A characterization of elliptical distributions and some optimality properties of principal components for functional data

Graciela Boente*[,a], Matías Salibián Barrera[b], David E. Tyler[c]

[a]*Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*
[b]*Department of Statistics, University of British Columbia, Canada*
[c]*Department of Statistics, Rutgers University, USA*

Abstract

As in the multivariate setting, the class of elliptical distributions on separable Hilbert spaces serves as an important vehicle and reference point for the development and evaluation of robust methods in functional data analysis. In this paper, we present a simple characterization of elliptical distributions on separable Hilbert spaces, namely we show that the class of elliptical distributions in the infinite–dimensional case is equivalent to the class of scale mixtures of Gaussian distributions on the space. Using this characterization, we establish a stochastic optimality property for the principal component subspaces associated with an elliptically distributed random element, which holds even when second moments do not exist. In addition, when second moments exist, we establish an optimality property regarding unitarily invariant norms of the residuals covariance operator.

Key words: Elliptical distributions, Functional Data Analysis, Principal Components

---

AMS Subject Classification: MSC 62H25; MSC 60G07

*Corresponding Author
Departamento de Matemáticas, FCEyN, UBA
Ciudad Universitaria, Pabellón 2
Buenos Aires, C1428EHA, Argentina
FAX 54-11-45763375
Email addresses: gboente@dm.uba.ar (Graciela Boente), matias@stat.ubc.ca (Matías Salibián Barrera), dtyler@rci.rutgers.edu (David E. Tyler)

Preprint submitted to Elsevier *May 4, 2014*

1. Introduction

When considering finite–dimensional random vectors, a natural and commonly used generalization of the family of multivariate normal distributions is given by the class of elliptical distributions. This class allows for heavy tail models while preserving many of the attractive properties of the multivariate normal model, such as regression being linear, as discussed for example in Muirhead (1982), Seber (1984), Bilodeau and Brenner (1999) and Frahm (2004). Multivariate elliptical distributions include the $t-$distributions, the symmetric generalized hyperbolic distribution, the multivariate Box-Tiao or power exponential family distributions and the sub-Gaussian $\alpha-$stable distributions, among others. From a practical point of view, Frahm (2004) has argued that the class of heavy tailed elliptical distributions offers a good alternative for modeling financial data in which the Gaussian assumption may not be reliable. Multivariate elliptical models have also been considered extensively within the area of robust statistics as a starting point for the development of the M-estimates of multivariate location and scatter, see e.g. Maronna (1976), and also as a class of models under which the asymptotic behavior and the influence functions of robust multivariate methods, such as robust principal components, can be evaluated and judged. (See, for instance, Hampel *et al.* (1986), Huber and Ronchetti, 2009 and Maronna *et al.*, 2006.)

In many areas of statistics, the collected data are more naturally represented as functions rather than finite-dimensional numerical vectors, as argued e.g. in Ramsay and Silverman (2005). Simplifying the functional model by discretizing the observations as sequences of numbers can often fail to capture some of its important characteristics, such as the smoothness and continuity of the underlying functions. For this reason, in the last decades different methods have been proposed to handle this type of "functional" data, which can be viewed as instances of random elements taking values in a space of functions such as $L^2(\mathcal{I})$, with $\mathcal{I} \subset \mathbb{R}$ a finite interval. A more general and inclusive framework is to view the observations as elements in a separable Hilbert space $\mathcal{H}$, which is not necessarily finite–dimensional.

The notion of principal components analysis, which is a fundamental concept in multivariate statistics, has been extended to the functional data setting and is commonly referred to as FPCA or functional principal components analysis. The first few principal components are typically used to explore the main characteristics of the data within a reduced dimensional space. In particular, exploring this lower dimensional principal components space can be useful for detecting atypical observations or outliers in the data set. The principal components subspace has the well known property that the first q principal components associated with the distribution of a random element with finite second moment provide the best q-dimensional linear approximation to the random element in terms of mean squared error. These linear approximations also minimize unitarily invariant norms of the covariance matrix of the residuals, in the finite–dimensional setting.

As in the multivariate setting, the class of elliptical distributions on separable Hilbert spaces can serve as an important vehicle and reference point for the development and evaluation of robust methods in functional data analysis. In addition, they allow for the development of FPCA even if the random elements do not possess second moments. An extension of the class of elliptical distributions to separable Hilbert spaces is given in the relatively recent paper by Bali and Boente (2009), while the Fisher–consistency of some robust estimates of principal directions for this class of elliptical distributions is established in Bali, *et al.* (2011). The main purpose of the present short paper is two-fold. First, in section 2, we present a simple characterization of elliptical distributions on separable Hilbert spaces, namely we show that the class of elliptical distributions in the infinite–dimensional case is equivalent to the class of scale mixtures of Gaussian distributions on the space, unless the random element is essentially finite–dimensional. Second, we then use this representation in section 3.1 to establish a stochastic best lower-dimensional approximation for elliptically distributed random elements and an optimality property for the scatter operator of the associated residuals. That is, we derive two optimality properties for the eigenfunctions associated with the largest eigenvalues of the scatter operator that hold even when second moments do not exist and which recover the same best lower dimensional approximation properties mentioned above when second moments do exist.

In section 3.2 we extend another known optimality property of principal components from Euclidean spaces to general Hilbert spaces. This optimality property holds not only for elliptical distributions, but for any distribution with finite second moments. As in the finite-dimensional case, when second moments exist a measure of closeness between a random element X and a predictor is the norm of the residuals covariance operator. Although many operator norms can be defined, in the principal components setting a reasonable requirement is that the operator norm be unitarily invariant. Under this assumption, we show that the optimal linear predictors are those obtained through the linear space spanned by the first q principal components. In Section 4 we use the theory developed in this paper to show that the spherical principal components of Locantore *et al.* (1999) and Gervini (2008) are Fisher consistent for elliptically distributed random elements. This result extends previous results obtained for random elements with a finite Karhunen-Loève expansion. Some mathematical concepts and proofs are presented in the Appendix.

2. Elliptical distributions over Hilbert spaces

There are a number of ways to define the class of elliptical distributions in the multivariate setting. An attractive constructive definition is to define them as the class of distributions generated by applying affine transformations to the class of spherical distributions. The properties of elliptical distributions then follow readily from the simpler class of spherical distributions.

Recall that a random vector $\mathbf{Z} \in \mathbb{R}^d$ is said to have a d–dimensional spherical distribution if its distribution is invariant under orthogonal transformations, i.e., if $\mathbf{QZ} \sim \mathbf{Z}$ for any $d \times d$ orthogonal matrix $\mathbf{Q}$. The classic example of a spherical distribution is the multivariate standard normal distribution. In general, if $\mathbf{Z}$ has a spherical distribution in $\mathbb{R}^d$ then $R = \|\mathbf{Z}\|_d$ and $\mathbf{U} = \mathbf{Z}/\|\mathbf{Z}\|_d$ are independent with $\mathbf{U}$ having a uniform distribution on the d–dimensional unit sphere. Here $\|\cdot\|_d$ refers to the Euclidean norm in $\mathbb{R}^d$. If $\mathbf{Z}$ is also absolutely continuous in $\mathbb{R}^d$, then it has a density of the form $f(\mathbf{z}) = g_d(\mathbf{z}^{\mathrm{T}}\mathbf{z})$ for some function $g_d(s) \geq 0$, i.e., it has spherical contours. The marginal density of a subset of the components of $\mathbf{Z}$ also has spherical contours, with the relationship between g_d and the $k-$dimensional density generator g_k, for $k < d$ being somewhat complicated. It turns out to be more convenient to denote a spherical distribution by its characteristic function. In general, the characteristic function of a spherically distributed $\mathbf{Z} \in \mathbb{R}^d$ is of the form $\psi_{\mathbf{Z}}(\mathbf{t}_d) = \varphi(\mathbf{t}_d^{\mathrm{T}}\mathbf{t}_d)$ for $\mathbf{t}_d \in \mathbb{R}^d$, and any distribution in $\mathbb{R}^d$ having a characteristic function of this form is a spherical distribution. Consequently, we express $\mathbf{Z} \sim \mathcal{S}_d(\varphi)$. This notation is convenient since, for $\mathbf{Z}^{\mathrm{T}} = (\mathbf{Z}_1^{\mathrm{T}}, \mathbf{Z}_2^{\mathrm{T}})$ with $\mathbf{Z}_1 \in \mathbb{R}^k$, the marginal $\mathbf{Z}_1$ is such that $\mathbf{Z}_1 \sim \mathcal{S}_k(\varphi)$. More generally, for any $k \times d$ matrix $\mathbf{Q}_k$ such that $\mathbf{Q}_k^{\mathrm{T}}\mathbf{Q}_k = I$, we have $\mathbf{Q}_k\mathbf{Z} \sim \mathcal{S}_k(\varphi)$. Note that if $\varphi(\mathbf{t}_d^{\mathrm{T}}\mathbf{t}_d)$ is a valid characteristic function in $\mathbb{R}^d$ then $\varphi(\mathbf{t}_k^{\mathrm{T}}\mathbf{t}_k)$, where $\mathbf{t}_k = (t_1, \ldots, t_k)$, is also a valid characteristic function in $\mathbb{R}^k$ for any $k < d$. For some families of spherical distributions defined across different dimensions, such as the multivariate power exponential family considered by Kuwana and Kariya (1991), the function φ may depend upon the dimension d. In such cases, the marginal distributions are not elements of the same family.

As already noted, the elliptical distributions in $\mathbb{R}^d$ correspond to those distributions arising from affine transformations of spherically distributed random vectors in $\mathbb{R}^d$. For a $d \times d$ matrix $\mathbf{B}$ and a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, the distribution of $\mathbf{X} = \mathbf{BZ} + \boldsymbol{\mu}$ when $\mathbf{Z} \sim \mathcal{S}_d(\varphi)$ is said to have an elliptical distribution, denoted by $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$, where $\boldsymbol{\Sigma} = \mathbf{BB}^{\mathrm{T}}$. Note that the distribution of $\mathbf{X}$ depends on $\mathbf{B}$ only through $\boldsymbol{\Sigma} = \mathbf{BB}^{\mathrm{T}}$. For a fixed φ, the family of elliptical distributions $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ forms a symmetric location–scatter class of distributions with location parameter $\boldsymbol{\mu}$ and symmetric positive semi–definite scatter parameter $\boldsymbol{\Sigma}$. If the first moment exists, then $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$, and if second moments exist then the variance covariance matrix of $\mathbf{X}$ is proportional to $\boldsymbol{\Sigma}$. If $\mathbf{Z}$ has the spherical density noted above and $\boldsymbol{\Sigma}$ is nonsingular then the density of $\mathbf{X}$ has elliptical contours and is given by $f(\mathbf{x}) = (\det \boldsymbol{\Sigma})^{-1/2} g_d\left((\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$ for $\mathbf{x} \in \mathbb{R}^d$. The characteristic function of $\mathbf{X}$ in general has the simple form $\psi_{\mathbf{X}}(\mathbf{t}) = \exp(i\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu})\varphi(\mathbf{t}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{t})$. Furthermore, the elliptical distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ can be characterized by its marginals, namely for any fixed $k < d$, $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ if and only if $\mathbf{AX} \sim \mathcal{E}_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}, \varphi)$ for all $k \times d$ matrices $\mathbf{A}$.

It is not possible to extend the definition of elliptical distributions from finite dimensional to infinite dimensional Hilbert spaces using a construction based on spherical distributions since such a spherical distribution cannot be defined in the latter case. Consequently, Bali and Boente (2009) proposed the following definition based on generalizing the characterizing relationship between an elliptical distribution and the distributions of its lower dimensional projections. Hereafter, $\mathcal{H}$ will denote a separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and corresponding induced norm $\| \cdot \|$.

Definition 2.1. *Let X be a random element in a separable Hilbert space $\mathcal{H}$. We will say that X has an elliptical distribution with parameters $\mu \in \mathcal{H}$ and $\boldsymbol{\Gamma} : \mathcal{H} \to \mathcal{H}$, where $\boldsymbol{\Gamma}$ is a self–adjoint, positive semi–definite and compact operator, if and only if, for any $d \geq 1$ and for any linear and bounded operator $A : \mathcal{H} \to \mathbb{R}^d$ we have that $AX \sim \mathcal{E}_d(A\mu,\, A\boldsymbol{\Gamma}A^*,\, \varphi)$ where $A^* : \mathbb{R}^d \to \mathcal{H}$ denotes the adjoint operator of A. We will write $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$.*

Since any multivariate distribution is determined by the distribution of its one dimensional projections, the above definition is equivalent to the case when one only considers $d = 1$. Consequently an equivalent definition for a random element in $\mathcal{H}$ to have the elliptical distribution $\mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$ can be stated as follows.

Definition 2.2. *For $X \in \mathcal{H}$, $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$ with parameters $\mu \in \mathcal{H}$ and $\boldsymbol{\Gamma} : \mathcal{H} \to \mathcal{H}$, where $\boldsymbol{\Gamma}$ is a self–adjoint, positive semi–definite and compact operator, if and only if $\langle a, X\rangle \sim \mathcal{E}_1(\langle a, \mu\rangle, \langle a, \boldsymbol{\Gamma}a\rangle, \varphi)$ for all $a \in \mathcal{H}$.*

As in the finite–dimensional case, if the covariance operator of a random element X, $\boldsymbol{\Gamma}_X$, exists then $\boldsymbol{\Gamma}_X = \gamma\boldsymbol{\Gamma}$ for some $\gamma > 0$. Also, as in the finite–dimensional setting, the scatter parameter $\boldsymbol{\Gamma}$ is confounded with the function φ in that for any $c > 0$, $\mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi) \sim \mathcal{E}(\mu, c\boldsymbol{\Gamma}, \varphi_c)$ where $\varphi_c(w) = \varphi(w/c)$. For a given φ, though, $\boldsymbol{\Gamma}$ is well defined.

It should be noted that if the range of $\boldsymbol{\Gamma}$ is q–dimensional, then it is only necessary that the function φ be such that $\psi_q(\mathbf{t}) = \varphi(\mathbf{t}^{\mathrm{T}}\mathbf{t})$ is a valid characteristic function over $\mathbf{t} \in \mathbb{R}^q$ in order for $\mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$ to be well defined. On the other hand, if the range of $\boldsymbol{\Gamma}$ is infinite–dimensional, then the function $\psi_d(\mathbf{t}) = \varphi(\mathbf{t}^{\mathrm{T}}\mathbf{t})$ must be a valid characteristic function over $\mathbf{t} \in \mathbb{R}^d$ for all finite d. Furthermore, when this latter condition holds, the family of distributions $\mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$ is defined over all $\mu \in \mathcal{H}$ and over all self–adjoint, positive semi–definite and compact operators $\boldsymbol{\Gamma} : \mathcal{H} \to \mathcal{H}$.

As defined, the class of infinite–dimensional elliptical distributions is not empty. It includes the family of Gaussian distributions on $\mathcal{H}$, which corresponds to choosing $\varphi(w) = \exp\{-w/2\}$. Other elliptical distributions can be obtained using the following construction. Let V be a Gaussian random element in $\mathcal{H}$ with zero mean and covariance operator $\boldsymbol{\Gamma}_V$, and let S be a non–negative real random variable independent of V. Given $\mu \in \mathcal{H}$, define $X = \mu + S\,V$, which is a scale mixture of the Gaussian element V. The resulting random element X then has the elliptical distribution $\mathcal{E}(\mu, \boldsymbol{\Gamma}, \varphi)$ with the operator $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_V$ and the function $\varphi(w) = \mathbb{E}(\exp\{-wS/2\})$. Note that $\boldsymbol{\Gamma}$ exists even when the second moment of X may not. Moreover, when $\mathbb{E}(S^2) < \infty$, the covariance operator of X, $\mathbf{\Gamma}_X = \mathbb{E}(S^2)\mathbf{\Gamma}_V$ (see Bali and Boente, 2009).

In the univariate, as well as in the multivariate setting, a scale mixture of a Gaussian distributions has a heavier tail than a Gaussian distribution. Hence, they are a often viewed as attractive longer tail alternatives to the Gaussian model. In the univariate case, the scale mixture distribution either does not have a fourth moment or, if it does, then its kurtosis, $\kappa = \mathbb{E}((X-\mu)^4)/(\mathbb{E}(X-\mu)^2)^2$, is greater than that of the Gaussian, which corresponds to $\kappa = 3$. For multivariate elliptical distributions in general, all univariate projections have distributions which lie in the same location-scale family, so the distribution of any univariate projection has the same kurtosis. It has been known for some time that, for higher dimensional elliptical distributions, this kurtosis parameter cannot be much smaller than that of a Gaussian vector. In particular, for a d–dimensional elliptical distribution, $\kappa \geq 3d/(d+2)$, see Tyler (1982), Bentler and Berkane (1986) or Anderson (2003). If this holds for all finite d, then $\kappa \geq 0$. Consequently, this implies that the distribution $\mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$, when the rank of $\mathbf{\Gamma}$ is not finite, cannot have much shorter tails than a Gaussian distribution, i.e., the kurtosis of $\mathcal{E}_1(\langle a, \mu\rangle, \langle a, \mathbf{\Gamma} a\rangle, \varphi)$ must be non–negative. Proposition 2.1 below gives an even stronger result, namely that such a distribution must be a scale mixture of Gaussians. The proof of the proposition is based on an application of the following lemma given in Kingman (1972).

Lemma 2.1. *Let $Y_1, Y_2, \ldots$ be an infinite sequence of real random variables with the property that, for any n, the distribution of the random vector $(Y_1, \ldots, Y_n)^{\mathrm{T}}$ has spherical symmetry. Then, there exists a random variable W, real and nonnegative, such that, conditional on W, $Y_1, Y_2, \ldots$ are mutually independent and normally distributed with mean zero and variance W.*

Proposition 2.1. *Let $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ be an elliptical element of a separable Hilbert space $\mathcal{H}$. If $\mathbf{\Gamma}$ does not have finite rank, then there exists a zero mean Gaussian element V and a random variable $S \geq 0$ independent of V, such that $X \sim \mu + S\,V$.*

PROOF. Without loss of generality assume $\mu = 0$. For $k \in \mathbb{N}$, let $\lambda_{i_k} > 0$ be the non-null eigenvalues of $\mathbf{\Gamma}$, with $\lambda_{i_1} \geq \lambda_{i_2} \geq \ldots$, and let ϕ_{i_k} denote an eigenfunction of $\mathbf{\Gamma}$ associated with λ_{i_k} chosen so that the set $\{\phi_{i_k}, k \in \mathbb{N}\}$ forms an orthonormal set in $\mathcal{H}$. We then have the representation $X = \sum_{k\geq 1} \langle \phi_{i_k}, X\rangle\, \phi_{i_k}$. For $k \geq 1$, let $Y_k = \lambda_{i_k}^{-1/2} \langle \phi_{i_k}, X\rangle$. Also, for each $n \in \mathbb{N}$ let $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$, and define the bounded linear operator $A_n : \mathcal{H} \to \mathbb{R}^n$ by $A_n x = (\lambda_{i_1}^{-1/2} \langle x, \phi_{i_1}\rangle, \ldots, \lambda_{i_n}^{-1/2} \langle x, \phi_{i_n}\rangle)^{\mathrm{T}}$. It then follows that $\mathbf{Y}_n = A_n\, X \sim \mathcal{E}_n\left(\mathbf{0}, A_n\, \mathbf{\Gamma}\, A_n^*, \varphi\right)$. Furthermore, $A_n \mathbf{\Gamma} A_n^* = \mathbf{I}_n$, so $\mathbf{Y}_n \sim \mathcal{S}_n(\varphi)$ for all $n \geq 1$. Thus, by Lemma 2.1, there then exists a non–negative random variable W such that, for all $n \in \mathbb{N}$, $\mathbf{Y}_n | W \sim \mathcal{N}_n(0, W\, \mathbf{I}_n)$. Let $S = W^{1/2}$, then $\mathbf{Y}_n$ has the same distribution as $S\, \mathbf{Z}_n$, where $\mathbf{Z}_n \sim \mathcal{N}_n(0, \mathbf{I}_n)$ and $\mathbf{Z}_n$ is independent of S. Hence, for all $k \geq 1$, $Y_k \sim S\, Z_k$, $k \geq 1$, with Z_k being i.i.d. standard normal random variables, independent of S. Consequently $X = \sum_{k\geq 1} \lambda_{i_k}^{1/2}\, Y_k\, \phi_{i_k} \sim SV$, with $V = \sum_{k\geq 1} \lambda_{i_k}^{1/2}\, Z_k\, \phi_{i_k}$ being a Gaussian element in $\mathcal{H}$. □

Remark 2.1. The above proof depends heavily on Kingman's result, which in turn depends heavily on arguments involving characteristic functions. An alternative simple self–contained proof for Proposition 2.1 can be given as follows. First note that the distribution of $X \sim \mathcal{E}(0, \mathbf{\Gamma}, \varphi)$ is determined by the distribution of any one–dimensional projection with a non–degenerate distribution, say $X_1 = \langle \phi_{i_1}, X\rangle$. Without loss of generality, assume X_1 is symmetrically distributed around zero. For each n, let $\mathbf{Y}_n$ be defined as above so that $Y_1 = X_1 \lambda_{i_1}^{-1/2}$. Since $\mathbf{Y}_n \sim \mathcal{S}_n(\varphi)$, its distribution can be represented as $\mathbf{Y}_n \sim R_n \mathbf{U}$, where $\mathbf{U}$ has a uniform distribution on the $n-$dimensional unit sphere, and R_n is a non–negative random variable independent from $\mathbf{U}$. The distribution of $\mathbf{U}$ itself can then be represented as $\mathbf{U} \sim \mathbf{Z}/\|\mathbf{Z}\|_n$ where $\mathbf{Z} \sim \mathcal{N}_n(0, \mathbf{I}_n)$ and independent of R_n. It then follows that $X_1 \sim \lambda_{i_1}^{1/2} D_n S_n Z_1$, where $D_n = \sqrt{n}/\|\mathbf{Z}\|_n$, $S_n = R_n/\sqrt{n}$ and $Z_1 \sim \mathcal{N}(0,1)$ independent of S_n. Now, by the weak law of large numbers, $D_n \xrightarrow{p} 1$. Since the distributions of X_1 and Z_1 do not depend on n, S_n must converge in distribution to a random variable S, with S being independent of Z_1. Thus, $X_1 = \lambda_{i_1}^{1/2} S Z_1$, which is a scale mixture of normals.

Remark 2.2. Proposition 2.1 only applies to the family $\mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ for which $\psi_d(\mathbf{t}) = \varphi(\mathbf{t}^{\mathrm{T}}\mathbf{t})$ is a valid characteristic function in $\mathbb{R}^d$ for all finite d. It is worth noting though that elliptical distributions on $\mathcal{H}$, as given in Definition 2.1 or 2.2, are also defined for the case in which $\psi_d(\mathbf{t})$ is a valid characteristic function in $\mathbb{R}^d$ for say $d = q$ but not for $d = q + 1$ for some finite q. In this case the family of distributions given by $\mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ is defined over all $\mu \in \mathcal{H}$ but only over those self–adjoint, positive semi–definite and compact operators $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$ with ranks no greater than q. In general, when the rank of $\mathbf{\Gamma}$ equals q, $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ admits a finite Karhunen-Loève expansion, namely $X = \mu + \sum_{j=1}^{q} \lambda_j^{1/2} Z_j \phi_j$ with $\mathbf{Z} = (Z_1, \ldots, Z_q)^{\mathrm{T}} \sim \mathcal{S}_q(\varphi)$ and where $\lambda_1 \geq \ldots \geq \lambda_q > 0$ and $\{\phi_j, j = 1, \ldots q\}$ are respectively the non–null eigenvalues and the associated orthonormal set of eigenfunctions of $\mathbf{\Gamma}$.

Note that Proposition 2.1 is not true if $\mathbf{\Gamma}$ has finite rank. In particular, it does not hold in the finite–dimensional setting, where examples of elliptical distributions that are not scale mixture of normals can easily be constructed. For example, a random vector $\mathbf{U}$ uniformly distributed on the d-dimensional unit sphere has an elliptical distribution but does not have a density function. However, a random vector $\mathbf{X} \in \mathbb{R}^d$ of the form $\mathbf{X} = S\mathbf{V}$ with $\mathbf{V} \sim N(\mathbf{0}, \mathbf{I}_q)$ and S independent of $\mathbf{V}$ is absolutely continuous. For more details, see for instance, Muirhead (1982) or Bilodeau and Brenner (1999).

3. Some optimality properties of principal components

3.1. Stochastic optimality of principal components for elliptical distributions

Principal components of finite–dimensional vectors were introduced by Hotelling (1933) as a tool to display data on a lower dimensional linear space that best preserves the variability of the data. Since then, principal components have been characterized by several other properties related to their optimality with respect to the expected value of some approximation or prediction error, see, for example, Rao (1964), Darroch (1965) and Okamoto and Kanazawa (1968). Furthermore, Proposition 1 in Boente (1983) shows that, given any proper linear space $\mathcal{L}$ of dimension q, the norm of the residuals of the orthogonal projection onto $\mathcal{L}$ of an elliptically distributed random vector is stochastically larger than that obtained when $\mathcal{L}$ is spanned by the first q principal components. In particular, when second moments exist, this result implies the well–known mean squared error property of the principal components. In this Section, we extend these results to elliptical random elements on separable Hilbert spaces.

In what follows $\pi(x, \mathcal{L})$ will denote the orthogonal projection of $x \in \mathcal{H}$ onto the linear closed space $\mathcal{L}$. Note that if $\mathcal{L}$ is a finite–dimensional linear space then it is closed. Let $X \in \mathcal{H}$ be an elliptically distributed random element $X \sim \mathcal{E}(0, \mathbf{\Gamma}, \varphi)$ as in Definition 2.1 or 2.2. Our first result shows that, for any fixed $q \geq 1$, the linear space $\mathcal{L}$ of dimension q that produces orthogonal projections $\pi(X, \mathcal{L})$ with (stochastically) largest norms is the one spanned by the first q eigenfunctions of $\mathbf{\Gamma}$. The proof relies on the well-known characterization of elliptical distributions $\mathcal{E}_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \varphi)$ on $\mathbb{R}^d$ as affine transformations of a spherically distributed d-dimensional random vector $\mathbf{Z} \sim \mathcal{S}_d(\varphi)$.

Proposition 3.1. *Let $X \in \mathcal{H}$ be an elliptically distributed random element $X \sim \mathcal{E}(0, \mathbf{\Gamma}, \varphi)$ with $\mathbf{\Gamma}$ a self–adjoint, positive semi–definite and compact operator. Denote by $\lambda_1 \geq \lambda_2 \geq \dots$ the eigenvalues of $\mathbf{\Gamma}$ with associated eigenfunctions ϕ_j, $j \geq 1$. Then, if $\lambda_q > \lambda_{q+1}$, for any linear space $\mathcal{L}$ of dimension $m \leq q$, $\|\pi(X, \mathcal{L})\| \leq_s \|\pi(X, \mathcal{L}_q)\|$, where $\mathcal{L}_q$ is the linear space spanned by $\phi_1 \dots \phi_q$ and $U \leq_s V$ means that the random variable U is stochastically smaller than V.*

Proof. First note that it is enough to show the result when $m = q$. To see this, let $\mathcal{L}$ be a linear space of dimension $m < q$ with orthonormal basis $\gamma_1, \dots, \gamma_m$ and find $\gamma_{m+1}, \dots, \gamma_q$ such that $\{\gamma_1, \dots, \gamma_q\}$ is an orthonormal set. If $\widetilde{\mathcal{L}}$ denotes the linear space spanned by $\gamma_1, \dots, \gamma_q$, then $\|\pi(X, \mathcal{L})\|^2 = \sum_{i=1}^m \langle X, \gamma_i \rangle^2 \leq_s \sum_{i=1}^q \langle X, \gamma_i \rangle^2 = \|\pi(X, \widetilde{\mathcal{L}})\|^2$. Thus, it is enough to show the result for spaces $\mathcal{L}$ of dimension q.

Let $B : \mathcal{H} \to \mathbb{R}^q$ and $B_0 : \mathcal{H} \to \mathbb{R}^q$ be linear and bounded operators defined by $BX = (\langle X, \gamma_1 \rangle, \dots, \langle X, \gamma_q \rangle)^{\mathrm{T}}$ and $B_0 X = (\langle X, \phi_1 \rangle, \dots, \langle X, \phi_q \rangle)^{\mathrm{T}}$ respectively. Define $\mathbf{Y} = BX$ and $\mathbf{Y}_0 = B_0 X$ and note that $\mathbf{Y} \sim \mathcal{E}_q(0, \mathbf{\Sigma}, \varphi)$ and $\mathbf{Y}_0 \sim \mathcal{E}_q(0, \mathbf{\Sigma}_0, \varphi)$, where $\mathbf{\Sigma} = B\mathbf{\Gamma}B^*$ and $\mathbf{\Sigma}_0 = \operatorname{diag}(\lambda_1, \dots, \lambda_q) = \mathbf{\Lambda}$. Furthermore we have that $\|\pi(X, \mathcal{L})\|^2 = \|\mathbf{Y}\|^2$ and $\|\pi(X, \mathcal{L}_q)\|^2 = \|\mathbf{Y}_0\|^2$. Write the spectral decomposition of $\mathbf{\Sigma} = \boldsymbol{\beta}\widetilde{\mathbf{\Lambda}}\boldsymbol{\beta}^{\mathrm{T}}$ where $\widetilde{\mathbf{\Lambda}} = \operatorname{diag}(\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_q)$ is the diagonal matrix of eigenvalues of $\mathbf{\Sigma}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q \times q}$ is orthonormal. Because $\mathbf{Y} \sim \mathcal{E}_q(0, \mathbf{\Sigma}, \varphi)$ and $\mathbf{Y}_0 \sim \mathcal{E}_q(0, \mathbf{\Sigma}_0, \varphi)$ we can find a random vector $\mathbf{Z} \sim \mathcal{S}_q(\varphi)$ such that $\mathbf{Y} \overset{\mathcal{D}}{\sim} \boldsymbol{\beta}\widetilde{\mathbf{\Lambda}}^{1/2}\mathbf{Z}$ and $\mathbf{Y}_0 \overset{\mathcal{D}}{\sim} \mathbf{\Lambda}^{1/2}\mathbf{Z}$, where $\mathbf{A} \overset{\mathcal{D}}{\sim} \mathbf{B}$ means that the vectors $\mathbf{A}$ and $\mathbf{B}$ have the same distribution. Then $\|\mathbf{Y}\|^2 \overset{\mathcal{D}}{\sim} \|\widetilde{\mathbf{\Lambda}}^{1/2}\mathbf{Z}\|^2$, $\|\mathbf{Y}_0\|^2 \overset{\mathcal{D}}{\sim} \|\mathbf{\Lambda}^{1/2}\mathbf{Z}\|^2$,

$$\mathbb{P}\left(\|\mathbf{Y}\|^2 \leq y\right) = \mathbb{P}\left(\sum_{i=1}^q \widetilde{\lambda}_i\, Z_i^2 \leq y\right) \text{ and } \mathbb{P}\left(\|\mathbf{Y}_0\|^2 \leq y\right) = \mathbb{P}\left(\sum_{i=1}^q \lambda_i\, Z_i^2 \leq y\right).$$

Now note that $BB^* = \mathbf{I}_q$, where $\mathbf{I}_q$ stands for the $q \times q$ identity matrix, i.e., B is a sub–unitary operator. Hence, using Proposition A.1 in the Appendix we have that

$$\widetilde{\lambda}_i = \lambda_i(B\mathbf{\Gamma}B^*) \leq \lambda_i(\mathbf{\Gamma}) = \lambda_i\ , \tag{1}$$

where $\lambda_i(\mathbf{\Upsilon})$ denotes the $i-$th largest eigenvalue of the operator $\mathbf{\Upsilon}$. It then follows that $\|\mathbf{Y}\|^2 \leq_s \|\mathbf{Y}_0\|^2$ which concludes the proof. □

It is easy to see that when the random element X has an elliptical distribution $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ and $\mathbf{\Gamma}$ has finite rank then the linear space spanned by the first q eigenfunctions of $\mathbf{\Gamma}$ provides the best q-dimensional approximation to $X - \mu$, in the sense of having stochastically smallest residual squared norms among all linear spaces of dimension q. More specifically, if $\mathbf{\Gamma}$ has rank p, then with probability one $X - \mu = \sum_{j=1}^p \lambda_j^{1/2}\, Z_j \phi_j$, where λ_j and ϕ_j denote the eigenvalues and eigenfunctions of $\mathbf{\Gamma}$, respectively, and $\mathbf{Z} = (Z_1, \dots, Z_p)^{\mathrm{T}} \sim \mathcal{S}_p(\varphi)$. The result now follows from the finite-dimensional result in Boente (1983) by identifying $X - \mu$ with the random vector $(\lambda_1^{1/2} Z_1, \dots, \lambda_p^{1/2} Z_p)^{\mathrm{T}}$. The arguments in the proof of Proposition 3.1 can be used to show that this also holds for linear spaces $\mathcal{L}$ of dimension $m \leq q$.

Together with Proposition 2.1, the following result shows that this property also holds when X is a random element with elliptical distribution $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ and $\mathbf{\Gamma}$ does not have finite rank. Note that this result also holds if the second moment of X is not finite, since $\mathbf{\Gamma}$ plays the role of a scatter operator. It is worth noting that the stochastic optimality provided in Proposition 3.2 is novel, even in the case of Gaussian processes. When second moments exist, Proposition 3.2 entails the well known "best q-term approximation" property mentioned, for instance, in Ramsay and Silverman (2005), which states that the best q-dimensional approximation is optimal in the sense of minimizing the mean squared error between $X-\mu$ and the resulting linear approximations $\sum_{j=1}^{q}\langle X-\mu,\psi_j\rangle\psi_j$, where $\langle \psi_j,\psi_\ell\rangle=0$ for $j\neq \ell$ and 1 otherwise.

Proposition 3.2. *Let $X\in\mathcal{H}$ be a random element such that $X=SV$, where $S\in\mathbb{R}$ is a non–negative real random variable independent of V and V is a zero mean Gaussian random element with covariance operator $\mathbf{\Gamma}$. Denote the eigenvalues and eigenfunctions of $\mathbf{\Gamma}$ by $\lambda_1\geq\lambda_2\geq\dots$ and ϕ_j, $j\geq 1$, respectively. If $\lambda_q>\lambda_{q+1}$, then for any linear space $\mathcal{L}$ of dimension q we have that $\|X-\pi(X,\mathcal{L}_q)\|\leq_s\|X-\pi(X,\mathcal{L})\|$, where $\mathcal{L}_q$ denotes the linear space spanned by $\phi_1\dots\phi_q$.*

Proof. First consider the case where $S\equiv 1$, i.e., $X=V$ is Gaussian. Let $\otimes$ denote the tensor product on $\mathcal{H}$, e.g., for $u,v\in\mathcal{H}$, the operator $u\otimes v:\mathcal{H}\to\mathcal{H}$ is defined as $(u\otimes v)w=\langle v,w\rangle u$. Let $\mathcal{L}$ be a linear space of dimension q with orthonormal basis $\gamma_1,\dots,\gamma_q$ and define the operators $P=\sum_{i=1}^{q}\gamma_i\otimes\gamma_i$ and $P_0=\sum_{i=1}^{q}\phi_i\otimes\phi_i$. Both P and P_0 are bounded operators with finite rank q. Let Y and Y_0 denote residuals of the orthogonal projections on $\mathcal{L}$ and $\mathcal{L}_q$ respectively, i.e. $Y=V-\pi(V,\mathcal{L})=(\mathbb{I}_{\mathcal{H}}-P)V$ and $Y_0=V-\pi(V,\mathcal{L}_q)=(\mathbb{I}_{\mathcal{H}}-P_0)V$, where $\mathbb{I}_{\mathcal{H}}$ denotes the identity operator in $\mathcal{H}$, i.e. $\mathbb{I}_{\mathcal{H}}x=x$ for all $x\in\mathcal{H}$.

Note that Y and Y_0 are mean zero Gaussian random elements with covariance operators $\widetilde{\mathbf{\Gamma}}=(\mathbb{I}_{\mathcal{H}}-P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}}-P)^*$ and $\mathbf{\Gamma}_0=(\mathbb{I}_{\mathcal{H}}-P_0)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}}-P_0)^*$, respectively. Furthermore, $\mathbf{\Gamma}_0$ has q zero eigenvalues related to $\phi_1,\dots,\phi_q$, the non–null ones being $\{\lambda_j\}_{j\geq q+1}$, so that $\lambda_j(\mathbf{\Gamma}_0)=\lambda_{j+q}=\lambda_{j+q}(\mathbf{\Gamma})$ and $\phi_j(\mathbf{\Gamma}_0)=\phi_{j+q}=\phi_{j+q}(\mathbf{\Gamma})$, for $j\geq 1$, where, in general, $\lambda_j(\mathbf{\Upsilon})$ and $\phi_j(\mathbf{\Upsilon})$ denote the eigenvalues and eigenfunctions of the operator $\mathbf{\Upsilon}$, respectively. Hence, we have the following representation for Y_0:

$$Y_0=\sum_{j\geq 1}\lambda_{j+q}^{\frac{1}{2}}f_j\phi_{j+q}\,, \tag{2}$$

where f_j are i.i.d $f_j\sim N(0,1)$. It is worth noting, that the convergence of the series in (2) is with respect to the strong topology in $\mathcal{H}$, that is, $\mathbb{P}(\lim_{M\to\infty}\|Y_0-\sum_{j=1}^{M}\lambda_{j+q}^{\frac{1}{2}}f_j\phi_{j+q}\|=0)=1$.

Similarly, $\widetilde{\mathbf{\Gamma}}$ has q zero eigenvalues related to $\gamma_1,\dots,\gamma_q$. Denote as $\{\widetilde{\lambda}_j\}_{j\geq 1}$, the non–null eigenvalues such that $\widetilde{\lambda}_1\geq\widetilde{\lambda}_2\geq\dots$ and by γ_{q+j}, $j\geq 1$ the orthonormal eigenfunctions of $\widetilde{\mathbf{\Gamma}}$ related to $\widetilde{\lambda}_j$ which form an orthonormal basis of $\mathcal{L}^{\perp}$. In other words, $\widetilde{\lambda}_j=\lambda_j(\widetilde{\mathbf{\Gamma}})$ and $\gamma_{q+j}=\phi_j(\widetilde{\mathbf{\Gamma}})$, and we have the following representation for Y:

$$Y=\sum_{j\geq 1}\widetilde{\lambda}_j^{\frac{1}{2}}f_j\gamma_{j+q}\,. \tag{3}$$

Thus, from (2) and (3) we have $\mathbb{P}\left(\|Y\|^2\leq y\right)=\mathbb{P}\left(\sum_{j\geq 1}\widetilde{\lambda}_jf_j^2\leq y\right)$ and $\mathbb{P}\left(\|Y_0\|^2\leq y\right)=\mathbb{P}\left(\sum_{j\geq 1}\lambda_{j+q}f_j^2\leq y\right)$. The result now follows from Proposition A.1 in the Appendix which shows that $\widetilde{\lambda}_j=\lambda_j((\mathbb{I}_{\mathcal{H}}-P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}}-P)^*)\geq\lambda_{j+q}(\mathbf{\Gamma})=\lambda_{j+q}$.

When S is a non-negative random variable independent from V we have $\|X-\pi(X,\mathcal{L})\|=|S|\,\|Y\|$ and $\|X-\pi(X,\mathcal{L}_q)\|=|S|\,\|Y_0\|$, where Y and Y_0 are as above. Therefore, the result follows easily from the previous case using that $\mathbb{P}\left(\|X-\pi(X,\mathcal{L})\|^2\leq y\right)=\mathbb{P}\left(S^2\|Y\|^2\leq y\right)=\mathbb{E}\left[\mathbb{P}\left(S^2\|Y\|^2\leq y|S\right)\right]$, and that for any $s\neq 0$,

$$\mathbb{P}\left(S^2\|Y\|^2\leq y|S=s\right)=\mathbb{P}\left(\|Y\|^2\leq ys^{-2}\right)\leq\mathbb{P}\left(\|Y_0\|^2\leq ys^{-2}\right).\ \square$$

As mentioned above, finite–dimensional counterparts of Propositions 3.1 and 3.2 are given in Proposition 1 in Boente (1983).

3.2. An optimality property of principal components related to the covariance operator

In this section, we extend another well-known optimality property of the principal components of random vectors on $\mathbb{R}^p$ to the more general setting of separable Hilbert spaces. More specifically, let $\mathbf{X}$ be a p-variate random vector with finite second moments and covariance matrix $\boldsymbol{\Sigma}$. Consider the problem of approximating $\mathbf{X}$ with q linear combinations of its own coordinates, and let $\boldsymbol{\Sigma}_{\text{RES}}$ be the covariance matrix of the vector of residuals. Then, for any orthogonal invariant matrix norm $|||\cdot|||$ (that satisfies $|||\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}||| = |||\boldsymbol{\Sigma}|||$ whenever $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices) we have that $|||\boldsymbol{\Sigma}_{\text{RES}}|||$ is minimized when one uses the projections onto the q largest principal components. This is a direct consequence of von Neumann's (1937) characterization of such matrix norms as those generated by symmetric gauge functions of the corresponding singular values of the matrices.

It is known (see Gohberg and Krein, 1969) that von Neumann's (1937) characterization also holds for finite-rank operators on infinite-dimensional Hilbert spaces. In what follows we show that this is also true for positive semi–definite self–adjoint compact operators, provided that the unitarily invariant norm $|||\cdot|||$ satisfies some relatively mild regularity conditions. This observation allows us to extend the above optimality property of principal components to random elements over separable Hilbert spaces.

Let $\mathcal{H}$ be a separable Hilbert space. A norm $|||\cdot|||$ for compact operators on $\mathcal{H}$ is said to be unitarily invariant norm if, for any compact operator $\boldsymbol{\Upsilon}$ we have that $|||U\boldsymbol{\Upsilon}V||| = |||\boldsymbol{\Upsilon}|||$ for arbitrary unitary operators U and V. Well-known unitarily invariant norms for linear bounded operators $\boldsymbol{\Upsilon}$ are the Schatten p-norms $\|\boldsymbol{\Upsilon}\|_{\text{SCH}} = \{\sum_{n\geq 1}\mu_n^p(\boldsymbol{\Upsilon})\}^{1/p}$, where $p \geq 1$ and $\mu_1(\boldsymbol{\Upsilon}) \geq \mu_2(\boldsymbol{\Upsilon}) \geq \cdots \geq 0$ denote the singular values of $\boldsymbol{\Upsilon}$, that is, the eigenvalues of the operator $|\boldsymbol{\Upsilon}| = (\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^*)^{1/2}$. In other words, $\|\boldsymbol{\Upsilon}\|_{\text{SCH}}^p = \text{trace}(|\boldsymbol{\Upsilon}|^p)$. The family of Schatten norms includes the Hilbert–Schmidt operator norm and also the trace norm. Furthermore, the usual operator norm $\|\boldsymbol{\Upsilon}\|_{\text{OP}} = \sup_{\|x\|=1}\|\boldsymbol{\Upsilon}x\|$ is also unitarily invariant. For a more detailed discussion on unitary norms we refer to Gohberg and Krein (1969).

As mentioned before (see also Gohberg and Krein, 1969), over the set of finite-rank operators, unitary norms are generated by symmetric gauge functions Φ defined over infinite sequences. More precisely, let ℓ_0 be the set of all infinite sequences $\mathbf{a} = \{a_n\}_{n\geq 1}$ converging to 0, and let $\mathcal{F}_0 \subset \ell_0$ be the subset of sequences with only finitely many non–zero elements. A function $\Phi : \mathcal{F}_0 \to \mathbb{R}$ is called a gauge function if it satisfies:

a) $\Phi(\mathbf{a}) > 0$ for any $\mathbf{a} \in \mathcal{F}_0$, $\mathbf{a} \neq 0$,
b) $\Phi(\alpha\mathbf{a}) = |\alpha|\,\Phi(\mathbf{a})$ for any $\mathbf{a} \in \mathcal{F}_0$, $\alpha \in \mathbb{R}$,
c) $\Phi(\mathbf{a}+\mathbf{b}) \leq \Phi(\mathbf{a}) + \Phi(\mathbf{b})$ for any $\mathbf{a}, \mathbf{b} \in \mathcal{F}_0$, and
d) $\Phi(1, 0, 0, \dots) = 1$.

It is said to be a symmetric gauge function if, in addition,

e) $\Phi(a_1, a_2, \dots, a_n, 0, 0, \dots) = \Phi(|a_{j_1}|, |a_{j_2}|, \dots, |a_{j_n}|, 0, 0, \dots)$ for any $\mathbf{a} \in \mathcal{F}_0$ and any permutation $j_1, j_2, \dots, j_n$ of the indices $1, \dots, n$.

In particular, Gohberg and Krein (1969) show that, if $\mathbf{a} \in \mathcal{F}_0$ and $\mathbf{b} \in \mathcal{F}_0$ are such that $a_i \geq 0$ and $b_i \geq 0$ and $\sum_{i=1}^k a_i \leq \sum_{i=1}^k b_i$, for any $k \geq 1$ (in particular, if $a_k \leq b_k$) then, $\Phi(\mathbf{a}) \leq \phi(\mathbf{b})$ (*increasing monotonicity*). Moreover, if $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$ then

$$a_1 \;\leq\; \Phi(a_1, a_2, \dots, a_n, 0, 0, \dots) \;\leq\; \sum_{i=1}^n a_i\,. \tag{4}$$

Furthermore, we have that for any unitarily invariant norm $|||\cdot|||$, there exists a symmetric gauge function Φ such that $|||\boldsymbol{\Upsilon}||| = \|\boldsymbol{\Upsilon}\|_\Phi = \Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}))$ for any finite-rank operator $\boldsymbol{\Upsilon}$. For example, for the Schatten p-norm $\|\boldsymbol{\Upsilon}\|_{\text{SCH}} = \{\sum_{j\geq 1}\mu_j^p(\boldsymbol{\Upsilon})\}^{1/p}$ the corresponding function Φ is the usual ℓ^p norm for sequences. Moreover, the singular values of semi–definite self–adjoint compact operators are equal to their eigenvalues, and hence $\|\boldsymbol{\Upsilon}\|_{\text{OP}} = \mu_1(\boldsymbol{\Upsilon})$ so that Φ is the ℓ^∞ norm for sequences.

To extend this result to positive semi-definite and self-adjoint compact operators, we need to introduce some notation. Given a symmetric gauge function Φ, its maximal space m_Φ contains those sequences $\mathbf{a} \in \ell_0$ such that $\lim_{n\to\infty}\Phi(a_1, a_2, \ldots, a_n, 0, 0, \ldots) \equiv \Phi(\mathbf{a})$ exists and is finite. Let $\mathcal{J}_\Phi$ denote the subset of compact operators $\boldsymbol{\Upsilon}$ such that the sequence of its singular values $\boldsymbol{\mu}(\boldsymbol{\Upsilon}) = \{\mu_j(\boldsymbol{\Upsilon})\}_{j\geq 1}$ belongs to m_Φ. We can define a norm $\|\cdot\|_\Phi$ over $\mathcal{J}_\Phi$ as $\|\boldsymbol{\Upsilon}\|_\Phi = \Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}))$, where each non-zero singular value appears repeated according to its multiplicity. Let $\mathcal{A}$ be the linear space of positive semi–definite, self–adjoint compact operators $\boldsymbol{\Upsilon}$. If the unitarily invariant norm $|||\cdot|||$ is continuous with respect to the weak convergence of operators and $|||\boldsymbol{\Upsilon}||| < \infty$, then $|||\boldsymbol{\Upsilon}||| = \lim_{n\to\infty}|||\boldsymbol{\Upsilon}_n|||$ where $\boldsymbol{\Upsilon}_n = \sum_{i=1}^n \lambda_i(\boldsymbol{\Upsilon})\phi_i(\boldsymbol{\Upsilon}) \otimes \phi_i(\boldsymbol{\Upsilon})$ and $\boldsymbol{\Upsilon} = \sum_{i=1}^\infty \lambda_i(\boldsymbol{\Upsilon})\phi_i(\boldsymbol{\Upsilon}) \otimes \phi_i(\boldsymbol{\Upsilon})$ is the spectral decomposition of $\boldsymbol{\Upsilon}$, with $\phi_i(\boldsymbol{\Upsilon})$ the orthonormal basis of eigenfunctions related to the eigenvalues $\lambda_1(\boldsymbol{\Upsilon}) \geq \lambda_2(\boldsymbol{\Upsilon}) \geq \ldots$. Since, for all $n \geq 1$, $\boldsymbol{\Upsilon}_n$ has finite rank, there exists a symmetric gauge function Φ such that $|||\boldsymbol{\Upsilon}||| = \lim_{n\to\infty}\Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}_n))$. Now note that $\boldsymbol{\mu}(\boldsymbol{\Upsilon}_n)$ equals to the first n entries of $\boldsymbol{\mu}(\boldsymbol{\Upsilon})$ and that, since $\lim_{n\to\infty}\Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}_n))$ exists (and equals $|||\boldsymbol{\Upsilon}|||$), we have that $\boldsymbol{\Upsilon} \in \mathcal{J}_\Phi$ and $\Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon})) \equiv \lim_{n\to\infty}\Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}_n))$. Thus, for this symmetric gauge function Φ we have $|||\boldsymbol{\Upsilon}||| = \Phi(\boldsymbol{\mu}(\boldsymbol{\Upsilon}))$, for any $\boldsymbol{\Upsilon}$ such that $|||\boldsymbol{\Upsilon}||| < \infty$. The same conclusion holds if the norm $|||\cdot|||$ satisfies the following weaker condition: for any $\boldsymbol{\Upsilon} \in \mathcal{A}$ such that $|||\boldsymbol{\Upsilon}||| < \infty$, we have $|||\boldsymbol{\Upsilon}_n - \boldsymbol{\Upsilon}||| \to 0$. Note that for any unitarily invariant norm $|||\cdot|||$, if the operator $\boldsymbol{\Upsilon}$ has finite trace (as covariance operators do), then the sequence $\{\boldsymbol{\Upsilon}_n\}_{n\geq 1}$ is Cauchy (because we have $|||\boldsymbol{\Upsilon}_{n+m} - \boldsymbol{\Upsilon}_n||| = \|\boldsymbol{\Upsilon}_{n+m} - \boldsymbol{\Upsilon}_n\|_\Phi = \Phi(\lambda_{n+1}(\boldsymbol{\Upsilon}), \lambda_{n+2}(\boldsymbol{\Upsilon}), \ldots, \lambda_{n+m}(\boldsymbol{\Upsilon}), 0, 0, \ldots) \leq \sum_{j=1}^m \lambda_{n+j}(\boldsymbol{\Upsilon})$, where the last inequality follows from (4)). Hence, if the set $\mathcal{B} = \{\boldsymbol{\Upsilon} : |||\boldsymbol{\Upsilon}||| < \infty\}$ is a Banach space a similar argument shows that there exists a symmetric gauge function Φ such that such that $|||\boldsymbol{\Upsilon}||| = \|\boldsymbol{\Upsilon}\|_\Phi$ for any $\boldsymbol{\Upsilon} \in \mathcal{A} \cap \mathcal{B}$.

Let X be a random element on a complete Hilbert space $\mathcal{H}$ with finite second moment. Without loss of generality assume that X has mean zero and let $\boldsymbol{\Gamma}$ denote its covariance operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ and associated eigenfunctions $\{\phi_j\}_{j\geq 1}$. Consider an arbitrary q-dimensional linear space $\mathcal{L}$ in $\mathcal{H}$, and let $\pi(X, \mathcal{L}) = P\,X$ be the orthogonal projection on $\mathcal{L}$, where P denotes the projection operator. Similarly, let $\mathcal{L}_0$ be the linear space spanned by $\phi_1, \phi_2, \ldots, \phi_q$, the first q eigenfunctions of $\boldsymbol{\Gamma}$ and let P_0 denote the orthogonal projection operator on $\mathcal{L}_0$, so that $\pi(X, \mathcal{L}_0) = P_0\,X$. If $Y = X - \pi(X, \mathcal{L})$ and $Y_0 = X - \pi(X, \mathcal{L}_0)$ are the corresponding residuals, then the covariance operators of Y and Y_0 are given by $\boldsymbol{\Gamma}_{\text{RES}} = (\mathbb{I}_\mathcal{H} - P)\boldsymbol{\Gamma}(\mathbb{I}_\mathcal{H} - P)^*$ and $\boldsymbol{\Gamma}_{\text{RES},0} = (\mathbb{I}_\mathcal{H} - P_0)\boldsymbol{\Gamma}(\mathbb{I}_\mathcal{H} - P_0)^*$, respectively. Similarly, those of $\pi(X, \mathcal{L})$ and $\pi(X, \mathcal{L}_0)$ are given by $\boldsymbol{\Gamma}_{\text{PROJ}} = P\boldsymbol{\Gamma}P^*$ and $\boldsymbol{\Gamma}_{\text{PROJ},0} = P_0\boldsymbol{\Gamma}P_0^*$, respectively.

As noted in the proof of Proposition 3.2, $\boldsymbol{\Gamma}_{\text{RES},0}$ has q zero eigenvalues related to $\phi_1, \ldots, \phi_q$, the non–null ones being $\{\lambda_j\}_{j\geq q+1}$, so that $\lambda_j(\boldsymbol{\Gamma}_{\text{RES},0}) = \lambda_{j+q} = \lambda_{j+q}(\boldsymbol{\Gamma})$ and $\phi_j(\boldsymbol{\Gamma}) = \phi_{j+q} = \phi_{j+q}(\boldsymbol{\Gamma})$, for $j \geq 1$. On the other hand, $\boldsymbol{\Gamma}_{\text{PROJ},0}$ is a finite–rank operator with eigenvalues $\lambda_1(\boldsymbol{\Gamma}) \geq \lambda_2(\boldsymbol{\Gamma}) \geq \cdots \geq \lambda_q(\boldsymbol{\Gamma})$, and Proposition A.1 in the Appendix shows that $\lambda_j(\boldsymbol{\Gamma}_{\text{RES}}) = \lambda_j((\mathbb{I}_\mathcal{H} - P)\boldsymbol{\Gamma}(\mathbb{I}_\mathcal{H} - P)^*) \geq \lambda_j(\boldsymbol{\Gamma}_{\text{RES},0}) = \lambda_{j+q}$, and $\lambda_j(\boldsymbol{\Gamma}_{\text{PROJ}}) \leq \lambda_j(\boldsymbol{\Gamma}_{\text{PROJ},0})$.

Using these results we immediately obtain the following proposition, which can be seen as an extension of the finite–dimensional one (see Okamoto and Kanazawa, 1968). Note that this proposition also holds when the random element X has an elliptical distribution $\mathcal{E}(0, \boldsymbol{\Gamma}, \varphi)$ with $\boldsymbol{\Gamma}$ a positive semi-definite self-adjoint and compact operator.

Proposition 3.3. *Let $X \in \mathcal{H}$ be a zero mean random element with covariance operator $\mathbf{\Gamma}$. Denote by $\lambda_1 \geq \lambda_2 \geq \dots$ the eigenvalues of $\mathbf{\Gamma}$ with associated eigenfunctions ϕ_j, $j \geq 1$. Then, if $\lambda_q > \lambda_{q+1}$, for any linear space $\mathcal{L}$ of dimension q, we have that*

(a) for any a symmetric gauge function Φ if $\mathbf{\Gamma} \in \mathcal{J}_\Phi$ we have that $\|\mathbf{\Gamma}_{\text{RES},0}\|_\Phi \leq \|\mathbf{\Gamma}_{\text{RES}}\|_\Phi$.

(b) $\|\mathbf{\Gamma}_{\text{RES},0}\|_{\text{OP}} \leq \|\mathbf{\Gamma}_{\text{RES}}\|_{\text{OP}}$ and equality is obtained only when $\mathcal{L} = \mathcal{L}_0$.

(c) $\|\mathbf{\Gamma}_{\text{PROJ}}\|_{\text{OP}} \leq \|\mathbf{\Gamma}_{\text{PROJ},0}\|_{\text{OP}}$.

(d) if, in addition, $\|\mathbf{\Gamma}\|_{\text{SCH}} < \infty$, then $\|\mathbf{\Gamma}_{\text{RES},0}\|_{\text{SCH}} \leq \|\mathbf{\Gamma}_{\text{RES}}\|_{\text{SCH}}$ and $\|\mathbf{\Gamma}_{\text{PROJ}}\|_{\text{SCH}} \leq \|\mathbf{\Gamma}_{\text{PROJ},0}\|_{\text{SCH}}$ and equality is obtained only when $\mathcal{L} = \mathcal{L}_0$.

(e) for any $|||\cdot|||$ unitarily invariant norm, $|||\mathbf{\Gamma}_{\text{PROJ}}||| \leq |||\mathbf{\Gamma}_{\text{PROJ},0}|||$.

PROOF. Recall that if $\mathbf{\Upsilon}$ is a positive semi–definite self–adjoint compact operator then $\mu_j(\mathbf{\Upsilon}) = \lambda_j(\mathbf{\Upsilon})$. Statement (a) follows from the fact that Φ is *increasing monotone*,

$$\Phi(\lambda_1(\mathbf{\Gamma}_{\text{RES}}), \lambda_2(\mathbf{\Gamma}_{\text{RES}}), \dots) = \lim_{n\to\infty} \Phi(\lambda_1(\mathbf{\Gamma}_{\text{RES}}), \lambda_2(\mathbf{\Gamma}_{\text{RES}}), \dots, \lambda_n(\mathbf{\Gamma}_{\text{RES}}), 0, 0, \dots),$$

and $\lambda_j(\mathbf{\Gamma}_{\text{RES}}) \geq \lambda_j(\mathbf{\Gamma}_{\text{RES},0})$.

Parts (b), (c) and (d) follow immediately from Proposition A.1 in the Appendix using that for any positive semi–definite self–adjoint compact operator $\mathbf{\Upsilon}$ we have $\|\mathbf{\Upsilon}\|_{\text{OP}} = \lambda_1(\mathbf{\Upsilon})$.

To prove (e) we use that $\mathbf{\Gamma}_{\text{PROJ}}$ and $\mathbf{\Gamma}_{\text{PROJ},0}$ are finite rank operators, so that $|||\mathbf{\Gamma}_{\text{PROJ}}||| = \|\mathbf{\Gamma}_{\text{PROJ}}\|_\Phi = \Phi(\lambda_1(\mathbf{\Gamma}_{\text{PROJ}}), \lambda_2(\mathbf{\Gamma}_{\text{PROJ}}), \dots)$ and $|||\mathbf{\Gamma}_{\text{PROJ},0}||| = \|\mathbf{\Gamma}_{\text{PROJ},0}\|_\Phi = \Phi(\lambda_1(\mathbf{\Gamma}_{\text{PROJ},0}), \lambda_2(\mathbf{\Gamma}_{\text{PROJ},0}), \dots)$ for some symmetric gauge function Φ so the result follows from Proposition A.1 and the above mentioned *increasing monotony* of Φ. $\square$

It is worth noting that we cannot ensure that in (c) and (e) above equality holds only when $\mathcal{L} = \mathcal{L}_0$. An example where this may happen for (e) is when using the Ky Fan's $k-$norm defined as $|||\mathbf{\Upsilon}||| = \sum_{i=1}^k \lambda_i(\mathbf{\Upsilon})$, for any self–adjoint positive semi–definite compact operator $\mathbf{\Upsilon}$. In particular, we have that $\|\cdot\|_{\text{OP}}$ corresponds to the choice $k = 1$. For these norms we have $|||\mathbf{\Gamma}_{\text{PROJ}}||| = \sum_{i=1}^k \lambda_i(\mathbf{\Gamma}_{\text{PROJ}})$, so that, if $k < q$, $|||\mathbf{\Gamma}_{\text{PROJ}}||| = |||\mathbf{\Gamma}_{\text{PROJ},0}|||$, but, since $\lambda_q > \lambda_{q+1}$, the linear space $\mathcal{L}$ spanned by $\phi_1, \phi_2, \dots, \phi_{q-1}, \phi_{q+1}$ is not equal to $\mathcal{L}_0$.

4. Applications and general comments

The results of the previous sections can be useful to study the properties of robust estimators for functional data in a framework more general than Gaussian random elements, without requiring finite moments. For example, as mentioned above, the robust estimates of principal directions defined in Bali, *et al.* (2011) are Fisher-consistent for elliptically distributed random elements. Proposition 2.1 was used in Bali and Boente (2013) to show the consistency of a projection-pursuit estimator for the first principal direction obtained by adapting the finite-dimensional algorithm of Croux and Ruiz–Gazen (1996). The optimality results in Section 3, in particular the stochastic inequality of Proposition 3.2, served as the foundation for the Sieves–based $S-$estimators for functional principal components in Boente and Salibian–Barrera (2013)

More in general, whenever considering elliptical random elements, the results in Section 2 allow us to consider two possible cases: either (a) $\mathbf{\Gamma}$ has finite rank, which reduces to the finite-dimensional setting; or (b) the rank of $\mathbf{\Gamma}$ is infinite and, by Proposition 2.1 the process is a scale mixture of a Gaussian element. This approach can be used to study the consistency of the spherical principal components estimators proposed in Gervini (2008). They are given by the eigenfunctions of the spherical covariance operator (Locantore *et al.*, 1999), defined as

$$\widetilde{\mathbf{\Gamma}} = \mathbb{E}\left\{ \frac{(X-\mu)\otimes(X-\mu)}{\|X-\mu\|^2} \right\} . \tag{5}$$

Gervini (2008) showed that these estimators are Fisher–consistent for the principal directions when the process admits a Karhunen–Loève expansion with only finitely many terms (i.e., assuming that the distribution is concentrated on a finite–dimensional subspace). The following Proposition, whose proof is included in the Appendix, shows that the spherical principal components are in fact Fisher–consistent for any elliptical distribution.

Proposition 4.1. *Let $X \in \mathcal{H}$ be an elliptically distributed random element $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ with $\mathbf{\Gamma}$ a self–adjoint, positive semi–definite and compact operator. Denote by $\lambda_1 \geq \lambda_2 \geq \dots$ the eigenvalues of $\mathbf{\Gamma}$ with associated eigenfunctions ϕ_j, $j \geq 1$, and let $\widetilde{\mathbf{\Gamma}}$ be as in* (5)*. Then, the eigenfunctions of $\widetilde{\mathbf{\Gamma}}$ are ϕ_j, $j \geq 1$ with eigenvalues $\widetilde{\lambda}_j$, $j \geq 1$, where $\widetilde{\lambda}_{j+1} \geq \widetilde{\lambda}_j$. Moreover, if $\lambda_{j+1} \geq \lambda_j$ we have $\widetilde{\lambda}_{j+1} > \widetilde{\lambda}_j$.*

Testing whether a random vector or random element has an elliptical distribution is a challenging problem. In the finite–dimensional setting tests have been proposed by Beran (1979), Koltchinskii and Sakhanenko (2000), Schott (2002), Zhu and Neuhaus (2003), Huffer and Park (2007) and more recently, Batsidis and Zografos (2013). Developing such a test for the distribution of an infinite–dimensional random element is beyond the scope of this paper. However, the random projections procedure used in Cuesta Albertos *et al.* (2014) to decide whether a stationary process has a Gaussian distribution may be generalized to build a consistent test for ellipticity. The above paper proposes tests based on testing if the projections are normal using univariate tests for normality. For ellipticity, the appropriate generalization of this method would be to replace the univariate tests for normality with univariate tests for detecting if two projections are location–scale transformations of each other, e.g., tests based on their relative kurtosis. Another possible and more exploratory approach would be to generalize the methods discussed in Tyler *et al.* (2009) for exploring whether and how a multivariate distribution deviates from an elliptical distribution. The method in Tyler *et al.* (2009) is based upon a comparison of different consistent estimates of the scatter matrix. In the infinite-dimensional setting, we are unaware of any consistent alternative to the sample covariance operator (presuming second moments). As an alternative approach one could consider comparing different consistent estimates of the principal components, such as those proposed in Bali *et al.*(2011), and if they vary greatly then this would indicate that the underlying distribution is not well modelled by an elliptical distribution. We leave these problems, however, as open topics for future research.

A. Appendix

In this Appendix, we recall some results related to the eigenvalues of compact self–adjoint operators. We first state a characterization of the eigenvalues, usually known as the min–max theorem or the Courant–Fisher principle, whose proof can be found in Dunford and Schwartz (1965), chapter 10.

Theorem A.1. *Let $\mathbf{\Gamma}$ be a compact, positive semi–definite and self adjoint operator over a separable Hilbert space $\mathcal{H}$ and $\lambda_1 \geq \lambda_2 \geq \dots$ its positive eigenvalues, each repeated a number of times equal* *to its multiplicity. Denote by ϕ_j the eigenfunction related to λ_j chosen so that $\{\phi_j,\ j \geq 1\}$ is an orthonormal set in $\mathcal{H}$. Then, we have that*

$$\begin{aligned}
\lambda_1 &= \max_{x \neq 0} \frac{\langle \mathbf{\Gamma} x, x\rangle}{\|x\|^2} \\
\lambda_{k+1} &= \min_{\mathcal{L}_k} \max_{x \in \mathcal{L}_k^{\perp}} \frac{\langle \mathbf{\Gamma} x, x\rangle}{\|x\|^2}\ ,
\end{aligned} \tag{A.1}$$

where $\mathcal{L}_k$ stands for any $k-$dimensional linear space and $\mathcal{L}_k^{\perp}$ for its orthogonal. The minimum in (A.1) is attained when $\mathcal{L}_k$ is spanned by $\phi_1, \ldots, \phi_k$.

Using Theorem A.1, we obtain the following Proposition whose part a) is known, in the finite–dimensional case, as the Poincaré's Theorem. As above let $\lambda_i(\mathbf{\Upsilon})$ be the $i-$th largest eigenvalue of the self–adjoint, positive semi-definite and compact operator $\mathbf{\Upsilon}$.

Proposition A.1. *Let $\mathbf{\Gamma}$ be a self–adjoint, positive semi-definite and compact operator over a separable Hilbert space $\mathcal{H}$ and $\lambda_1 \geq \lambda_2 \geq \ldots$ its positive eigenvalues, each repeated a number of times equal to its multiplicity. Let ϕ_j denote the eigenfunction associated to λ_j, $j \geq 1$. Then:*

a) If $B : \mathcal{H} \to \mathcal{H}_1$ is a linear sub-unitary operator, that is, $BB^* = \mathbb{I}_{\mathcal{F}}$, where $\mathcal{F} = \text{range}(B)$ and $\mathcal{H}_1$ is a separable Hilbert space, then $\lambda_{k+1}(B\mathbf{\Gamma}B^*) \leq \lambda_{k+1}(\mathbf{\Gamma}) = \lambda_{k+1}$.

b) Let $\mathcal{L}$ be a linear space of dimension q and assume that $\mathcal{L}$ is spanned by $\gamma_1, \ldots, \gamma_q$ with γ_j orthonormal elements of $\mathcal{H}$. Denote by P the bounded projection operator over $\mathcal{L}$, i.e., $Px = \sum_{j=1}^q \langle x, \gamma_j\rangle \gamma_j$. Then, we have that $\lambda_k((\mathbb{I}_{\mathcal{H}} - P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^*) \geq \lambda_{k+q}(\mathbf{\Gamma})$.

Proof. a) The proof follows the same lines as in the finite–dimensional setting (see, Okamoto, 1969). First note that, since $BB^* = I$, B^* and B are both bounded operators, so $B\mathbf{\Gamma}B^*$ is a compact operator. From Theorem A.1 we have that

$$\begin{aligned}
\lambda_{k+1} &= \max_{\substack{\langle x, \phi_j\rangle = 0 \\ 1 \leq j \leq k}} \frac{\langle \mathbf{\Gamma} x, x\rangle}{\langle x, x\rangle} \\
&\geq \max_{\substack{\langle x, \phi_j\rangle = 0 \\ 1 \leq j \leq k, x = B^*\mathbf{y}}} \frac{\langle \mathbf{\Gamma} x, x\rangle}{\langle x, x\rangle} = \max_{\substack{\langle B^*\mathbf{y}, \phi_j\rangle = 0 \\ 1 \leq j \leq k}} \frac{\langle \mathbf{\Gamma} B^*\mathbf{y}, B^*\mathbf{y}\rangle}{\langle B^*\mathbf{y}, B^*\mathbf{y}\rangle} = \max_{\substack{\langle B^*\mathbf{y}, \phi_j\rangle = 0 \\ 1 \leq j \leq k}} \frac{\langle B\mathbf{\Gamma} B^*\mathbf{y}, \mathbf{y}\rangle_1}{\langle BB^*\mathbf{y}, \mathbf{y}\rangle_1} \\
&\geq \max_{\substack{\langle \mathbf{y}, B\phi_j\rangle_1 = 0 \\ 1 \leq j \leq k}} \frac{\langle B\mathbf{\Gamma} B^*\mathbf{y}, \mathbf{y}\rangle_1}{\langle \mathbf{y}, \mathbf{y}\rangle_1}\ ,
\end{aligned}$$

where $\langle \cdot, \cdot\rangle_1$ denotes the inner product in $\mathcal{H}_1$. Denote by $\mathcal{L}$ the linear space in $\mathcal{H}_1$ spanned by $B\phi_1, \ldots, B\phi_k$, then $\dim(\mathcal{L}) = k_o \leq k$, and so, if $\mathcal{S}_{k_o}$ stands for any linear space of dimension k_o

$$\max_{\substack{\langle \mathbf{y}, B\phi_j\rangle_1 = 0 \\ 1 \leq j \leq k}} \frac{\langle B\mathbf{\Gamma} B^*\mathbf{y}, \mathbf{y}\rangle_1}{\langle \mathbf{y}, \mathbf{y}\rangle_1} \geq \inf_{\mathcal{S}_{k_o}} \max_{y \perp \mathcal{S}_{k_o}} \frac{\langle B\mathbf{\Gamma} B^*\mathbf{y}, \mathbf{y}\rangle_1}{\langle \mathbf{y}, \mathbf{y}\rangle_1} = \lambda_{k_o+1}(B\mathbf{\Gamma}B^*)\ ,$$

concluding the proof of a) since $\lambda_{k_o+1}(B\mathbf{\Gamma}B^*) \geq \lambda_{k+1}(B\mathbf{\Gamma}B^*)$.

b) It is worth noting that since $P = \sum_{j=1}^q \gamma_j \otimes \gamma_j$, then $\mathbb{I}_{\mathcal{H}} - P = (\mathbb{I}_{\mathcal{H}} - P)^*$. Let ψ_j be the eigenfunction associated to the eigenvalue $\lambda_j((\mathbb{I}_{\mathcal{H}} - P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^*)$, $j \geq 1$ chosen so that $\{\psi_j\ ,\ j \geq 1\}$

form an orthonormal set. Denote $\mathcal{S}_{k-1}$ the linear space of dimension k spanned by $\psi_1, \ldots, \psi_{k-1}$. Then,

$$\begin{aligned}
\lambda_k((\mathbb{I}_{\mathcal{H}} - P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^*) &= \max_{x \in \mathcal{S}_{k-1}^{\perp}} \frac{\langle (\mathbb{I}_{\mathcal{H}} - P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^* x, x\rangle}{\|x\|^2} \\
&= \max_{x \in \mathcal{S}_{k-1}^{\perp}} \frac{\langle \mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^* x, (\mathbb{I}_{\mathcal{H}} - P)^* x\rangle}{\|x\|^2} \\
&\geq \max_{x \in \mathcal{S}_{k-1}^{\perp} \cap \mathcal{L}^{\perp}} \frac{\langle \mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^* x, (\mathbb{I}_{\mathcal{H}} - P)^* x\rangle}{\|x\|^2} \\
&= \max_{x \in \mathcal{S}_{k-1}^{\perp} \cap \mathcal{L}^{\perp}} \frac{\langle \mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^* x, (\mathbb{I}_{\mathcal{H}} - P)^* x\rangle}{\langle (\mathbb{I}_{\mathcal{H}} - P)^* x, (\mathbb{I}_{\mathcal{H}} - P)^* x\rangle} \,.
\end{aligned}$$

Note that for any $x \in \mathcal{H}$, $(\mathbb{I}_{\mathcal{H}} - P)^* x$ is orthogonal to $\mathcal{L}$. Moreover, if $x \in \mathcal{S}_{k-1}^{\perp} \cap \mathcal{L}^{\perp}$ then $(\mathbb{I}_{\mathcal{H}} - P)^* x = x$ is also orthogonal to $(\mathbb{I}_{\mathcal{H}} - P)\psi_j$, $j = 1, \ldots, k-1$. Let $\mathcal{D}$ be the linear subspace spanned by $(\mathbb{I}_{\mathcal{H}} - P)\psi_1, \ldots (\mathbb{I}_{\mathcal{H}} - P)\psi_{k-1}, \gamma_1, \gamma_2, \ldots, \gamma_q$, and note that $\mathcal{D}^{\perp} = \mathcal{S}_{k-1}^{\perp} \cap \mathcal{L}^{\perp}$ and $\dim(\mathcal{D}) = m \leq k + q - 1$. Then, it follows that

$$\lambda_k((\mathbb{I}_{\mathcal{H}} - P)\mathbf{\Gamma}(\mathbb{I}_{\mathcal{H}} - P)^*) \geq \max_{y \in \mathcal{D}^{\perp}} \frac{\langle \mathbf{\Gamma}\, y, y\rangle}{\langle y, y\rangle} \geq \lambda_{m+1}(\mathbf{\Gamma}) \geq \lambda_{k+q}(\mathbf{\Gamma})\,. \qquad \square$$

It is worth noting that when $\mathcal{H}_1 = \mathcal{H}$ and $B : \mathcal{H} \to \mathcal{H}$ is the projection operator over a closed space, the results in Proposition A.1 can also be derived from Theorem 2.2 of Pousse and Téchené (1997). However, we keep our proof due to its simplicity.

Proof of Proposition 4.1. When $\mathbf{\Gamma}$ has finite rank, the result follows immediately from Theorem 3 in Gervini (2008) since in this case X admits a finite–dimensional expansion. On the other hand, if $\mathbf{\Gamma}$ does not have finite rank, then by Proposition 2.1. there exists a zero mean Gaussian element V and a random variable $S \geq 0$, independent of V, such that $X \sim \mu + S\, V$. Hence, the spherical scatter operator $\widetilde{\mathbf{\Gamma}}$ defined in (5) equals $\widetilde{\mathbf{\Gamma}}_V$ where

$$\widetilde{\mathbf{\Gamma}}_V = \mathbb{E}\left\{\frac{V \otimes V}{\|V\|^2}\right\} \,.$$

Furthermore, as mentioned in Section 2, the covariance operator of the Gaussian process V, $\mathbf{\Gamma}_V$, satisfies $\mathbf{\Gamma}_V = \mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is the scatter operator of X. Thus, we only need to show that Proposition 4.1 holds for Gaussian processes.

Assume, for the sake of simplicity, that the non-null eigenvalues of $\mathbf{\Gamma}$ are $\lambda_1 \geq \lambda_2 \geq \ldots$. Then, V has the following Karhunen–Loève expansion, $V = \sum_{j \geq 1} \lambda_j^{1/2} \xi_j \phi_j$, with $\xi_j \sim N(0,1)$, independent of each other. Hence, $(V \otimes V)\, \phi_k = \lambda_k^{1/2} \xi_k V$. Since ξ_k has a symmetric distribution, we have that $\mathbb{E}\left(\xi_k \sum_{j \neq k} \lambda_j^{1/2} \xi_j \phi_j / \|V\|^2\right) = 0$, and hence

$$\widetilde{\mathbf{\Gamma}}_V \phi_k = \lambda_k^{1/2}\, \mathbb{E}\left\{\frac{\xi_k V}{\|V\|^2}\right\} = \lambda_k\, \mathbb{E}\left\{\frac{\xi_k^2}{\|V\|^2}\right\} \phi_k = \widetilde{\lambda}_k \phi_k \ ,$$

where $\widetilde{\lambda}_k = \lambda_k\, \mathbb{E}(\xi_k^2 / \|V\|^2)$. Thus, we have that ϕ_k, $k \geq 1$, are the eigenfunctions of $\widetilde{\mathbf{\Gamma}}_V$ with associated eigenvalues $\widetilde{\lambda}_k$.

To show that the order among the eigenvalues is preserved we use a similar argument to that in the proof of Theorem 3 in Gervini (2008). Define $g_k : [0,\infty) \to \mathbb{R}$ as $g_k(\lambda) = \lambda\xi_k^2 / \left(\sum_{j\neq k} \lambda_j \xi_j^2 + \lambda\xi_k^2\right)$ and note that $\widetilde{\lambda}_k = \mathbb{E} g_k(\lambda_k)$ and that g_k is strictly increasing as a function of λ (because $\mathbb{P}(\xi_k \neq 0) = 1$). Hence, $\widetilde{\lambda}_k = \mathbb{E} g_k(\lambda_k) \geq \mathbb{E} g_k(\lambda_{k+1})$, with strict inequality if $\lambda_k > \lambda_{k+1}$. On the other hand, we have that

$$\mathbb{E} g_k(\lambda_{k+1}) = \mathbb{E}\left\{\frac{\lambda_{k+1}\xi_k^2}{\sum_{j\neq k,k+1} \lambda_j \xi_j^2 + \lambda_{k+1}\xi_{k+1}^2 + \lambda_{k+1}\xi_k^2}\right\} = \mathbb{E}\left\{\frac{\lambda_{k+1}\xi_{k+1}^2}{\sum_{j\neq k,k+1} \lambda_j \xi_j^2 + \lambda_{k+1}\xi_k^2 + \lambda_{k+1}\xi_{k+1}^2}\right\},$$

because $\xi_j \sim N(0,1)$ are independent of each other, and hence $(\sum_{j\neq k,k+1} \lambda_j \xi_j^2, \lambda_{k+1}\xi_{k+1}^2, \lambda_{k+1}\xi_k^2)$ has the same distribution as $(\sum_{j\neq k,k+1} \lambda_j \xi_j^2, \lambda_{k+1}\xi_k^2, \lambda_{k+1}\xi_{k+1}^2)$. Finally, note that $\lambda_k \geq \lambda_{k+1}$ implies that

$$\begin{aligned}
\widetilde{\lambda}_k = \mathbb{E} g_k(\lambda_k) &\geq \mathbb{E} g_k(\lambda_{k+1}) = \mathbb{E}\left\{\frac{\lambda_{k+1}\xi_{k+1}^2}{\sum_{j\neq k,k+1} \lambda_j \xi_j^2 + \lambda_{k+1}\xi_{k+1}^2 + \lambda_{k+1}\xi_k^2}\right\} \\
&\geq \mathbb{E}\left\{\frac{\lambda_{k+1}\xi_{k+1}^2}{\sum_{j\neq k,k+1} \lambda_j \xi_j^2 + \lambda_{k+1}\xi_{k+1}^2 + \lambda_k\xi_k^2}\right\} = \widetilde{\lambda}_{k+1} \,. \square
\end{aligned}$$

Acknowledgements. This research was partially supported by Grants from Universidad of Buenos Aires, CONICET and ANPCYT, Argentina (G. Boente), Discovery Grant of the Natural Sciences and Engineering Research Council of Canada (M. Salibián Barrera) and NSF grant DMS-0906773 (D. E. Tyler). The authors thank the Associate Editor and two anonymous reviewers for their constructive comments that helped to improve the quality of the paper.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley & Sons, New Jersey.

[2] Bali, J. L. and Boente, G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statistics and Probability Letters*, **79**, 1858-1865.

[3] Bali, J. L. and Boente, G. (2013). Consistency of a numerical approximation to the first principal component projection pursuit estimator. Submitted.

[4] Bali, J. L., Boente, G., Tyler, D. E. and Wang, J-L. (2011). Robust Functional Principal Components: a projection-pursuit approach. *Annals of Statistics*, **39**,2852-2882.

[5] Batsidis, A. and Zografos, K. (2013). A necessary test of fit of specific elliptical distributions based on an estimator of Song's measure. *Journal of Multivariate Analysis*, **113**, 91–105.

[6] Bentler, P. M. and Berkane, M. (1986). Greatest lower bound to the elliptical theory kurtosis parameter. *Biometrika*, **73**, 240-241.

[7] Beran, R. (1979). Testing for elliptical symmetry of a multivariate density. *Annals of Statistics*, **7**, 150–162.

[8] Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer, New York.

[9] Boente, G. (1983). *Robust Methods for Principal Components*, unpublished doctoral thesis, University of Buenos Aires [in Spanish].

[10] Boente, G. and Salibian–Barrera, M. (2013). $S-$estimators for functional principal component analysis. Submitted. Available at `http://www.stat.ubc.ca/~matias/Boente_Salibian_TR_06_06_2013.pdf`

[11] Cuesta–Albertos, J. A., Gamboa, F. and Nieto–Reyes, A. (2014). A random-projection based procedure to test if a stationary process is Gaussian. To appear in *Computational Statistics and Data Analysis.*

[12] Darroch, J. N. (1965). An optimality property of principal components. *Annals of Mathematical Statistics*, **36**, 1579-1582.

[13] Frahm, G. (2004). *Generalized Elliptical Distributions: Theory and Applications.* PhD. thesis from the University of Köln, Germany.

[14] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.

[15] Gohberg, I. C. and Krein, M. G. (1969). *Introduction to the Theory of Linear Nonselfadjoint Operators*, Volume 18, Translations of Mathematical Monographs.

[16] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[17] Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics.* Wiley, Hoboken, New Jersey.

[18] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441; 498-520.

[19] Huffer, F. and Park, C. (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis*, **98**, 256–281.

[20] Kano, Y. (1994). Consistency Property of Elliptic Probability Density Functions. *Journal of Multivariate Analysis*, **51**, 139-147.

[21] Kingman, J. F. C. (1972). On random sequences with spherical symmetry on random sequences with spherical symmetry. *Biometrika*, **59**, 492-494.

[22] Koltchinskii, V. and Sakhanenko, L. (2000). Testing for ellipsoidal symmetry of a multivariate distribution. In: *High Dimensional Probability II*, Eds. Giné, E., Mason, D. and Wellner, J., pp. 493–510.

[23] Kuwana, Y. and Kariya, T. (1991). LBI tests for multivariate normality in exponential power distributions. *Journal of Multivariate Analysis*, **39**, 117-134.

[24] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., and Cohen, K. L. (1999). Robust principal components for functional data. *Test*, **8**, 1-28.

[25] Maronna, R., Martin, R. D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons.

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

[27] Okamoto, M. and Kanazawa, M. (1968). Minimization of eigenvalues of a matrix and optimality of principal components. *Annals of Mathematical Statistics*, **39**, 859-863.

[28] Pousse, A. and Téchené, J. J. (1997). Quelques propriétés extrémales des valeurs singulières d'un opérateur compact et leur applications aux analyses factorielles d' une probablité ou d'une fonction aléatoire. *Probability and Mathematical Statistics*, **17**, 197-221.

[29] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.

[30] Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya, Ser. A*, **26**, 329-358.

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

[32] Schott, J. R. (2002). Testing for elliptical symmetry in covariance-matrix-based analyses.*Statist. Probab. Lett.*, **60**, 395–404.

[33] Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika*, **69**, 429-436.

[34] Tyler, D. E., Critchley, F., Dumbgen, L. and Oja, H. (2009). Invariant coordinate selection (with discussion). *J. Roy. Statist. Soc. Ser. B*, **71**, 549-592.

[35] von Neumann, J. (1937). Some matrix-inequalities and metrization of matrix-space. *Tomsk University Review*, **1**, 286-300.