Supplement to “ S –estimators for functional principal component analysis” *

Graciela Boente and Matías Salibián Barrera

Abstract

This supplementary file contains a discussion and proof of the Fisher–consistency of the Sieves–approach for S –estimators for functional principal components, and the proofs of Propositions 2.1 and 2.2. It also includes the analysis of the French mortality data set. In this supplement, equations are numbered (S.1), (S.2), ..., while Definitions and Propositions are numbered S.1.1, References to Propositions 2.1 and 2.2 (without the capital “S”) refer to those on the main body of the paper.

S.1 Fisher consistency of S –estimators for functional data

In this Section we derive the Fisher–consistency of the Sieves–approach for observations generated by an elliptically distributed random object, which is a natural generalization of elliptical random vectors to an infinite–dimensional setup. The following definition was given in Bali and Boente (2009).

*Graciela Boente (Email: gboente@dm.uba.ar) is Full Professor at the Departamento de Matemáticas from the Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires 1428, Argentina (Tel./FAX: 54-11-45763375). She also has a researcher position at the CONICET. Matías Salibián Barrera (Email: matias@stat.ubc.ca) is Associate Professor at the Department of Statistics from the University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, BC, V6T 1Z4, Canada (Tel./FAX: 1-604-8223410).

Definition S.1.1. Let X be a random element in a separable Hilbert space $\mathcal{H}$. We will say that X has an elliptical distribution with parameters $\mu_{\mathcal{H}} \in \mathcal{H}$ and $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$, where Γ is a self-adjoint, positive semi-definite and compact operator, if and only if for any linear and bounded operator $A : \mathcal{H} \rightarrow \mathbb{R}^d$ we have that the vector AX has a d -variate elliptical distribution with location parameter $A\mu_{\mathcal{H}}$, shape matrix $A\Gamma A^*$ and characteristic generator ϕ , that is, $AX \sim \mathcal{E}_d(A\mu_{\mathcal{H}}, A\Gamma A^*, \phi)$ where $A^* : \mathbb{R}^d \rightarrow \mathcal{H}$ denotes the adjoint operator of A . We write $X \sim \mathcal{E}(\mu_{\mathcal{H}}, \Gamma, \phi)$.

To study the Fisher-consistency of our Sieves approach, we need to introduce some notation in order to define the corresponding functional. Let $\otimes$ denote the tensor product in $\mathcal{H}$, i.e., for any two elements $u, v \in \mathcal{H}$ the operator $u \otimes v : \mathcal{H} \rightarrow \mathcal{H}$ is defined as $(u \otimes v)w = \langle v, w \rangle u$ for $w \in \mathcal{H}$. To simplify the presentation, assume that the location parameter $\mu_{\mathcal{H}}$ equals 0. Let $\mathbf{x} \in \mathbb{R}^p$ be the random vector defined by $\mathbf{x} = A_p X$ with $A_p : \mathcal{H} \rightarrow \mathbb{R}^p$ defined by

$$A_p = \sum_{j=1}^p \mathbf{e}_j \otimes \delta_j, \quad (\text{S.1})$$

where $\mathbf{e}_j$, $1 \leq j \leq p$, denote the elements of the canonical basis of $\mathbb{R}^p$ and $\{\delta_i\}_{i \geq 1}$ is a fixed orthonormal basis of $\mathcal{H}$. In other words, $A_p X$ consists of the p coefficients of X on the basis $\delta_1, \dots, \delta_p$. For a matrix $\mathbf{B} \in \mathbb{R}^{p \times q}$ with $\mathbf{B}^T \mathbf{B} = \mathbf{I}_q$, let $\mathcal{L}_{\mathbf{B}}$ denote the linear space spanned by its columns. As in Section 2 of the paper, define the objective function

$$\Psi_p(\mathcal{L}_{\mathbf{B}}) = \sum_{j=1}^p \sigma_{j, \mathcal{L}_{\mathbf{B}}}^2, \quad (\text{S.2})$$

where $\sigma_{j, \mathcal{L}_{\mathbf{B}}} = \sigma_{\mathbf{R}}(F_j(\mathcal{L}_{\mathbf{B}}))$ and $F_j(\mathcal{L}_{\mathbf{B}})$ denotes the distribution of the j th coordinate $r_j(\mathcal{L}_{\mathbf{B}})$ of the vector of residuals $\mathbf{r}(\mathcal{L}_{\mathbf{B}}) = \mathbf{x} - \pi(\mathbf{x}, \mathcal{L}_{\mathbf{B}}) = (\mathbf{I} - \mathbf{B}\mathbf{B}^T)\mathbf{x} = (r_1(\mathcal{L}_{\mathbf{B}}), \dots, r_p(\mathcal{L}_{\mathbf{B}}))^T$. The subscript p in the function defined in (S.2) emphasizes the fact that we have transformed the random object X into the p -dimensional random vector $\mathbf{x}$. Let $\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(q)}$ denote the columns of the matrix $\mathbf{B}$ and let $\phi_{\ell}(\mathbf{B}) \in \mathcal{H}$ be given by

$$\phi_{\ell}(\mathbf{B}) = \sum_{j=1}^p b_{\ell j} \delta_j = \left(\sum_{j=1}^p \delta_j \otimes \mathbf{e}_j \right) \mathbf{b}^{(\ell)}, \quad 1 \leq \ell \leq q. \quad (\text{S.3})$$

We denote as $\mathcal{H}_{\mathbf{B}}$ the linear space spanned by the orthonormal elements $\phi_1(\mathbf{B}), \dots, \phi_q(\mathbf{B})$.

In what follows, and without loss of generality, we will assume that $\mu_{\mathcal{H}} = 0$. Let X be an elliptical random element $X \sim \mathcal{E}(0, \Gamma, \phi)$, with Γ the self-adjoint, positive semi-definite and compact scale operator. Consider the spectral value decomposition of the scale operator $\Gamma = \sum_{j=1}^{\infty} \lambda_j \phi_j \otimes \phi_j$, where λ_j denotes the j th largest eigenvalue with associated eigenfunction ϕ_j , $j \geq 1$. The next proposition shows that, as p tends to infinity, the lowest value of $\Psi_p(\mathcal{L}_{\mathbf{B}})$ converges to $\sum_{j \geq q+1} \lambda_j$, the trace of the operator $(\mathbb{I}_{\mathcal{H}} - P)\Gamma(\mathbb{I}_{\mathcal{H}} - P)^*$ where $P = \sum_{j=1}^q \phi_j \otimes \phi_j$, and $\mathbb{I}_{\mathcal{H}}$ is the identity operator in $\mathcal{H}$. This is the infinite-dimensional counterpart of the classical optimal property of principal components for random vectors. Together with Proposition A.1 in Boente *et al.* (2012), the following result shows that the proposed estimators are Fisher-consistent for elliptically distributed random elements on a separable Hilbert space $\mathcal{H}$.

Proposition S.1.1. *Let $X \sim \mathcal{E}(0, \Gamma, \phi)$ be an elliptically distributed random element on a separable Hilbert space $\mathcal{H}$ with location 0 and positive semi-definite, self-adjoint and compact scale operator Γ . Let $\lambda_1 \geq \lambda_2 \geq \dots$ be the eigenvalues of Γ with associated eigenfunctions ϕ_j , $j \geq 1$. If $\sum_{j \geq 1} \lambda_j < \infty$ and $\lambda_q > \lambda_{q+1}$, then*

$$\lim_{p \rightarrow \infty} \min_{\mathbf{B} \in \mathbb{R}^{p \times q}, \mathbf{B}^T \mathbf{B} = \mathbf{I}_q} \Psi_p(\mathcal{L}_{\mathbf{B}}) = \sum_{j \geq q+1} \lambda_j. \quad (\text{S.4})$$

Let $\mathbf{B}_{0,p}$ be the minimizer of (S.2) over $\{\mathbf{B} \in \mathbb{R}^{p \times q}, \mathbf{B}^T \mathbf{B} = \mathbf{I}_q\}$. Then, as $p \rightarrow \infty$, the sequence of linear spaces $\mathcal{H}_{\mathbf{B}_{0,p}}$ converges to the linear space spanned by the eigenfunctions $\phi_1, \dots, \phi_q$ associated with the q largest eigenvalues of Γ .

S.2 Example: Mortality data

In this example we explore human mortality data, available on-line from the Human Mortality Database (Human Mortality Database, 2013). We restrict our attention to death rates per age group for men in France. For each year, we use the logarithm of the death rate of people between the ages of 0 and 99. Panel (a) in Figure 1 shows the

mortality curves for the years between 1816 to 2010. Dark lines correspond to years after 1945. We observe a clear difference in the pattern of male mortality curves in France

(a) Mortality data for the years 1816–2010 (b) Robust predictions for the years 1816–1948

Figure 1: Mortality data. Panel (a) contains the curves for the years 1816 to 2010. Darker lines correspond to years after 1945. Panel (b) depicts the predicted trajectories corresponding to the 2–dimensional subspace that best approximates the curves before the post–war years (1816–1948), estimated using our S –estimator. The black line is the estimated central curve.

before and after the Second World War. This phenomenon is sometimes attributed to technological advances and quality of life changes in Europe after 1945. We also note that there is a 3–year transitional period (1946–1948) in which the mortality curves lie between the two main groups. In this analysis we focus on the period 1816–1948, that includes the pre–war and the “transition” periods. The purpose of this analysis is to detect years in which the pattern of mortality across age groups is noticeably different from the majority of curves in the data. We computed an S –estimator with tuning constant $c = 3$ to find the best 2–dimensional subspace approximating these curves. We used the approach described in Section 3 of the paper with a cubic B –spline basis of dimension $p = 20$. A two–dimensional fit was also considered in Hyndman and Ullah (2007) and Hyndman and Shang (2010). Figure 1 (b) contains the estimated central curve plotted over the original data, and also over the robustly predicted curves.

As described in Section 4 of the paper, we looked for outlying curves by means of the adjusted boxplot of Hubert and Vandervieren (2008) on the squared prediction errors

using the LS and S —estimators. With the robust fit we identified the following years as atypical: 1855, 1871, 1914–1919, 1940, and 1942–1948, while the LS fit only identifies the periods 1914–1915 and 1943–1948. It is interesting to note that in 1855 France was involved in the Crimean War, and in 1871 in the Prussian War. The period 1914 to 1919 corresponds to World War I and the Spanish Flu epidemic. France falls to German occupation in 1940 and after a relatively calm year in 1941, sees more action in the period 1942 to 1944. Figure 2 contains the curves corresponding to these four events (the Crimean and Prussian Wars, and the 2 World Wars), along with the predictions resulting from the S and LS estimators. Note that the predicted curves based on the robust estimator do not fit well the mortality profiles for these atypical years, which allows us to detect them as outlying. On the other hand, it is interesting to note that the LS estimator is not able to detect the Crimean and Prussian Wars, neither the early World War II casualties in France (1940 and 1942). Both estimators properly identify the post-war years as atypical.

S.3 Proofs

S.3.1 Proof of Proposition 2.1

Recall that, if $\mathbf{x} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, its characteristic function equals $\varphi_{\mathbf{x}}(\mathbf{t}) = \exp(it^T \boldsymbol{\mu}) \phi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$.

Note that since $\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}, \phi)$, $\mathbf{z} = \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\beta}^T \mathbf{x}$ is spherically distributed, so that all its components have the same distribution G . Without loss of generality, assume that $\sigma_R(G) = 1$. Let $\mathcal{L}$ be a linear space of dimension q , with orthonormal basis $\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(q)}$. If we arrange this basis as columns of a matrix $\mathbf{B} \in \mathbb{R}^{p \times q}$ we have that $\mathbf{r}(\mathcal{L}) = (r_1(\mathcal{L}), \dots, r_p(\mathcal{L}))^T = \mathbf{x} - \pi(\mathbf{x}, \mathcal{L}) = (\mathbf{I} - \mathbf{B}\mathbf{B}^T)\mathbf{x} \sim \mathcal{E}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{L}}, \phi)$, with $\boldsymbol{\Sigma}_{\mathcal{L}} = (\mathbf{I} - \mathbf{B}\mathbf{B}^T)\boldsymbol{\Sigma}(\mathbf{I} - \mathbf{B}\mathbf{B}^T)^T = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$, with $\mathbf{C} = (\mathbf{I} - \mathbf{B}\mathbf{B}^T)$. Since $\mathbf{x} = \boldsymbol{\beta}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$, we see that $\mathbf{r}(\mathcal{L})$ can be written as $\mathbf{C}\boldsymbol{\beta}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$. Therefore, the characteristic function of $\mathbf{r}(\mathcal{L})$ is given by $\varphi_{\mathbf{r}(\mathcal{L})}(\mathbf{t}) = \varphi_{\mathbf{z}}(\boldsymbol{\Lambda}^{1/2}\boldsymbol{\beta}^T\mathbf{C}^T\mathbf{t}) = \phi(\mathbf{t}^T\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T\mathbf{t})$, where ϕ denotes the generator of the characteristic function of $\mathbf{z}$. Hence, for the j th coordinate of the vector of residuals we have $\varphi_{r_j(\mathcal{L})}(t) = \varphi_{\mathbf{r}(\mathcal{L})}(t\mathbf{e}_j) = \phi(t^2\mathbf{e}_j^T\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T\mathbf{e}_j) = \phi(t^2\mathbf{c}_j^T\boldsymbol{\Sigma}\mathbf{c}_j) = \varphi_{z_1}(t^2\mathbf{c}_j^T\boldsymbol{\Sigma}\mathbf{c}_j)$. It follows

Figure 2: Years 1855 and 1871 were detected as outliers by the robust estimator, during the Crimean and Prussian War, respectively (plots a) to d)). Curves for the years 1914–1919 (plots e) and f)) and 1940–1945 (plots g) and h) row). The curves of mortality rates for these periods, corresponding to both World Wars, were identified as outliers by the robust estimator. Solid thick gray lines correspond to the observed curve, while the black dashed lines correspond to the predicted trajectory.

that $r_j(\mathcal{L}) \sim \xi_j z_1$ where $z_1 \sim G$ and $\xi_j^2 = \mathbf{c}_j^T \Sigma \mathbf{c}_j$. This implies that $\sigma_{j,\mathcal{L}}^2 = \sigma_R^2(F_j(\mathcal{L})) = \mathbf{c}_j^T \Sigma \mathbf{c}_j$. Hence, $\sum_{j=1}^p \sigma_{j,\mathcal{L}}^2 = \sum_{j=1}^p \mathbf{c}_j^T \Sigma \mathbf{c}_j = \text{tr}(\mathbf{C} \Sigma \mathbf{C}^T)$. This last expression is minimized when $\mathbf{B} = (\beta_1, \dots, \beta_q)$ (see Seber, 1984, Theorem 5.3) and the solution is unique since $\lambda_q > \lambda_{q+1}$. $\square$

S.3.2 Proof of Proposition 2.2

Let $a_n = \sup_{\dim(\mathcal{L})=q} |\widehat{\Psi}_n(\mathcal{L}) - \Psi(\mathcal{L})|$ and note that $\widehat{\Psi}_n(\widehat{\mathcal{L}}) \leq \widehat{\Psi}_n(\mathcal{L}(P)) = \Psi(\mathcal{L}(P)) + a_n$ and similarly $\Psi(\mathcal{L}(P)) \leq \Psi(\widehat{\mathcal{L}}) \leq \widehat{\Psi}_n(\widehat{\mathcal{L}}) + a_n$. Hence $\widehat{\Psi}_n(\widehat{\mathcal{L}}) \geq \Psi(\mathcal{L}(P)) - a_n$ and we obtain $\widehat{\Psi}_n(\widehat{\mathcal{L}}) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \Psi(\mathcal{L}(P))$ and $\Psi(\widehat{\mathcal{L}}) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \Psi(\mathcal{L}(P))$. Standard arguments now imply the convergence of the linear spaces since $\mathcal{L}(P)$ is unique. Hence, one can choose an orthonormal basis of $\widehat{\mathcal{L}}$ converging with probability one to a basis of $\mathcal{L}(P)$. $\square$

S.3.3 Proof of Proposition S.1.1

To illustrate the main idea of the proof, we start with the (easy) case where the orthonormal basis $\{\delta_j\}$ is the basis ϕ_j of eigenfunctions of Γ . Assume that $m = m_n$ is such that $m_n > q$ and $\{\phi_1, \dots, \phi_q\} \subset \{\delta_1, \delta_2, \dots, \delta_{m_n}\}$. Without loss of generality, assume that $\delta_j = \phi_j$, for $1 \leq j \leq q$ and that $\delta_j = \phi_{\ell_j}$ for $q+1 \leq j \leq m_n$ with $q < \ell_{q+1} < \dots < \ell_{m_n}$. Then, $\mathbf{x} = A\mathbf{X} \sim \mathcal{E}_p(\mathbf{0}, \Sigma, \phi)$ where A is defined in (S.1) and $\Sigma = A\Gamma A^* = \text{diag}(\lambda_1, \dots, \lambda_q, \lambda_{\ell_{q+1}}, \dots, \lambda_{\ell_{m_n}})$ where $\lambda_q > \lambda_{\ell_{q+1}} > \dots > \lambda_{\ell_{m_n}}$. Then, using Proposition 2.1, for any $\mathbf{B} \in \mathbb{R}^{m \times q}$ such that $\mathbf{B}^T \mathbf{B} = \mathbf{I}_q$, we have $\Psi_m(\mathcal{L}_{\mathbf{B}}) = \sum_{j=1}^m \sigma_{j,\mathcal{L}_{\mathbf{B}}}^2 \geq \sum_{j=1}^m \sigma_{j,\mathcal{L}_{\mathbf{B}_{0,m}}}^2 = \sum_{s=q+1}^m \lambda_{\ell_s}$, where $\mathbf{B}_{0,m} = (\mathbf{e}_1, \dots, \mathbf{e}_q)$. Hence, using that $\lim_{m \rightarrow \infty} \sum_{s=q+1}^m \lambda_{\ell_s} = \sum_{s \geq q+1} \lambda_s = \text{tr}(\Gamma) - \sum_{j=1}^q \lambda_j = \text{tr}((\mathbb{I}_{\mathcal{H}} - P)\Gamma(\mathbb{I}_{\mathcal{H}} - P)^*)$. Note that, in this case, $\phi_j(\mathbf{B}_{0,m}) = \phi_j$, where $\phi_j(\mathbf{B})$ is defined in (S.3). Hence, $\mathcal{H}_{\mathbf{B}_{0,m}}$ is the linear space spanned by $\phi_1, \dots, \phi_q$, which shows Fisher-consistency.

Let us now consider the general situation. As before, we have $\mathbf{x} = A\mathbf{X} \sim \mathcal{E}_p(\mathbf{0}, \Sigma, \phi)$ where A is defined in (S.1) and $\Sigma = A\Gamma A^*$. Recall that $A^* = \sum_{j=1}^m \delta_j \otimes \mathbf{e}_j$, so that for any $\mathbf{y} \in \mathbb{R}^m$, $A^* \mathbf{y} = \sum_{j=1}^m y_j \delta_j$. Let $\mathcal{H}_m$ be the linear subspace spanned by $\{\delta_1, \dots, \delta_m\}$ and $\Pi_m : \mathcal{H} \rightarrow \mathcal{H}_m$ be the projection operator over $\mathcal{H}_m$, that is, $\Pi_m = \sum_{j=1}^m \delta_j \otimes \delta_j$. We

have that Π_m is self-adjoint and $\Pi_m \nu = \nu$ for $\nu \in \mathcal{H}_m$. Moreover, $\Pi_m \rightarrow \mathbb{I}_{\mathcal{H}}$ in the strong operator topology, where $\mathbb{I}_{\mathcal{H}}$ is the identity operator in $\mathcal{H}$, that is, $\Pi_m x \rightarrow x$ for any $x \in \mathcal{H}$, as $m \rightarrow \infty$. It follows that for any compact operator Υ , $\Pi_m \Upsilon \rightarrow \Upsilon$ as $m \rightarrow \infty$ in the norm operator topology.

It is easy to show that, if $\mathbf{u} \in \mathbb{R}^m$ is an eigenvector of Σ related to an eigenvalue α , then $\nu = A^* \mathbf{u}$ is an eigenfunction of $\Upsilon_m = \Pi_m \Gamma \Pi_m^*$ associated to α . Similarly, if ν is an eigenfunction of Υ_m with eigenvalue α , then $A\nu$ is an eigenvector of Σ with the same eigenvalue α . Hence, the m -largest eigenvalues of Υ_m are those of Σ with the relation among eigenvectors and eigenfunctions just described. Note that since the range of Υ_m is m , Υ_m has at most m non-null eigenvalues. Let $\mathbf{B}_{0,m} \in \mathbb{R}^{m \times q}$ be a matrix containing the eigenvectors of Σ related to its m largest eigenvalues as columns. In other words, $\mathbf{B}_{0,m} = (\beta_1, \dots, \beta_q)$ where β_j is the eigenvector of Σ related to its j th largest eigenvalue denoted α_j . Then, $\alpha_j = \lambda_j(\Upsilon_m)$, where $\lambda_j(\Upsilon)$ denotes the j th largest eigenvalue of the operator Υ .

Using Proposition 2.1, we get that, for any $\mathbf{B} \in \mathbb{R}^{m \times q}$ such that $\mathbf{B}^T \mathbf{B} = \mathbf{I}_q$, $\Psi_m(\mathcal{L}_{\mathbf{B}}) = \sum_{j=1}^m \sigma_{j,\mathcal{L}_{\mathbf{B}}}^2 \geq \sum_{j=1}^m \sigma_{j,\mathcal{L}_{\mathbf{B}_{0,m}}}^2 = \sum_{s=q+1}^m \alpha_s = \sum_{s=q+1}^m \lambda_s(\Upsilon_m)$. Noting that $\text{tr}(\Sigma) = \text{tr}(\Upsilon_m) = \sum_{s=1}^m \lambda_s(\Upsilon_m)$, we obtain the bound $\Psi_m(\mathcal{L}_{\mathbf{B}}) \geq \Psi_m(\mathcal{L}_{\mathbf{B}_{0,m}}) = \text{tr}(\Upsilon_m) - \sum_{s=1}^q \lambda_s(\Upsilon_m)$, that is

$$\min_{\mathbf{B} \in \mathbb{R}^{m \times q}, \mathbf{B}^T \mathbf{B} = \mathbf{I}_q} \Psi_m(\mathcal{L}_{\mathbf{B}}) = \Psi_m(\mathcal{L}_{\mathbf{B}_{0,m}}) = \text{tr}(\Upsilon_m) - \sum_{s=1}^q \lambda_s(\Upsilon_m). \quad (\text{S.5})$$

As noted above, we have $\|\Upsilon_m - \Gamma\| \rightarrow 0$ as $m \rightarrow \infty$. By the continuity of the eigenvalues with respect to the operators norm (see for instance, Osborn, 1975), we have that, for each fixed k , $\lambda_k(\Upsilon_m) \rightarrow \lambda_k(\Gamma) = \lambda_k$ as $m \rightarrow \infty$. Hence, $\lim_{m \rightarrow \infty} \sum_{s=1}^q \lambda_s(\Upsilon_m) = \sum_{s=1}^q \lambda_s$.

It remains to show that $\lim_{m \rightarrow \infty} \text{tr}(\Upsilon_m) = \text{tr}(\Gamma)$. First note that, Proposition A.1 in Boente *et al.* (2012) shows that $\lambda_k(\Upsilon_m) \leq \lambda_k$, hence $\text{tr}(\Upsilon_m) = \sum_{s=1}^m \lambda_s(\Upsilon_m) \leq \sum_{s=1}^m \lambda_s \leq \text{tr}(\Gamma)$. Therefore, we only have to show that, for any $\epsilon > 0$, there exists m_0 such that for $m \geq m_0$, we have $\text{tr}(\Upsilon_m) \geq \text{tr}(\Gamma) - \epsilon$. Since $\text{tr}(\Gamma) < \infty$, there exists $N \in \mathbb{N}$ such that $N > q$ and $0 \leq \text{tr}(\Gamma) - \sum_{j=1}^N \lambda_j < \epsilon/2$. Using that $\lim_{m \rightarrow \infty} \sum_{j=1}^N \lambda_j(\Upsilon) = \sum_{j=1}^N \lambda_j$, choose m_0 such that for $m \geq m_0$, $|\sum_{j=1}^N \lambda_j(\Upsilon_m) - \sum_{j=1}^N \lambda_j| \leq \epsilon/2$. Now, for $m \geq$

$\max\{m_0, N\}$ we have $\text{tr}(\Upsilon_m) = \sum_{j=1}^m \lambda_j(\Upsilon_m) \geq \sum_{j=1}^N \lambda_j(\Upsilon_m) \geq \sum_{j=1}^N \lambda_j - \epsilon/2 \geq \text{tr}(\Gamma) - \epsilon$, as desired. Hence, from (S.5), we have that $\lim_{m \rightarrow \infty} \min_{\mathbf{B} \in \mathbb{R}^{m \times q}, \mathbf{B}^T \mathbf{B} = \mathbf{I}_q} \Psi_m(\mathcal{L}_{\mathbf{B}}) = \lim_{m \rightarrow \infty} \Psi_m(\mathcal{L}_{\mathbf{B}_{0,m}}) = \text{tr}(\Gamma) - \sum_{s=1}^q \lambda_s$, concluding the proof of (S.4). Finally, note that the linear space $\mathcal{H}_{\mathbf{B}_{0,m}}$ is spanned by $\phi_1(\mathbf{B}_{0,m}), \dots, \phi_q(\mathbf{B}_{0,m})$, where $\phi_j(\mathbf{B}_{0,m}) = A^* \beta_j$. Then, we have $\phi_j(\mathbf{B}_{0,m}) = \phi_j(\Upsilon_m)$. Using again that, $\|\Upsilon_m - \Gamma\| \rightarrow 0$ as $m \rightarrow \infty$ and the fact that $\lambda_q > \lambda_{q+1}$, we see that the linear space spanned by $\phi_1(\Upsilon_m), \dots, \phi_q(\Upsilon_m)$ converges to that spanned by $\phi_1, \dots, \phi_q$, (see for instance, Osborn 1975 or Dunford and Schwartz, 1963), concluding the proof. $\square$

References

- Bali, J. L. and Boente, G. (2009). “Principal points and elliptical distributions from the multivariate setting to the functional case”. *Statist. Probab. Lett.*, **79**, 1858-1865.
- Boente, G., Salibian-Barrera, M. and Tyler, D. (2012). “A characterization of elliptical distributions and some optimality properties of principal components for functional data”. Technical report. Available at <http://www.stat.ubc.ca/~matias/pubs.html>
- Dunford, N. and Schwartz, J. (1963). *Linear Operators. II: Spectral Theory, Selfadjoint operators in Hilbert spaces*. Interscience, New York.
- Hubert, M. and Vandervieren, E. (2008). “An adjusted boxplot for skewed distributions”. *Computational Statistics & Data Analysis*, **52**, 5186-5201.
- Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 8 Feb 2013).
- Hyndman, R. J. and Ullah, S. (2007). “Robust forecasting of mortality and fertility rates: A functional data approach”. *Computational Statistics and Data Analysis*, **51**, 4942-4956.
- Hyndman, R. J. and Shang, H.L. (2010) “Rainbow plots, bagplots, and boxplots for functional data”. *Journal of Computational and Graphical Statistics*, **19**, 29-45.

Osborn, J. (1975). “Spectral approximation for compact operators”. *Mathematics of Computation*, **29**, 712-725.

Seber, G. (1984). *Multivariate Observations*. Wiley, New York.