ISSN: 1017-060X (Print)

ISSN: 1735-8515 (Online)

Bulletin of the

Iranian Mathematical Society

Vol. 43 (2017), No. 2, pp. 467–475

Title:

On convergence of sample and population Hilbertian functional principal components

Author(s):

A. R. Soltani, A. R. Nematollahi and R. Nasirzadeh

Published by Iranian Mathematical Society
http://bims.ims.ir

Bull. Iranian Math. Soc.
Vol. 43 (2017), No. 2, pp. 467–475
Online ISSN: 1735-8515

ON CONVERGENCE OF SAMPLE AND POPULATION HILBERTIAN FUNCTIONAL PRINCIPAL COMPONENTS

A. R. SOLTANI*, A. R. NEMATOLLAHI AND R. NASIRZADEH

(Communicated by Hamid Pezeshk)

Abstract. In this article we consider the sequences of sample and population covariance operators for a sequence of arrays of Hilbertian random elements. Then, under the assumptions that sequences of the covariance operators norm are uniformly bounded and the sequences of the principal component scores are uniformly summable, we prove that the convergence of the sequences of covariance operators would imply the convergence of the corresponding sequences of the sample and population eigenvalues and eigenvectors, and vice versa. In particular we prove that the principal component scores converge in distribution in certain family of Hilbertian elliptically contoured distributions.
Keywords: Hilbertian random elements, functional data analysis, functional principal component analysis, covariance operators, operator convergence.
MSC(2010): Primary: 46S50; Secondary: 62H25, 46A32.

1. Introduction

We let $\mathcal{H}$ denote a (real) separable Hilbert space. We consider a sequence of arrays of Hilbertian, $\mathcal{H}$-valued, random elements $\mathbb{X}(\ell) = [\mathbf{X}_1(\ell), \dots, \mathbf{X}_n(\ell)]$, $\ell = 1, 2, 3, ..., n$. We also let $\mathbf{C}_{\mathbb{X}(\ell)}$ and $\hat{\mathbf{C}}_{\mathbb{X}(\ell)}$, $\ell = 1, 2, 3, ...$ denote the corresponding sequences of the population and sample covariance operators. In this article we provide sufficient conditions under which the convergence in norm of the sample and the population covariance operators gives rise to the convergence of the corresponding sequences of the principal values and principal components, these terms are defined in the next section.

In theory, the convergence of eigenvalues and eigenvectors of sequences of operators is addressed in functional analysis and perturbation theory. Kato in [9] gave a counter example exhibiting that weak convergence of the bounded

Article electronically published on 30 April, 2017.
Received: 4 July 2015, Accepted: 1 December 2015.
*Corresponding author.

operators does not necessarily leads to the convergence of the corresponding eigenvalues and eigenvectors. An interesting example was also given in [2] for sequences of variance-covariance matrices, see also [15]. In the infinite dimensional case this issue is more complicated. There is also variety in the type of convergence.

In application, specially in dealing with longitudinal data and repeated measurement, the convergence of sequences of certain statistics appear to be informative in [1]. The sequences of arrays of Hilbertian random elements are used to model repeated measurements functional data. The index ℓ stands for the number of measurements in time, while n stands for the number of subjects. The convergence of the eigenvalues and eigenvectors shows eventual consistency in direction and magnitude of variation in data. This is important in analysis of repeated measurements.

In this article we mostly consider the theoretical aspect of the problem. The key tool in our analysis is the Bosq inequalities concerning the norm deviations between corresponding eigenvalues and eigenvectors of two nuclear operators, Lemma 3.1 and Lemma 3.2, and the following key assumptions that are introduced in this article.
Let $\{\mathbf{C}_\ell\}$ be a sequence of nuclear operators, $\|\mathbf{C}_\ell\|_{\mathbb{N}}$ the nuclear norm, and $\{\lambda_j(\ell)\}_j$,
$\ell = 0, 1, 2, ...$ the corresponding sequences of eigenvalues.

Assumption (A): The sequence $\{\|\mathbf{C}_\ell\|_{\mathbb{N}}\}_{\ell\geq 0}$ is uniformly bounded, i.e., there is a constant M such that

$$\| \mathbf{C}_\ell \| < M, \text{ for } \ell = 1, 2, ...$$

Assumption (B): The sequences of eigenvalues $\{\lambda_j(\ell)\}_j,\ \ell = 1, 2, 3, ...$ are uniformly sumable, i.e., for given $\epsilon > 0$ there is N depending only on ϵ such that

$$\sum_{j=N}^{\infty} |\lambda_j(\ell)| < \epsilon, \text{ for every } \ell \geq 1.$$

Then we prove that the under conditions (A) and (B), $\{\mathbf{C}_\ell\} \to \mathbf{C}_0$ weakly if and only if the convergence takes place in every pair of the corresponding eigenvalue eigenvectors. More details are given in Sections 3 & 4.
The Condition (B) is of practical interest. It concerns two important issues in the principal component analysis in repeated measurements, PCA in abbreviation, namely, reduction of dimension and measurements repeats. This issue is even more important in FDA. Condition (B) states that among infinite independent factors that produce a random element, only finite number of them are significant, and the number of significant factors remains to be uniformly the same when the number of repeated measurements exceeds certain level for each subject. Customary, among other techniques, the Fraction of Variation

Explained (FVE) is a reliable statistical method to detect significant underlying factors in a single Hilbertian random element. To be more precise, FVE states that when the eigenvalues are in decreasing order, then the minimum number m satisfying $\sum_{i=1}^{m} \lambda_i / \sum_{i=1}^{\infty} \lambda_i \geq \alpha$, gives the sufficient number of significant factors, where α is a positive number less than but close to one. Equivalently,

$$\frac{\sum_{i=m+1}^{\infty} \lambda_i(\ell)}{\sum_{i=1}^{\infty} \lambda_i(\ell)} < 1-\alpha, \quad \ell = 1, 2, 3, ..., \tag{1.1}$$

which, under (A), is equivalent to (B).

As an application of our derivations, We deduce that the principal component scores converge in distribution in certain family of Hilbertian elliptically contoured distributions, discussed in [10].

The literature on principal values, principal components and their applications is quite rich. For more information and related references, we refer the readers to [4, 5, 8, 11–13].

This article is organized as follows. In Section 2, we provide basic notations and preliminaries. In Section 3, we consider the convergence of the sample eigenvalues and eigenvectors. In Section 4, we provide sufficient conditions for the continuity of the population eigenvalues and eigenvectors. We conclude this section by proving that the PC scores of sequences of certain Hilbertian elliptically contoured distributed random elements converge in distribution, Theorem 4.4.

2. Notations and preliminary

We let $(\Omega, \mathcal{A}, P)$ denote a probability space and $\mathcal{H}$ denote a (real) separable Hilbert space. A random element $\mathbf{X}$ is a Borel measurable mapping defined on $(\Omega, \mathcal{A}, P)$ taking its values in $\mathcal{H}$. We say a random element is weakly second order if $E\left|\langle \mathbf{X}, \mathbf{x}\rangle\right|^2 < \infty$, for every $\mathbf{x} \in \mathcal{H}$, where $\langle ., . \rangle$ is the inner product of $\mathcal{H}$, and strongly second order if $E\left\|\mathbf{X}\right\|^2 < \infty$, where $\left\|\mathbf{x}\right\|^2 = \langle \mathbf{x}, \mathbf{x}\rangle$ for all $x \in \mathcal{H}$.

We let $\mathbb{N}, \mathbb{HS}$ and $\mathbb{B}$ stand for nuclear, Hilbert Schmidt and bounded linear operators on $\mathcal{H}$, equipped with the corresponding norms, respectively.

It is well known that a nuclear operator $\mathbf{C}$ on $\mathcal{H}$ admits the spectral representation $\mathbf{C}(\mathbf{x}) = \sum_{j=1}^{\infty} \lambda_j \langle \mathbf{x}, \mathbf{e}_j\rangle \mathbf{e}_j$, $\mathbf{x} \in \mathcal{H}$, where $(\mathbf{e}_j)$ is an orthonormal bases for $\mathcal{H}$, called eigenvectors of $\mathbf{C}$, and (λ_j) is a sequence of real numbers, called eigenvalues of $\mathbf{C}$, such that $\sum_{j=1}^{\infty} |\lambda_j| < \infty$. The nuclear norm of $\mathbf{C}$ is $\|\mathbf{C}\|_{\mathbb{N}} = \sum_{j=1}^{\infty} |\lambda_j|$. If $\mathbf{C}$ is self adjoint and positive, i.e., $\langle \mathbf{C}(\mathbf{x}), \mathbf{x}\rangle > 0$, then $\lambda_j > 0$ for every j . A Hilbert Schmidt operator $\mathbf{C}$ admits similar spectral representation, but the coefficients λ_j are square sumable; then $\|\mathbf{C}\|_{\mathbb{HS}} = \sum_{j=1}^{\infty} \lambda_j^2 < \infty$. It is well known that $\mathbb{N} \subset \mathbb{HS}$, Gelfand and Vilenkin [7].

For an $\mathcal{H}$-valued random element $\mathbf{X}$, the population covariance operator $\mathbf{C_X}$ is defined by

$$\mathbf{C_X}(\mathbf{x}) = E(\mathbf{X} \otimes \mathbf{X})(\mathbf{x}) = E\langle \mathbf{X}, \mathbf{x} \rangle \mathbf{X}, \qquad \mathbf{x} \in \mathcal{H}. \tag{2.1}$$

A population covariance operator is nuclear and positive; $\mathbf{C_X} = \sum_{j=1}^{\infty} \lambda_j \pi\left[\mathbf{e}_j\right]$, $\pi\left[\mathbf{e}\right] : \mathcal{H} \to \{\alpha \mathbf{e},\ \alpha \in \mathbb{R}\}$ is the one *dimensional projection* on $\mathcal{H}$ onto the one dimensional subspace generated by $\mathbf{e}$. For a random element $\mathbf{X}$, the expectation is defined in the sense of Bochner, $E\mathbf{X} = \int \mathbf{X} dP$.

Throughout this paper, we assume that $\mathbf{X}$ is a centered $\mathcal{H}$-valued strongly second order random element and any finite number of random elements $\mathbf{X}_1, \ldots,$ $\mathbf{X}_n$ are linearly independent with probability one. This will be true if $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent.

If $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is a finite set of random elements, in particular a random sample for a random element $\mathbf{X}$, a natural estimator for the covariance operator $\mathbf{C_X}$ is the so-called empirical covariance operator denoted by $\widehat{\mathbf{C}}$ and is given by

$$\widehat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \otimes \ \mathbf{X}_i. \tag{2.2}$$

The sample covariance operator is finite with probability one, w.p.1 in abbreviation, in the sense that its range has finite dimension. Under the linear independence assumption, the dimension of the range, $R\left[\mathbf{X}_1, \ldots, \mathbf{X}_n\right] =$ spanclosure$[\mathbf{X_1}, \ldots, \mathbf{X_n}]$ in $\mathcal{H}$, is constant w.p.1; $\widehat{\mathbf{C}} : \mathcal{H} \to R\left[\mathbf{X}_1, \ldots, \mathbf{X}_n\right]$. The sample covariance operator $\widehat{\mathbf{C}}$ admits the representation

$$\widehat{\mathbf{C}} = \sum_{j=1}^{n} \hat{\lambda}_j \pi\left[\hat{\mathbf{e}}_j\right]. \tag{2.3}$$

Interesting, random elements $\hat{\mathbf{e}}_j$, $j = 1, ..., n$ form a basis for $R\left[\mathbf{X}_1, \ldots, \mathbf{X}_n\right]$; moreover $\hat{\lambda}_j$ are distinct w.p.1.

Let us denote $\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)$, $\ell = 1, 2, ..., L$, repeated samples as in repeated measurements or in analysis of variance; L, number of measurements, could be finite or infinite. For fix ℓ, $\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)$ are readings for n subjects; for fix j, $\mathbf{X}_j(1), \ldots, \mathbf{X}_j(\ell)$ are repeated measurements for the subject j.

In the next section we investigate the convergence of $\widehat{\mathbf{C}}_\ell$ (introduced in Theorem 3), as ℓ increases. Upon the convergence of $\widehat{\mathbf{C}}_\ell$, repeats of measurements become less informative in time, and the experimenter may decide stop further sampling at a certain time epoch ℓ.

3. Sample covariance behavior

In this section we provide proofs for the convergence of sequence of the sample covariance operator and corresponding principal values and principal components. Let us first provide the following two basic approximation results that are of independent interests, as well.

Lemma 3.1 ([4, p. 103])**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two compact linear operators on* $\mathcal{H}$ *with spectral decompositions*

$$\begin{aligned}\mathbf{A} &= \sum_{j=1}^{\infty} \lambda_j(A)\mathbf{e}_j(A) \otimes \mathbf{e}_j(A);\\ \mathbf{B} &= \sum_{j=1}^{\infty} \lambda_j(B)\mathbf{e}_j(B) \otimes \mathbf{e}_j(B).\end{aligned}$$

Then we have

$$|\lambda_j(B) - \lambda_j(A)| \leq \|\mathbf{A} - \mathbf{B}\|, \ j \geq 1.$$

Bosq [4] gives the lemma for the population covariances operators. It is indeed true for positive nuclear operators. The proof is similar to the proof of Bosq [2, P. 103].

Lemma 3.2. *If* $(\mathbf{e}_j(A), \lambda_j(A))$ *and* $(\mathbf{e}_j(B), \lambda_j(B))$ *are corresponding pairs of eigenvectors and eigenvalues of* $\mathbf{A}$ *and* $\mathbf{B}$ *as in Lemma 3.1, respectively, and if* $\mathbf{e}_j(A)$ *and* $\mathbf{e}_j(B)$ *are one dimensional, then*

$$\|\mathbf{e}_j(B) - \mathbf{e}_j(A)\| \leq \alpha_j(A, B) \|\mathbf{A} - \mathbf{B}\|, \quad j \geq 1,$$

where

$$\alpha_j(A) = 2\sqrt{2} \max\left[(\lambda_{j-1}(A) - \lambda_j(A))^{-1}, (\lambda_j(A) - \lambda_{j+1}(A))^{-1}\right], \quad j \geq 2,$$

$$\alpha_j(B) = 2\sqrt{2} \max\left[(\lambda_{j-1}(B) - \lambda_j(B))^{-1}, (\lambda_j(B) - \lambda_{j+1}(B))^{-1}\right], \ j \geq 2,$$

and

$$\begin{gathered}\alpha_1(A) = 2\sqrt{2}\,(\lambda_1(A) - \lambda_2(A))^{-1},\\ \alpha_1(B) = 2\sqrt{2}\,(\lambda_1(B) - \lambda_2(B))^{-1},\\ \alpha_j(A, B) = \max\left[\alpha_j(A), \alpha_j(B)\right], \ j \geq 1.\end{gathered}$$

The following theorem is the main result of this section that shows the convergence for the sample eigenvalues and eigenvectors is induced from the convergence of the sample covariance operators and vice versa.

Theorem 3.3. *Let* $\mathbb{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$ *and* $\mathbb{X}(\ell) = [\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)]$*, and let* $\widehat{\mathbf{C}}$ *and* $\widehat{\mathbf{C}}_\ell$ *denote corresponding sample covariance operators, respectively,* $\ell = 1, 2, \ldots L$*. Also let* $\left\{\widehat{\lambda}_j, \hat{\mathbf{e}}_j\right\}$ *and* $\left\{\widehat{\lambda}_j(\ell), \hat{\mathbf{e}}_j(\ell)\right\}$ *be the corresponding {eigenvalue, eigenvector} pairs for* $\widehat{\mathbf{C}}$ *and* $\widehat{\mathbf{C}}_\ell$*, respectively. Then the followings are satisfied:*

(i) *If* $\mathbb{X}(\ell)$ *converges to* $\mathbb{X}$ *in norm w.p.1, then* $\widehat{\mathbf{C}}_\ell$ *converges to* $\widehat{\mathbf{C}}$ *in norm w.p.1.*

(ii) *If* $\widehat{\mathbf{C}}_\ell$ *converges to* $\widehat{\mathbf{C}}$ *in norm w.p.1, then* $\widehat{\lambda}_j(\ell) \to \widehat{\lambda}_j$*; if* $\widehat{\lambda}_j$*'s are distinct then* $\hat{\mathbf{e}}_j(\ell) \to \hat{\mathbf{e}}_j$ *in norm w.p.1.*

(iii) *Assume for* $j = 1, ..., n$, $\widehat{\lambda}_j(\ell) \to \widehat{\lambda}_j$ *and* $\hat{\mathbf{e}}_j(\ell) \to \hat{\mathbf{e}}_j$ *in norm w.p.1, as* $\ell \to L$*, and every* $\widehat{\lambda}_j > 0$ *w.p.1. Then* $\widehat{\mathbf{C}}_\ell$ *converges to a nuclear operator, say* $\widehat{\mathbf{C}}$*, in norm, w.p.1. which is a sample covariance operator with eigenvalue, eigenvector pairs* $\left\{\widehat{\lambda}_j, \hat{\mathbf{e}}_j\right\}$.

Proof. (i) This is immediate due to the fact that the projection operator $\pi[\mathbf{x}]$ is continuous in $\mathbf{x}$. (ii) Every empirical covariance operator is a compact linear operator. It follows from Lemma 3.1 that

$$\left|\widehat{\lambda}_j(l) - \widehat{\lambda}_j\right| \le \left\|\widehat{\mathbf{C}}_l - \widehat{\mathbf{C}}\right\|; \quad j \ge 1, \tag{3.1}$$

giving that the convergence of eigenvalues follows from the convergence of the corresponding covariance operators. If the eigenvalues of $\widehat{\mathbf{C}}$ are distinct w.p.1, then the eigenspace for each eigenvector will be one dimensional. Therefore by using Lemma 3.2, we obtain

$$\|\hat{\mathbf{e}}_j(l) - \hat{\mathbf{e}}_j\| \le \alpha_j \left\|\widehat{\mathbf{C}}_l - \widehat{\mathbf{C}}\right\|; \quad j \ge 1, \tag{3.2}$$

where

$$\begin{aligned}
\alpha_j &= 2\sqrt{2}\max\left[\left(\widehat{\lambda}_{j-1} - \widehat{\lambda}_j\right)^{-1}, \left(\widehat{\lambda}_j - \widehat{\lambda}_{j+1}\right)^{-1}\right]; \quad j \ge 2 \\
\alpha_1 &= 2\sqrt{2}\left(\widehat{\lambda}_1 - \widehat{\lambda}_2\right)^{-1}.
\end{aligned}$$

But $\widehat{\mathbf{C}}$ has finite number of distinct eigenvalues, therefore α_j is bounded from below. Consequently, the convergence of the covariance operators implies the convergence of the corresponding eigenvectors. (iii) $\{\widehat{\mathbf{e}}_j\}$ is an orthonormal basis. Indeed $\delta_{jk} = \langle \widehat{\mathbf{e}}_j(l), \widehat{\mathbf{e}}_k(l)\rangle \to \langle \widehat{\mathbf{e}}_j, \widehat{\mathbf{e}}_k\rangle$ as l converges to L. Therefore $\langle \widehat{\mathbf{e}}_j, \widehat{\mathbf{e}}_k\rangle = \delta_{jk}$. On the other hand $\widehat{\mathbf{C}}_l \to \widehat{\mathbf{C}}$ in norm, because

$$\begin{aligned}
\left\|\widehat{\mathbf{C}}_l(\mathbf{x}) - \widehat{\mathbf{C}}(\mathbf{x})\right\|^2 &= \left\|\sum_{j=1}^n \widehat{\lambda}_j(l)\langle \mathbf{x}, \widehat{\mathbf{e}}_j(l)\rangle \widehat{\mathbf{e}}_j(l) - \sum_{j=1}^n \widehat{\lambda}_j \langle \mathbf{x}, \widehat{\mathbf{e}}_j\rangle \widehat{\mathbf{e}}_j\right\|^2 \\
&= \sum_{i=1}^n \widehat{\lambda}_i^2 \langle \mathbf{x}, \widehat{\mathbf{e}}_i\rangle^2 + \sum_{i=1}^n \widehat{\lambda}_i^2(l)\langle \mathbf{x}, \widehat{\mathbf{e}}_i(l)\rangle^2 \\
&\quad - \sum_{j=1}^n \sum_{i=1}^n \widehat{\lambda}_j(l)\widehat{\lambda}_i \langle \mathbf{x}, \widehat{\mathbf{e}}_j(l)\rangle \langle \widehat{\mathbf{e}}_i, \mathbf{x}\rangle \langle \widehat{\mathbf{e}}_j(l), \widehat{\mathbf{e}}_i\rangle \\
&\quad - \sum_{j=1}^n \sum_{i=1}^n \widehat{\lambda}_j(l)\widehat{\lambda}_i \langle \mathbf{x}, \widehat{\mathbf{e}}_i\rangle \langle \mathbf{x}, \widehat{\mathbf{e}}_j(l)\rangle \langle \widehat{\mathbf{e}}_i, \widehat{\mathbf{e}}_j(l)\rangle \\
&\to 0.
\end{aligned}$$

Since $\widehat{\mathbf{C}}_l - \widehat{\mathbf{C}}$ is of finite dimension, it follows that $\left\| \widehat{\mathbf{C}}_l - \widehat{\mathbf{C}} \right\| \to 0$ as well. Note that the resulting operator $\widehat{\mathbf{C}}$ is symmetric and positive, so $\widehat{\mathbf{C}}$ is a covariance operator, Gelfand and Vilenkin [(1964), Theorem 10 page 50]. □

4. **Population covariance behavior**

In this section we investigate whether the convergence of the sequence of population covariance operators implies the convergence of the corresponding sequences of the eigenvalues and eigenvectors. The following theorem is the main result of this section.

Theorem 4.1. *Let $\mathbf{X}$ and $\mathbf{X}(\ell)$, $l = 1, 2, ...$ be random elements in $\mathcal{H}$. Also let $\mathbf{C}$ and $\mathbf{C}_\ell$ denote the corresponding covariance operators, respectively, and $\{\lambda_j, \mathbf{e}_j\}$ and $\{\lambda_j(\ell), \mathbf{e}_j(\ell)\}$ be the corresponding {eigenvalue, eigenvector} pairs for $\mathbf{C}$ and $\mathbf{C}_\ell$. Then the followings are satisfied*

(i) Assume the assumption (A), given in the Introduction, is satisfied and $\mathbf{C}_\ell$ converges in norm to $\mathbf{C}$. Then $\mathbf{C}$ is a covariance operator; moreover if $\lambda_j(\ell) \to \lambda_j$, as $\ell \to L$, and if $\lambda_j s$ are distinct, then $\mathbf{e}_j(\ell) \to \mathbf{e}_j$ in norm.

(ii) Assume the assumption (B) is satisfied, and $\lambda_j(\ell) \to \lambda_j > 0$, and $\mathbf{e}_j(\ell) \to \mathbf{e}_j$ in norm. Then $\mathbf{C} = \sum_j \lambda_j \, \boldsymbol{\pi}[\mathbf{e}_j]$ will be a covariance operator, and $\mathbf{C}_\ell$ converges to $\mathbf{C}$ weakly.

Proof. (i) Let $\mathbf{C} = \sum_j \lambda_j \boldsymbol{\pi}[\mathbf{e}_j]$. Then the resulting operator $\mathbf{C}$ is symmetric and positive, therefore by Gelfand and Vilenkin [7, Theorem 10 page 50], $\mathbf{C}$ is a covariance operator. Every empirical covariance operator is a compact linear operator. It follows from Lemma 3.1 that

$$|\lambda_j(\ell) - \lambda_j| \le \|\mathbf{C}_\ell - \mathbf{C}\| \, ; \quad j \ge 1, \tag{4.1}$$

giving the result. For the eigenvectors, by Lemma 3.2,

$$\|\mathbf{e}_j(\ell) - \mathbf{e}_j\| \le \alpha_j \, \|\mathbf{C}_\ell - \mathbf{C}\| \, ; \quad j \ge 1, \tag{4.2}$$

where

$$\begin{aligned} \alpha_j &= 2\sqrt{2} \max \left[(\lambda_{j-1} - \lambda_j)^{-1}, (\lambda_j - \lambda_{j+1})^{-1} \right]; \quad j \ge 2 \\ \alpha_1 &= 2\sqrt{2} \, (\lambda_1 - \lambda_2)^{-1}. \end{aligned}$$

If the eigenvalues of $\mathbf{C}$ are distinct, then $\alpha_1 > 0$ and $\alpha_j > 0$. Therefore the convergence of operators will imply the convergence of corresponding eigenvectors.
(ii) Under the assumption (B), by using a classical argument, it can be proved that $\sum_{j=1}^{\infty} \lambda_j(\ell) \to \sum_{j=1}^{\infty} \lambda_j$ and $\sum_{j=1}^{\infty} \lambda_j < \infty$. Also since every $\{\mathbf{e}_j(\ell), j = 1, 2, ...\}$ is orthonormal, the $\{\mathbf{e}_j, j = 1, 2, ...\}$ is orthonormal as well. Thus $\mathbf{C} = \sum \lambda_j \pi \, [\mathbf{e}_j]$

is a covariance operator. We proceed on to show $\mathbf{C}_\ell \to \mathbf{C}$ weakly. Fix $f \in \mathcal{H}$, then $\mathbf{C}_\ell f = \sum_{j=1} \lambda_j(\ell)\langle f, \mathbf{e}_j(\ell)\rangle \mathbf{e}_j(\ell)$. Since

$$\|\lambda_j(\ell)\langle f, \mathbf{e}_j(\ell)\rangle \mathbf{e}_j(\ell)\| \leq \|f\|\,\lambda_j(\ell); \quad j, \ell = 1, 2, ...$$

it follows that under (B), $\{\mathbf{x}_j(\ell)\}_j$, $\ell = 1, 2, ...$, where $\mathbf{x}_j(\ell) = \lambda_j(\ell)\langle f, \mathbf{e}_j(\ell)\rangle \mathbf{e}_j(\ell)$ is uniformly sumable, i.e., for $\varepsilon > 0$ there is N_ε such that

$$\sum_{j=N}^{\infty} \|\lambda_j(\ell)\langle f, \mathbf{e}_j(\ell)\rangle \mathbf{e}_j(\ell)\| < \varepsilon.$$

Also $\mathbf{x}_j(\ell) \to \mathbf{x}_j = \lambda_j\langle f,\ \mathbf{e}_j\rangle \mathbf{e}_j$, as $\ell \to \infty$. Therefore $\mathbf{C}_\ell f \to \mathbf{C}f$. □

Remark 4.2. The convergence of a sequence of covariance operators, in general, does not imply the convergence of its eigenvectors in $\mathbb{B}(\mathcal{H})$. In fact the distinction of λ_j is necessary. There are counter examples in Kato (1995) page 111 and in Alqallaf, Soltani and Alkandari (2011).

Remark 4.3. For fix ℓ, the random elements $\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)$, $\ell = 1, 2, ..., L$ in $\mathcal{H}$ can be considered as a random element $\mathbf{X}_\ell^n = [\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)]$ in $\mathcal{H}^n = \mathcal{H} \times \mathcal{H} \times .... \times \mathcal{H}$, the Hilbert space formed by the n-fold Cartesian product of $\mathcal{H}$ by itself. Therefore under the assumptions (A) and (B) for every $u = 1, ..., n$, the implications in Theorem 4 are easily extended to the sequences of random elements $[\mathbf{X}_1(\ell), \ldots, \mathbf{X}_n(\ell)]$.

Let us conclude this section with an interesting application of Theorem 4. A random element $\mathbf{X}$ is said to have a Hilbertian elliptically contoured distribution, $\mathbf{X} \sim HEC_{\mathcal{H}}(\mu, \mathbf{C}, \phi)$, if the characteristic function of $\mathbf{X} - \mu$ assume the form $\phi_{\mathbf{X}-\mu}(\mathbf{x}) = \phi_0(\langle \mathbf{x}, \mathbf{C}\mathbf{x}\rangle)$, $\mathbf{x} \in \mathcal{H}$, where $\phi_0 : \mathbb{R}^+ \to \mathbb{R}$, and $\phi_0(\langle \mathbf{x}, \mathbf{C}\mathbf{x}\rangle)$ is positive on $\mathcal{H}$. Equivalently, $\mathbf{X} \sim HEC_{\mathcal{H}}(\mu, \mathbf{C}, \phi)$, if and only if $\mathbf{X} \stackrel{d}{=} \mu + \mathbf{R}\mathbf{U}$, where $\mathbf{U} \sim HEC_{\mathcal{H}}(0, \mathbf{C})$ and R is a nonnegative random variable independent of $\mathbf{U}$, [10].

Let $\mathbf{X}$ be a random element in $\mathcal{H}$, and let $\{\lambda_j, \mathbf{e}_j\}_j$ be the corresponding sequence of its principal component score-value pairs. Let $\xi[j] = \langle \mathbf{e}_j, \mathbf{X}\rangle$ be the jth PC score of $\mathbf{X}$. The following theorem concerns the convergence in distribution of PC scores.

Theorem 4.4. *Assume* $\mathbf{X}_1, \mathbf{X}_2, ...$ *are random elements in* $\mathbf{H}$ *with covariance operators* $\mathbf{C}_1, \mathbf{C}_2, ...$, *respectively. Assume* $\mathbf{X}_1 \sim HEC_{\mathcal{H}}(0, \mathbf{C}_1, \phi_0), \mathbf{X}_2 \sim HEC_{\mathcal{H}}(0, \mathbf{C}_2, \phi_0), ...$, *and* ϕ_0 *is continuous.*

(i) If $\mathbf{C}_\ell$ *converges weakly to an operator* $\mathbf{C}$ *and condition (A) is satisfied, then* $\mathbf{X}_\ell$ *converges in distribution to a random element* $\mathbf{X} \sim HEC_{\mathcal{H}}(0, \mathbf{C}, \phi_0)$.

(ii) If $\mathbf{C}_\ell$ *converges weakly to an operator* $\mathbf{C}$ *and condition (A) is satisfied, then* $(\xi_\ell[j_1], ..., \xi_\ell[j_n])$ *converges in distribution to* $(\xi[j_1], ..., \xi[j_n])$.

(iii) Assume the assumption (B) is satisfied, and $\lambda_j(\ell) \to \lambda_j$, $\mathbf{e}_j(\ell) \to \mathbf{e}_j$ *in norm. Then* $\mathbf{X}_\ell \xrightarrow{d} \mathbf{X}$.

Proof. (i) Apply Theorem 4.1(i) and the continuity theorem. (ii) Apply Theorem 4.1(i) and the result of Li in the cited reference that $(\xi_\ell[j_1], ..., \xi_\ell[j_n]) \sim HEC_{\mathcal{H}}(0, diag(\lambda_\ell[j_1], ..., \lambda_\ell[j_n]), \phi_0)$ and $(\xi[j_1], ..., \xi[j_n]) \sim HEC_{\mathcal{H}}(0, diag(\lambda[j_1]$ $, ..., \lambda[j_n]),$
$\phi_0)$. (iii) Apply Theorem 4.1(ii) and then part (i). The proof is complete. □

Acknowledgements

The authors express their sincere thanks to a reviewer for encouraging comments.

References

[1] R.A. Al-Jarallah, A.R. Soltani and A.N. Al-Kandari, On continuity of the Pearson statistic and sample quantiles, *Ann. Inst. Statist. Math.* **58** (2006), no. 3, 527–535 .

[2] F. Alqallaf, A.R. Soltani and N. Al-Kandar, Continuity and analysis of sequences of principal components, *Comm. Statist. Theory Methods* **39** (2010), no. 19, 3558–3567.

[3] T.W. Anderson, Asymptotic theory for principal component analysis, *Ann. Math. Statist.* **34** (1963) 122–148.

[4] D. Bosq, Linear Process in Function Spaces, Theory and Applications, Springer-Verlag, Berlin, 2000.

[5] L. Debnath and P. Mikusinski, Introduction to Hilbert Spaces with Applications, Academic Press, 1990.

[6] P.J. Diggle, P. Heagerty, K.Y. Liang and S.L. Zeger, Analysis of Longitudinal Data, Oxford Univ. Press, New York, 2002.

[7] I.M. Gelfand and N. Ya. Vilenkin, Generalized Functions, Applications of Harmonic Analysis, Vol. 4, Academic press, 1964.

[8] I.T. Jolliffe and N. Ya. Vilenkin, Principal Component Analysis, Springer, 2nd edition, New York, 2002.

[9] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1970.

[10] Y. Li, A note on Hilbertian elliptically contoured distribution, Technical Report, Iowa State University, 2007.

[11] Y. Li, N. Wang and R.J. Caroll, Selecting the number of principal components in functional data, *J. Amer. Statist. Assoc.* **108** (2013), no. 504, 1284–1294.

[12] J.O. Ramsay and B.W. Silverman, Applied Functional Data Analysis, Springer, New York, 2002.

[13] J.O. Ramsay and B.W. Silverman, Applied Functional Data Analysis, Springer, New York, 2005.

[14] J.O. Ramsay, G. Hooker and S. Graves, Functional Data Analysis with R and MATLAB, Springer, New York, 2009.

[15] J.R. Schott, Matrix Analysis for Statistics, John Wiley, 2nd edition, New York, 2005.

(Ahmad Reza Soltani) Department of Statistics, Shiraz University and Department of Statistics and Operations Research, Kuwait University, State of Kuwait.
E-mail address: `asoltanir@yahoo.com`

(Ali Reza Nematollahi) Department of Statistics, Shiraz University, Shiraz, Iran.
E-mail address: `nematollahi@susc.ac.ir`

(Roya Nasirzadeh) Department of Statistics, Shiraz University, Shiraz, Iran.
E-mail address: `nasirzadeh_roya@yahoo.com`