Orthogonality of matrices in the Ky Fan k-norms

Priyanka Grover *†
Department of Mathematics, School of Natural Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh 201314, India

Abstract

We obtain necessary and sufficient conditions for a matrix A to be Birkhoff-James orthogonal to another matrix B in the Ky Fan k-norms. A characterization for A to be Birkhoff-James orthogonal to any subspace $\mathscr{W}$ of $\mathbb{M}(n)$ is also obtained.

AMS classification: 15A60, 47A12, 47A30, 15A18
Keywords: Birkhoff-James orthogonality, Subdifferential, Singular value decomposition, Ky Fan norms, k-numerical range, Hausdorff-Toeplitz Theorem, Separating Hyperplane theorem, Norm parallelism

1 Introduction

Let $\mathbb{M}(n)$ be the space of $n \times n$ complex matrices. Let $\|\cdot\|$ be any norm on $\mathbb{M}(n)$. Let $A, B \in \mathbb{M}(n)$. Then A is said to be (Birkhoff-James) orthogonal to B in $\|\cdot\|$ if

$$\|A + \lambda B\| \geq \|A\| \text{ for all } \lambda \in \mathbb{C}. \tag{1.1}$$

In [5], Bhatia and Šemrl obtained a characterization for A to be orthogonal to B in the operator norm (also known as the spectral norm) $\|\cdot\|_\infty$. They showed that A is orthogonal to B in $\|\cdot\|_\infty$ if and only if there exists a unit vector $x \in \mathbb{C}^n$ such that $\|Ax\| = \|A\|_\infty$ and $\langle Ax, Bx \rangle = 0$. (All inner products in this note are conjugate linear in the first component and linear in the second component.) Different proofs for this result have been studied in [7, 11, 12]. This result can be restated as follows. If $A = U|A|$ is a polar decomposition of A, then A is orthogonal to B in $\|\cdot\|_\infty$ if and only if there exists a unit vector $x \in \mathbb{C}^n$ such that $|A|x = \|A\|_\infty x$ and $\langle x, U^*Bx \rangle = 0$. In [5], it was also showed that if $\operatorname{tr} U^*B = 0$ then A is orthogonal to B in the trace norm $\|\cdot\|_1$. And the converse is true if A is taken to be invertible. Later, Li and Schneider [12] gave a characterization for orthogonality in $\|\cdot\|_1$ when A need

*Email: priyanka.grover@snu.edu.in
†The author is supported by the research grant of INSPIRE Faculty Award of Department of Science and Technology, India.

not be necessarily invertible. They showed the following. Let the number of zero singular values of A be ℓ. Let $A = USV^*$ be a singular value decomposition of A. Let $B = U \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} V^*$, where $B_{11} \in \mathbb{M}(n-\ell), B_{22} \in \mathbb{M}(\ell)$. Then $\|A + \lambda B\|_1 \geq \|A\|_1$ for all $\lambda \in \mathbb{C}$ if and only if $|\operatorname{tr} B_{11}| \leq \|B_{22}\|_1$.

The trace norm and the operator norm are special cases of two classes of norms, namely the Schatten p-norms $\|\cdot\|_p$ and the Ky Fan k-norms $\|\cdot\|_{(k)}$. In [5] and [12], the authors have investigated the problem of finding necessary and sufficient conditions for orthogonality of matrices in $\|\cdot\|_p$, $1 \leq p \leq \infty$. In this note, we obtain characterizations for orthogonality of matrices in $\|\cdot\|_{(k)}$, $1 \leq k \leq n$. Let $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A) \geq 0$ be the singular values of A. Then $\|A\|_{(k)}$ is defined as

$$\|A\|_{(k)} = s_1(A) + s_2(A) + \cdots + s_k(A). \tag{1.2}$$

The cases $k = 1$ and $k = n$ correspond to the operator norm $\|\cdot\|_\infty$ and the trace norm $\|\cdot\|_1$, respectively. We show the following.

Theorem 1.1. *Let $A = U|A|$ be a polar decomposition of A. If there exist k orthonormal vectors $u_1, u_2, \ldots, u_k$ such that*

$$|A|\ u_i = s_i(A) u_i \text{ for all } 1 \leq i \leq k \tag{1.3}$$

and

$$\sum_{i=1}^{k} \langle u_i, U^* B u_i \rangle = 0, \tag{1.4}$$

then A is orthogonal to B in $\|\cdot\|_{(k)}$. If $s_k(A) > 0$, then the converse is also true.

The next theorem gives a more general characterization.

Theorem 1.2. *Let $A = USV^*$ be a singular value decomposition of A. Let the multiplicity of $s_k(A)$ be $r + q$, where $r \geq 0$ and $q \geq 1$, such that*

$$s_{k-q+1}(A) = \cdots = s_{k+r}(A).$$

Let $B = U \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} V^*$, where $B_{11} \in \mathbb{M}(k-q), B_{22} \in \mathbb{M}(r+q), B_{33} \in \mathbb{M}(n-k-r)$.

(a) Let $s_k(A) > 0$. Then A is orthogonal to B in $\|\cdot\|_{(k)}$ if and only if there exists a positive semidefinite matrix $T \in \mathbb{M}(r+q)$ with $\lambda_1(T) \leq 1$ and $\sum_{j=1}^{r+q} \lambda_j(T) = q$ such that $\operatorname{tr} B_{11} + \operatorname{tr}(T^* B_{22}) = 0$.

(b) Let $s_k(A) = 0$. Then A is orthogonal to B in $\|\cdot\|_{(k)}$ if and only if there exists $T \in \mathbb{M}(n-k+q, r+q)$ with $s_1(T) \leq 1$, and $\sum_{j=1}^{r+q} s_j(T) \leq q$ such that $\operatorname{tr} B_{11} + \operatorname{tr}\left(T^* \begin{bmatrix} B_{22} \\ B_{32} \end{bmatrix}\right) = 0$.

Let $\mathscr{W}$ be any subspace of $\mathbb{M}(n)$. Then A is said to be orthogonal to $\mathscr{W}$ (in the Birkhoff-James sense) in a given norm $\|\cdot\|$ on $\mathbb{M}(n)$ if

$$\|A+W\| \geq \|A\| \text{ for all } W \in \mathscr{W}. \tag{1.5}$$

In [10], we obtained a necessary and sufficient condition for A to be orthogonal to $\mathscr{W}$ in the operator norm. Our next theorem gives a characterization for A to be orthogonal to $\mathscr{W}$ in $\|\cdot\|_{(k)}$.

Theorem 1.3. *Let $A = U|A|$ be a polar decomposition of A. Let $\mathscr{W}$ be any subspace of $\mathbb{M}(n)$. If there exist density matrices $P_1, P_2, \ldots, P_k$ such that $\|\sum_{i=1}^k P_i\|_\infty \leq 1$, $|A|P_i = s_i(A)P_i$ $(1 \leq i \leq k)$ and $U\sum_{i=1}^k P_i \in \mathscr{W}^\perp$, then A is orthogonal to $\mathscr{W}$ in $\|\cdot\|_{(k)}$. If $s_k(A) > 0$, then the converse is also true.*

If $m_i(A)$ is the multiplicity of $s_i(A)$, then the condition $|A|P_i = s_i(A)P_i$ implies that the range of P_i is a subspace of the eigenspace of $|A|$ corresponding to $s_i(A)$. So rank P_i is at most $m_i(A)$.

The problem of finding characterizations of orthogonality of a matrix to a subspace $\mathscr{W}$ of $\mathbb{M}(n)$ is closely related to the best approximation problems [18]. A specific question is when is the zero matrix a best approximation to A from $\mathscr{W}$? This is the same as asking when is A orthogonal to $\mathscr{W}$?

In [12], the authors studied a characterization for orthogonality in the induced matrix norms. Benítez, Fernández and Soriano [6] showed that a necessary and sufficient condition for the norm of a real finite dimensional normed space $\mathscr{X}$ to be induced by an inner product is that for any bounded linear operators A, B from $\mathscr{X}$ into itself, A is orthogonal to B if and only if there exists a unit vector $x \in X$ such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$. More results in this direction have been obtained recently in [15, 16]. Characterizations of orthogonality on Hilbert C*-modules have been studied in [1, 2, 3, 7].

To obtain the proofs of the above theorems, we use methods that we had introduced in [7] and [10]. We first obtain some new expressions for the subdifferential of the map taking a matrix A to its Ky Fan k-norm $\|A\|_{(k)}$ in Section 2. The proofs of the above theorems are given in Section 3 followed by some remarks in Section 4.

2 Subdifferentials of the Ky Fan k-norm

Let $\mathscr{X}$ be a Banach space and let $f : \mathscr{X} \to \mathbb{R}$ be a convex function.

Definition 2.1. A subgradient of f at $a \in \mathscr{X}$ is an element φ of the dual space $\mathscr{X}^*$ such that

$$f(y) - f(a) \geq \operatorname{Re} \varphi(y-a) \quad \text{for all } y \in \mathscr{X}. \tag{2.1}$$

The *subdifferential* of f at a is the set of bounded linear functionals $\varphi \in \mathscr{X}^*$ satisfying (2.1) and is denoted by $\partial f(a)$. It is a non-empty weak* compact convex subset of $\mathscr{X}^*$. For more details, see [9, Chapter D] and [21, Chapter 2]. The following proposition is a direct consequence of the definition of the subdifferential. It is one of the most useful tools that we require in Section 3.

Proposition 2.2. A continuous convex function $f : \mathscr{X} \to \mathbb{R}$ attains its minimum value at a if and only if $0 \in \partial f(a)$.

An equivalent definition of the subdifferential of a continuous convex function can be given in terms of $f'_+(a, x)$, the right directional derivative of f at a in the direction x:

$$\partial f(a) = \{\varphi \in \mathscr{X}^* : \text{Re } \varphi(x) \le f'_+(a, x) \text{ for all } x \in \mathscr{X}\}. \tag{2.2}$$

Moreover, for each $x \in \mathscr{X}$,

$$f'_+(a, x) = \max\{\text{Re}\, \varphi(x) : \varphi \in \partial f(a)\}. \tag{2.3}$$

The following rule of subdifferential calculus will be helpful in our analysis later.

Proposition 2.3. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces. Let $S : \mathscr{X} \to \mathscr{Y}$ be a bounded linear map and let $L : \mathscr{X} \to \mathscr{Y}$ be the continuous affine map defined by $L(x) = S(x) + y_0$, for some $y_0 \in \mathscr{Y}$. Let $g : \mathscr{Y} \to \mathbb{R}$ be a continuous convex function. Then

$$\partial(g \circ L)(a) = S^* \partial g(L(a)) \text{ for all } a \in \mathscr{X}, \tag{2.4}$$

where S^* denotes the real or complex adjoint of S (depending on whether $\mathscr{X}$ and $\mathscr{Y}$ are both real or both complex Banach spaces.)

For any norm $\|\cdot\|$ on the space $\mathbb{M}(n)$, it is well known that

$$\partial\|A\| = \{G \in \mathbb{M}(n) : \|A\| = \text{Re}\,\text{tr}(G^*A), \|G\|^* \le 1\}, \tag{2.5}$$

where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$, and

$$\|T\| = \sup_{\|X\|^*=1} |\,\text{tr}(T^*X)| = \sup_{\|X\|^*=1} \text{Re}\,\text{tr}(T^*X). \tag{2.6}$$

The subdifferentials of some classes of matrix norms, namely unitarily invariant norms and induced norms, have been computed by Watson [19]. The following expression for the subdifferential of the Ky Fan k-norms was also given by him in [20]. Let $1 \le k \le n$. Let the multiplicity of $s_k(A)$ be $r + q$, where $r \ge 0$ and $q \ge 1$, such that

$$s_{k-q+1}(A) = \cdots = s_{k+r}(A).$$

Let $g : \mathbb{M}(n) \to \mathbb{R}$ be the function defined as $g(A) = \|A\|_{(k)}$.

Theorem 2.4 ([20])**.** *Let $A = USV^*$ be a singular value decomposition of A and let the matrices U, V be partitioned as $U = [U_1 : U_2 : U_3]$ and $V = [V_1 : V_2 : V_3]$ where $U_1, V_1 \in \mathbb{M}(n, k-q); U_2, V_2 \in \mathbb{M}(n, r+q); U_3, V_3 \in \mathbb{M}(n, n-k-r)$. If $s_k(A) > 0$, then $G \in \partial g(A)$ if and only if there exists a positive semidefinite matrix $T \in \mathbb{M}(r+q)$ with $\lambda_1(T) \le 1$ and $\sum_{j=1}^{r+q} \lambda_j(T) = q$ such that $G = U_1V_1^* + U_2TV_2^*$. If $s_k(A) = 0$, then $G \in \partial g(A)$ if and only if there exists $T \in \mathbb{M}(n-k+q, r+q)$ with $s_1(T) \le 1$ and $\sum_{j=1}^{r+q} s_j(T) \le q$ such that $G = U_1V_1^* + [U_2 : U_3]TV_2^*$.*

We obtain new formulas for $\partial g(A)$ that can be used more easily in our problem. The computations are similar to the ones in [19]. To do so, we first calculate $g'_+(A,\cdot)$. For this, an important thing to observe is that the Ky Fan k-norm of a matrix A is also given by

$$\|A\|_{(k)} = \max_{\substack{U,V\in\mathbb{M}(n,k)\\ U^*U=V^*V=I_k}} \operatorname{Re}\operatorname{tr} U^*AV = \max_{\substack{U,V\in\mathbb{M}(n,k)\\ U^*U=V^*V=I_k}} |\operatorname{tr} U^*AV|. \tag{2.7}$$

(See [13, p. 791].) If A is positive semidefinite, then

$$\|A\|_{(k)} = \max_{\substack{U\in\mathbb{M}(n,k)\\ U^*U=I_k}} \operatorname{tr} U^*AU. \tag{2.8}$$

Theorem 2.5. *For $X \in \mathbb{M}(n)$,*

$$g'_+(A,X) = \max_{\substack{u_1,\dots,u_k \text{ o.n.}\\ v_1,\dots,v_k \text{ o.n.}\\ Av_i=s_i(A)u_i}} \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xv_i\rangle. \tag{2.9}$$

Proof. From (2.7), we have

$$\|A\|_{(k)} = \max_{\substack{u_1,\dots,u_k \text{ o.n.}\\ v_1,\dots,v_k \text{ o.n.}}} \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Av_i\rangle. \tag{2.10}$$

For any sets of k orthonormal vectors $u_1,\dots,u_k$ and $v_1,\dots,v_k$ satisfying $Av_i = s_i(A)u_i, 1 \le i \le k$, we have

$$\begin{aligned}
\|A+tX\|_{(k)} &\geq \sum_{i=1}^{k} \operatorname{Re}\langle u_i, (A+tX)v_i\rangle \\
&= \sum_{i=1}^{k} s_i(A) + t\sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xv_i\rangle \\
&= \|A\|_{(k)} + t\sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xv_i\rangle.
\end{aligned}$$

This gives for $t > 0$,

$$\frac{\|A+tX\|_{(k)} - \|A\|_{(k)}}{t} \geq \max_{\substack{u_1,\dots,u_k \text{ o.n.}\\ v_1,\dots,v_k \text{ o.n.}\\ Av_i=s_i(A)u_i}} \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xv_i\rangle. \tag{2.11}$$

Now for any sets of k orthonormal vectors $u_1(t),\dots,u_k(t)$ and $v_1(t),\dots,v_k(t)$ satisfying

$$(A+tX)v_i(t) = s_i(A+tX)u_i(t), \quad 1 \le i \le k, \tag{2.12}$$

we have

$$\begin{aligned} \|A\|_{(k)} &\geq \sum_{i=1}^{k} \mathrm{Re}\langle u_i(t), Av_i(t)\rangle \\ &= \sum_{i=1}^{k} s_i(A+tX) - t\sum_{i=1}^{k} \mathrm{Re}\langle u_i(t), Xv_i(t)\rangle \\ &= \|A+tX\|_{(k)} - t\sum_{i=1}^{k} \mathrm{Re}\langle u_i(t), Xv_i(t)\rangle. \end{aligned}$$

So for each $t > 0$, we obtain

$$\frac{\|A+tX\|_{(k)} - \|A\|_{(k)}}{t} \leq \sum_{i=1}^{k} \mathrm{Re}\langle u_i(t), Xv_i(t)\rangle. \tag{2.13}$$

Consider a sequence $\{t_n\}$ of positive real numbers converging to zero as $n \to \infty$. Since the unit ball in $\mathbb{C}^n$ is compact, there exists a subsequence $\{t_{n_m}\}$ of $\{t_n\}$ such that for each $1 \leq i \leq k$, there exist u_i' and v_i' such that $\{u_i(t_{n_m})\}$ and $\{v_i(t_{n_m})\}$ converge to u_i' and v_i', respectively, as $m \to \infty$. Then the sets of vectors $u_1', \ldots, u_k'$ and $v_1', \ldots, v_k'$ are orthonormal. By continuity of singular values, we also know that

$$s_i(A + t_{n_m}B) \to s_i(A) \text{ as } m \to \infty. \tag{2.14}$$

Hence we obtain $Av_i' = s_i(A)u_i'$ for all $1 \leq i \leq k$. By (2.13), we get that

$$g_+'(A,X) = \lim_{m\to\infty} \frac{\|A+t_{n_m}X\|_{(k)} - \|A\|_{(k)}}{t_{n_m}} \leq \max_{\substack{u_1,\ldots,u_k \text{ o.n.} \\ v_1,\ldots,v_k \text{ o.n.} \\ Av_i = s_i(A)u_i}} \sum_{i=1}^{k} \mathrm{Re}\langle u_i, Xv_i\rangle. \tag{2.15}$$

Combining this with (2.11), we obtain the required result. □

The above proof works equally well if the maximum in (2.9) is taken over the sets of orthonormal vectors $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ such that for each $1 \leq i \leq k$, u_i and v_i are left and right singular vectors of A, respectively, corresponding to the ith singular value $s_i(A)$ of A. We note here that for each $t > 0$, if along with (2.12), we also have

$$(A+tX)^* u_i(t) = s_i(A+tX)v_i(t),$$

then by passing onto a subsequence $\{t_{n_m}\}$ as in the above proof, and taking the limit as $m \to \infty$, we obtain

$$A^* u_i' = s_i(A)v_i'.$$

So for each $X \in \mathbb{M}(n)$, we get

$$g'_+(A,X) = \max_{\substack{u_1,\dots,u_k \text{ o.n.} \\ v_1,\dots,v_k \text{ o.n.} \\ Av_i = s_i(A)u_i \\ A^*u_i = s_i(A)v_i}} \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xv_i\rangle. \tag{2.16}$$

Corollary 2.6. Let A be positive semidefinite. Let $\lambda_1(A) \geq \cdots \geq \lambda_n(A) \geq 0$ be the eigenvalues of A, with $\lambda_k(A) > 0$. Then

$$g'_+(A,X) = \max_{\substack{u_1,\dots,u_k \text{ o.n.} \\ Au_i = \lambda_i(A)u_i}} \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Xu_i\rangle. \tag{2.17}$$

Proof. We know that if $Av = \lambda u$ and $Au = \lambda v$, where $\lambda > 0$, then $u = v$. Using this, the required result follows from (2.16). $\square$

Theorem 2.7. *Let $A \in \mathbb{M}(n)$. Then*

$$\begin{aligned} \partial g(A) \quad = \quad & \operatorname{conv}\Bigg\{ \sum_{i=1}^{k} u_i v_i^* : u_1,\dots,u_k, v_1,\dots,v_k \in \mathbb{C}^n, u_1,\dots,u_k \text{ o.n.}, v_1,\dots,v_k \text{ o.n.}, \\ & \qquad Av_i = s_i(A)u_i \textit{ for all } 1 \leq i \leq k \Bigg\} \qquad (2.18) \\ = \quad & \operatorname{conv}\Bigg\{ \sum_{i=1}^{k} u_i v_i^* : u_1,\dots,u_k, v_1,\dots,v_k \in \mathbb{C}^n, u_1,\dots,u_k \text{ o.n.}, v_1,\dots,v_k \text{ o.n.}, \\ & \qquad Av_i = s_i(A)u_i, A^*u_i = s_i(A)v_i \textit{ for all } 1 \leq i \leq k \Bigg\}. \qquad (2.19) \end{aligned}$$

Proof. Denote the set on the right hand side of (2.18) by $\mathbb{H}(A)$. Let $G \in \mathbb{H}(A)$. Then

$$G = \sum_{i=1}^{k} u_i v_i^*,$$

where $u_1,\dots,u_k$ and $v_1,\dots,v_k$ are orthonormal sets of vectors such that $Av_i = s_i(A)u_i$ for all $1 \leq i \leq k$. So

$$\begin{aligned} \operatorname{Re}\operatorname{tr}(G^*A) &= \sum_{i=1}^{k} \operatorname{Re}\langle u_i, Av_i\rangle \\ &= \sum_{i=1}^{k} s_i(A) \\ &= \|A\|_{(k)}, \end{aligned}$$

and

$$\begin{aligned}\operatorname{Re}\operatorname{tr}(G^*X) &= \sum_{i=1}^{k}\operatorname{Re}\langle u_i, Xv_i\rangle \\ &\leq \|X\|_{(k)}.\end{aligned}$$

Thus

$$\|G\|^* \leq 1.$$

So we get by (2.5) that $\mathbb{H}(A) \subseteq \partial g(A)$, and therefore $\operatorname{conv}\mathbb{H}(A) \subseteq \partial g(A)$.

Now let $G \in \partial g(A)$. Suppose $G \notin \operatorname{conv}\mathbb{H}(A)$. The set $\mathbb{H}(A)$ is compact, and so is its convex hull. By the Separating Hyperplane Theorem, there exists $X \in \mathbb{M}(n)$ such that for all sets of k orthonormal vectors $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ satisfying $Av_i = s_i(A)u_i$ for $1 \leq i \leq k$, we have

$$\operatorname{Re}\operatorname{tr}\left(X^*\left(\sum_{i=1}^{k} u_i v_i^* - G\right)\right) < 0.$$

This implies

$$\max_{\substack{u_1,\ldots,u_k \text{ o.n.}\\ v_1,\ldots,v_k \text{ o.n.}\\ Av_i = s_i(A)u_i}} \sum_{i=1}^{k}\operatorname{Re}\langle u_i, Xv_i\rangle < \max_{G\in\partial g(A)} \operatorname{Re}\operatorname{tr}(X^*G).$$

By (2.3), the right hand side is $g'_+(A, X)$. By (2.9), this should be equal to the left hand side. This gives a contradiction. Thus we obtain (2.18).

The expression (2.19) can be proved similarly by using (2.16), instead of (2.9).

□

Corollary 2.8. Let A be a positive semidefinite matrix, with eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_n(A) \geq 0$ such that $\lambda_k(A) > 0$. Then

$$\partial g(A) = \operatorname{conv}\left\{\sum_{i=1}^{k} u_i u_i^* : u_1, \ldots, u_k \in \mathbb{C}^n, u_1, \ldots, u_k \text{ o.n.}, Au_i = \lambda_i(A)u_i \text{ for all } 1 \leq i \leq k\right\}. \tag{2.20}$$

3 Proofs

To prove Theorem 1.1, we require the following lemma.

Lemma 3.1. Let $X, Y \in \mathbb{M}(n)$ and let Y be positive semidefinite. Let $\lambda_1(Y) \geq \cdots \geq \lambda_n(Y) \geq 0$ be the eigenvalues of Y. For $1 \leq r \leq n$, let

$$\mathcal{W}(X, Y) = \left\{\sum_{i=1}^{r}\langle u_i, Xu_i\rangle : u_1, \ldots, u_r \in \mathbb{C}^n, u_1, \ldots, u_r \text{ o.n.}, Yu_i = \lambda_i(Y)u_i \text{ for all } 1 \leq i \leq r\right\}.$$

Then $\mathcal{W}(X, Y)$ is a convex set.

Proof. Let the number of distinct eigenvalues of Y be ℓ and let $\mathcal{H}_1, \ldots, \mathcal{H}_\ell$ be the respective eigenspaces. Let $m_1, \ldots, m_\ell$ be the dimensions of $\mathcal{H}_1, \ldots, \mathcal{H}_\ell$, respectively. Let $1 \leq \ell' \leq \ell$ be such that $m_1 + \cdots + m_{\ell'-1} < r \leq m_1 + \cdots + m_{\ell'}$. Let $m = r - (m_1 + \cdots + m_{\ell'-1})$. Set

$$\mathcal{W}_j(X) = \left\{ \sum_{i=1}^{m_j} \langle u_i, X u_i \rangle : u_1, \ldots, u_{m_j} \in \mathcal{H}_j, u_1, \ldots, u_{m_j} \text{ o.n.} \right\} \text{ for } 1 \leq j \leq \ell' - 1,$$

and

$$\mathcal{W}_{\ell'}(X) = \left\{ \sum_{i=1}^{m} \langle u_i, X u_i \rangle : u_1, \ldots, u_m \in \mathcal{H}_{\ell'}, u_1, \ldots, u_m \text{ o.n.} \right\}.$$

Since $\mathcal{H}_1, \ldots, \mathcal{H}_\ell$ are mutually orthogonal, we have

$$\mathcal{W}(X, Y) = \sum_{j=1}^{\ell'} \mathcal{W}_j(X). \tag{3.1}$$

Note that $\mathcal{W}_j(X)$ is a singleton set for $1 \leq j \leq \ell' - 1$. Hence it is sufficient to show that $\mathcal{W}_{\ell'}(X)$ is convex. Let $\mathscr{P}_{\ell'}$ be the orthogonal projection from $\mathbb{C}^n$ onto $\mathcal{H}_{\ell'}$, and let $\iota_{\ell'}$ denote its adjoint (which is the inclusion map of $\mathcal{H}_{\ell'}$ into $\mathbb{C}^n$). Then $\mathcal{W}_{\ell'}(X)$ is the m-numerical range of $\mathscr{P}_{\ell'} X \iota_{\ell'}$, which is convex (see [8, p. 315]). □

We now state and prove a real version of Theorem 1.1.

Theorem 3.2. *Let $A = U|A|$ be a polar decomposition of A. If there exist k orthonormal vectors $u_1, u_2, \ldots, u_k$ such that*

$$|A| \; u_i = s_i(A) u_i \text{ for all } 1 \leq i \leq k \tag{3.2}$$

and

$$\sum_{i=1}^{k} \mathrm{Re}\langle u_i, U^* B u_i \rangle = 0, \tag{3.3}$$

then

$$\|A + tB\|_{(k)} \geq \|A\|_{(k)} \text{ for all } t \in \mathbb{R}. \tag{3.4}$$

If $s_k(A) > 0$, then the converse is also true.

Proof. First suppose that there exist k orthonormal vectors $u_1, u_2, \ldots, u_k$ such that $|A| \; u_i = s_i(A) \; u_i$ for all $1 \leq i \leq k$ and $\sum_{i=1}^{k} \mathrm{Re}\langle u_i, U^* B u_i \rangle = 0$. We have

$$\|A + tB\|_{(k)} = \||A| + tU^* B\|_{(k)}$$

and by (2.7),

$$\||A| + tU^* B\|_{(k)} \geq \sum_{i=1}^{k} \mathrm{Re}\langle u_i, (|A| + tU^* B) u_i \rangle.$$

So we get

$$\begin{aligned}\|A+tB\|_{(k)} &\geq \sum_{i=1}^{k}\langle u_i, |A|u_i\rangle + t\sum_{i=1}^{k}\mathrm{Re}\langle u_i, U^*Bu_i\rangle \\ &= \sum_{i=1}^{k} s_i(A) \\ &= \|A\|_{(k)}.\end{aligned}$$

Now suppose that $s_k(A) > 0$ and

$$\|A+tB\|_{(k)} \geq \|A\|_{(k)} \text{ for all } t \in \mathbb{R}.$$

This can also be written as

$$\||A| + tU^*B\|_{(k)} \geq \||A|\|_{(k)} \text{ for all } t \in \mathbb{R}. \tag{3.5}$$

Let $S : \mathbb{R} \to \mathbb{M}(n)$ be the map given by $S(t) = tU^*B$, $L : \mathbb{R} \to \mathbb{M}(n)$ be the map defined as $L(t) = |A| + tU^*B$ and $g : \mathbb{M}(n) \to \mathbb{R}_+$ be the map defined by $g(X) = \|X\|_{(k)}$. Then we have that $g \circ L$ attains its minimum at zero. By Proposition 2.2, we obtain that $0 \in \partial(g \circ L)(0)$. Using Proposition 2.3, we obtain

$$0 \in S^*\partial g(|A|). \tag{3.6}$$

By Corollary 2.8, this is equivalent to saying that

$$0 \in \mathrm{conv}\left\{\mathrm{Re}\sum_{i=1}^{k}\langle u_i, U^*Bu_i\rangle : u_1,\ldots,u_k \in \mathbb{C}^n, u_1,\ldots,u_k \text{ o.n.}, |A|u_i = \lambda_i(|A|)u_i \text{ for all } 1 \leq i \leq k\right\}.$$

The set in the above equation is $\mathrm{conv}(\mathrm{Re}\,\mathcal{W}(U^*B, |A|))$. By Lemma 3.1, $\mathrm{Re}\,\mathcal{W}(U^*B, |A|)$ is a convex set. So there exist k orthonormal vectors $u_1, \ldots, u_k$ such that

$$|A|u_i = s_i(A)u_i$$

and

$$\mathrm{Re}\sum_{i=1}^{k}\langle u_i, U^*Bu_i\rangle = 0.$$

□

Proof of Theorem 1.1 Suppose that there exist k orthonormal vectors $u_1, u_2, \ldots, u_k$ satisfying (1.3) and (1.4). Let $\lambda \in \mathbb{C}$. Then similar to the argument in the proof of Theorem 3.2, we get

$$\begin{aligned} \|A+\lambda B\|_{(k)} &= \||A|+\lambda U^*B\|_{(k)} \\ &\geq \left|\sum_{i=1}^{k}\langle u_i, (|A|+\lambda U^*B)u_i\rangle\right| \\ &= \left|\sum_{i=1}^{k}\langle u_i, |A|u_i\rangle + \lambda\sum_{i=1}^{k}\langle u_i, U^*Bu_i\rangle\right| \\ &= \sum_{i=1}^{k} s_i(A) \\ &= \|A\|_{(k)}. \end{aligned}$$

So A is orthogonal to B in $\|\cdot\|_{(k)}$.

Conversely, let $s_k(A)>0$ and A is orthogonal to B in $\|\cdot\|_{(k)}$. So

$$\||A|+re^{i\theta}U^*B\|_{(k)} \geq \|A\|_{(k)} \text{ for all } r,\theta\in\mathbb{R}.$$

For $\theta\in\mathbb{R}$, let $B^{(\theta)}=e^{i\theta}B$. Then we get

$$\||A|+rU^*B^{(\theta)}\|_{(k)} \geq \|A\|_{(k)} \text{ for all } r\in\mathbb{R}.$$

By Theorem 3.2, there exist k orthonormal vectors $u_1^{(\theta)},\ldots,u_k^{(\theta)}$ such that

$$|A|u_j^{(\theta)} = s_j(A)u_j^{(\theta)} \text{ for all } 1\leq j\leq k$$

and

$$\operatorname{Re}\sum_{j=1}^{k}\langle u_j^{(\theta)}, U^*B^{(\theta)}u_j^{(\theta)}\rangle = 0, \text{ that is, } \operatorname{Re} e^{i\theta}\sum_{j=1}^{k}\langle u_j^{(\theta)}, U^*Bu_j^{(\theta)}\rangle = 0. \tag{3.7}$$

Now by Lemma 3.1, the set $\mathcal{W}(U^*B,|A|)$ is convex in $\mathbb{C}$. It is also compact in $\mathbb{C}$. If $0\notin\mathcal{W}(U^*B,|A|)$, then by the Separating Hyperplane Theorem, there exists a θ_0 such that

$$\operatorname{Re} e^{i\theta_0}\sum_{j=1}^{k}\langle u_j, U^*Bu_j\rangle > 0 \text{ for all } u_1,\ldots,u_k \text{ o.n., } |A|u_j = s_j(A)u_j \text{ for } 1\leq j\leq k.$$

This is a contradiction to (3.7). Thus $0\in\mathcal{W}(U^*B,|A|)$, and so there exist k orthonormal vectors $u_1,\ldots,u_k$ such that

$$|A|u_i = s_i(A)u_i \text{ for all } 1\leq i\leq k$$

and

$$\sum_{i=1}^{k}\langle u_i, U^*Bu_i\rangle = 0.$$

Proof of Theorem 1.2 Let $S, L : \mathbb{C} \to \mathbb{M}(n)$ and $g : \mathbb{M}(n) \to \mathbb{R}_+$ be the maps defined as $S(\lambda) = \lambda B$, $L(\lambda) = A + \lambda B$ and $g(X) = \|X\|_{(k)}$. Then we get $\|A + \lambda B\|_{(k)} \geq \|A\|_{(k)}$ for all $\lambda \in \mathbb{C}$ if and only if $g \circ L$ attains its minimum at 0. By Proposition 2.2 and Proposition 2.3, a necessary and sufficient condition for this is that $0 \in S^* \partial g(A)$. Let the matrices U, V be partitioned as $U = [U_1 : U_2 : U_3]$ and $V = [V_1 : V_2 : V_3]$, where $U_1, V_1 \in \mathbb{M}(n, k-q); U_2, V_2 \in \mathbb{M}(n, r+q); U_3, V_3 \in \mathbb{M}(n, n-k-r)$. If $s_k(A) > 0$, then by Theorem 2.4, we get that $0 \in S^* \partial g(A)$ if and only if there exists a positive semidefinite matrix $T \in \mathbb{M}(r+q)$ with $\lambda_1(T) \leq 1$ and $\sum_{j=1}^{r+q} \lambda_j(T) = q$ such that $\operatorname{tr} B^*(U_1 V_1^* + U_2 T V_2^*) = 0$. Similarly, when $s_k(A) = 0$, we get that $0 \in S^* \partial g(A)$ if and only if there exists $T \in \mathbb{M}(n-k+q, r+q)$ with $s_1(T) \leq 1$ and $\sum_{j=1}^{r+q} s_j(T) \leq q$ such that $\operatorname{tr} B^*(U_1 V_1^* + [U_2 : U_3] T V_2^*) = 0$. A calculation shows that

$$\operatorname{tr} B^*(U_1 V_1^* + U_2 T V_2^*) = \operatorname{tr} B_{11}^* + \operatorname{tr}(B_{22}^* T)$$

and

$$\operatorname{tr} B^*(U_1 V_1^* + [U_2 : U_3] T V_2^*) = \operatorname{tr} B_{11}^* + \operatorname{tr}([B_{22}^* : B_{32}^*]\, T).$$

This gives the required result.

Proof of Theorem 1.3 First suppose that there exist density matrices $P_1, \ldots, P_k$ such that $\| \sum_{i=1}^k P_i \|_\infty \leq 1$,

$$|A| P_i = s_i(A) P_i \quad \text{for all } 1 \leq i \leq k \tag{3.8}$$

and $U \sum_{i=1}^k P_i \in \mathscr{W}^\perp$. Let $Q = \sum_{i=1}^k P_i$. Then Q is a positive semidefinite matrix such that $\|Q\|_\infty \leq 1$, $\frac{1}{k}\|Q\|_1 = \frac{1}{k}\sum_{i=1}^k \operatorname{tr} P_i = 1$ and

$$\operatorname{tr}(W^* U Q) = 0 \text{ for all } W \in \mathscr{W}. \tag{3.9}$$

So by using (2.6) and the fact that $\|X\|_{(k)}^* = \max\{\|X\|_\infty, \frac{1}{k}\|X\|_1\}$ [4, Ex. IV.2.12]., we get that for any $W \in \mathscr{W}$,

$$\begin{aligned}
\|A + W\|_{(k)} &= \||A| + U^* W\|_{(k)} \\
&\geq \operatorname{tr}(|A| Q + U^* W Q) \\
&= \operatorname{tr}(|A| Q) \qquad \text{(by (3.9))} \\
&= \sum_{i=1}^k \operatorname{tr} |A| P_i \\
&= \|A\|_{(k)} \qquad \text{(by (3.8))}.
\end{aligned}$$

Conversely, suppose A is orthogonal to $\mathscr{W}$ in $\|\cdot\|_{(k)}$ and $s_k(A) > 0$. Define $S : \mathscr{W} \to \mathbb{M}(n)$ as $S(W) = U^* W$. Then $S^* : \mathbb{M}(n) \to \mathscr{W}$ is given by $S^*(T) = \mathscr{P}_{\mathscr{W}}(UT)$, where $\mathscr{P}_{\mathscr{W}}$ is the orthogonal projection onto the subspace $\mathscr{W}$. Let $L : \mathscr{W} \to \mathbb{M}(n)$ be the map defined as $L(W) = |A| + U^* W$ and let $g : \mathbb{M}(n) \to \mathbb{R}_+$ be the map defined as $g(X) = \|X\|_{(k)}$. Then by Proposition 2.2 and Proposition 2.3, we have that $\|A + W\|_{(k)} \geq \|A\|_{(k)}$ for all $W \in \mathscr{W}$ if and only if $0 \in S^* \partial g(|A|)$. By Corollary 2.8, there exist numbers $t_1, \ldots, t_m$ such that $0 \leq t_j \leq$

1, $\sum_{j=1}^m t_j = 1$ and for each $1 \le j \le m$, there exist k orthonormal vectors $u_1^{(j)}, \dots, u_k^{(j)}$ such that

$$|A|u_i^{(j)} = s_i(A)u_i^{(j)} \text{ for all } 1 \le i \le k \tag{3.10}$$

and

$$S^*\left(\sum_{i=1}^k \sum_{j=1}^m t_j u_i^{(j)} u_i^{(j)*}\right) = 0. \tag{3.11}$$

Let $P_i = \sum_{j=1}^m t_j u_i^{(j)} u_i^{(j)*}$. Then each P_i is a density matrix. Also, by (3.10), we get $|A|P_i = s_i(A)P_i$. Equation (3.11) says that $S^*(\sum_{i=1}^k P_i) = 0$, which is equivalent to saying that $U\sum_{i=1}^k P_i \in \mathscr{W}^\perp$. For each $1 \le j \le m$, the matrix $\sum_{i=1}^k u_i^{(j)} u_i^{(j)*}$ is an orthogonal projection of rank k onto the linear span of $\{u_i^{(j)} : 1 \le i \le k\}$. In particular $\|\sum_{i=1}^k u_i^{(j)} u_i^{(j)*}\|_\infty \le 1$. Thus

$$\begin{aligned} \|\sum_{i=1}^k P_i\|_\infty &= \|\sum_{j=1}^m t_j \sum_{i=1}^k u_i^{(j)} u_i^{(j)*}\|_\infty \\ &\le \sum_{j=1}^m t_j \|\sum_{i=1}^k u_i^{(j)} u_i^{(j)*}\|_\infty \\ &\le 1. \end{aligned}$$

4 Remarks

1. Another necessary and sufficient condition for A to be orthogonal to B in $\|\cdot\|_1$ given in [12] is that there exists a matrix $G \in \mathbb{M}(n)$ such that $\|G\|_\infty \le 1$, $\operatorname{tr}(G^*A) = \|A\|_1$ and $\operatorname{tr}(G^*B) = 0$. One can derive an analogous characterization for orthogonality in $\|\cdot\|_{(k)}$ using (2.5). We can show that A is orthogonal to B in $\|\cdot\|_{(k)}$ if and only if there exists a matrix $G \in \mathbb{M}(n)$ such that $\|G\|_\infty \le 1$, $\|G\|_1 \le k$, $\operatorname{tr}(G^*A) = \|A\|_{(k)}$ and $\operatorname{tr}(G^*B) = 0$. Let S, L, g be the maps as defined above in the proof of Theorem 1.2. Then Proposition 2.2, Proposition 2.3 and (2.5) gives that A is orthogonal to B in $\|\cdot\|_{(k)}$ if and only if there exists a matrix $G \in \mathbb{M}(n)$ such that $\|G\|_\infty \le 1$, $\|G\|_1 \le k$, $\operatorname{Re}\operatorname{tr}(G^*A) = \|A\|_{(k)}$ and $\operatorname{tr}(G^*B) = 0$. We observe that if $\|G\|_{(k)}^* \le 1$ then $\operatorname{Re}\operatorname{tr}(G^*A) = \|A\|_{(k)}$ if and only if $\operatorname{tr}(G^*A) = \|A\|_{(k)}$. This is because if $\operatorname{Re}\operatorname{tr}(G^*A) = \|A\|_{(k)}$, then

   $$\|A\|_{(k)} \le |\operatorname{tr}(G^*A)| \le \|G\|_{(k)}^* \|A\|_{(k)} \le \|A\|_{(k)}.$$

   So $\operatorname{Im}\operatorname{tr}(G^*A) = 0$ and hence $\operatorname{tr}(G^*A) = \operatorname{Re}\operatorname{tr}(G^*A) = \|A\|_{(k)}$. Thus we obtain the required result.

2. The characterizations for Birkhoff-James orthogonality are closely related to the recent work in norm parallelism [17, 22, 14]. In a normed linear space, an element x is said to be *norm-parallel* to another element

y (denoted as $x||y$) if $\|x+\lambda y\| = \|x\|+\|y\|$ for some $\lambda \in \mathbb{C}, |\lambda| = 1$. Let $A = U|A|$ be a polar decomposition of A and $s_k(A) > 0$. Then by Theorem 2.4 in [14] and Theorem 1.1, we get that $A||B$ in $\|\cdot\|_{(k)}$ if and only if there exists $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ and k orthonormal vectors $u_1, u_2, \ldots, u_k$ such that $|A| \; u_i = s_i(A)u_i$ for all $1 \leq i \leq k$ and $\sum_{i=1}^{k} \langle u_i, U^*(\|B\|_{(k)}A + \lambda\|A\|_{(k)}B)u_i\rangle = 0$. Simplifying the expressions and using the fact that $|\lambda| = 1$, we obtain that $A||B$ in $\|\cdot\|_{(k)}$ if and only if there exist k orthonormal vectors $u_1, u_2, \ldots, u_k$ such that $|A| \; u_i = s_i(A)u_i$ for all $1 \leq i \leq k$ and $|\sum_{i=1}^{k} \langle u_i, U^*Bu_i\rangle| = \|B\|_{(k)}$. For $k = 1$, this is just Corollary 2.15 of [14].

Acknowledgement I would like to thank the referee for several valuable comments and suggestions.

References

[1] L. Arambašić, R. Rajić, The Birkhoff-James orthogonality in Hilbert C*-modules, *Linear Algebra Appl.* 437 (2012) 1913–1929.

[2] L. Arambašić, R. Rajić, A strong version of the Birkhoff-James orthogonality in Hilbert C*-modules, *Ann. Funct. Anal.* 5 (2014) 109–120.

[3] L. Arambašić, R. Rajić, On three concepts of orthogonality in Hilbert C*-modules, *Linear Multilinear Algebra* 63 (2015) 1485–1500.

[4] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.

[5] R. Bhatia, P. Šemrl, Orthogonality of Matrices and Some Distance Problems, *Linear Algebra Appl.* 287 (1999) 77-86.

[6] C. Benítez, M. Fernández, M.L. Soriano, Orthogonality of matrices, *Linear Algebra Appl.* 422 (2007) 155–163.

[7] T. Bhattacharyya, P. Grover, Characterization of Birkhoff-James orthogonality, *J. Math. Anal. Appl.* 407 (2013) 350–358.

[8] P.R. Halmos, *A Hilbert Space Problem Book*, Narosa Publishing House, 1978.

[9] J.B. Hiriart-Urruty, C. Lemarèchal, *Fundamentals of Convex Analysis*, Springer, 2000.

[10] P. Grover, Orthogonality to matrix subspaces, and a distance formula, *Linear Algebra Appl.* 445 (2014) 280–288.

[11] D.J. Kečkiċ, Gateaux derivative of B(H) norm, *Proc. Amer. Math. Soc.* 133 (2005) 2061-2067.

[12] C.K. Li, H. Schneider, Orthogonality of Matrices, *Linear Algebra Appl.* 347 (2002) 115–122.

[13] A.W. Marshall, I. Olkin, B. C. Arnold *Inequalities: Theory of Majorization and Its Applications*, Springer, 2011.

[14] M.S. Moslehian, A. Zamani, Norm-parallelism in the geometry of Hilbert C^*-modules, *Indag. Math.* 27 (2016) 266–281.

[15] D. Sain, K. Paul, Operator norm attainment and inner product spaces, *Linear Algebra Appl.* 439 (2013) 2448–2452.

[16] D. Sain, K. Paul, S. Hait, Operator norm attainment and Birkhoff-James orthogonality, *Linear Algebra Appl.* 476 (2015) 85–97.

[17] A. Seddik, Rank one operators and norm of elementary operators, *Linear Algebra Appl.* 424 (2007) 177–183.

[18] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Springer, 1970.

[19] G.A. Watson, Characterization of the Subdifferential of Some Matrix Norms, *Linear Algebra Appl.* 170 (1992) 33-45.

[20] G.A. Watson, On matrix approximation problems with Ky Fan k norms, *Numer. Algo.* 5 (1993) 263–272.

[21] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002.

[22] A. Zamani, M.S. Moslehian, Exact and approximate operator parallelism, *Canad. Math. Bull.* 58 (2015) 207–224.