J. Math. Anal. Appl. 407 (2013) 350–358

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Characterization of Birkhoff–James orthogonality

Tirthankar Bhattacharyya [a], Priyanka Grover [b,*]

[a] Department of Mathematics, Indian Institute of Science, Bangalore-560012, India

[b] Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi Centre, 7, S.J.S. Sansanwal Marg, New Delhi-110016, India

ARTICLE INFO

Article history:
Received 7 February 2013
Available online 16 May 2013
Submitted by R.M. Aron

Keywords:
Birkhoff–James orthogonality
Subdifferential
Hausdorff–Toeplitz theorem
Numerical range
Separating Hyperplane theorem
Singular value decomposition
Operator tuples
C^*-algebra
Hilbert C^*-module
State
Variance

ABSTRACT

The Birkhoff–James orthogonality is a generalization of Hilbert space orthogonality to Banach spaces. We investigate this notion of orthogonality when the Banach space has more structures. We start by doing so for the Banach space of square matrices moving gradually to all bounded operators on any Hilbert space, then to an arbitrary C^*-algebra and finally a Hilbert C^*-module.

1. Introduction

Let X be a complex Banach space. An element $x \in X$ is said to be *Birkhoff–James orthogonal* to another element $y \in X$ if $\|x + \lambda y\| \geq \|x\|$ for all complex numbers λ. It is easy to see that Birkhoff–James orthogonality is equivalent to the usual orthogonality in case X is a Hilbert space. When $X = \mathbb{M}(n)$, the Banach space of all $n \times n$ complex square matrices, a very tractable condition of Birkhoff–James orthogonality was found by Bhatia and Šemrl in [4]. They showed that an $n \times n$ matrix A is Birkhoff–James orthogonal to an $n \times n$ matrix B if and only if there is a unit vector $x \in \mathbb{C}^n$ such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$. Here $\|A\|$ denotes the operator norm of A. Later Benitez, Fernandez and Soriano [3] showed that a necessary and sufficient condition for the norm of a real finite dimensional normed space X to be induced by an inner product is that for any $A, B \in \mathscr{B}(X)$, A is Birkhoff–James orthogonal to B if and only if there exists a unit vector $x \in X$ such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$.

Motivated by these, in this note we explore Birkhoff–James orthogonality in the setting of Hilbert C^*-modules. All inner products in this note are conjugate linear in the first component and linear in the second component. We start by giving a new proof of the Bhatia–Šemrl theorem using tools of convex analysis. This is more illuminating because it involves minimization of a certain convex function and is therefore a geometric approach. This is a new way of looking at the theorem and might be useful elsewhere. Another method was given by Kečkič in [7]. He first computes the φ-Gateaux derivative $D_{\varphi,A}(B)$ of the norm at A, in the B and φ directions, which is defined by $D_{\varphi,A}(B) = \lim_{t\to 0^+} \frac{\|A+te^{i\varphi}B\|-\|A\|}{t}$. Then he uses the property that A

* Corresponding author.
E-mail addresses: tirtha@member.ams.org (T. Bhattacharyya), priyanka.grover17@gmail.com, pgrover8r@isid.ac.in (P. Grover).

0022-247X/$ – see front matter
http://dx.doi.org/10.1016/j.jmaa.2013.05.022

is Birkhoff–James orthogonal to B if and only if $\inf_\varphi \mathrm{D}_{\varphi,A}(B) \geq 0$ to prove the Bhatia–Šemrl theorem. The unit vector x that crops up naturally suggests that this theorem should be generalizable to Hilbert C^*-modules with x replaced by a state of the underlying C^*-algebra. We handle the special module $\mathscr{B}(\mathscr{H}, \mathscr{K})$ in Section 3 with several applications to operator tuples. Section 4 is on arbitrary Hilbert C^*-modules.

These results can be applied to obtain some distance formulas in $\mathbb{M}(n)$ and other C^*-algebras. These are also important in problems related to derivations and operator approximations. In approximation theory the condition that A is Birkhoff–James orthogonal to B can be interpreted as follows. Suppose $A \in \mathbb{M}(n)$ is not in $\mathbb{C}B$, the subspace spanned by the matrix B. Then the zero matrix is the best approximation to A among all matrices in $\mathbb{C}B$.

Recently, Birkhoff–James orthogonality in Hilbert C^*-modules has been studied in [1] as our work was in progress. Some of our results overlap with them. Our approach is very different from that in [1]. We proceed gradually from square matrices to Hilbert C^*-modules. This is a natural development.

2. Bhatia–Šemrl theorem

The statement of the theorem is as follows.

Theorem 2.1. *Let $A, B \in \mathbb{M}(n)$. Then $\|A + \lambda B\| \geq \|A\|$ for all $\lambda \in \mathbb{C}$ if and only if there is a unit vector x such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$.*

We first deal with $\|A + tB\| \geq \|A\|$ for all $t \in \mathbb{R}$. Our approach will revolve around the function $f(t) = \|A + tB\|$ mapping $\mathbb{R}$ into $\mathbb{R}_+$. To say that $\|A + tB\| \geq \|A\|$ for all $t \in \mathbb{R}$ is to say that f attains its minimum at the point 0. Since f is a convex function, the tools of convex analysis are available. The crux of our argument lies in calculating the *subdifferential* $\partial f(t)$ of f, and then showing that the point 0 is in the set $\partial f(0)$.

Definition 2.2. Let $f : X \to \mathbb{R}$ be a convex function. The subdifferential of f at a point $x \in X$ is the set $\partial f(x)$ of continuous linear functionals $v^* \in X^*$ such that

$$f(y) - f(x) \geq \operatorname{Re} v^*(y - x) \quad \text{for all } y \in X.$$

It is a convex subset of X^*. Its importance lies in the following proposition.

Proposition 2.3. *A convex function $f : X \to \mathbb{R}$ attains its minimum value at $x \in X$ if and only if $0 \in \partial f(x)$.*

We want to apply this to the function $f(t) = \|A + tB\|$ which we shall realize as the composition of two functions. The first of them is $t \to A + tB$ from $\mathbb{R}$ into $\mathbb{M}(n)$. The second is from $\mathbb{M}(n)$ to $\mathbb{R}_+$, sending any $T \in \mathbb{M}(n)$ to $\|T\|$. Thus we need to find subdifferentials of compositions. The subdifferential of the norm function has been calculated in [11]. We need it at a positive semidefinite matrix.

Proposition 2.4. *Let A be a positive semidefinite matrix. Then*

$$\partial\|A\| = \text{convex hull of } \{uu^* : \|u\| = 1, Au = \|A\|u\}. \tag{2.1}$$

To handle the composition maps, we need a chain rule.

Proposition 2.5. *Consider the composite map*

$$\mathbb{R}^n \xrightarrow{L} \mathbb{M}(n) \xrightarrow{g} \mathbb{R},$$

where g is a convex map and $L(x) = A + S(x)$ for all $x \in \mathbb{R}^n$, with $S : \mathbb{R}^n \to \mathbb{M}(n)$ being a linear map. Then the subdifferential of $g \circ L$ at a point $x \in \mathbb{R}^n$ is given by

$$\partial(g \circ L)(x) = S^*\partial g(L(x)) \quad \text{for all } x \in \mathbb{R}^n, \tag{2.2}$$

where $S^* : \mathbb{M}(n) \to \mathbb{R}^n$ is the adjoint of S satisfying

$$(S^*(T))'y = \operatorname{Re} \operatorname{tr} T^*S(y) \quad \text{for all } T \in \mathbb{M}(n),\ y \in \mathbb{R}^n.$$

(Here $(S^*(T))'$ means the transpose of the vector $S^*(T)$.)

These elementary facts can be found in [6]. We are now ready to prove a real version of the Bhatia–Šemrl theorem using these concepts.

Theorem 2.6. *Let $A, B \in \mathbb{M}(n)$. Then $\|A + t B\| \geq \|A\|$ for all $t \in \mathbb{R}$ if and only if there exists a unit vector x such that $\|Ax\| = \|A\|$ and $\operatorname{Re}\langle Ax, Bx\rangle = 0$.*

Proof. If such a unit vector x exists, then for $t \in \mathbb{R}$

$$\begin{aligned}\|A + t\,B\|^2 &\geq \|(A + t\,B)x\|^2\\ &= \|Ax\|^2 + t^2\|Bx\|^2 + 2t\,\mathrm{Re}\langle Ax, Bx\rangle\\ &= \|Ax\|^2 + t^2\|Bx\|^2\\ &\geq \|Ax\|^2\\ &= \|A\|^2.\end{aligned}$$

Conversely let

$$\|A + t\,B\| \geq \|A\| \quad \text{for all } t \in \mathbb{R}. \tag{2.3}$$

First we note that it is enough to show that if A is a positive semidefinite matrix and $B \in \mathbb{M}(n)$ such that (2.3) holds then there exists a unit vector y such that

$$Ay = \|A\|y \quad \text{and} \quad \mathrm{Re}\langle Ay, By\rangle = 0. \tag{2.4}$$

The general case may be reduced to this by using a singular value decomposition of A. Let $A = UA^{\dagger}V$ be a singular value decomposition of A. Then (2.3) implies

$$\|A^{\dagger} + t\,U^*BV^*\| \geq \|A^{\dagger}\| \quad \text{for all } t \in \mathbb{R}. \tag{2.5}$$

If there exists a unit vector y such that

$$A^{\dagger}y = \|A^{\dagger}\|y \quad \text{and} \quad \mathrm{Re}\langle A^{\dagger}y, U^*BV^*y\rangle = 0,$$

then for $x = V^*y$ we have

$$\|Ax\| = \|A\| \quad \text{and} \quad \mathrm{Re}\langle Ax, Bx\rangle = 0.$$

Thus assume that A is a positive semidefinite matrix in (2.3). Let $S : \mathbb{R} \to \mathbb{M}(n)$ be the linear map defined as

$$S(t) = t\,B$$

and let $L : \mathbb{R} \to \mathbb{M}(n)$ be the affine map

$$L(t) = A + S(t).$$

Let $g : \mathbb{M}(n) \to \mathbb{R}$ be the convex map given by

$$g(T) = \|T\|.$$

Then (2.3) can be rewritten as

$$(g \circ L)(t) \geq (g \circ L)(0).$$

By Proposition 2.3 we get

$$0 \in \partial(g \circ L)(0). \tag{2.6}$$

Using Propositions 2.4 and 2.5, we obtain

$$\partial(g \circ L)(0) = \text{ convex hull of } \{\mathrm{Re}\langle u, Bu\rangle : \|u\| = 1, Au = \|A\|u\}. \tag{2.7}$$

The set $\{\langle u, Bu\rangle : \|u\| = 1, Au = \|A\|u\}$ is the image of the set $\{u : \|u\| = 1, Au = \|A\|u\}$ under the quadratic form $u \to \langle u, Bu\rangle$. We call this the restriction of the numerical range of B to the eigenspace of A corresponding to the maximum eigenvalue $\|A\|$. By the Hausdorff–Toeplitz theorem, this is a convex set. Therefore the set $\{\mathrm{Re}\langle u, Bu\rangle : \|u\| = 1, Au = \|A\|u\}$ is convex. So from (2.6) and (2.7) we get that there exists a unit vector y such that

$$Ay = \|A\|y \quad \text{and} \quad \mathrm{Re}\langle y, By\rangle = 0.$$

These together imply that

$$\mathrm{Re}\langle Ay, By\rangle = 0. \quad \square \tag{2.8}$$

Remark. Another necessary and sufficient condition for a matrix A to be Birkhoff–James orthogonal to another matrix B is given in [8, Theorem 3.1(c)]. In the case when A is positive semidefinite and k is the multiplicity of the largest eigenvalue $\|A\|$, it says that for any $n \times k$ matrix U with orthonormal columns that form a basis for the eigenspace of A corresponding to $\|A\|$, we have 0 belongs to the numerical range of U^*BAU. This can also be interpreted as 0 belongs to the numerical range of A, restricted to the eigenspace of A corresponding to the largest eigenvalue $\|A\|$. An analogous condition for $\|A + t\,B\| \geq \|A\|$

for all $t \in \mathbb{R}$ is established through our proof. We obtain that if A is positive semidefinite and $\|A + t\,B\| \geq \|A\|$ for all $t \in \mathbb{R}$ then 0 belongs to the real part of the numerical range of B, restricted to the eigenspace of A corresponding to $\|A\|$. The general case, when A is not necessarily positive semidefinite, can be obtained from this by using singular value decomposition.

We now go to complex scalars.

Theorem 2.7. *Let $A, B \in \mathbb{M}(n)$. Then $\|A + \lambda\,B\| \geq \|A\|$ for all $\lambda \in \mathbb{C}$ if and only if there exists a unit vector x such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$.*

Proof. If such a unit vector x exists, then by a similar argument as in the proof of Theorem 2.6 we get that

$$\|A + \lambda\,B\| \geq \|A\| \quad \text{for all } \lambda \in \mathbb{C}. \tag{2.9}$$

Now suppose (2.9) holds, that is,

$$\|A + re^{i\theta}\,B\| \geq \|A\| \quad \text{for all } r, \theta \in \mathbb{R}.$$

Fix θ and let $B_\theta = e^{i\theta}\,B$. Then we have

$$\|A + r\,B_\theta\| \geq \|A\| \quad \text{for all } r \in \mathbb{R}.$$

We can assume A to be positive semidefinite as in the proof of Theorem 2.6. By (2.4), there exists a unit vector y_θ such that

$$Ay_\theta = \|A\|y_\theta \quad \text{and} \quad \operatorname{Re} e^{i\theta}\langle Ay_\theta, By_\theta\rangle = 0. \tag{2.10}$$

Consider the set $\{\langle B^*Ay, y\rangle : \|y\| = 1,\ Ay = \|A\|y\}$. This is the restriction of the numerical range of B^*A to the eigenspace of A corresponding to the maximum eigenvalue $\|A\|$ and therefore is a compact convex set in $\mathbb{C}$. If 0 does not belong to this set, then there exists $z \in \mathbb{C}$ such that for all unit vectors y satisfying $Ay = \|A\|y$, we have

$$\operatorname{Re} \bar{z}\langle B^*Ay, y\rangle > 0. \tag{2.11}$$

This is a consequence of the Separating Hyperplane Theorem. A proof of this can be found in [10, p. 13]. Putting $z = |z|e^{i\theta_0}$ in (2.11) we get

$$\operatorname{Re} e^{-i\theta_0}\langle B^*Ay, y\rangle > 0 \quad \text{for all } y \text{ such that } \|y\| = 1,\ Ay = \|A\|y.$$

This is a contradiction to (2.10). Thus we get that

$$0 \in \{\langle B^*Ay, y\rangle : \|y\| = 1,\ Ay = \|A\|y\},$$

that is, there exists a unit vector y such that

$$Ay = \|A\|y \quad \text{and} \quad \langle Ay, By\rangle = 0. \quad \square$$

3. The infinite dimensional case

Theorem 3.1. *Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces. Let $A, B \in \mathscr{B}(\mathcal{H}, \mathcal{K})$. Then A is Birkhoff–James orthogonal to B if and only if there exists a sequence of unit vectors $\{x_n\}$ in $\mathcal{H}$ such that $\|Ax_n\| \to \|A\|$ and $\langle Ax_n, Bx_n\rangle \to 0$ as $n \to \infty$.*

Proof. First suppose that such a sequence $\{x_n\}$ of unit vectors in $\mathcal{H}$ exists. Then for every $\lambda \in \mathbb{C}$

$$\begin{aligned}\|A + \lambda B\|^2 &\geq \|(A + \lambda B)x_n\|^2 \\ &= \|Ax_n\|^2 + |\lambda|^2\|Bx_n\|^2 + 2\operatorname{Re}\bar{\lambda}\,\langle Ax_n, Bx_n\rangle\,.\end{aligned}$$

This holds for all n. Taking lim inf on both the sides as $n \to \infty$ we get

$$\|A + \lambda B\|^2 \geq \|A\|^2 + |\lambda|^2 \liminf_{n\to\infty} \|Bx_n\|^2 \geq \|A\|^2.$$

Conversely let A be Birkhoff–James orthogonal to B. For any operator $T : \mathcal{H} \to \mathcal{K}$ we denote by $\tilde{T}$, the operator on $\mathcal{H} \oplus \mathcal{K}$ given by

$$\tilde{T} = \begin{bmatrix} 0 & 0 \\ T & 0 \end{bmatrix}.$$

Note that $\|\tilde{T}\| = \|T\|$. Therefore we have $\|\tilde{A} + \lambda\tilde{B}\| \geq \|\tilde{A}\|$ for all $\lambda \in \mathbb{C}$. By Remark 3.1 in [4], we get a sequence $\{h_n \oplus k_n\}$ of unit vectors in $\mathcal{H} \oplus \mathcal{K}$ such that

$$\|\tilde{A}(h_n \oplus k_n)\| \to \|\tilde{A}\| \quad \text{and} \quad \left\langle \tilde{A}(h_n \oplus k_n), \tilde{B}(h_n \oplus k_n)\right\rangle \to 0 \quad \text{as } n \to \infty. \tag{3.1}$$

The first equation gives

$$\|Ah_n\| \to \|A\| \quad \text{as } n \to \infty. \tag{3.2}$$

Now

$$\|A\| = \lim_{n\to\infty} \|Ah_n\| \le \|A\| \liminf_{n\to\infty} \|h_n\|.$$

Therefore $\liminf_{n\to\infty} \|h_n\| \ge 1$. Since $\|h_n\| \le 1$ for every n, we obtain $\limsup_{n\to\infty} \|h_n\| \le 1$. So

$$\lim_{n\to\infty} \|h_n\| = 1.$$

Consider $x_n = \begin{cases} \frac{h_n}{\|h_n\|} & \text{if } h_n \neq 0 \\ 0 & \text{if } h_n = 0 \end{cases}$. Passing onto a subsequence, if necessary, we get a sequence $\{x_n\}$ of unit vectors in $\mathscr{H}$ such that

$$\|Ax_n\| \to \|A\| \quad \text{and} \quad \langle Ax_n, Bx_n\rangle \to 0 \quad \text{as } n \to \infty. \quad \square$$

Corollary 3.2. *Let $\mathscr{H}$ and $\mathscr{K}$ be two finite dimensional Hilbert spaces. Then A is orthogonal to B in the Birkhoff–James sense if and only if there exists a unit vector $x \in \mathscr{H}$ such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$.*

Proof. If A is Birkhoff–James orthogonal to B then from Theorem 3.3 we obtain a sequence $\{x_n\}$ of unit vectors such that $\|Ax_n\| \to \|A\|$ and $\langle Ax_n, Bx_n\rangle \to 0$ as $n \to \infty$. Since $\{x_n\}$ is a bounded sequence therefore it has a convergent subsequence converging to a vector x. This x is the required unit vector. $\square$

Corollary 3.3. *Let $\mathscr{H}$ and $\mathscr{K}$ be two Hilbert spaces. Let $A, B \in \mathscr{B}(\mathscr{H}, \mathscr{K})$. Then A is Birkhoff–James orthogonal to B if and only if there exists a state φ on $\mathscr{B}(\mathscr{H})$ such that $\varphi(A^*A) = \|A\|^2$ and $\varphi(A^*B) = 0$. (Note that this state φ may not be of the form $\varphi(T) = \langle x, Tx\rangle$ for any x.)*

Proof. First suppose that such a state exists. For every $\lambda \in \mathbb{C}$ we have

$$\begin{aligned} \|A + \lambda B\|^2 &\ge |\varphi((A + \lambda B)^*(A + \lambda B))| \\ &= |\varphi(A^*A) + \bar{\lambda}\, \varphi(B^*A) + \lambda\, \varphi(A^*B) + |\lambda|^2\, \varphi(B^*B)| \\ &\ge \|A\|^2. \end{aligned} \tag{3.3}$$

Conversely let A be Birkhoff–James orthogonal to B. Then by Theorem 3.3 there exists a sequence $\{x_n\}$ of unit vectors in $\mathscr{H}$ such that $\|Ax_n\| \to \|A\|$ and $\langle Ax_n, Bx_n\rangle \to 0$ as $n \to \infty$. Define $\varphi_n : \mathscr{B}(\mathscr{H}) \to \mathbb{C}$ as

$$\varphi_n(T) = \langle x_n, Tx_n\rangle.$$

Then φ_n is a state on $\mathscr{B}(\mathscr{H})$. Note that $\varphi_n(A^*A) = \langle Ax_n, Ax_n\rangle \to \|A\|^2$ and $\varphi_n(A^*B) = \langle Ax_n, Bx_n\rangle \to 0$ as $n \to \infty$. Since the collection of all states on any C^*-algebra is weak* compact, $\{\varphi_n\}$ has a convergent subnet $\{\psi_\alpha\}$ converging to a state φ in weak* topology. We have

$$\varphi(A^*A) = \lim_\alpha \psi_\alpha(A^*A) = \|A\|^2$$

and

$$\varphi(A^*B) = \lim_\alpha \psi_\alpha(A^*B) = 0. \quad \square$$

In the following three corollaries, we reformulate the result of Corollary 3.2 to show what it looks like in three specific situations. In these corollaries, all the Hilbert spaces considered are finite-dimensional.

Corollary 3.4. *Let $A_j \in \mathscr{B}(\mathscr{H}, \mathscr{K}_j)$ for $j = 1, \ldots, d$. Consider the column operator $\begin{pmatrix} A_1 \\ \vdots \\ A_d \end{pmatrix} : \mathscr{H} \to \mathscr{K}_1 \oplus \cdots \oplus \mathscr{K}_d$ which takes $x \in \mathscr{H}$ to $\begin{pmatrix} A_1x \\ \vdots \\ A_dx \end{pmatrix}$. Then*

$$\left\| \begin{pmatrix} A_1 + \lambda B_1 \\ \vdots \\ A_d + \lambda B_d \end{pmatrix} \right\| \ge \left\| \begin{pmatrix} A_1 \\ \vdots \\ A_d \end{pmatrix} \right\| \quad \textit{for all } \lambda \in \mathbb{C}$$

if and only if there exists a unit vector $x \in \mathcal{H}$ such that

$$\sum_{j=1}^{d} \|A_j x\|^2 = \left\| \sum_{j=1}^{d} A_j^* A_j \right\| \quad \text{and} \quad \sum_{j=1}^{d} \left\langle A_j x, B_j x \right\rangle = 0.$$

Corollary 3.5. *Let $A_j \in \mathscr{B}(\mathcal{H}_j, \mathcal{K})$ for $j = 1, \ldots, d$. Consider the row operator $(A_1, \ldots, A_d) : \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_d \to \mathcal{K}$ which takes $\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$ to $A_1 x_1 + \cdots + A_d x_d$. Then*

$$\|(A_1 + \lambda B_1, \ldots, A_d + \lambda B_d)\| \geq \|(A_1, \ldots, A_d)\| \quad \text{for all } \lambda \in \mathbb{C}$$

if and only if there exists a unit vector $\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_d$ such that

$$\|(A_1, \ldots, A_d)\|^2 = \sum_{j=1}^{d} \|A_j x_j\|^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{d} \left\langle A_i x_i, A_j x_j \right\rangle \quad \text{and} \quad \sum_{i,j=1}^{d} \left\langle A_i x_i, B_j x_j \right\rangle = 0.$$

Corollary 3.6. *Let $A_j \in \mathscr{B}(\mathcal{H}_j, \mathcal{K}_j)$ for $j = 1, \ldots, d$. Consider the "diagonal" operator*

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_d \end{pmatrix} : \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_d \to \mathcal{K}_1 \oplus \cdots \oplus \mathcal{K}_d.$$

Then

$$\left\| \begin{pmatrix} A_1 + \lambda B_1 & & \\ & \ddots & \\ & & A_d + \lambda B_d \end{pmatrix} \right\| \geq \left\| \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_d \end{pmatrix} \right\| \quad \text{for all } \lambda \in \mathbb{C}$$

if and only if there exists a unit vector $\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_d$ such that

$$\max_{1 \leq k \leq d} \|A_k\|^2 = \sum_{j=1}^{d} \|A_j x_j\|^2 \quad \text{and} \quad \sum_{j=1}^{d} \left\langle A_j x_j, B_j x_j \right\rangle = 0.$$

As an application of this, we have the following.

Corollary 3.7. *Let $n_1, \ldots, n_k$ be a partition of a positive integer n, that is, $\sum n_j = n$. Let $A, B \in \mathbb{M}(m, n)$. Let $A = [A_1, \ldots, A_k]$, where each A_j is an $m \times n_j$ matrix. Define*

$$\|A\|_{\text{col}} = \max_{1 \leq j \leq k} \|A_j\|.$$

Suppose this maximum is attained at d indices, say $j_1, \ldots, j_d$. Then A is Birkhoff–James orthogonal to B in $\|\cdot\|_{\text{col}}$ if and only if $[A_{j_1}, \ldots, A_{j_d}]$ is Birkhoff–James orthogonal to $[B_{j_1}, \ldots, B_{j_d}]$ in $\|\cdot\|_{\text{col}}$.

Proof. If $[A_{j_1}, \ldots, A_{j_d}]$ is Birkhoff–James orthogonal to $[B_{j_1}, \ldots, B_{j_d}]$ in $\|\cdot\|_{\text{col}}$, then for all $\lambda \in \mathbb{C}$

$$\begin{aligned} \|A + \lambda B\|_{\text{col}} &= \max_{1 \leq j \leq k} \|A_j + \lambda B_j\| \\ &\geq \max_{1 \leq p \leq d} \|A_{j_p} + \lambda B_{j_p}\| \\ &= \|[A_{j_1}, \ldots, A_{j_d}] + \lambda [B_{j_1}, \ldots, B_{j_d}]\|_{\text{col}} \\ &\geq \|[A_{j_1}, \ldots, A_{j_d}]\|_{\text{col}} \\ &= \|A_{j_p}\| \quad \text{for all } 1 \leq p \leq d \\ &= \|A\|_{\text{col}}. \end{aligned}$$

For the converse, first note that, by virtue of the norm on A being maximum of the norms of A_j, the matrix $[A_1, \ldots, A_k]$ being Birkhoff–James orthogonal to the matrix $[B_1, \ldots, B_k]$ in $\|\cdot\|_{\text{col}}$ is same as saying that

$$\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{bmatrix} \text{ is Birkhoff–James orthogonal to } \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_k \end{bmatrix} \text{ in } \|\cdot\|. \tag{3.4}$$

Assume, without loss of generality, that $j_p = p$ for all $1 \le p \le d$, that is,

$$\|A\|_{\text{col}} = \|A_1\| = \cdots = \|A_d\|.$$

Thus we have to prove that

$$\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_d \end{bmatrix} \text{ is Birkhoff–James orthogonal to } \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_d \end{bmatrix} \text{ in } \|\cdot\|.$$

Now we use Eq. (3.4) and Corollary 3.6 to conclude that there exist $x_j \in \mathbb{C}^{n_j},\ j = 1, \ldots, k$, satisfying

$$\sum_{j=1}^{k} \|x_j\|^2 = 1, \qquad \sum_{j=1}^{k} \|A_j x_j\|^2 = \|A_1\|^2 = \cdots = \|A_d\|^2 \quad \text{and} \quad \sum_{j=1}^{k} \langle A_j x_j, B_j x_j \rangle = 0.$$

Now

$$\|A_1\|^2 = \sum_{j=1}^{k} \|A_j x_j\|^2 \le \sum_{j=1}^{k} \|A_j\|^2 \|x_j\|^2 \le \|A_1\|^2.$$

This gives

$$\sum_{j=1}^{k} \|A_j\|^2 \|x_j\|^2 = \|A_1\|^2 = \cdots = \|A_d\|^2.$$

Therefore we get $x_{d+1} = \cdots = x_k = 0$. So now we have $x_1, \ldots, x_d$ in $\mathbb{C}^{n_1}, \ldots, \mathbb{C}^{n_d}$ respectively such that

$$\sum_{j=1}^{d} \|x_j\|^2 = 1, \qquad \sum_{j=1}^{d} \|A_j x_j\|^2 = \|A_1\|^2 = \cdots = \|A_d\|^2 \quad \text{and} \quad \sum_{j=1}^{d} \langle A_j x_j, B_j x_j \rangle = 0.$$

Again by Corollary 3.6, we are done. □

4. Birkhoff–James orthogonality in Hilbert C^*-modules

Proposition 4.1. *Let $\mathcal{A}$ be a C^*-algebra. Let $a, b \in \mathcal{A}$. Then $\|a + \lambda b\| \ge \|a\|$ for all $\lambda \in \mathbb{C}$ if and only if there exists a state φ on $\mathcal{A}$ such that*

$$\varphi(a^*a) = \|a\|^2 \quad \textit{and} \quad \varphi(a^*b) = 0.$$

Proof. If such a state exists, then a similar proof as in (3.3) shows that $\|a + \lambda b\| \ge \|a\|$ for all $\lambda \in \mathbb{C}$. For the converse let $\pi : \mathcal{A} \to \mathscr{B}(\mathcal{H})$ be a faithful representation. Let $A = \pi(a)$ and $B = \pi(b)$. Then we have $\|A + \lambda B\| \ge \|A\|$ for all $\lambda \in \mathbb{C}$. By Corollary 3.3 there exists a state ψ on $\mathscr{B}(\mathcal{H})$ such that

$$\psi(A^*A) = \|A\|^2 \quad \text{and} \quad \psi(A^*B) = 0. \tag{4.1}$$

Let $\varphi : \mathcal{A} \to \mathbb{C}$ be defined as $\varphi(a) = \psi(\pi(a))$. Then φ is a state on $\mathcal{A}$. Eq. (4.1) implies that

$$\varphi(a^*a) = \|a\|^2 \quad \text{and} \quad \varphi(a^*b) = 0. \quad \square$$

For any given $a \in \mathcal{A}$ and state φ on $\mathcal{A}$, let *variance* of a with respect to φ, denoted by $\text{var}_\varphi(a)$, be defined as

$$\text{var}_\varphi(a) = \varphi(a^*a) - |\varphi(a)|^2.$$

Let $\Delta(a) = \min\{\|a - \lambda 1\| : \lambda \in \mathbb{C}\}$ be the distance of a from $\mathbb{C}1$. The following corollary describes $\Delta(a)$ in terms of $\text{var}_\varphi(a)$. This is a generalization of a special case of Theorem 9 in [2] and has been described in Theorem 3.10 in [9]. We provide a proof of it using the above theorem.

Corollary 4.2. *With the notations as above, we have for any* $a \in \mathcal{A}$

$$\Delta(a)^2 = \max\{\mathrm{var}_\varphi(a) : \varphi \in S(\mathcal{A})\}, \tag{4.2}$$

where $S(\mathcal{A})$ *denotes the state space of* $\mathcal{A}$.

Proof. First note that for any $\varphi \in S(\mathcal{A})$,

$$\varphi(a^*a) \le \|a\|^2.$$

Therefore

$$\mathrm{var}_\varphi(a) = \varphi(a^*a) - |\varphi(a)|^2 \le \|a\|^2.$$

Let $\lambda \in \mathbb{C}$. Changing a to $a + \lambda 1$ in the above equation, we see

$$\mathrm{var}_\varphi(a+\lambda 1) = \varphi((a+\lambda 1)^*(a+\lambda 1)) - |\varphi(a+\lambda 1)|^2 \le \|a+\lambda 1\|^2.$$

The left hand side is invariant under the translation $a \to a + \lambda 1$, that is,

$$\mathrm{var}_\varphi(a+\lambda 1) = \mathrm{var}_\varphi(a).$$

This gives

$$\max\{\mathrm{var}_\varphi(a) : \varphi \in S(\mathcal{A})\} \le \Delta(a)^2. \tag{4.3}$$

Now $\Delta(a) = \|a - \lambda_0\|$, for some $\lambda_0 \in \mathbb{C}$. We denote $a - \lambda_0$ by a_0. Then $\|a_0 + \lambda 1\| \ge \|a_0\|$ for all $\lambda \in \mathbb{C}$. By Proposition 4.1, there exists a state ψ on $\mathcal{A}$ such that

$$\psi(a_0^* a_0) = \|a_0\|^2 \quad \text{and} \quad \psi(a_0^*) = 0. \tag{4.4}$$

So from the first equation in (4.4) we have

$$\Delta(a)^2 = \|a_0\|^2 = \psi(a_0^*a_0) = \psi(a^*a) - \overline{\lambda_0}\psi(a) - \lambda_0\overline{\psi(a)} + |\lambda_0|^2.$$

From the second equation in (4.4) we get $\psi(a) = \lambda_0$. Using this we obtain

$$\Delta(a)^2 = \psi(a^*a) - |\psi(a)|^2 = \mathrm{var}_\psi(a) \le \max\{\mathrm{var}_\varphi(a) : \varphi \in S(\mathcal{A})\}.$$

This together with (4.3) completes the proof. □

We now obtain a characterization of Birkhoff–James orthogonality in Hilbert C^*-modules. For this we require the following lemma, which is a reinterpretation of Theorem 3.4 in [5].

Lemma 4.3. *Let* $\mathcal{E}$ *be a Hilbert* C^**-module over a* C^**-algebra* $\mathcal{A}$. *Then* $\mathcal{E}$ *can be isometrically embedded in* $\mathcal{B}(\mathcal{H}, \mathcal{K})$ *for some Hilbert spaces* $\mathcal{H}, \mathcal{K}$. *Here* $\mathcal{H}$ *is a Hilbert space such that there exists a faithful representation* $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ *and the isometric embedding* $L : \mathcal{E} \to \mathcal{B}(\mathcal{H}, \mathcal{K})$ *satisfies*

$$\langle L(e_1)h_1, L(e_2)h_2\rangle = \langle h_1, \pi(\langle e_1, e_2\rangle)h_2\rangle \quad \textit{for all } e_1, e_2 \in \mathcal{E} \textit{ and } h_1, h_2 \in \mathcal{H}.$$

Our next theorem is the same as Theorem 2.7 in [1]. We give a simple proof of that using the above lemma. This makes the characterization more natural.

Theorem 4.4. *Let* $\mathcal{E}$ *be a Hilbert* C^**-module over a* C^**-algebra* $\mathcal{A}$. *Let* $e_1, e_2 \in \mathcal{E}$. *Then* e_1 *is orthogonal to* e_2 *in the Banach space* $\mathcal{E}$ *in the Birkhoff–James sense, that is,*

$$\|e_1 + \lambda e_2\| \ge \|e_1\| \quad \textit{for all } \lambda \in \mathbb{C} \tag{4.5}$$

if and only if there exists a state φ *on* $\mathcal{A}$ *such that*

$$\varphi(\langle e_1, e_1\rangle) = \|e_1\|^2 \quad \textit{and} \quad \varphi(\langle e_1, e_2\rangle) = 0.$$

Proof. First suppose that such a state exists. Then for every $\lambda \in \mathbb{C}$

$$\begin{aligned}\|e_1 + \lambda e_2\|^2 &= \|\langle e_1 + \lambda e_2, e_1 + \lambda e_2\rangle\| \\ &\ge |\varphi(\langle e_1, e_1\rangle) + \bar{\lambda}\,\varphi(\langle e_2, e_1\rangle) + \lambda\,\varphi(\langle e_1, e_2\rangle) + |\lambda|^2\ \varphi(\langle e_2, e_2\rangle)| \\ &\ge \|e_1\|^2.\end{aligned}$$

Now suppose (4.5) holds. Let $L : \mathcal{E} \to \mathcal{B}(\mathcal{H}, \mathcal{K})$ be the isometric embedding of $\mathcal{E}$ into $\mathcal{B}(\mathcal{H}, \mathcal{K})$ as given in the previous lemma. Then (4.5) gives

$$\|L(e_1) + \lambda L(e_2)\| \ge \|L(e_1)\| \quad \text{for all } \lambda \in \mathbb{C}.$$

By Theorem 3.3 there exists a sequence of unit vectors $\{x_n\}$ in $\mathscr{H}$ such that $\|L(e_1)x_n\| \to \|L(e_1)\|$ and $\langle L(e_1)x_n, L(e_2)x_n\rangle \to 0$ as $n \to \infty$. Define $\varphi_n : \mathcal{A} \to \mathbb{C}$ as

$$\varphi_n(a) = \langle x_n, \pi(a)x_n\rangle,$$

where $\pi : \mathcal{A} \to \mathscr{B}(\mathscr{H})$ is a faithful representation, as in the previous lemma. Then φ_n is a state on $\mathcal{A}$. Note that $\varphi_n(\langle e_1, e_1\rangle) = \langle L(e_1)x_n, L(e_1)x_n\rangle \to \|L(e_1)\|^2$ and $\varphi_n(\langle e_1, e_2\rangle) = \langle L(e_1)x_n, L(e_2)x_n\rangle \to 0$ as $n \to \infty$. Since the collection of all states on $\mathcal{A}$ is a weak* compact subset of $\mathcal{A}^*$, $\{\varphi_n\}$ has a convergent subnet $\{\psi_\alpha\}$ which converges to some φ in weak* topology. We have

$$\varphi(\langle e_1, e_1\rangle) = \lim_\alpha \psi_\alpha(\langle e_1, e_1\rangle) = \|e_1\|^2$$

and

$$\varphi(\langle e_1, e_2\rangle) = \lim_\alpha \psi_\alpha(\langle e_1, e_2\rangle) = 0. \quad \square$$

Acknowledgments

The authors express their gratitude to Rajendra Bhatia for many useful discussions. We also thank Partha Sarathi Chakraborty for adding to our understanding of Hilbert C^*-modules.

The first named author's research is supported by Department of Science and Technology, India through the project numbered SR/S4/MS:766/12 and University Grants Commission, India.

References

[1] L. Arambašić, R. Rajić, The Birkhoff–James orthogonality in Hilbert C^*-modules, Linear Algebra Appl. 437 (2012) 1913–1929.
[2] K.M.R. Audenaert, Variance bounds, with an application to norm bounds for commutators, Linear Algebra Appl. 432 (2010) 1126–1143.
[3] C. Benitez, M. Fernandez, M.L. Soriano, Orthogonality of matrices, Linear Algebra Appl. 422 (2007) 155–163.
[4] R. Bhatia, P. Šemrl, Orthogonality of matrices and some distance problems, Linear Algebra Appl. 287 (1999) 77–86.
[5] D.P. Blecher, A new approach to Hilbert C^*-modules, Math. Ann. 307 (1997) 253–290.
[6] J.B. Hiriart-Urruty, C. Lemarèchal, Fundamentals of Convex Analysis, Springer, 2000.
[7] D.J. Kečkiċ, Gateaux derivative of $B(H)$ norm, Proc. Amer. Math. Soc. 133 (2005) 2061–2067.
[8] C.K. Li, H. Schneider, Orthogonality of matrices, Linear Algebra Appl. 347 (2002) 115–122.
[9] M.A. Rieffel, Standard deviation is a strongly Leibniz seminorm. arXiv:1208.4072v2.
[10] G.A. Watson, Approximation Theory and Numerical Methods, John Wiley & Sons, 1980.
[11] G.A. Watson, Characterization of the subdifferential of some matrix norms, Linear Algebra Appl. 170 (1992) 33–45.