same will be true for $A_1\mathbf{\Gamma}A_2^* : \mathcal{H}_{\infty} \to \mathbb{R}^q$.

Define the projection operators $P_{\mathcal{W}} = A_1^*A_1 : \mathcal{H} \to \mathcal{W}$ (the projection over $\mathcal{W}$) and $P_{\mathcal{W}^{\perp}} = A_2^*A_2 : \mathcal{H} \to \mathcal{W}^{\perp}$ (the projection over $\mathcal{W}^{\perp}$). Then, $P_{\mathcal{W}^{\perp}}\mathbf{\Gamma}P_{\mathcal{W}} = 0$ and $P_{\mathcal{W}}\mathbf{\Gamma}P_{\mathcal{W}^{\perp}} = 0$ and so,

$\mathbf{\Gamma}\mathcal{W} \subset \mathcal{W}$ and $\mathbf{\Gamma}\mathcal{W}^{\perp}) \subset \mathcal{W}^{\perp}$ which implies that $\mathcal{W}$ and $\mathcal{W}^{\perp}$ are $\mathbf{\Gamma}-$invariant, i.e., $\mathcal{W}$ decomposes $\mathbf{\Gamma}$.

Then, the restriction of the covariance operator $\mathbf{\Gamma}|_{\mathcal{W}} : \mathcal{W} \to \mathcal{W}$ will be well defined. $\mathbf{\Gamma}$ is compact and self-adjoint and so, diagonalizable. Besides, $\mathbf{\Gamma}$ restricted to $\mathcal{W}$ will also be compact and self-adjoint and thus, it will be diagonalizable with the same eigenfunctions. We will then have that $\mathcal{W}$, the domain of $\mathbf{\Gamma}|_{\mathcal{W}}$ is spanned by a set of eigenfunctions of $\mathbf{\Gamma}$. Using that $\mathcal{W}$ has dimension q, we get that $\mathcal{W}$ is spanned by q eigenfunctions of $\mathbf{\Gamma}$ concluding the proof. □

Proof of Theorem 3.3. As in the proof of Theorem 3.2, we will avoid the index V in $\mathbf{\Gamma}_V$. Let $\lambda_1 \geq \lambda_2 \ldots$ the ordered eigenvalues of $\mathbf{\Gamma}$, with its corresponding eigenfunctions $\phi_1, \phi_2, \ldots$. Let $\{y_1, \ldots, y_k\}$ be a set of k principal points that span $\mathcal{W}$. Theorem 3.2 entails that $\mathcal{W}$ is spanned by q eigenfunctions of $\mathbf{\Gamma}$. Let r be an integer such that $\{\phi_1, \ldots, \phi_r\}$ contains the q eigenfunctions that generate $\mathcal{W}$.

Denote, for each principal point y_j, $a_{ji} =< \phi_i, y_j >$, so that, $a_{ji} = 0$ for $j > r$ and $y_j = \sum_{i=1}^r a_{ji}\phi_i$. Define $\mathbf{a}_j = (a_{j1}, \ldots, a_{jr})^{\mathrm{T}}$ and $\mathbf{X} = (< \phi_1, V >, \ldots, < \phi_r, V >)^{\mathrm{T}}$.

The eigenfunctions will conform an orthonormal basis, then, we have that

$$\begin{aligned} \|V - y_j\|^2 &= \|\sum_{i=1}^r (< \phi_i, V > -a_{ji})\phi_i\|^2 + \|\sum_{i=r+1}^{\infty} < \phi_i, V > \phi_i\|^2 \\ &= \sum_{i=1}^r (< \phi_i, V > -a_{ji})^2 + \sum_{i=r+1}^{\infty} < \phi_i, V >^2 = \|\mathbf{X} - \mathbf{a}_j\|_r^2 + \sum_{i=r+1}^{\infty} < \phi_i, V >^2 , \end{aligned}$$

where $\| \cdot \|_r$ is the euclidean norm in $\mathbb{R}^r$. Note that the k principal points are the k points that minimize $MSE(V, \{\xi_1, \ldots, xi_k\})E(\min_{1\leq j\leq k} \|V - \xi_j\|^2)$ over the sets of k points. Besides,

$$MSE(V, \{y_1, \ldots, y_k\}) = E(\min_{1\leq j\leq k} \|V - y_j\|^2) = E(\min_{1\leq j\leq k} \|\mathbf{X} - \mathbf{a}_j\|_r^2) + \sum_{i=r+1}^{\infty} \lambda_i.$$

and so, $MSE(\mathbf{X}, \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}) = MSE(V, \{y_1, \ldots, y_k\}) - \sum_{i=r+1}^{\infty} \lambda_i$.

Using that $\{y_1, \ldots, y_k\}$ minimize $MSE(V, \cdot)$, it is easy to obtain that $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ minimize $MSE(\mathbf{X}, \{\mathbf{b}_1, \ldots, \mathbf{b}_k\})$, over the sets of k points in $\mathbb{R}^r$ which entails that $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\})$ is a set of k principal points for $\mathbf{X}$. On the other hand, $\mathbf{X}$ has an elliptical distribution since V has an elliptical distribution. Thus, using the result in Tarpey, Li and Flury [20], we obtain that the k principal points of $\mathbf{X}$ lie in the linear space related to the q largest eigenvalues of the covariance, $\mathbf{\Sigma_X}$, of $\mathbf{X}$.

Define $A : \mathcal{H} \to \mathbb{R}^r$ as $A(v) = (< \phi_1, v >, \ldots, < \phi_r, v >)^{\mathrm{T}}$. A is a linear and bounded operator. Denote by $A^* : \mathbb{R}^r \to \mathcal{H}$ the adjoint operator, i.e., $A^*(\mathbf{x}) = \sum_{i=1}^r x_i\phi_i$. Noticing that $\mathbf{X} = AV$ and that $\mathbf{a}_j = Ay_j$, we get easily that $\mathbf{\Sigma_X} = A\mathbf{\Gamma}A^*$. Therefore, the q largest eigenvalues of $\mathbf{\Sigma_X}$ will be equal to the q largest eigenvalues of $\mathbf{\Gamma}$. Moreover, the eigenvectors of $\mathbf{\Sigma_X}$ can be written as $A\phi_i$.

In conclusion, $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \{Ay_1, \ldots, Ay_k\}$ are the k principal points of $\mathbf{X} = AV$ and that the linear space spanned by them is spanned by $A\phi_i$ with $i = 1, \ldots, q$. By restricting A to the space linear space $\mathcal{M}$ spanned by $\{\phi_1, \ldots, \phi_r\}$ we have a surjective isometry and so, if we define

$\tilde{A} : \mathcal{M} \to \mathbb{R}^r$ as the restriction of A to the subspace $\mathcal{M}$, its inverse will be given by $\tilde{A}^* : \mathbb{R}^r \to \mathcal{M}$ which is essentially equal to A^* except for the codomain. The proof follows now easily by noticing that since $y_j \in \mathcal{M}$, $\{Ay_1, \ldots, Ay_k\} = \{\tilde{A}y_1, \ldots, \tilde{A}y_k\} \subset \tilde{\mathcal{M}}$ with $\tilde{\mathcal{M}}$ the linear space spanned by $\{\tilde{A}\phi_1, \ldots, \tilde{A}\phi_q\}$. Hence, spplying $\tilde{A}^{-1} = \tilde{A}^*$, we get that $\{y_1, \ldots, y_k\}$ is included in the linear space spanned by $\{\phi_1, \ldots, \phi_q\}$. □

Proof of Theorem 3.4. Without loss of generality we can assume that V has mean $\mu = 0$. Let $c_1, \ldots, c_k \in \mathcal{H}$ be arbitrary elements and define $b_i = c_i - c_k$, $1 \le i \le k-1$. Denote by $m \le k-1$ the dimension of the linear space $\mathcal{M}_1$ spanned by $b_1, \ldots, b_{k-1}$ and by $\mathcal{M}_2 = \mathcal{M}_1^{\perp}$. Let $a_1, \ldots, a_m$ be an orthonormal basis of $\mathcal{M}_1$ so that, $\mathcal{M}_2$ will be spanned by $a_{m+1}, a_{m+2}, \ldots$, being $\{a_i\}_{i \ge 1}$ an orthonormal basis of $\mathcal{H}$.

As in (2), define $A_1 : \mathcal{H} \to \mathcal{H}_\infty \subset \mathbb{R}^{\mathbb{N}}$ as

$$A_1(a_i) = \begin{cases} (e_i)_{\mathbb{R}^{\mathbb{N}}} & i = 1, 2, \ldots, m \\ 0 & i > m \end{cases}$$

and $A_2 : \mathcal{H} \to \mathcal{H}_\infty \subset \mathbb{R}^{\mathbb{N}}$ as

$$A_2(a_i) = \begin{cases} 0 & i = 1, 2, \ldots, m \\ (e_i)_{\mathbb{R}^{\mathbb{N}}} & i > m \end{cases}$$

Furthermore, let $A : \mathcal{H} \to \mathcal{H}_\infty \subset \mathbb{R}^{\mathbb{N}}$ be $A(h) = A_1(h) + A_2(h)$. Notice that $A(a_i) = e_i$ so that A is a surjective isometry and so, since it is an unitary application, its inverse will be its adjoint, the $A^* : \mathcal{H}_\infty \to \mathcal{H}$ such that $A^*(e_i) = a_i$. Define $d_i = Ac_i = A_1c_i + A_2c_i = d_i^{(1)} + d_i^{(2)}$. Then, using that $c_i - c_j = b_i - b_j \in \mathcal{M}_1$, we get that $d_i^{(2)} - d_j^{(2)} = A_2c_i - A_2c_j = A_2(c_i - c_j) = 0$, that is, all the values $d_i^{(2)}$, $1 \le i \le k$, are equal to a value that we will denote by $d^{(2)}$. Moreover, we have that $d_i = d_i^{(1)} + d^{(2)}$, with both terms orthogonal between themselves.

Define $W_1 = A_1V$, $W_2 = A_2V$, $W = W_1 + W_2 = AV$ and W_1 and W_2 are orthogonal. Using that A is unitary, we get that

$$E(d^2(V, \{c_1, \ldots, c_k\})) = E(d^2(AV, \{Ac_1, \ldots, Ac_k\})) = E(d^2(W, \{d_1, \ldots, d_k\})). \quad \text{(A.3)}$$

Then, $\|W - d_i\|^2 = \|W_1 - d_i^{(1)}\|^2 + \|W_2 - d^{(2)}\|^2$, and so that (A.3) can be written as

$$E(d^2(V, \{c_1, \ldots, c_k\})) = E(d^2(W_1, \{d_1^{(1)}, \ldots, d_k^{(1)}\})) + E(d^2(W_2, \{d^{(2)}, \ldots, d^{(2)}\})).$$

The second term on the right hand side equals $E(d^2(W_2, \{d^{(2)}\}))$ which is minimized when $d^{(2)} = E(W_2) = A_2\mu = 0$. Therefore,

$$E(d^2(V, \{c_1, \ldots, c_k\})) \ge E(d^2(W_1, \{d_1^{(1)}, \ldots, d_k^{(1)}\})) + E(d^2(W_2, \{0, \ldots, 0\})),$$

reaching the equality when $d^{(2)} = 0$. Define $c_i^* = A_1^* d_i^{(1)} = A_1^* A_1 c_i = P_{\mathcal{M}_1} c_i$, $V_1 = A_1^* W_1$ and $V_2 = A_2^* W_2$ ($V = V_1 + V_2$), then

$$\begin{aligned}
E(d^2(V, \{c_1, \ldots, c_k\})) &\geq E(d^2(W_1, \{d_1^{(1)}, \ldots, d_k^{(1)}\})) + E(d^2(W_2, \{0, \ldots, 0\})) \\
&= E(d^2(A_1^* W_1, \{A_1^* d_1^{(1)}, \ldots, A_1^* d_k^{(1)}\})) + E(d^2(A_2^* W_1, \{0, \ldots, 0\})) \\
&= E(d^2(V_1, \{A_1^* d_1^{(1)}, \ldots, A_1^* d_k^{(1)}\})) + E(d^2(V_2, \{0, \ldots, 0\})) \\
&= E(d^2(V_1, \{A_1^* A_1 c_1, \ldots, A_1^* A_1 c_k\})) + E(d^2(V_2, \{0, \ldots, 0\})) \\
&= E(d^2(V, \{A_1^* A_1 c_1 + 0, \ldots, A_1^* A_1 c_k + 0\})) = E(d^2(V, \{c_1^*, \ldots, c_k^*\})) \, .
\end{aligned}$$

where the last equality follows from the orthogonality of the decomposition. Summarizing $E(d^2(V, \{c_1^*, \ldots, c_k^*\})) \leq E(d^2(V, \{c_1, \ldots, c_k\}))$, where the equality holds if $c_i = A_1^* A_1 c_i = P_{\mathcal{M}_1} c_i$. At the principal points we will get the equality since by definition principal points minimize $E(d^2(V, \{c_1, \ldots, c_k\}))$, hence, if c_i correspond to the principal points ξ_i, then $c_i \in \mathcal{M}_1$. Using that $\mathcal{M}_1$ has dimension lower or equal than $k-1$, we obtain the desired result. □

The following result, which we state for completeness, can be found in Flury [2].

Lemma A.2. *Let Y_1 e Y_2 be two real random variables such that Y_2 has the same distribution as ρY_1 for some value of ρ. Then,*

$$D_{Y_1}(k)/Var(Y_1) = D_{Y_2}(k)/Var(Y_2) \, .$$

Proof of Theorem 3.5 Without loss of generality, we sill assume that $\mu = 0$. So as to reduce notation burden, define $D_V(c_1, c_2) = E(d^2(V, \{c_1, c_2\}))$.

We will first show that D_V is minimized if the two elements $c_1, c_2 \in \mathcal{H}$ lie on a straight line with direction $c_2 - c_1$. Theorem 3.4 allows us to do this. Effectively, in the proof of Theorem 3.4, we derived that each principal point (assuming existence) y_i belongs to the linear space $\mathcal{M}_1$ spanned by $y_2 - y_1$. That is, both elements lie in a straight line with direction $a_1 = (y_2 - y_1)/\|y_2 - y_1\|$.

Take $c_1, c_2 \in \mathcal{H}$, $c_1 \neq c_2$ and let $\mathcal{M}_1$ the linear space of dimension 1 spanned by $a_1 = (c_2 - c_1)/\|c_2 - c_1\|$ and $\mathcal{M}_2 = \mathcal{M}_1^{\perp}$ with orthonormal base $\{a_j : j \geq 2\}$. So, using the same notation as in Theorem 3.4, we consider $A_1 : \mathcal{H} \to \mathcal{H}_\infty$ defined as $A_1(a_j) = 0$ if $j \geq 2$ y $A_1(a_1) = e_1$ with e_j the element of $\mathcal{H}_\infty$ with its j coordinate equal to 1 and all the others equal to 0, and $A_2 : \mathcal{H} \to \mathcal{H}_\infty$ defined as $A_2(a_1) = 0$, $A_2(a_i) = e_i$ if $i > 1$. Let $W_1 = A_1 V =< a_1, V > e_1$, $W_2 = A_2 V$, $W = W_1 + W_2$. Using that $\mathcal{M}_1$ is a one dimensional subspace, we get that W_1 has the same distribution as the random variable $Y_1 =< a_1, V >$, which is elliptic and so symmetric around 0, since V is elliptic. Let us remember that if $d_i = A c_i = A_1 c_i + A_2 c_i = d_i^{(1)} + d_i^{(2)}$, then $d_1^{(2)} = d_2^{(2)} = d^{(2)}$ and

$$\begin{aligned}
E(d^2(V, \{c_1, c_2\})) &= E(d^2(W, \{d_1, d_2\})) = E(d^2(W_1, \{d_1^{(1)}, d_2^{(1)}\})) + E(d^2(W_2, \{d^{(2)}, d^{(2)}\})) \\
&\geq E(d^2(W_1, \{d_1^{(1)}, d_2^{(1)}\})) + E(d^2(W_2, \{0, 0\})).
\end{aligned}$$

Then, for fixed c_1 y c_2, $E(d^2(W_1,\{d_1^{(1)},d_2^{(1)}\}))$ can be minimized taking $d_1^{(1)}$ y $d_2^{(1)}$ as the principal points of $Y_1 =< a_1, V >$. Let ξ_1 and ξ_2 be the principal points of Y_1. Define $d_1^* = \xi_1 e_1$, $d_2^* = \xi_2 e_1$. It follows that $E(d^2(W,\{d_1^*,d_2^*\})) \le E(d^2(W,\{d_1,d_2\}))$, with equality if $d_1 = d_1^*$ y $d_2 = d_2^*$.

Using thta $W = AV$, we get $V = A^* W$ and so,

$$c_1^* = A^* d_1^* = A^* \xi_1 e_1 = \xi_1 A_1^* e_1 = \xi_1 a_1 = \xi_1 (c_2 - c_1)/\|c_2 - c_1\|.$$

Analogously, $c_2^* = \xi_2 (c_2 - c_1)/\|c_2 - c_1\|$. Hence,

$$E(d^2(V,\{\xi_1 a_1, \xi_2 a_1\})) = E(d^2(V,\{c_1^*,c_2^*\})) \le E(d^2(V,\{c_1,c_2\}))\ .$$

Given $a \in \mathcal{H}$ such that $\|a\| = 1$, for each pair c_1, c_2 such that $c_2 - c_1$ is proportional to the element a, we will have that $E(d^2(V,\{\xi_1 a, \xi_2 a\})) = E(d^2(V,\{c_1^*,c_2^*\})) \le E(d^2(V,\{c_1,c_2\}))$, therefore it is possible to determine the principal points of V by considering those of $W_1 = A_1 V$, defining $c_1^* = \xi_1 a$ and $c_2^* = \xi_2 a$ and then minimizing over a.

Therefore, it only remains to obtain $a \in \mathcal{H}$. Remember that the operator A_1 depends on that element a, since it is defined using the normalization of $c_2 - c_1$ which is equal to a. To make explicit the dependence, we will denote it as $A_1^{(a)}$, and also $W_1^{(a)} = A_1^{(a)} V =< a, V > e_1 = Y_1^{(a)} e_1$. Note that $\Sigma_{Y_1^{(a)}} = Var(< a, V >) =< a, \mathbf{\Gamma} a >$. Since the principal points will lie in a straight line with normalized direction a, which we are trying to find, they can be written as λa, with $\lambda \in \mathbb{R}$.

Using Lemma A.2, we get that $D_{Y_1^{(a)}}(2) =< a, \mathbf{\Gamma} a > D_{\lambda Y_1^{(a)}}(2)/Var(\lambda Y_1^{(a)})$. On the other hand, we have that

$$D_{\lambda Y_1^{(a)}}(2) = \min_{\eta_1,\eta_2 \in \mathbb{R}} E\left(\min_{i=1,2}\left\{|\lambda Y_1^{(a)} - \eta_1|^2, |\lambda Y_1^{(a)} - \eta_2|^2\right\}\right) \le E(|\lambda Y_1^{(a)}|^2) = Var\left(\lambda Y_1^{(a)}\right)$$

which implies that $D_{\lambda Y_1^{(a)}}(2)/Var(\lambda Y_1^{(a)}) < 1$. Note that by Lemma A.2 we have that

$$\frac{D_{\lambda Y_1^{(a)}}(2)}{Var(\lambda Y_1^{(a)})} = \frac{D_{Y_1^{(a)}}(2)}{Var(Y_1^{(a)})}$$

and so the ratio does not depend on λ. Furthermore, we will show that it does not depend on a.

Using that V is elliptic, we get that for any linear and bounded operator $B : \mathcal{H} \to \mathbb{R}^p$, $\mathbf{Y} = BV$ has an elliptical distribution with parameters $B\mu = 0$ and $\mathbf{\Sigma} = B\mathbf{\Gamma}B^*$. So, its characteristic function can be written as $\varphi_{\mathbf{Y}}(\mathbf{y}) = \phi(\mathbf{y}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{y})$ with ϕ independent of B. In particular, for any $a \in \mathcal{H}$, we have that $\varphi_{Y_1^{(a)}}(b) = \phi(b^2 < a, \mathbf{\Gamma} a >) = \phi(b^2 Var(Y_1^{(a)}))$ which implies that $Z_a = Y_1^{(a)}/\sqrt{Var(Y_1^{(a)})}$ has the same distribution for any element $a \in \mathcal{H}$. Therefore,

$$\frac{D_{Y_1^{(a)}}(2)}{Var(Y_1^{(a)})} = \frac{\min_{\eta_1,\eta_2 \in \mathbb{R}} E(\min\{|Y_1^{(a)} - \eta_1|^2, |Y_1^{(a)} - \eta_2|^2\})}{Var(Y_1^{(a)})}$$

$$
\begin{aligned}
&= \min_{\eta_1,\eta_2 \in \mathbb{R}} E(\min\left\{ \left(\frac{|Y_1^{(a)} - \eta_1|}{\sqrt{Var(Y_1^{(a)})}} \right)^2, \left(\frac{|Y_1^{(a)} - \eta_2|}{\sqrt{Var(Y_1^{(a)})}} \right)^2 \right\}) \\
&= \min_{\eta_1^\star,\eta_2^\star \in \mathbb{R}} E(\min\{|Z_a - \eta_1^\star|^2, |Z_a - \eta_2^\star|^2\})
\end{aligned}
$$

does not depend on a. Hence, we can write $D_{\lambda Y_1^{(a)}}(2)/Var(\lambda Y_1^{(a)}) = g < 1$ with g independent of a and so $D_{Y_1^{(a)}}(2) = g < a, \mathbf{\Gamma} a >$.

Note that $V = Y_1^{(a)}\, a + (V - Y_1^{(a)}\, a) = P_{<a>}V + P_{<a>^\perp}V$. Then, denoting by ξ_i^a the principal points of $Y_1^{(a)}$ and using that that $\|a\| = 1$, we obtain

$$
\|V - \xi_i^{(a)}\, a\|^2 = \|Y_1^{(a)}\, a - \xi_i^{(a)}\, a + (V - Y_1^{(a)}\, a)\|^2 = \|a \cdot (Y_1^{(a)} - \xi_i^a) + P_{<a>^\perp}V\|^2 = (Y_1^{(a)} - \xi_i^a)^2 + \|P_{<a>^\perp}V\|^2.
$$

Taking minimum for $i = 1, 2$ and then applying expectation, we obtain

$$
\begin{aligned}
E(d^2(V, \{\xi_1^{(a)}\, a, \xi_2^{(a)}\, a\})) &= E(\min_{i=1,2} \|V - \xi_i^{(a)}\, a\|^2) = D_{Y_1^{(a)}}(2) + E(\|P_{<a>^\perp}V\|^2) \\
&= g < a, \mathbf{\Gamma} a > + E(\|P_{<a>^\perp}V\|^2).
\end{aligned}
$$

Denote $Z = P_{<a>^\perp}V$ and let ϕ_j be the orthonormal base of $\mathcal{H}$ obtained by the eigenfunctions of $\mathbf{\Gamma}$ related to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$, then,

$$
\begin{aligned}
E(\|P_{<a>^\perp}V\|^2) &= E(\|Z\|^2) = E(\|V\|^2) - E(\|P_{<a>}V\|^2) = Var(V) - Var(Y_1^{(a)}) \\
&= \sum_{n=1}^{\infty} < \mathbf{\Gamma} \phi_n, \phi_n > - Var(Y_1^{(a)}) = \sum_{j \geq 1} \lambda_j - Var(Y_1^{(a)}) \\
&= tr(\mathbf{\Gamma}) - Var(Y_1^{(a)}) = tr(\mathbf{\Gamma}) - < a, \mathbf{\Gamma} a > .
\end{aligned}
$$

Therefore, we obtain

$$
E(d^2(V, \{a \cdot \xi_1^a, a \cdot \xi_2^a\})) = g < a, \mathbf{\Gamma} a > + tr(\mathbf{\Gamma}) - < a, \mathbf{\Gamma} a > = tr(\mathbf{\Gamma}) - (1 - g) < a, \mathbf{\Gamma} a > .
$$

To minimize the left hand side of the above equality it is enough to maximize $< a, \mathbf{\Gamma} a >$ over the elements $a \in \mathcal{H}$ with norm equal to 1. Using the compactness of the covariance operator $\mathbf{\Gamma}$, we obtain the maximum is reached if we choose a as the eigenfunction related to the largest eigenvalue of $\mathbf{\Gamma}$, concluding the proof. □

References

[1] Conway J. B. (1990) *A Course in Functional Analysis*, Springer Graduate Texts in Mathematics

[2] Flury, B.A. (1990). Principal Points, *Biometrika*, **77**, 33-41.

[3] Flury B.A. (1993) Estimation of Principal Points, *Appl. Statist.*, **42**, 139-151.

[4] Flury, B. y Tarpey, T. (1993). Representing a Large Collection of Curves: A Case for Principal Points. *The American Statistician*, **47**, 304-306.

[5] Flury, B. y Tarpey, T. (1998). Principal Points. In: *Encyclopedia of Statistical Science*, Update Volume 2, Kotz, S., Read, C., and Banks, D. Eds. Wiley: New York, pp. 545-548.

[6] Flury, B. y Tarpey, T. (1999). Self–Consistency (Update). In: *Encyclopedia of Statistical Science*, Update Volume 3, Kotz, S., Read, C., and Banks, D. Eds. Wiley: New York, 655-659.

[7] Frahm, G. (2004). *Generalized Elliptical Distributions: Theory and Applications*. Tesis doctoral de la Universidad de Colonia, Alemania.

[8] Kelker, D. (1970). Distribution Theory of Spherical Distributions and a Location–Scale Parameter Generalization. *Sankhya A*, **32**, 419-430.

[9] Muirhead R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley and Sons Canada.

[10] Ramsay, J.O. and Silverman, B.W. (1997) *Functional Data Analysis*, Springer-Verlag.

[11] Seber G.A.F. (1984). *Multivariate Observations*, John Wiley and Sons.

[12] Tarpey, T. (1994). Two Principal Points of Symmetric, Strongly Unimodal Distributions. *Statistics and Probability Letters*, **20**, 253-257.

[13] Tarpey, T. (1995). Principal Points and Self–Consistent Points of Symmetric Multivariate Distributions. *Journal of Multivariate Analysis*, **53**, 39-51.

[14] Tarpey, T. (1997). Estimating Principal Points of Univariate Distributions. *Journal of Applied Statistics*, **24**, 483-496.

[15] Tarpey, T. (1998). Self–Consistent Patterns for Symmetric Multivariate Distributions. *The Journal of Classification*, **15**, 57-79.

[16] Tarpey, T. (1999). Self–Consistency and Principal Component Analysis. *Journal of the American Statistical Association*, **94**, 456-467.

[17] Tarpey, T. (1999). Self–Consistency Algorithms. *Journal of Computational and Graphical Statistics*, **8**, 889-905.

[18] Tarpey T. and Flury B. (1996). Self-Consistency: A Fundamental Concept in Statistics, *Statistical Science*, **11**, 229-243.

[19] Tarpey T. and Kinateder K. (2003). Clustering Functional Data, *Journal of Classification*, **20**, 93-114.

[20] Tarpey T., Li L. and Flury B. (1995) Principal Points and Self-Consistent Points of Elliptical Distributions. *The Annals of Statistics*, **23**, 103-112.

[21] Tarpey, T., Petkova, E., and Ogden, R. T. (2003). Profiling placebo responders by self–consistent partitions of functional data. *Journal of the American Statistical Association*, **98**, 850-858.