Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(1): 1937–1958.
DOI:10.3934/math.2023100
Received: 10 July 2022
Revised: 29 September 2022
Accepted: 08 October 2022
Published: 26 October 2022

Research article

Supporting vectors vs. principal components

Almudena P. Márquez[1,*], Francisco Javier García-Pacheco[1], Míriam Mengibar-Rodríguez[2] and Alberto Sánchez-Alzola[3]

[1] Department of Mathematics, College of Engineering, University of Cadiz, 11519 Puerto Real, Spain
[2] Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain
[3] Department of Statistics and Operation Research, University of Cadiz, 11519 Puerto Real, Spain

* **Correspondence:** Email: almudena.marquez@uca.es.

Abstract: Let $T : X \to Y$ be a bounded linear operator between Banach spaces X, Y. A vector $x_0 \in \mathsf{S}_X$ in the unit sphere S_X of X is called a supporting vector of T provided that $\|T(x_0)\| = \sup\{\|T(x)\| : \|x\| = 1\} = \|T\|$. Since matrices induce linear operators between finite-dimensional Hilbert spaces, we can consider their supporting vectors. In this manuscript, we unveil the relationship between the principal components of a matrix and its supporting vectors. Applications of our results to real-life problems are provided.

Keywords: bounded linear operator; Hilbert space; mean operator; principal components; supporting vector
Mathematics Subject Classification: 51F30, 54E35, 54E45

1. Introduction

Supporting Vector Analysis (SVA) is a relatively recent technique that allows one to solve analytically many real-life problems that used to be tackled by means of Heuristic methods. The lack of mathematical formalism of Heuristic methods resulted many times in unpredictable solutions, that is, mathematical solutions whose real-life interpretations make no sense. Supporting vectors came into play to overcome this issue. This way, supporting vectors were used in a successful way to solve multiobjective optimization problems coming from different disciplines, such as Bioengineering, Physics, and Statistics [4, 6–9, 15, 22], improving considerably the results achieved by other methods like, for instance, Heuristic techniques [10, 11, 20, 21].

In [4, 6, 15], it was proven that Singular Value Decomposition (SVD) can be seen as a particular case of SVA. This fact triggered the new trend of restating Statistical notions from the perspective of

Functional Analysis and Operator Theory. The main objective of this manuscript is to study Principal Component Analysis (PCA) by means of SVA.

2. Materials and methods

We will review several basic notions from Operator Theory that will turn out to be crucial for the development of this manuscript.

2.1. Centering and standardizing

If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then the mean of x is defined as $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$, and its standard deviation is given by $s_x := \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$. Notice that

$$\sqrt{n}s_x = \left\|x - \overline{\mathbf{x}}\right\|_2, \tag{2.1.1}$$

where $\overline{\mathbf{x}} := (\overline{x}, \overset{n}{\ldots}, \overline{x})$ denotes the constant vector of term $\overline{x}$ (in general, if $a \in \mathbb{R}$, then $\mathbf{a} := (a, \overset{n}{\ldots}, a)$ denotes the constant vector of term a).

We say that $x \in \mathbb{R}^n$ is centered provided that $\overline{x} = 0$, and it is standardized provided that $\overline{x} = 0$ and $s_x = 1$. In the latter situation, $\|x\|_2 = \sqrt{n}$, in view of (2.1.1). The subset of centered vectors of $\mathbb{R}^n$ is usually denoted by $\mathrm{cen}(\mathbb{R}^n)$, that is, $\mathrm{cen}(\mathbb{R}^n) := \{x \in \mathbb{R}^n : \overline{x} = 0\}$. The subset of standardized vectors of $\mathbb{R}^n$ is usually denoted by $\mathrm{stan}(\mathbb{R}^n)$, that is,

$$\mathrm{stan}(\mathbb{R}^n) := \{x \in \mathbb{R}^n : \overline{x} = 0 \text{ and } s_x = 1\}.$$

According to (2.1.1), $\mathrm{stan}(\mathbb{R}^n) \subseteq \sqrt{n}\mathsf{S}_{\ell_2^n}$, where $\mathsf{S}_{\ell_2^n}$ stands for the unit sphere of $\ell_2^n := (\mathbb{R}^n, \|\cdot\|_2)$. In Topology, $\mathsf{S}_{\ell_2^n}$ is denoted as S^{n-1}.

2.2. Principal component analysis

The covariance of two vectors $x, y \in \mathbb{R}^n$ is defined as

$$s_{x,y} := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

Notice that $s_{x,x} = s_x^2$, that is, the variance of x. The covariance matrix of a given matrix $A \in \mathcal{M}_{m\times n}$ is defined by $s_{\mathbf{a_1},\ldots,\mathbf{a_n}} := \left(s_{\mathbf{a_i},\mathbf{a_j}}\right)_{i,j=1,\ldots,n}$, where $\mathbf{a_1}, \ldots, \mathbf{a_n}$ stand for the column vectors of A.

Consider a matrix $A \in \mathcal{M}_{m\times n}$. The principal components of A are defined as $Ax_1, \ldots, Ax_n$, where $\{x_1, \ldots, x_n\}$ is an ordered orthonormal basis of eigenvectors of $s_{\mathbf{a_1},\ldots,\mathbf{a_n}}$, sorting the eigenvalues of $s_{\mathbf{a_1},\ldots,\mathbf{a_n}}$ decreasingly.

We refer the reader to [24] for a wider perspective on PCA. Interesting applications of PCA to certain Engineering fields, such as video processing and Big Data, have been provided in [3, 12].

2.3. Supporting vector analysis

Let X, Y be Banach spaces. Let $T : X \to Y$ be a bounded linear operator. The operator norm of T is given by

$$\|T\| := \sup\{\|T(x)\| : \|x\| = 1\}. \tag{2.3.1}$$

The vector space $\mathscr{CL}(X, Y)$ of continuous linear operators from X to Y becomes a Banach space when endowed with the operator norm (2.3.1). In the case $X = Y$, $\mathscr{CL}(X, Y)$ is simply denoted as $\mathscr{CL}(X)$. If $Y = \mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$), then $\mathscr{CL}(X, Y)$ is denoted as X^*, that is, the dual space of X. It is also common to denote $\mathscr{CL}(X, Y)$ by $\mathcal{B}(X, Y)$ and $\mathscr{CL}(X)$ by $\mathcal{B}(X)$.

The supporting vector notion was formally posed for the first time in [5]. However, this concept can be found implicitly and scattered throughout the literature of Banach Space Theory [1, 2, 18, 19].

The set of supporting vectors of a bounded linear operator $T : X \to Y$ between Banach spaces X, Y is defined by

$$\mathrm{suppv}(T) := \{x \in \mathsf{S}_X : \|T(x)\| = \|T\|\} = \arg\max_{\|x\|=1} \|T(x)\|. \tag{2.3.2}$$

Here, S_X stands for the unit sphere of X, and B_X denotes the (closed) unit ball of X. In the infinite-dimensional setting, it may occur that (2.3.2) is empty. Note that $\mathrm{suppv}(T) = \mathrm{suppv}(\lambda T)$ for all $\lambda \in \mathbb{K} \setminus \{0\}$, and $\mathrm{suppv}(T) = \mathsf{S}_{\mathbb{K}}\mathrm{suppv}(T)$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. For a topological and geometrical analysis of the above set, we strongly refer the reader to [13, 14, 23].

For linear functionals, a special subset of supporting vectors is worth regarding. Consider a continuous linear functional $f \in X^*$ in the dual X^* of a Banach space X. We define the set of 1-supporting vectors of f by

$$\mathrm{suppv}_1(f) := \{x \in \mathsf{S}_X : f(x) = \|f\|\}. \tag{2.3.3}$$

Notice that 1-supporting vectors are particular cases of supporting vectors; in other words, $\mathrm{suppv}_1(f) \subseteq \mathrm{suppv}(f)$. In the upcoming sections, 1-supporting vectors will be very much relied on. The following remark highlights a standard geometrical property satisfied by 1-supporting vectors.

Remark 2.1. *Consider a Banach space X and a nonzero linear functional $f \in X^* \setminus \{0\}$. For every $x, y \in \mathrm{suppv}_1(f)$ and every $\lambda \in [0, 1]$, we have that $\lambda x + (1 - \lambda)y \in \mathrm{suppv}_1(f)$, that is, $\mathrm{suppv}_1(f)$ is a convex subset of the unit sphere S_X of X.*

A direct consequence of Remark 2.1 is that $\mathrm{suppv}_1(f)$ is either empty or a singleton in strictly convex Banach spaces, like, for instance, Hilbert spaces.

2.4. Hilbert space theory

Representation Theory is one of the most important theories in Mathematics. A major result in Representation Theory is undoubtedly the Riesz Representation Theorem. This is a key result in Functional Analysis and is crucial for working with self-adjoint operators on Hilbert spaces.

Riesz Representation Theorem. *In a Hilbert space H, for every $h^* \in H^*$ there exists a unique $h \in H$ satisfying $h^* = (\cdot|h)$. This assignment between H and H^* is a surjective linear isometry.*

In view of Remark 2.1 and under the settings of the Riesz Representation Theorem, for every $h \in H \setminus \{0\}$, we have that $\mathrm{suppv}_1(h^*) = \left\{\frac{h}{\|h\|}\right\}$, that is, $\frac{h}{\|h\|}$ is the only 1-supporting vector of h^*.

On the other hand, the orthogonal subspace of a closed subspace V of a Hilbert space H is denoted by $V^\perp$. The orthogonal projection of H onto V is usually denoted by p_V. Observe that $H = V \oplus_2 V^\perp$, that is, for each $h \in H$,

$$h = p_V(h) + p_{V^\perp}(h), \text{ and } \|h\|^2 = \|p_V(h)\|^2 + \|p_{V^\perp}(h)\|^2. \tag{2.4.1}$$

The adjoint of a bounded linear operator $T : H \to K$ between Hilbert spaces H, K is defined as the unique bounded linear operator $T^* : K \to H$ satisfying $(T(h)|k) = (h|T^*(k))$ for each $h \in H$ and each $k \in K$. Basic properties satisfied by the adjoint operator are $\|T^*\| = \|T\|$, $(T^*)^* = T$, $(T+S)^* = T^* + S^*$, $(T \circ S)^* = S^* \circ T^*$ and $(\lambda T)^* = \overline{\lambda} T^*$.

Lemma 2.2. *Every bounded linear operator $T : H \to K$ between Hilbert spaces H, K verifies that* $\mathrm{cl}(T(H)) = \ker(T^*)^\perp$ *and* $T(H)^\perp = \mathrm{cl}(T(H))^\perp = \ker(T^*)$.

Proof. First off, it is a trivial observation that $T(H)^\perp = \mathrm{cl}(T(H))^\perp$. Fix an arbitrary $h \in H$. For every $k \in \ker(T^*)$, $(T(h)|k) = (h|T^*(k)) = (h|0) = 0$. This shows that $T(h) \in \ker(T^*)^\perp$. The arbitrariness of $h \in H$ means that $T(H) \subseteq \ker(T^*)^\perp$. By taking orthogonal complements, we obtain that $\ker(T^*) \subseteq T(H)^\perp$. Conversely, fix an arbitrary $k \in T(H)^\perp$. For every $h \in H$, $0 = (T(h)|k) = (h|T^*(k))$. This shows that $T^*(k) = 0$, and hence $k \in \ker(T^*)$. The arbitrariness of $k \in T(H)^\perp$ means that $T(H)^\perp \subseteq \ker(T^*)$. By taking orthogonal complements, we finally obtain that $\ker(T^*)^\perp \subseteq \mathrm{cl}(T(H))$. ■

A bounded operator $T \in \mathcal{B}(H)$ is said to be self-adjoint provided that $T^* = T$. If H is complex, then $T \in \mathcal{B}(H)$ is self-adjoint if and only if $(T(h)|h) \in \mathbb{R}$ for all $h \in H$. A self-adjoint operator is called positive provided that $(T(h)|h) \geq 0$ for all $h \in H$.

For every $T \in \mathcal{B}(H)$, $\sigma(T) := \{\lambda \in \mathbb{C} : \lambda I - T \notin \mathcal{U}(\mathcal{B}(H))\}$ is the spectrum of T, where $\mathcal{U}(\mathcal{B}(H))$ is the multiplicative group of invertible operators on H. Among other spectral properties, the spectrum is compact and nonempty, and $\|T\| \geq \max|\sigma(T)|$. A special subset of the spectrum, called the point spectrum, $\sigma_p(T) := \{\lambda \in \mathbb{C} : \ker(\lambda I - T) \neq \{0\}\}$, whose elements are called the eigenvalues of T, will be very much employed. Note that $\sigma_p(T) \subseteq \sigma(T)$. Furthermore, for each $\lambda \in \sigma_p(T)$, $V_T(\lambda) := \{h \in H : T(h) = \lambda h\}$ stands for the subspace of eigenvectors associated with λ. In case there is no confusion with T, we will simply denote $V_T(\lambda)$ by $V(\lambda)$.

Suppose next that $\|T\|$ is an eigenvalue of T, that is, $\|T\| \in \sigma_p(T)$. In this situation, since $\|T\| \geq \max|\sigma(T)|$, we conclude that $\|T\|$ is the maximum of $|\sigma(T)|$; in other words, $\|T\| = \max|\sigma(T)|$. In this case, we write $\|T\| = \lambda_{\max}(T)$. Observe also that $V(\|T\|) \cap \mathsf{S}_X \subseteq \mathrm{suppv}(T)$. Indeed, if $x \in V(\|T\|) \cap \mathsf{S}_X$, then $T(x) = \|T\|x$, so $\|T(x)\| = \|T\|$, and hence $x \in \mathrm{suppv}(T)$.

Nevertheless, in general, $\|T\| \notin \sigma_p(T)$, unless, for instance, T is compact, self-adjoint and positive. This is why we have to rely on the adjoint T^* and on the strongly positive operator $T^* \circ T$. It is straightforward to check that the eigenvalues of a self-adjoint operator are real, and the eigenvalues of a self-adjoint positive operator are positive. When T is compact, it holds that $T^* \circ T$ is compact, self-adjoint and positive.

The following result, on which we will strongly rely later on, can be found in [6, Theorem 4], which is itself a refinement of [14, Theorem 9].

Theorem 2.3. *Let H, K be Hilbert spaces. Let $T \in \mathcal{B}(H, K)$. Then, $\|T\|^2 = \|T^* \circ T\|$, and* $\mathrm{suppv}(T) \subseteq \mathrm{suppv}(T^* \circ T)$. *Furthermore,* $\mathrm{suppv}(T) \neq \varnothing$ *if and only if $\|T^* \circ T\| \in \sigma_p(T^* \circ T)$. In this situation, $\|T\| = \sqrt{\lambda_{\max}(T^* \circ T)}$, and* $\mathrm{suppv}(T) = V(\lambda_{\max}(T^* \circ T)) \cap \mathsf{S}_H$.

3. Results

In this section, we will state and prove all the novel theorems of this work. This section is divided into four subsections, in which we will deal with supporting vectors, principal components and the topological structure of the subsets of centered and standardized vectors.

3.1. Topological structure of the subsets of centered and standardized vectors

This subsection begins unveiling the topological structure of the subsets $\mathrm{cen}(\mathbb{R}^n)$ and $\mathrm{stan}(\mathbb{R}^n)$ of centered and standardized vectors, respectively. Notice that $\mathrm{cen}(\mathbb{R}) = \{0\}$ and $\mathrm{stan}(\mathbb{R}) = \varnothing$.

Theorem 3.1. *If $n \geq 1$, then $\mathrm{cen}(\mathbb{R}^n)$ is linearly isometric, and hence homeomorphic, to ℓ_2^{n-1}. If $n \geq 2$, then $\mathrm{stan}(\mathbb{R}^n)$ is linearly isometric, and hence homeomorphic, to S^{n-2}.*

Proof. Notice that $\overline{x+y} = \overline{x} + \overline{y}$, and $\overline{tx} = t\overline{x}$ for all $x \in \mathbb{R}^n$ and all $t \in \mathbb{R}$. As a consequence,

$$\begin{array}{rccl} \bar{\cdot}: & \mathbb{R}^n & \to & \mathbb{R} \\ & x & \mapsto & \overline{x} = \frac{1}{n}\sum_{i=1}^n x_i \end{array} \tag{3.1.1}$$

is a linear functional (usually called the *mean functional*). Next, simply observe that $\mathrm{cen}(\mathbb{R}^n) = \ker(\bar{\cdot})$. Thus, by bearing in mind (2.1.1), we immediately obtain that $\mathrm{stan}(\mathbb{R}^n) = \ker(\bar{\cdot}) \cap \sqrt{n}\mathsf{S}_{\ell_2^n} = \sqrt{n}\mathsf{S}_{\ker(\bar{\cdot})}$. In other words, $\mathrm{stan}(\mathbb{R}^n)$ is precisely a multiple of the unit sphere of the Hilbert subspace $\ker(\bar{\cdot})$ of ℓ_2^n. Since $\dim(\ker(\bar{\cdot})) = n-1$, we have that $\ker(\bar{\cdot})$ is linearly isometric to ℓ_2^{n-1}. As a consequence, the unit sphere of $\ker(\bar{\cdot})$, $\frac{1}{\sqrt{n}}\mathrm{stan}(\mathbb{R}^n)$, is linearly isometric to the unit sphere of ℓ_2^{n-1}, $\mathsf{S}_{\ell_2^{n-1}} = \mathsf{S}^{n-2}$. ■

The following lemma aims at computing the norm of the mean functional as an element of the dual space of ℓ_2^n as well as its only 1-supporting vector.

Lemma 3.2. *In $\left(\ell_2^n\right)^*$, $\|\bar{\cdot}\| = \frac{1}{\sqrt{n}}$ and $\mathrm{suppv}_1(\bar{\cdot}) = \left\{\frac{\mathbf{1}}{\sqrt{n}}\right\}$.*

Proof. Hölder's Inequality ensures that

$$|\overline{x}| = \left|\frac{1}{n}\sum_{i=1}^n x_i\right| \leq \sum_{i=1}^n \frac{1}{n}|x_i| \leq \left(\sum_{i=1}^n \frac{1}{n^2}\right)^{\frac{1}{2}} \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} = \frac{1}{\sqrt{n}}\|x\|_2$$

for every $x \in \ell_2^n$. This shows that $\|\bar{\cdot}\| \leq \frac{1}{\sqrt{n}}$. On the other hand, $\|\mathbf{1}\|_2 = \sqrt{n}$, that is, $\frac{\mathbf{1}}{\sqrt{n}} \in \mathsf{S}_{\ell_2^n}$. Finally,

$$\overline{\frac{\mathbf{1}}{\sqrt{n}}} = \frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{n}}.$$

As a consequence, $\|\bar{\cdot}\| = \frac{1}{\sqrt{n}}$ and $\mathrm{suppv}_1(\bar{\cdot}) = \left\{\frac{\mathbf{1}}{\sqrt{n}}\right\}$. ■

As a direct consequence of Lemma 3.2, the standard deviation of a vector $x \in \mathbb{R}^n$ can be rewritten as the following (2.1.1):

$$s_x = \frac{1}{\sqrt{n}}\left\|x - \overline{\mathbf{x}}\right\|_2 = \|\bar{\cdot}\|\left\|x - \overline{\mathbf{x}}\right\|_2. \tag{3.1.2}$$

By bearing in mind the Riesz Representation Theorem, Lemma 3.2 assures that if $h := \frac{\mathbf{1}}{\mathbf{n}}$, then $h^* = \bar{\cdot}$ in $H := \ell_2^n$. Notice that $h := \frac{\mathbf{1}}{\mathbf{n}} = \left(\frac{1}{n}, \overset{n}{\ldots}, \frac{1}{n}\right)$ can be seen as a (finite) convex series. This fact will help us generalize these concepts to an infinite-dimensional separable Hilbert space setting later on in the Discussion.

Definition 3.3. *For $n \geq 1$, the mean operator is defined by*

$$\begin{array}{rccl} \overline{\bullet}: & \mathbb{R}^n & \to & \mathbb{R}^n \\ & x & \mapsto & \overline{\mathbf{x}} = (\overline{x}, \overset{n}{\ldots}, \overline{x}), \end{array} \tag{3.1.3}$$

and the centering operator is defined as

$$\begin{array}{rccl} \mathrm{cen}: & \mathbb{R}^n & \to & \mathbb{R}^n \\ & x & \mapsto & \mathrm{cen}(x) := x - \overline{\mathbf{x}} = (x_1 - \overline{x}, \overset{n}{\ldots}, x_n - \overline{x}). \end{array} \tag{3.1.4}$$

For $n \geq 2$, the standardizing operator is defined as

$$\begin{array}{rccl} \mathrm{stan}: & \mathbb{R}^n \setminus \mathbb{R}^n\mathbf{1} & \to & \mathbb{R}^n \\ & x & \mapsto & \mathrm{stan}(x) := \sqrt{n}\frac{x-\overline{\mathbf{x}}}{\|x-\overline{\mathbf{x}}\|_2}. \end{array} \tag{3.1.5}$$

It is a trivial observation that

$$\mathrm{stan}(x) = \sqrt{n}\frac{\mathrm{cen}(x)}{\|\mathrm{cen}(x)\|_2} = \frac{\mathrm{cen}(x)}{\|\bar{\cdot}\|\,\|\mathrm{cen}(x)\|_2} \tag{3.1.6}$$

for all $x \in \mathbb{R}^n$.

Recall that a projection on a Banach space X is a continuous, linear, and idempotent map $P : X \to X$. Its dual operator $P^* : X^* \to X^*$ is also a projection. The complementary projection of P is defined as $I_X - P$, which is also a projection. Every non-zero projection has norm greater than or equal to 1. A 1-projection is a projection of norm 1 (also called a contractive projection), and a $(1, 1)$-projection is a 1-projection whose complementary projection is also a 1-projection (also called bicontractive). Orthogonal projections in Hilbert spaces are the most representative examples of bicontractive projections.

The final result of this first subsection serves to show that both the mean operator and the centering operator are complementary projections to each other of norm 1, that is, bicontractive, for the Euclidean norm. Even more, the mean operator and the centering operator are orthogonal projections to each other. This will allow us to directly obtain the König-Huygens Theorem (3.1.7) as a direct consequence of the Pythagorean Theorem in Hilbert spaces. The König-Huygens Theorem provides the classical decomposition of the 2-norm of a vector of $\mathbb{R}^n$ in terms of the mean and the standard deviation.

Theorem 3.4. *The centering operator is a linear projection on $\mathbb{R}^n$ whose kernel is* $\ker(\mathrm{cen}) = \mathbb{R}^n\mathbf{1}$*, whose range is* $\mathrm{cen}(\mathbb{R}^n) = \ker(\bar{\cdot})$*, and whose complementary projection is the mean operator. Furthermore, if we consider $\mathbb{R}^n$ endowed with the Euclidean norm, then $\|\overline{\bullet}\| = \|\mathrm{cen}\| = 1$, and $\overline{\bullet}$ and* cen *are complementary orthogonal projections. As a consequence, for every $x \in \mathbb{R}^n$,*

$$\|x\|_2^2 = \|\overline{\mathbf{x}}\|_2^2 + \|x - \overline{\mathbf{x}}\|_2^2 = n\left(|\overline{x}|^2 + s_x^2\right). \tag{3.1.7}$$

Proof. In the first place, notice that $\overline{\bullet}$ is clearly linear, since $\overline{\bullet} = \bar{\cdot}\mathbf{1}$; in other words, $\overline{\mathbf{x}} = (\overline{x}, \overset{n}{\ldots}, \overline{x}) = \overline{x}(1, \overset{n}{\ldots}, 1) = \overline{x}\mathbf{1}$ for all $x \in \mathbb{R}^n$. As a consequence, cen is linear as well because $\mathrm{cen} = I_{\mathbb{R}^n} - \overline{\bullet}$, that is, the difference of two linear operators on $\mathbb{R}^n$. On the other hand, observe that if $x \in \mathbb{R}^n$ is a constant vector, that is, $x = t\mathbf{1} = \mathbf{t}$ for some $t \in \mathbb{R}$, then $\mathrm{cen}(x) = 0$. Conversely, if $\mathrm{cen}(x) = 0$, then $x = \overline{\mathbf{x}} = \overline{x}\mathbf{1}$, and hence $x \in \mathbb{R}^n\mathbf{1}$ is a constant vector. This shows that $\ker(\mathrm{cen}) = \mathbb{R}^n\mathbf{1}$. From Theorem 3.1, we already know that $\mathrm{cen}(\mathbb{R}^n) = \ker(\bar{\cdot})$. Next, let us prove that cen is a projection. Fix an arbitrary $x \in \mathbb{R}^n$. Notice that

$$\mathrm{cen}(\mathrm{cen}(x)) = \mathrm{cen}(x - \overline{\mathbf{x}}) = \mathrm{cen}(x) - \mathrm{cen}(\overline{\mathbf{x}}) = x - \overline{\mathbf{x}} - 0 = x - \overline{\mathbf{x}} = \mathrm{cen}(x).$$

Since $\bar{\bullet} = I_{\mathbb{R}^n} - \mathrm{cen}$, we conclude that the mean operator is the complementary projection to the centering operator. Finally, let us compute $\|\bar{\bullet}\|$ and $\|\mathrm{cen}\|$. In accordance with Lemma 3.2, for every $x \in \mathbb{R}^n$ we have that

$$\|\overline{\mathbf{x}}\|_2 = \sqrt{n}|\overline{x}| \leq \sqrt{n}\frac{1}{\sqrt{n}}\|x\|_2 = \|x\|_2,$$

meaning that $\|\bar{\bullet}\| \leq 1$. Next,

$$\left\|\frac{\mathbf{1}}{\sqrt{n}}\right\|_2 = \left\|\left(\frac{1}{\sqrt{n}}, \overset{n}{\ldots}, \frac{1}{\sqrt{n}}\right)\right\|_2 = \sqrt{n}\frac{1}{\sqrt{n}} = 1.$$

This shows that $\|\bar{\bullet}\| = 1$. In order to prove that $\bar{\bullet}$ and cen are orthogonal projections, it only suffices to realize that $\mathbb{R}\mathbf{1}$ and $\mathrm{cen}(\mathbb{R}^n)$ are orthogonal subspaces. Indeed, for every $t \in \mathbb{R}$ and every $x \in \mathbb{R}^n$ with $\overline{x} = 0$, we have that $(t\mathbf{1}|x) = t(\mathbf{1}|x) = t\sum_{i=1}^n x_i = tn\overline{x} = 0$, and hence $\mathbb{R}\mathbf{1} \subseteq \mathrm{cen}(\mathbb{R}^n)^\perp$, or equivalently, $\mathrm{cen}(\mathbb{R}^n) \subseteq (\mathbb{R}\mathbf{1})^\perp$. Furthermore, the Pythagorean Theorem in ℓ_2^n allows that

$$\|t\mathbf{1} + x\|_2^2 = \|t\mathbf{1}\|_2^2 + \|x\|_2^2$$

for every $t \in \mathbb{R}$ and every $x \in \mathbb{R}^n$ with $\overline{x} = 0$. Next, if $y \in (\mathbb{R}\mathbf{1})^\perp$, then $0 = \frac{1}{n}(\mathbf{1}|y) = \frac{1}{n}\sum_{i=1}^n y_i = \overline{y}$, meaning that $y \in \ker(\bar{\cdot}) = \mathrm{cen}(\mathbb{R}^n)$. As a consequence, $\mathrm{cen}(\mathbb{R}^n)^\perp = \mathbb{R}\mathbf{1}$, resulting in

$$\|x\|_2^2 = \|\overline{\mathbf{x}}\|_2^2 + \|x - \overline{\mathbf{x}}\|_2^2 = n\left(|\overline{x}|^2 + s_x^2\right)$$

for every $x \in \mathbb{R}^n$. It only remains to show that $\|\mathrm{cen}\| = 1$, but this is a direct consequence of the fact that cen is an orthogonal projection. ■

3.2. Supporting vectors and first principal component

In this subsection, we will provide sufficient conditions for the supporting vectors to coincide with the first principal component. First, we will need some definitions.

Definition 3.5. *A matrix is said to be centered (standardized) provided that all of its column vectors are centered (standardized) vectors.*

The following lemma displays a simple characterization of centered matrices.

Lemma 3.6. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The following conditions are equivalent:*

1) A is centered.
2) Ax is centered for all $x \in \mathbb{R}^n$.

Proof. Let $\{e_1, \ldots, e_n\}$ denote the canonical basis of $\mathbb{R}^n$. Notice that the columns of A are precisely Ae_j for $j = 1, \ldots, n$. Suppose first that A is centered. Fix an arbitrary $x \in \mathbb{R}^n$. The linearity of the mean functional allows that

$$\overline{Ax} = \overline{\sum_{j=1}^n x_j A e_j} = \sum_{j=1}^n x_j \overline{Ae_j} = 0.$$

As a consequence, Ax is centered for all $x \in \mathbb{R}^n$. Conversely, suppose now that Ax is centered for all $x \in \mathbb{R}^n$. In particular, $\overline{Ae_j} = 0$ for $j = 1, \ldots, n$, meaning that the columns of A are centered vectors, that is, A is centered.

■

The following characterization of centered matrices is a bit more sophisticated. First, a technical lemma is needed.

Lemma 3.7. *Consider $x, y \in \mathbb{R}^m$. Then,*

1) *$ms_{x,y} = x \bullet y$ if and only if either x or y is centered.*
2) *x is standardized if and only if x is centered and $x \bullet x = m$.*

Proof.

1) Let us observe that

$$\begin{aligned} ms_{x,y} &= \sum_{i=1}^{m} (x_i - \overline{x})(y_i - \overline{y}) \\ &= \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} \overline{x} y_i - \sum_{i=1}^{m} x_i \overline{y} + \sum_{i=1}^{m} \overline{x}\,\overline{y} \\ &= \sum_{i=1}^{m} x_i y_i - \overline{x} \sum_{i=1}^{m} y_i - \overline{y} \sum_{i=1}^{m} x_i + \overline{x}\,\overline{y} \sum_{i=1}^{m} 1 \\ &= \sum_{i=1}^{m} x_i y_i - \overline{x} m \overline{y} - \overline{y} m \overline{x} + \overline{x}\,\overline{y} \\ &= x \bullet y - m\overline{x}\,\overline{y}. \end{aligned}$$

As a consequence, $ms_{x,y} = x \bullet y$ if and only if either x or y is centered.

2) By definition, x is standardized if and only if x is centered and $s_x = 1$. We know that then $\|x - \overline{x}\|_2 = \sqrt{m} s_x$. In the context of centered vectors, the previous expression becomes $\sqrt{x \bullet x} = \|x\|_2 = \sqrt{m} s_x$. Therefore, x is standardized if and only if x is centered and $x \bullet x = m$.

■

Recall that the covariance matrix of a given matrix $A \in \mathcal{M}_{m\times n}$ is defined by $s_{\mathbf{a_1},\ldots,\mathbf{a_n}} := \left(s_{\mathbf{a_i},\mathbf{a_j}}\right)_{i,j=1,\ldots,n}$, where $\mathbf{a_1}, \ldots, \mathbf{a_n}$ stand for the column vectors of A. Also, recall that if B is a square matrix, then $\operatorname{diag}(B)$ stands for the diagonal of B.

Proposition 3.8. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The following conditions are equivalent:*

1) *A is centered.*
2) *$A^t A = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$.*
3) *$\operatorname{diag}(A^t A) = \operatorname{diag}(ms_{\mathbf{a_1},\ldots,\mathbf{a_n}})$.*

Proof. Notice that $A^tA = \left(\mathbf{a_i} \bullet \mathbf{a_j}\right)_{i,j=1,\ldots,n}$. Suppose first that A is centered. In view of Lemma 3.7, $ms_{\mathbf{a_i},\mathbf{a_j}} = \mathbf{a_i} \bullet \mathbf{a_j}$ for all $i, j \in \{1, \ldots, n\}$, meaning that $A^tA = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$. Next, if $A^tA = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$, then we trivially have that $\operatorname{diag}(A^tA) = \operatorname{diag}(ms_{\mathbf{a_1},\ldots,\mathbf{a_n}})$. Finally, assume that $\operatorname{diag}(A^tA) = \operatorname{diag}(ms_{\mathbf{a_1},\ldots,\mathbf{a_n}})$. Then, $ms_{\mathbf{a_i},\mathbf{a_i}} = \mathbf{a_i} \bullet \mathbf{a_i}$ for all $i = 1, \ldots, n$. In accordance with Lemma 3.7, $\mathbf{a_i}$ is centered for all $i = 1, \ldots, n$, meaning that A is centered. ■

As an immediate consequence of Lemma 3.7 and Proposition 3.8, we obtain the following corollary.

Corollary 3.9. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The following conditions are equivalent:*

1) *A is standardized.*
2) *$A^tA = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$ and* $\operatorname{diag}(A^tA) = (m, \overset{n}{\ldots}, m)$.

Finally, we have gathered all the necessary tools to prove our main results of this subsection. Simply keep in mind the small observation that if $B \in \mathcal{M}_n(\mathbb{R})$, then $\sigma_p(\alpha B) = \alpha\sigma_p(B)$ and $V_{\alpha B}(\alpha\lambda) = V_B(\lambda)$ for all $\alpha \in \mathbb{R} \setminus \{0\}$ and all $\lambda \in \sigma_p(B)$.

Theorem 3.10. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. If A is centered, then*

$$\operatorname{suppv}(A) = \left\{x \in \mathsf{S}_{\ell_2^n} : Ax \textit{ is the first principal component of } A\right\},$$

where A is seen as a linear operator

$$\begin{array}{rccc} A: & \ell_2^n & \to & \ell_2^m \\ & x & \mapsto & Ax. \end{array} \tag{3.2.1}$$

Proof. According to Theorem 2.3, $\operatorname{suppv}(A) = V\left(\lambda_{\max}(A^* \circ A)\right) \cap \mathsf{S}_{\ell_2^n}$. Notice that the adjoint of A, A^*, coincides with its transpose, A^t. On the other hand, since A is centered, Proposition 3.8 allows that $A^tA = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$. Finally,

$$\begin{aligned}
\operatorname{suppv}(A) &= V_{A^*\circ A}\left(\lambda_{\max}(A^* \circ A)\right) \cap \mathsf{S}_{\ell_2^n} \\
&= V_{A^tA}\left(\lambda_{\max}(A^tA)\right) \cap \mathsf{S}_{\ell_2^n} \\
&= V_{ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}}\left(\lambda_{\max}(ms_{\mathbf{a_1},\ldots,\mathbf{a_n}})\right) \cap \mathsf{S}_{\ell_2^n} \\
&= V_{ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}}\left(m\lambda_{\max}(s_{\mathbf{a_1},\ldots,\mathbf{a_n}})\right) \cap \mathsf{S}_{\ell_2^n} \\
&= V_{s_{\mathbf{a_1},\ldots,\mathbf{a_n}}}\left(\lambda_{\max}(s_{\mathbf{a_1},\ldots,\mathbf{a_n}})\right) \cap \mathsf{S}_{\ell_2^n} \\
&= \left\{x \in \mathsf{S}_{\ell_2^n} : Ax \text{ is the first principal component of } A\right\}.
\end{aligned}$$

■

We will conclude this subsection with an example of a centered matrix A whose last principal component has at least two dimensions.

Remark 3.11. *Let $T : H \to K$ be a bounded linear operator between Hilbert spaces H, K. Then,* $\ker(T) \subseteq \ker(T^* \circ T)$. *As a consequence, if* $0 \in \sigma_p(T)$, *then* $0 \in \sigma_p(T^* \circ T)$.

Notice that if $T \in \mathcal{B}(H)$ is a self-adjoint positive operator on a Hilbert space H such that $\ker(T) \neq \{0\}$, then 0 is the minimum of $\sigma_p(T)$ since all the eigenvalues of T are positive.

Example 3.12. *Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ be any centered matrix with three equal columns. Then,* $\ker(A) \subseteq \ker(A^tA) = \ker(s_{\mathbf{a_1},\ldots,\mathbf{a_n}})$ *since $A^tA = ms_{\mathbf{a_1},\ldots,\mathbf{a_n}}$ in accordance with Proposition 3.8. Notice that* $\ker(A)$ *has at least dimension* 2. *Observe that* $0 \in \sigma_p(s_{\mathbf{a_1},\ldots,\mathbf{a_n}})$. *As a consequence, the last principal component of A corresponds to $Ax_1, \ldots, Ax_p$, where $\{x_1, \ldots, x_p\}$ is an orthonormal basis of* $V_{s_{\mathbf{a_1},\ldots,\mathbf{a_n}}}(0)$.

3.3. Second principal component as supporting vector

This subsection is aimed at showing the second principal component of a centered matrix can be obtained via a supporting vector of a derived matrix from the original one.

Theorem 3.13. *Let H, K be Hilbert spaces and $T : H \to K$ be a continuous linear operator. Let H_1 be a closed subspace of H. Then,*

1) $\left\|T|_{H_1}\right\| = \left\|T \circ p_{H_1}\right\|$.
2) $\operatorname{suppv}(T|_{H_1}) = \operatorname{suppv}(T \circ p_{H_1})$.
3) $(T|_{H_1})^* = p_{H_1} \circ T^*$.
4) *If H_1 is an invariant subspace of $T^* \circ T$, then* $(T|_{H_1})^* \circ T|_{H_1} = (T^* \circ T)|_{H_1}$.

Proof.

1) In the first place, p_{H_1} is an orthogonal projection. Thus, $\left\|p_{H_1}\right\| = 1$, and therefore $\left\|T \circ p_{H_1}\right\| = \left\|T|_{H_1} \circ p_{H_1}\right\| \leq \left\|T|_{H_1}\right\| \left\|p_{H_1}\right\| = \left\|T|_{H_1}\right\|$. On the other hand, for every $h_1 \in \mathsf{B}_{H_1}$, $\left\|T|_{H_1}(h_1)\right\| = \|T(h_1)\| = \left\|T(p_{H_1})\right\| = \left\|(T \circ p_{H_1})(h_1)\right\| \leq \left\|T \circ p_{H_1}\right\|$. As a consequence, $\left\|T|_{H_1}\right\| \leq \left\|T \circ p_{H_1}\right\|$, meaning that $\left\|T|_{H_1}\right\| = \left\|T \circ p_{H_1}\right\|$.
2) Fix an arbitrary $h_1 \in \operatorname{suppv}(T|_{H_1})$. Then, $\left\|(T \circ p_{H_1})(h_1)\right\| = \left\|T|_{H_1}(h_1)\right\| = \left\|T|_{H_1}\right\| = \left\|T \circ p_{H_1}\right\|$, meaning that $h_1 \in \operatorname{suppv}(T \circ p_{H_1})$. Conversely, take $h \in \operatorname{suppv}(T \circ p_{H_1})$. We will show first that $p_{H_1}(h) \in \operatorname{suppv}(T|_{H_1})$. Indeed, $\left\|T|_{H_1}(p_{H_1}(h))\right\| = \left\|(T \circ p_{H_1})(h)\right\| = \left\|T \circ p_{H_1}\right\| = \left\|T|_{H_1}\right\|$. Since $\left\|p_{H_1}(h)\right\| \leq \|h\| = 1$, we conclude that $\left\|p_{H_1}(h)\right\| = 1$, and hence $p_{H_1}(h) \in \operatorname{suppv}(T|_{H_1})$. Now, notice that
$$1 = \|h\|^2 = \left\|p_{H_1}(h)\right\|^2 + \left\|p_{H_1^\perp}(h)\right\|^2 = 1 + \left\|p_{H_1^\perp}(h)\right\|^2,$$
meaning that $p_{H_1^\perp}(h) = 0$ and $h = p_{H_1}(h) \in \operatorname{suppv}(T|_{H_1})$.
3) Fix arbitrary elements $h_1 \in H_1$ and $k \in K$. Notice that
$$\begin{aligned}(h_1|(p_{H_1} \circ T^*)(k)) &= (h_1|p_{H_1}(T^*(k))) \\ &= (h_1|p_{H_1}(T^*(k))) + \left(h_1|p_{H_1^\perp}(T^*(k))\right) \\ &= \left(h_1|p_{H_1}(T^*(k)) + p_{H_1^\perp}(T^*(k))\right) \\ &= (h_1|T^*(k)) \\ &= (T(h_1)|k) \\ &= (T|_{H_1}(h_1)|k).\end{aligned}$$
By the uniqueness of the adjoint, we conclude that $(T|_{H_1})^* = p_{H_1} \circ T^*$.
4) Finally, let us show that $(T|_{H_1})^* \circ T|_{H_1} = (T^* \circ T)|_{H_1}$ whenever H_1 is an invariant subspace of $T^* \circ T$. Indeed, fix an arbitrary $h_1 \in H_1$. Then,
$$(T^* \circ T)|_{H_1}(h_1) = T^*(T|_{H_1}(h_1))$$

$$
\begin{aligned}
&= p_{H_1}\left(T^*(T|_{H_1}(h_1))\right) \\
&= \left((p_{H_1} \circ T^*) \circ T|_{H_1}\right)(h_1) \\
&= \left((T|_{H_1})^* \circ T|_{H_1}\right)(h_1).
\end{aligned}
$$

■

If X is a normed space, $T : X \to X$ is a continuous linear operator, and $\lambda \in \sigma_p(T)$, then it is clear that $T\left(\mathsf{V}(\lambda)\right) \subseteq \mathsf{V}(\lambda)$. The following lemma is also well known, yet we include the proof for the sake of completeness.

Lemma 3.14. *Let H be a Hilbert space, and $T : H \to H$ is a self-adjoint continuous linear operator. Let $\lambda \in \sigma_p(T)$. Then,*

1) $T\left(\mathsf{V}(\lambda)^\perp\right) \subseteq \mathsf{V}(\lambda)^\perp$.
2) $\sigma_p\left(T|_{\mathsf{V}(\lambda)^\perp}\right) = \sigma_p(T) \setminus \{\lambda\}$.

Proof.

1) Fix arbitrary elements $w \in \mathsf{V}(\lambda)^\perp$ and $v \in \mathsf{V}(\lambda)$. Observe that

$$(T(w)|v) = (w|T(v)) = (w|\lambda v) = \overline{\lambda}(w|v) = 0.$$

This shows that $T(w) \in \mathsf{V}(\lambda)^\perp$. The arbitrariness of $w \in \mathsf{V}(\lambda)^\perp$ serves to assure that $T\left(\mathsf{V}(\lambda)^\perp\right) \subseteq \mathsf{V}(\lambda)^\perp$.

2) Take any $\gamma \in \sigma_p\left(T|_{\mathsf{V}(\lambda)^\perp}\right)$. There exists $w \in \mathsf{V}(\lambda)^\perp \setminus \{0\}$ such that $T(w) = \gamma w$. It is clear that $\gamma \in \sigma_p(T)$. If $\gamma = \lambda$, then $w \in \mathsf{V}(\lambda)$, meaning the contradiction that $w = 0$. Conversely, take any $\gamma \in \sigma_p(T) \setminus \{\lambda\}$. There exists $w \in H \setminus \{0\}$ such that $T(w) = \gamma w$. It suffices to show that $w \in \mathsf{V}(\lambda)^\perp$. Indeed, since $\lambda \neq \gamma$, either λ or γ is not 0. We can assume without any loss of generality that $\gamma \neq 0$. Then, for every $v \in \mathsf{V}(\lambda)$, we have that

$$(w|v) = \frac{1}{\gamma}(\gamma w|v) = \frac{1}{\gamma}(T(w)|v) = \frac{1}{\gamma}(w|T(v)) = \frac{1}{\gamma}(w|\lambda v) = \frac{\lambda}{\gamma}(w|v).$$

Since $\lambda \neq \gamma$, the only possibility is that $(w|v) = 0$. As a consequence, $w \in \mathsf{V}(\lambda)^\perp$.

■

Theorem 3.15. *Let H, K be Hilbert spaces and $T : H \to K$ be a compact linear operator. If $\lambda \in \sigma_p(T^* \circ T)$ is the second largest eigenvalue of $T^* \circ T$, then $\lambda = \left\| T|_{\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp} \right\|^2$ and*

$$\mathsf{V}_{T^*\circ T}(\lambda) \cap \mathsf{S}_H = \operatorname{suppv}\left(T|_{\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp}\right) = \operatorname{suppv}\left(T \circ p_{\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp}\right).$$

Proof. In the first place, $T^* \circ T : H \to H$ is self-adjoint, positive, and compact. According to Theorem 2.3, $\|T\|^2 = \|T^* \circ T\| \in \sigma_p(T^* \circ T)$, and $\|T\|^2$ is the largest eigenvalue of $T^* \circ T$. By applying Lemma 3.14, we have that $(T^* \circ T)\left(\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp\right) \subseteq \mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp$ and

$$\sigma_p\left((T^* \circ T)|_{\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp}\right) = \sigma_p(T^* \circ T) \setminus \left\{\|T\|^2\right\}.$$

As a consequence, λ is the largest eigenvalue of $(T^* \circ T)|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}$. Next, since $\mathsf{V}_{T^*\circ T}\left(\|T\|^2\right)^\perp$ is an invariant subspace of $T^* \circ T$, Theorem 3.13(4) allows that

$$(T^* \circ T)|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp} = \left(T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right)^* \circ T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}.$$

This means that λ is the largest eigenvalue of $\left(T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right)^* \circ T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}$. By relying again on Theorem 2.3 and on the fact that $T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}$ is also compact, we obtain that $\lambda = \left\|T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right\|^2$ and

$$\mathsf{V}_{T^*\circ T}(\lambda) \cap \mathsf{S}_H = \operatorname{suppv}\left(T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right).$$

Finally, Theorem 3.13(2) assures that

$$\mathsf{V}_{T^*\circ T}(\lambda) \cap \mathsf{S}_H = \operatorname{suppv}\left(T|_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right) = \operatorname{suppv}\left(T \circ p_{\mathsf{V}_{T^*\circ T}(\|T\|^2)^\perp}\right).$$

■

3.4. Supporting vectors of quotient operators

Let X be a Banach space and $M \subseteq X$ be a closed subspace. The quotient space of X by M is defined by $X/M := \{x + M : x \in X\}$ and becomes a Banach space endowed with the distance-to-M norm $\|x + M\| := d(x, M) := \inf\{\|x - m\| : m \in M\}$. The canonical projection of X onto X/M is given by

$$\begin{aligned} \pi_M : \quad X &\to X/M \\ x &\mapsto \pi_M(x) := x + M, \end{aligned} \tag{3.4.1}$$

and it is clearly a linear operator of norm $\|\pi_M\| = 1$. The closed subspace M is called proximinal provided that for every $x \in X$, the distance from x to M is attained at some $m_0 \in M$, that is, $d(x, M) = \|x - m_0\|$.

Let X, Y be Banach spaces and $T : X \to Y$ be a continuous linear operator. Let $M \subseteq \ker(T)$ be a closed subspace. The quotient operator of T

$$\begin{aligned} T_M : \quad X/M &\to Y \\ x + M &\mapsto T_M(x + M) := T(x) \end{aligned} \tag{3.4.2}$$

is a well-defined continuous linear operator. Notice that when M is chosen to be $\ker(T)$, we obtain the First Isomorphism Theorem. The following theorem relates the supporting vectors of T with those of T_M.

Theorem 3.16. *Let X, Y be Banach spaces and $T : X \to Y$ be a continuous linear operator. Let $M \subseteq \ker(T)$ be a closed subspace. Then,*

1) $\|T_M\| = \|T\|$.
2) $\pi_M(\operatorname{suppv}(T)) \subseteq \operatorname{suppv}(T_M)$.
3) If M is proximinal, then $\pi_M(\operatorname{suppv}(T)) = \operatorname{suppv}(T_M)$.

Proof.

1) Indeed, if $\|x\| \leq 1$, then $\|x + M\| := d(x, M) \leq \|x - 0\| \leq 1$. Therefore, $\|T(x)\| = \|T_M(x + M)\| \leq \|T_M\|$. This shows that $\|T\| \leq \|T_M\|$. Conversely, if $\|x + M\| = 1$, then we can find a sequence $(m_k)_{k\in\mathbb{N}} \subseteq M$ such that $\|x - m_k\| \to d(x, M) = \|x + M\| = 1$ as $k \to \infty$. Notice that $\left\|T\left(\frac{x-m_k}{\|x-m_k\|}\right)\right\| \leq \|T\|$ for all $k \in \mathbb{N}$, that is, $\|T_M(x + M)\| = \|T(x)\| = \|T(x - m_k)\| \leq \|T\|\|x - m_k\|$ for all $k \in \mathbb{N}$. Since $\|x - m_k\| \to 1$ as $k \to \infty$, we conclude that $\|T_M(x + M)\| \leq \|T\|$. The arbitrariness of $x + M \in \mathsf{S}_{X/M}$ assures that $\|T_M\| \leq \|T\|$.
2) Fix an arbitrary $x \in \operatorname{suppv}(T)$. Notice that $\|x\| = 1$, so $\|x + M\| = d(x, M) \leq \|x - 0\| = \|x\| = 1$. This shows that $x + M \in \mathsf{B}_{X/M}$. Next, $\|T_M(x + M)\| = \|T(x)\| = \|T\| = \|T_M\|$. As a consequence, $\|x + M\| \in \operatorname{suppv}(T_M)$.
3) Fix an arbitrary $x + M \in \operatorname{suppv}(T_M)$. Since M is proximinal, there exists $m_0 \in M$ such that $1 = \|x + M\| = d(x, M) = \|x - m_0\|$. We will show that $x - m_0 \in \operatorname{suppv}(T)$. Indeed, $\|x - m_0\| = 1$, and $\|T(x - m_0)\| = \|T(x)\| = \|T_M(x + M)\| = \|T_M\| = \|T\|$. Finally, $\pi_M(x - m_0) = x + M$.

■

3.5. Applications of SVA/PCA to real-life situations

In this section, we present the application of our SVA/PCA theorems to solve a real-life problem focused on the distribution of political sensitivities with several economic variables. For that, we have used the results of the 2018 Andalusian (Spain) elections. The data analyzed is provided by the Institute of Statistics and Cartography of Andalusia [17] and combines information from 153 municipalities with more than ten thousand inhabitants and 8 measured variables. The data include unemployment rate, aging rate and 6 generalist policy options ranging from left to right wing. Note that the variables in columns were previously centered.

A Graphical User Interface (GUI) has been developed in Python code and aims to compute the eigenvectors, including the supporting vector (first eigenvector), and the principal components. Figure 1 shows the GUI of the PCA input, where the data matrix is introduced. It is important to note that the input data can be imported by a CSV file. This option is very useful when we are working with large matrices. The results obtained are shown in Figure 2, which can also be exported as a CSV file.

First of all, we note that the $A \in \mathcal{M}_{153\times 8}(\mathbb{R})$ data matrix is centered in the origin by subtracting the mean. Then, we calculate the covariance matrix for the centered matrix. Afterwards, we compute the eigenvalues and eigenvectors. In particular, as mentioned in this paper, the eigenvector associated with the maximum eigenvalue corresponds to the supporting vector. These eigenvalues and eigenvectors are sorted in descending order, so the first element corresponds to the supporting vector and the first principal component. Finally, the products of the initial matrix and the eigenvectors are done, obtaining the principal components.

In particular, focusing on the example of the Andalusian elections, the results show the correspondence of the first principal component with the classic political sensitivities associated with old-aging municipalities. On the other hand, the second principal component is related to the differentiation between left-wing and right-wing political options.

In Figure 3, the 8 variables used in the example are shown in the eigenvector reference system with the supporting vector and the second eigenvector. The representation of the 153 municipalities in the principal component reference system (Ax) is shown in Figure 4. This 2D image with the first and the second principal component group the municipalities according to the political sensitivities and their demographic characteristics.

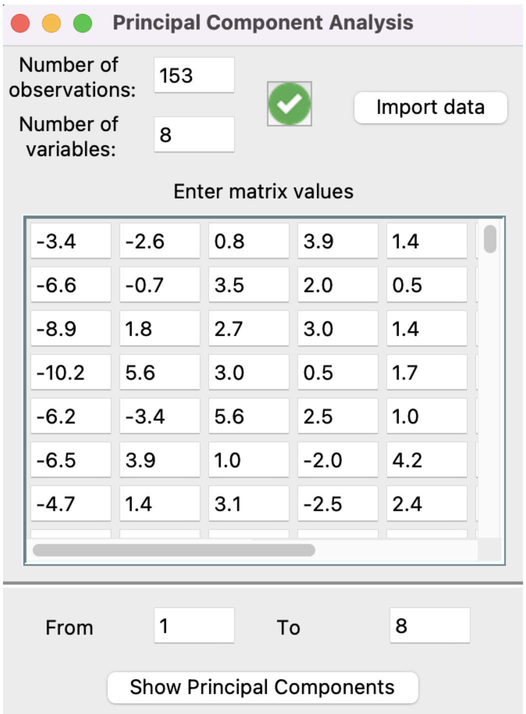
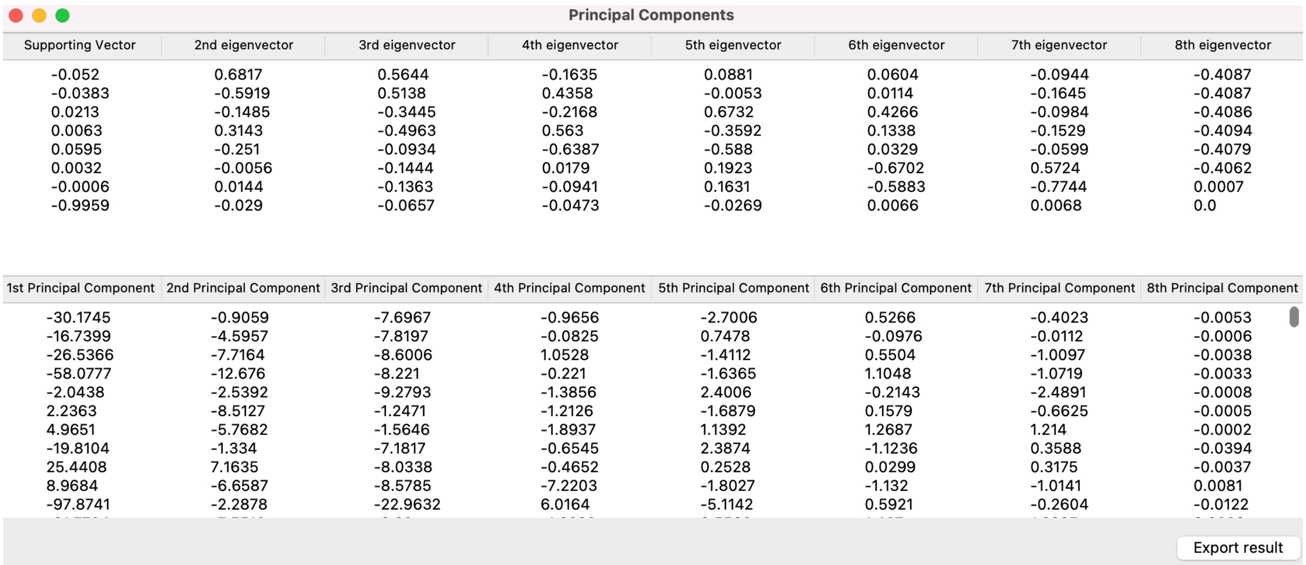


Figure 1. GUI of the PCA input for a matrix $A \in \mathcal{M}_{153\times 8}(\mathbb{R})$.



Principal Components

Supporting Vector	2nd eigenvector	3rd eigenvector	4th eigenvector	5th eigenvector	6th eigenvector	7th eigenvector	8th eigenvector
-0.052	0.6817	0.5644	-0.1635	0.0881	0.0604	-0.0944	-0.4087
-0.0383	-0.5919	0.5138	0.4358	-0.0053	0.0114	-0.1645	-0.4087
0.0213	-0.1485	-0.3445	-0.2168	0.6732	0.4266	-0.0984	-0.4086
0.0063	0.3143	-0.4963	0.563	-0.3592	0.1338	-0.1529	-0.4094
0.0595	-0.251	-0.0934	-0.6387	-0.588	0.0329	-0.0599	-0.4079
0.0032	-0.0056	-0.1444	0.0179	0.1923	-0.6702	0.5724	-0.4062
-0.0006	0.0144	-0.1363	-0.0941	0.1631	-0.5883	-0.7744	0.0007
-0.9959	-0.029	-0.0657	-0.0473	-0.0269	0.0066	0.0068	0.0

1st Principal Component	2nd Principal Component	3rd Principal Component	4th Principal Component	5th Principal Component	6th Principal Component	7th Principal Component	8th Principal Component
-30.1745	-0.9059	-7.6967	-0.9656	-2.7006	0.5266	-0.4023	-0.0053
-16.7399	-4.5957	-7.8197	-0.0825	0.7478	-0.0976	-0.0112	-0.0006
-26.5366	-7.7164	-8.6006	1.0528	-1.4112	0.5504	-1.0097	-0.0038
-58.0777	-12.676	-8.221	-0.221	-1.6365	1.1048	-1.0719	-0.0033
-2.0438	-2.5392	-9.2793	-1.3856	2.4006	-0.2143	-2.4891	-0.0008
2.2363	-8.5127	-1.2471	-1.2126	-1.6879	0.1579	-0.6625	-0.0005
4.9651	-5.7682	-1.5646	-1.8937	1.1392	1.2687	1.214	-0.0002
-19.8104	-1.334	-7.1817	-0.6545	2.3874	-1.1236	0.3588	-0.0394
25.4408	7.1635	-8.0338	-0.4652	0.2528	0.0299	0.3175	-0.0037
8.9684	-6.6587	-8.5785	-7.2203	-1.8027	-1.132	-1.0141	0.0081
-97.8741	-2.2878	-22.9632	6.0164	-5.1142	0.5921	-0.2604	-0.0122

Export result

Figure 2. GUI of the PCA output for a matrix $A \in \mathcal{M}_{153\times 8}(\mathbb{R})$.

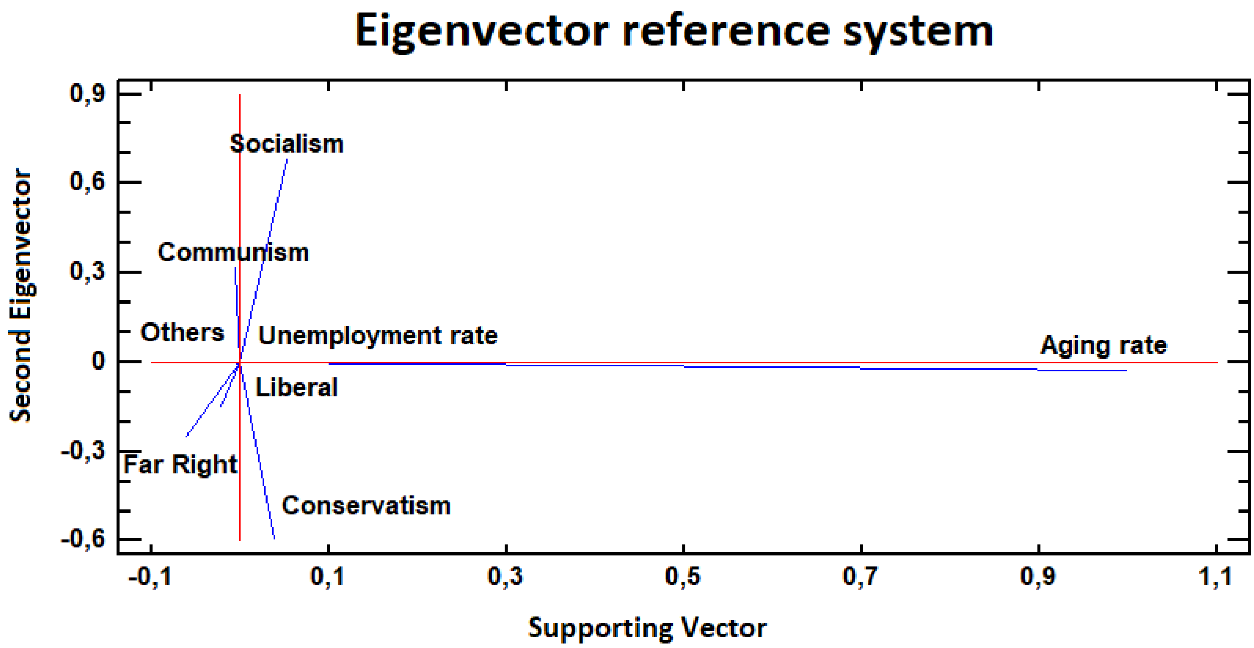


Figure 3. Variables represented in the eigenvector reference system with the supporting vector and the second eigenvector.

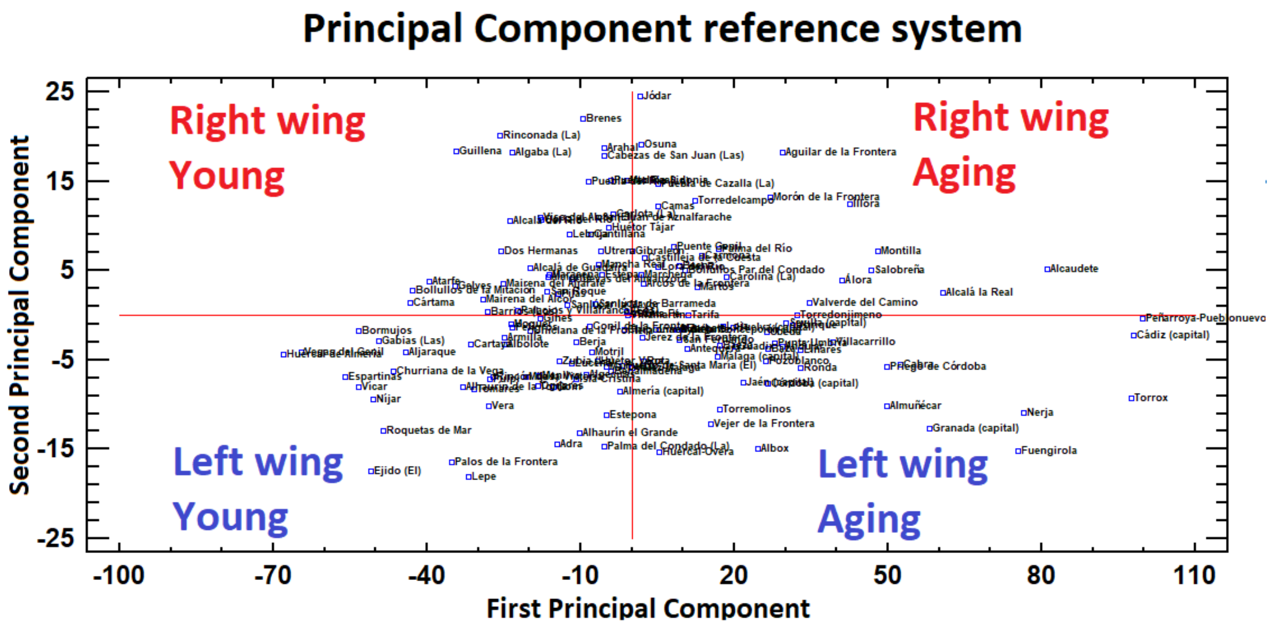


Figure 4. Municipalities represented in the principal component reference system (first and second PC).

4. Discussion

We will discuss how to transport the mean functional, the mean operator and the centering operator to abstract settings, such as Hilbert spaces, Banach algebras and probability spaces.

4.1. Generalization to Hilbert spaces

As mentioned in Lemma 3.2, in view of the Riesz Representation Theorem, if $h_0 := \frac{\mathbf{1}}{\mathbf{n}} = \left(\frac{1}{n}, \overset{n}{\ldots}, \frac{1}{n}\right)$, then $h_0^* = \overline{\cdot}$ is precisely the mean functional in $H := \ell_2^n$. Actually, $\|h_0\|_2 = \frac{1}{\sqrt{n}} = \|h_0^*\|$. At this stage, the key is to realize that h_0 can be seen as a (finite) convex series. Recall that a convex series is a convergent series $\sum_{n=1}^{\infty} t_n$ such that $\sum_{n=1}^{\infty} t_n = 1$ and $t_n \geq 0$ for all $n \in \mathbb{N}$. Notice that if $\sum_{n=1}^{\infty} t_n$ is a convex series, then $\sum_{n=1}^{\infty} t_n^2 \leq \sum_{n=1}^{\infty} t_n = 1$, and hence $(t_n)_{n\in\mathbb{N}} \in \ell_2$. Now, we are in the right position to define the notions of mean and standard deviation on separable Hilbert spaces.

Let H be a separable Hilbert space with an orthonormal basis $(e_n)_{n\in\mathbb{N}}$. Fix a convex series $\sum_{n=1}^{\infty} t_n$. Let $h \in H$ and write $h = \sum_{n=1}^{\infty} (h|e_n)\, e_n$. The mean of h, with respect to $(t_n)_{n\in\mathbb{N}}$, is defined as

$$\overline{h} := \sum_{n=1}^{\infty} t_n\, (h|e_n)\,. \tag{4.1.1}$$

The mean functional, with respect to $(t_n)_{n\in\mathbb{N}}$, is given by

$$\begin{array}{rccl} \overline{\cdot}: & H & \to & \mathbb{K} \\ & h & \mapsto & \overline{h} := \sum_{n=1}^{\infty} t_n\, (h|e_n)\,. \end{array} \tag{4.1.2}$$

By relying on Hölder's Inequality, it is not hard to check that the mean functional is an element of H^* whose norm is precisely $\sqrt{\sum_{n=1}^{\infty} t_n^2} = \|(t_n)_{n\in\mathbb{N}}\|_2$. In accordance with the Riesz Representation Theorem, there exists $h_0 \in H$ such that $(h|h_0) = \overline{h}$ for all $h \in H$. If we let $h = \sum_{n=1}^{\infty} (h|e_n)\, e_n$, then we obtain that

$$\sum_{n=1}^{\infty} (h|e_n)\,(e_n|\, h_0) = \left(\sum_{n=1}^{\infty} (h|e_n)\, e_n \,\middle|\, h_0 \right) = \sum_{n=1}^{\infty} t_n\, (h|e_n)\,. \tag{4.1.3}$$

By taking $h = e_n$ for every $n \in \mathbb{N}$ in (4.1.3), we conclude that $(e_n|h_0) = t_n$ for every $n \in \mathbb{N}$. In particular, $h_0 = \sum_{n=1}^{\infty} t_n e_n$. As expected,

$$\|h_0\| = \|(t_n)_{n\in\mathbb{N}}\|_2 = \sqrt{\sum_{n=1}^{\infty} t_n^2} = \|\overline{\cdot}\|.$$

In view of the Riesz Representation Theorem, $h_0^* = (\cdot|h_0) = \overline{\cdot}$. Notice that, in order to conclude that $h_0^* = (\cdot|h_0) = \overline{\cdot}$, the Riesz Representation Theorem is really not needed. Let us get back for a second to ℓ_2^n. For every $x \in \ell_2^n$,

$$\overline{\mathbf{x}} = (\overline{x}, \overset{n}{\ldots}, \overline{x}) = \overline{x}\mathbf{1} = \frac{\overline{x}}{\frac{1}{\sqrt{n}}} \frac{\frac{\mathbf{1}}{\mathbf{n}}}{\frac{1}{\sqrt{n}}} = \frac{\overline{\cdot}}{\frac{1}{\sqrt{n}}}(x) \frac{\frac{\mathbf{1}}{\mathbf{n}}}{\frac{1}{\sqrt{n}}}.$$

Then, going back to a general separable Hilbert space H, the mean operator is defined as

$$\begin{array}{rcl} H & \to & H \\ h & \mapsto & \overline{\mathbf{h}} := \frac{h_0^*}{\|h_0^*\|}(h)\frac{h_0}{\|h_0\|}. \end{array} \tag{4.1.4}$$

It is not hard to check that the mean operator is an orthogonal projection on H whose complementary projection is, precisely, the centering operator:

$$\begin{array}{rcl} H & \to & H \\ h & \mapsto & \operatorname{cen}(h) := h - \overline{\mathbf{h}}. \end{array} \tag{4.1.5}$$

4.2. Generalization to Banach algebras

A Banach algebra is a real or complex algebra A endowed with a complete vector norm that is also a ring norm, that is, $\|ab\| \leq \|a\|\|b\|$ for all $a, b \in A$. We say that A is unital if it is unitary, that is, it has a unity $\mathbf{1} \in A$, and $\|\mathbf{1}\| = 1$. In this situation, according to the Hahn-Banach Theorem, there exists a continuous linear functional, which we will denote by $\mathbf{1}^* \in A^*$, such that $\|\mathbf{1}^*\| = 1$ and $\mathbf{1}^*(\mathbf{1}) = 1$. Then, the mean functional is precisely $\mathbf{1}^*$, that is, defined by

$$\begin{array}{rcl} A & \to & \mathbb{K} \\ a & \mapsto & \bar{a} := \mathbf{1}^*(a). \end{array} \tag{4.2.1}$$

The mean operator is defined as

$$\begin{array}{rcl} A & \to & A \\ a & \mapsto & \overline{\mathbf{a}} := \mathbf{1}^*(a)\mathbf{1}. \end{array} \tag{4.2.2}$$

Finally, the centering operator is defined as

$$\begin{array}{rcl} A & \to & A \\ a & \mapsto & \operatorname{cen}(a) := a - \overline{\mathbf{a}}. \end{array} \tag{4.2.3}$$

Following (3.1.2), the variance of an element $a \in A$ can be defined by

$$s_a^2 := \mathbf{1}^*\left((a - \overline{\mathbf{a}})^2\right) = \mathbf{1}^*(a^2) - \mathbf{1}^*(a)^2. \tag{4.2.4}$$

Notice that this generalization to unital Banach algebras presents a weakness: The existence of the functional $\mathbf{1}^* \in A^*$ is guaranteed by the Hahn-Banach Theorem, but its uniqueness is not guaranteed. In fact, in many unital Banach algebras, such as $\ell_\infty(\Lambda)$ for instance, $\mathbf{1}$ is not a smooth point of $\mathsf{B}_{\ell_\infty(\Lambda)}$ [16, Theorem 2.9], and thus there are infinitely many functionals of norm 1 attaining their norm at $\mathbf{1}$. It seems not trivial to overcome this issue. Maybe an option is to try to renorm equivalently the unital Banach algebra in such a way that $\mathbf{1}$ becomes a smooth point of the new unit ball of the algebra or, at least, to find another smooth point in the unit ball. According to [16, Theorem 2.1], the canonical unit vector e_λ is a smooth point of $\mathsf{B}_{\ell_\infty(\Lambda)}$ for each $\lambda \in \Lambda$. Another possibility may rely on constructing the mean functional in a C^*-algebra. Recall that a C^*-algebra is a Banach algebra A endowed with an antimultiplicative and antilinear involution $* : A \to A$ satisfying that $\|aa^*\| = \|a\|^2$ for all $a \in A$.

4.3. Generalization to probability spaces

A probability space is a 3-tuple $(\Omega, \Sigma, \mathbb{P})$ where (Ω, Σ) is a measurable space and $\mathbb{P} : \Sigma \to [0, 1]$ is a probability measure, that is, a countably additive positive measure such that $\mathbb{P}(\Omega) = 1$. If X is a Banach space, by $L^1((\Omega, \Sigma, \mathbb{P}), X)$ we denote the Banach space of all absolutely integrable functions, that is,

$$L^1((\Omega, \Sigma, \mathbb{P}), X) := \left\{ f \in X^\Omega : f \text{ is measurable and } \int_\Omega \|f\| d\mathbb{P} < \infty \right\},$$

endowed with the norm

$$\|f\|_1 := \int_\Omega \|f\| d\mathbb{P}.$$

For each $f \in L^1((\Omega, \Sigma, \mathbb{P}), X)$, the mean of f is defined as

$$\mu(f) := \int_\Omega f d\mathbb{P}. \tag{4.3.1}$$

The mean functional is given by

$$\begin{array}{rccl} \mu: & L^1((\Omega, \Sigma, \mathbb{P}), X) & \to & X \\ & f & \mapsto & \mu(f) := \int_\Omega f d\mathbb{P}. \end{array} \tag{4.3.2}$$

Notice that the mean functional is actually an operator, but we keep calling it functional not to mistake it with the mean operator, which will be defined next. It can be easily shown that the mean functional has norm equal to 1. Indeed,

$$\|\mu(f)\| = \left\| \int_\Omega f d\mathbb{P} \right\| \leq \int_\Omega \|f\| d\mathbb{P} = \|f\|_1$$

for every $f \in L^1((\Omega, \Sigma, \mathbb{P}), X)$, meaning that $\|\mu\| \leq 1$. Now, if we choose any $x \in \mathsf{S}_X$, then $x\chi_\Omega \in \mathsf{S}_{L^1((\Omega,\Sigma,\mathbb{P}),X)}$, and

$$\mu(x\chi_\Omega) = \int_\Omega x\chi_\Omega d\mathbb{P} = x\mathbb{P}(\Omega) = x.$$

Hence,

$$\|\mu(x\chi_\Omega)\| = \left\| \int_\Omega x\chi_\Omega d\mathbb{P} \right\| = \|x\|\mathbb{P}(\Omega) = 1 = \|x\chi_\Omega\|_1 .$$

This proves that $\|\mu\| = 1$ and $x\chi_\Omega \in \operatorname{suppv}(\mu)$. The mean operator is then defined as

$$\begin{array}{rcl} L^1((\Omega, \Sigma, \mathbb{P}), X) & \to & L^1((\Omega, \Sigma, \mathbb{P}), X) \\ f & \mapsto & \mu(f)\chi_\Omega, \end{array} \tag{4.3.3}$$

and the centering operator is

$$\begin{array}{rcl} L^1((\Omega, \Sigma, \mathbb{P}), X) & \to & L^1((\Omega, \Sigma, \mathbb{P}), X) \\ f & \mapsto & f - \mu(f)\chi_\Omega. \end{array} \tag{4.3.4}$$

Finally, if X is a unital Banach algebra, then the natural way of defining the variance of $f \in L^1((\Omega, \Sigma, \mathbb{P}), X)$ is

$$\sigma(f) := \mu\left((f - \mu(f)\chi_\Omega)^2\right) = \int_\Omega (f - \mu(f)\chi_\Omega)^2 \, d\mathbb{P}. \tag{4.3.5}$$

5. Conclusions

It is well known in the literature of the Geometry of Banach Spaces that Hilbert spaces are transitive Banach spaces, meaning that every two elements of the unit sphere of a Hilbert space can be transported one into another by means of a surjective linear isometry. This fact confers Hilbert spaces with a certain freedom when it comes to choosing the convex series that defines the mean functional, the mean operator, the centering operator and the standard deviation. Theorem 3.1, Lemma 3.2 and Theorem 3.4 can be transported to the more general scope provided by separable Hilbert spaces discussed in the previous section. In fact, in the Discussion we unveiled how to extend the mean functional, the mean operator, the centering operator and the variance to spaces of absolutely integrable functions defined on a probability space and valued on a unital Banach algebra.

On the other hand, the study of Principal Component Analysis through Supporting Vector Analysis is a revolutionary trend that allows one to look at these Statistical concepts from a Functional Analysis viewpoint, which is more general and works in infinite dimensional environments, making possible applications in very specific settings such as, for instance, Quantum Mechanical Systems.

Acknowledgments

This work has been partially supported by the Research Grant PGC-101514-B-I00 awarded by the Ministry of Science, Innovation and Universities of Spain and co-financed by the 2014-2020 ERDF Operational Programme and by the Department of Economy, Knowledge, Business and University of the Regional Government of Andalusia under Project reference FEDER-UCA18-105867. The APCs have been paid by the Department of Mathematics of the University of Cadiz.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. E. Bishop, R. R. Phelps, A proof that every Banach space is subreflexive, *Bull. Amer. Math. Soc.*, **67** (1961), 97–98. https://doi.org/10.1090/S0002-9904-1961-10514-4
2. E. Bishop, R. R. Phelps, The support functionals of a convex set, In: *Proceedings of Symposia in Pure Mathematics, Vol. VII*, Providence, R.I.: Amer. Math. Soc., 1963, 27–35.
3. T. Bouwmans, S. Javed, H. Zhang, Z. Lin, R. Otazo, On the applications of robust PCA in image and video processing, *Proc. IEEE*, **106** (2018), 1427–1457. https://doi.org/10.1109/JPROC.2018.2853589
4. C. Cobos-Sánchez, F. J. Garcia-Pacheco, J. M. Guerrero Rodriguez, J. R. Hill, An inverse boundary element method computational framework for designing optimal TMS coils, *Eng. Anal. Bound. Elem.*, **88** (2018), 156–169. https://doi.org/10.1016/j.enganabound.2017.11.002
5. C. Cobos-Sánchez, F. J. García-Pacheco, S. Moreno-Pulido, S. Sáez-Martínez, Supporting vectors of continuous linear operators, *Ann. Funct. Anal.*, **8** (2017), 520–530. https://doi.org/10.1215/20088752-2017-0016
6. C. Cobos-Sánchez, J. A. Vilchez-Membrilla, A. Campos-Jiménez, F. J. García-Pacheco, Pareto optimality for multioptimization of continuous linear operators, *Symmetry*, **13** (2021), 661. https://doi.org/10.3390/sym13040661

7. C. Cobos-Sánchez, M. R. Cabello, Á. Q. Olozábal, M. F. Pantoja, Design of TMS coils with reduced lorentz forces: application to concurrent TMS-fMRI, *J. Neural Eng.*, **17** (2020), 016056. https://doi.org/10.1088/1741-2552/ab4ba2

8. C. Cobos-Sánchez, J. J. J. García, M. R. Cabello, M. F. Pantoja, Design of coils for lateralized TMS on mice, *J. Neural Eng.*, **17** (2020), 036007. https://doi.org/10.1088/1741-2552/ab89fe

9. C. Cobos-Sánchez, F. J. Garcia-Pacheco, J. M. Guerrero-Rodriguez, L. Garcia-Barrachina, Solving an IBEM with supporting vector analysis to design quiet TMS coils, *Eng. Anal. Bound. Elem.*, **117** (2020), 1–12. https://doi.org/10.1016/j.enganabound.2020.04.013

10. C. Cobos-Sánchez, J. M. Guerrero-Rodriguez, Á. Q. Olozábal, D. Blanco-Navarro, Novel TMS coils designed using an inverse boundary element method, *Phys. Med. Biol.*, **62** (2016), 73–90. https://doi.org/10.1088/1361-6560/62/1/73

11. C. M. Epstein, E. Wassermann, U. Ziemann, *Oxford Handbook of Transcranial Stimulation*, New York: Oxford University Press, 2008. https://doi.org/10.1093/oxfordhb/9780198568926.001.0001

12. J. Fan, Q. Sun, W.-X. Zhou, Z. Zhu, Principal component analysis for big data, *Wiley StatsRef: Statistics Reference Online*, in press. https://doi.org/10.1002/9781118445112.stat08122

13. F. J. García-Pacheco, E. Naranjo-Guerra, Supporting vectors of continuous linear projections, *International Journal of Functional Analysis, Operator Theory and Applications*, **9** (2017), 85–95.

14. F. J. García-Pacheco, Lineability of the set of supporting vectors, *RACSAM*, **115** (2021), 41, https://doi.org/10.1007/s13398-020-00981-6

15. F. J. Garcia-Pacheco, C. Cobos-Sánchez, S. Moreno-Pulido, A. Sanchez-Alzola, Exact solutions to $\max_{\|x\|=1} \sum_{i=1}^{\infty} \|T_i(x)\|^2$ with applications to Physics, Bioengineering and Statistics, *Commun. Nonlinear Sci. Numer. Simul.*, **82** (2020), 105054. https://doi.org/10.1016/j.cnsns.2019.105054

16. F.-K. Garsiya-Pacheko, The cardinality of the set Λ determines the geometry of the spaces $B_{\ell_\infty(\Lambda)}$ and $B_{\ell_\infty(\Lambda)^*}$, (Russian), *Funktsional. Anal. i Prilozhen.*, **52** (2018), 62–71. https://doi.org/10.4213/faa3534

17. Instituto de Estadística y Cartografía de Andalucía. Available from: `https://www.juntadeandalucia.es/institutodeestadisticaycartografia`.

18. R. C. James, Characterizations of reflexivity, *Stud. Math.*, **23** (1964), 205–216. https://doi.org/10.4064/sm-23-3-205-216

19. J. Lindenstrauss, On operators which attain their norm, *Israel J. Math.*, **1** (1963), 139–148. https://doi.org/10.1007/BF02759700

20. L. Marin, H. Power, R. W. Bowtell, C. Cobos-Sánchez, A. A. Becker, P. Glover, et al., Numerical solution of an inverse problem in magnetic resonance imaging using a regularized higher-order boundary element method, In: *Boundary elements and other mesh reduction methods XXIX*, Southampton: WIT Press, 2007, 323–332. https://doi.org/10.2495/BE070311

21. L. Marin, H. Power, R. W. Bowtell, C. Cobos-Sánchez, A. A. Becker, P. Glover, et al., Boundary element method for an inverse problem in magnetic resonance imaging gradient coils, *CMES Comput. Model. Eng. Sci.*, **23** (2008), 149–173. https://doi.org/10.3970/cmes.2008.023.149

22. S. Moreno-Pulido, F. J. Garcia-Pacheco, C. Cobos-Sánchez, A. Sanchez-Alzola, Exact solutions to the maxmin problem $\max \|ax\|$ subject to $\|bx\| \leq 1$, *Mathematics*, **8** (2020), 85. http://doi.org/10.3390/math8010085

23. A. Sánchez-Alzola, F. J. García-Pacheco, E. Naranjo-Guerra, S. Moreno-Pulido, Supporting vectors for the ℓ_1-norm and the ℓ_∞-norm and an application, *Math. Sci.*, **15** (2021), 173–187. https://doi.org/10.1007/s40096-021-00400-w

24. L. Surhone, M. Timpledon, S. Marseken, *Principal component analysis: Karhunen-Loève Theorem, Harold Hotelling, Karl Pearson, Exploratory Data Analysis, Eigendecomposition of a Matrix, Covariance Matrix, Singular Value Decomposition, Factor Analysis*, Betascript Publishing, 2010.

Supplementary: Python GUI code

Let $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ be a general matrix. The algorithm to compute the supporting vector with its corresponding first principal component, the second eigenvector with its corresponding second principal component and the next eigenvectors and principal components of A is the following:

```
def PCA(X, num_components):
    X_meaned = X - np.mean(X, axis=0)
    cov_mat = np.cov(X_meaned, rowvar=False)
    eigen_values, eigen_vectors = np.linalg.eigh(cov_mat)
    sorted_index = np.argsort(eigen_values)[::-1]
    sorted_eigenvectors = eigen_vectors[:, sorted_index]
    eigenvector_subset = sorted_eigenvectors[:, 0:num_components]
    X_new_values = np.dot(eigenvector_subset.transpose(),
                   X_meaned.transpose()).transpose()
    return X_new_values, eigenvector_subset
```

One of the novelties of this work is to apply the PCA method with a different procedure specifically, using an algorithm based on the mathematical idea that the second eigenvector, with its associated second principal component, is the supporting vector of the original points projected to the orthogonal complement of the original supporting vector. Consequently, all the principal components can be computed in an iterative process via a supporting vector.

Let $x \in \mathbb{R}^n$ be a vector. First, we show the following function to calculate the orthogonal complement of x:

```
def calculate_orthogonal_complement(x, normalize=True, threshold=1e-15):
    x = np.asarray(x)
    r, c = x.shape
    if r < c:
        import warnings
        warnings.warn('fewer rows than columns', UserWarning)
    s, v, d = np.linalg.svd(x)
    rank = (v > threshold).sum()
```

```
        oc = s[:, rank:]
        if normalize:
            k_oc = oc.shape[1]
            oc = oc.dot(np.linalg.inv(oc[:k_oc, :]))
        return oc
```

Next, by using the previous function, we calculate the orthogonal complement of the supporting vector, representing a hyperplane with dimension $n-1$. Afterwards, the points of the original matrix, used for the initial supporting vector, are projected to this hyperplane. Hence, a new matrix is derived, whose first principal component corresponds to the second principal component of the original matrix. Thus, by applying this iterative algorithm, we can calculate all the principal components via the orthogonal complement of a supporting vector.

It is easier to understand this algorithm considering a special case. Let $A \in \mathcal{M}_{3\times 3}(\mathbb{R})$ be a matrix. In an intuitive manner, each row of the matrix represents points of $\mathbb{R}^3$. Now, we present a process of computing the first eigenvector (supporting vector) and its corresponding first principal component as before, but with a different procedure to obtain the followings. As mentioned, we notice that the first principal component via a supporting vector of a derived matrix from the original one coincides with the second principal component of the original matrix. This derived matrix is constructed projecting the original points of the original matrix to a plane formed by the orthogonal complement (two vectors) of the initial supporting vector (one vector) and the origin point. Therefore, we obtain the same result by applying the PCA presented before and this algorithm determining the next eigenvector as a supporting vector of the original supporting vector.

```
    supporting_vector = PCA(input_matrix, 1)
    orthogonal_complement = calculate_orthogonal_complement(supporting_vector)
    v1 = Vector(orthogonal_complement.T[0, :])
    v2 = Vector(orthogonal_complement.T[1, :])
    point = Point([0, 0, 0])
    plane = Plane.from_vectors(point, v1, v2)
    projected_points = np.array([plane.project_point(x) for x in input_matrix])
    supporting_vector_projected_points = PCA(projected_points, 1)
```

AIMS Press