Linear Algebra and its Applications 445 (2014) 280–288

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Orthogonality to matrix subspaces, and a distance formula

Priyanka Grover

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 6 August 2013
Accepted 26 November 2013
Available online 25 December 2013
Submitted by P. Semrl

MSC:
15A60
15A09
47A12

Keywords:
Birkhoff–James orthogonality
Subdifferential
Singular value decomposition
Moore–Penrose inverse
Pinching
Variance

ABSTRACT

We obtain a necessary and sufficient condition for a matrix A to be Birkhoff–James orthogonal to any subspace $\mathscr{W}$ of $\mathbb{M}_n(\mathbb{C})$. Using this we obtain an expression for the distance of A from any unital C^*-subalgebra of $\mathbb{M}_n(\mathbb{C})$.

1. Introduction

Let $\mathbb{M}_n(\mathbb{C})$ be the space of $n \times n$ complex matrices and let $\mathscr{W}$ be any subspace of $\mathbb{M}_n(\mathbb{C})$. For any $A \in \mathbb{M}_n(\mathbb{C})$, let

$$\|A\| = \max_{x \in \mathbb{C}^n, \|x\|=1} \|Ax\|$$

be the *operator norm* of A. Then A is said to be (*Birkhoff–James*) *orthogonal* to $\mathscr{W}$ if

$$\|A + W\| \geqslant \|A\| \quad \text{for all } W \in \mathscr{W}. \tag{1}$$

E-mail address: pgrover8r@isid.ac.in.

0024-3795/$ – see front matter
http://dx.doi.org/10.1016/j.laa.2013.11.040

The space $\mathbb{M}_n(\mathbb{C})$ is a complex Hilbert space under the inner product $\langle A, B\rangle_c = \operatorname{tr}(A^*B)$ and a real Hilbert space under the inner product $\langle A, B\rangle_r = \operatorname{Re}\operatorname{tr}(A^*B)$. Let $\mathscr{W}^\perp$ be the orthogonal complement of $\mathscr{W}$, where the orthogonal complement is taken with respect to the usual Hilbert space orthogonality in $\mathbb{M}_n(\mathbb{C})$ with the inner product $\langle\cdot,\cdot\rangle_c$ or $\langle\cdot,\cdot\rangle_r$, depending upon whether $\mathscr{W}$ is a complex or a real subspace.

Bhatia and Šemrl [6] obtained an interesting characterization of orthogonality when $\mathscr{W} = \mathbb{C}B$, where B is any matrix in $\mathbb{M}_n(\mathbb{C})$. They showed that A is orthogonal to $\mathbb{C}B$ if and only if there exists a unit vector x such that $\|Ax\| = \|A\|$ and $\langle Ax, Bx\rangle = 0$. In other words, A is orthogonal to $\mathbb{C}B$ if and only if there exists a positive semidefinite matrix P of rank one such that $\operatorname{tr} P = 1$, $\operatorname{tr} A^*AP = \|A\|^2$ and $AP \in (\mathbb{C}B)^\perp$. Positive semidefinite matrices with trace 1 are called *density matrices*. We use the notation $P \geqslant 0$ to mean P is positive semidefinite.

Let $\mathscr{W} = \mathbb{D}_n(\mathbb{R})$, the subspace of all diagonal matrices with real entries, and let A be any Hermitian matrix. Then A is called *minimal* if $\|A + D\| \geqslant \|A\|$ for all $D \in \mathbb{D}_n(\mathbb{R})$. Andruchow, Larotonda, Recht, and Varela [1, Theorem 1] showed that a Hermitian matrix A is minimal if and only if there exists a density matrix P such that $PA^2 = \|A\|^2P$ and all diagonal entries of PA are zero. In our notation, A is minimal is the same as saying that A is orthogonal to the subspace $\mathbb{D}_n(\mathbb{R})$. If A is Hermitian, then note that A is orthogonal to $\mathbb{D}_n(\mathbb{R})$ if and only if A is orthogonal to $\mathbb{D}_n(\mathbb{C})$. Now $\mathbb{D}_n(\mathbb{C})^\perp$ is the subspace of all matrices such that their diagonal entries are zero. The condition $PA^2 = \|A\|^2P$ is the same as $A^2P = \|A\|^2P$ and diagonal entries of PA are the same as diagonal entries of AP. Therefore Theorem 1 in [1] can be interpreted as follows. A Hermitian matrix A is orthogonal to $\mathbb{D}_n(\mathbb{C})$ if any only if $A^2P = \|A\|^2P$ and $AP \in \mathbb{D}_n(\mathbb{C})^\perp$. The following theorem is a generalization of this result as well as Bhatia–Šemrl theorem.

Theorem 1. *Let $A \in \mathbb{M}_n(\mathbb{C})$ and let $m(A)$ be the multiplicity of the maximum singular value $\|A\|$ of A. Let $\mathscr{W}$ be any (real or complex) subspace of $\mathbb{M}_n(\mathbb{C})$. Then A is orthogonal to $\mathscr{W}$ if and only if there exists a density matrix P of complex rank at most $m(A)$ such that $A^*AP = \|A\|^2P$ and $AP \in \mathscr{W}^\perp$. (If rank $P = \ell$, then P has the form $P = \sum_{i=1}^{\ell} t_i v_{(i)} v_{(i)}^*$ where $v_{(i)}$ $(1 \leqslant i \leqslant \ell)$ are unit vectors such that $A^*Av_{(i)} = \|A\|^2 v_{(i)}$ and t_i $(1 \leqslant i \leqslant \ell)$ are such that $0 \leqslant t_i \leqslant 1$ and $\sum_{i=1}^{\ell} t_i = 1$.)*

Here, $m(A)$ is the best possible upper bound on rank P. This will be illustrated later in Remark 4 in Section 4. When $\mathscr{W} = \mathbb{C}B$, the above theorem says that A is orthogonal to $\mathbb{C}B$ if and only if there exists a density matrix P of the form $P = \sum_{i=1}^{\ell} t_i v_{(i)} v_{(i)}^*$, where $0 \leqslant t_i \leqslant 1$, $\sum_{i=1}^{\ell} t_i = 1$, $\|v_{(i)}\| = 1$, $A^*Av_{(i)} = \|A\|^2 v_{(i)}$ and $\sum_{i=1}^{\ell} t_i \langle B^*Av_{(i)}, v_{(i)}\rangle = 0$. By the Hausdorff–Toeplitz theorem, we get a unit vector v such that $A^*Av = \|A\|^2 v$ and $\langle B^*Av, v\rangle = 0$. The first condition is stronger than that in [6, Theorem 1.1].

In approximation theory, the condition that A is orthogonal to $\mathscr{W}$ is the same as saying that the zero matrix is a best approximation to A from $\mathscr{W}$. Problems of finding a best approximation to a matrix A from a subspace $\mathscr{W}$ of $\mathbb{M}_n(\mathbb{C})$ have been studied earlier (see for example [12,15,16]).

Let $\operatorname{dist}(A, \mathscr{W})$ denote the distance of a matrix A from the subspace $\mathscr{W}$, defined as

$$\operatorname{dist}(A, \mathscr{W}) = \min\{\|A - W\|: W \in \mathscr{W}\}.$$

Audenaert [2] showed that when $\mathscr{W} = \mathbb{C}I$, then

$$\operatorname{dist}(A, \mathbb{C}I)^2 = \max\{\operatorname{tr}(A^*AP) - |\operatorname{tr}(AP)|^2: P \geqslant 0,\ \operatorname{tr} P = 1\}. \tag{2}$$

Further the maximization over P on the right hand side of (2) can be restricted to density matrices of rank 1. The quantity $\operatorname{tr}(A^*AP) - |\operatorname{tr}(AP)|^2$ is called the *variance* of A with respect to the density matrix P. Bhatia and Sharma [7] showed that if $\Phi : \mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ is any positive unital linear map, then

$$\Phi(A^*A) - \Phi(A)^*\Phi(A) \leqslant \operatorname{dist}(A, \mathbb{C}I)^2.$$

By choosing $\Phi(A) = \operatorname{tr}(AP)$ for different density matrices P, they obtained various interesting bounds on $\operatorname{dist}(A, \mathbb{C}I)^2$.

It would be interesting to have a generalization of (2) with $\mathbb{C}I$ replaced by any unital C^*-subalgebra of $\mathbb{M}_n(\mathbb{C})$. (This problem has also been raised by Rieffel in [13].) Let $\mathscr{B}$ be any unital C^*-subalgebra of $\mathbb{M}_n(\mathbb{C})$. Let $\mathcal{C}_{\mathscr{B}} : \mathbb{M}_n(\mathbb{C}) \to \mathscr{B}$ denote the projection of $\mathbb{M}_n(\mathbb{C})$ onto $\mathscr{B}$. We note that $\mathcal{C}_{\mathscr{B}}$ is a bimodule map:

$$\mathcal{C}_{\mathscr{B}}(BX) = B\mathcal{C}_{\mathscr{B}}(X) \quad \text{and} \quad \mathcal{C}_{\mathscr{B}}(XB) = \mathcal{C}_{\mathscr{B}}(X)B \quad \text{for all } B \in \mathscr{B},\ X \in \mathbb{M}_n(\mathbb{C}). \tag{3}$$

In particular, when $\mathscr{B}$ is a subalgebra of block diagonal matrices, the matrix $\mathcal{C}_{\mathscr{B}}(X)$ is called a *pinching* of X and is denoted by $\mathcal{C}(X)$. It is defined as follows. If $X = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots \\ X_{k1} & \cdots & X_{kk} \end{bmatrix}$, then

$$\mathcal{C}(X) = \begin{bmatrix} X_{11} & & & \\ & X_{22} & & \\ & & \ddots & \\ & & & X_{kk} \end{bmatrix}. \tag{4}$$

Properties of pinchings are studied in detail in [3] and [4]. Our next result provides a generalization of (2) for the distance of A to any unital C^*-subalgebra of $\mathbb{M}_n(\mathbb{C})$.

Theorem 2. *Let $\mathscr{B}$ be any unital C^*-subalgebra of $\mathbb{M}_n(\mathbb{C})$. Let $\mathcal{C}_{\mathscr{B}} : \mathbb{M}_n(\mathbb{C}) \to \mathscr{B}$ denote the projection of $\mathbb{M}_n(\mathbb{C})$ onto $\mathscr{B}$. Then*

$$\operatorname{dist}(A, \mathscr{B})^2 = \max\bigl\{\operatorname{tr}\bigl(A^*AP - \mathcal{C}_{\mathscr{B}}(AP)^*\mathcal{C}_{\mathscr{B}}(AP)\mathcal{C}_{\mathscr{B}}(P)^{-1}\bigr):\ P \geqslant 0,\ \operatorname{tr} P = 1\bigr\}, \tag{5}$$

where $\mathcal{C}_{\mathscr{B}}(P)^{-1}$ denotes the Moore–Penrose inverse of $\mathcal{C}_{\mathscr{B}}(P)$. Further the maximization on the right hand side of (5) can be restricted to density matrices P with rank $P \leqslant m(A)$.

We prove Theorem 1 using ideas of subdifferential calculus. A brief summary of these is given in Section 2. The proofs of the theorems are given in Section 3, followed by several remarks in Section 4.

2. Preliminaries

Let X be a complex Hilbert space. Let $f : X \to \mathbb{R}$ be a convex function. Then the *subdifferential* of f at any point $x \in X$, denoted by $\partial f(x)$, is the set of $v^* \in X^*$ such that

$$f(y) - f(x) \geqslant \operatorname{Re} v^*(y - x) \quad \text{for all } y \in X. \tag{6}$$

It follows from (6) that f is minimized at x if and only if $0 \in \partial f(x)$.

We use an idea similar to the one in [8, Theorem 2.6]. Let $f(W) = \|A + W\|$. This is the composition of two functions: namely, $W \mapsto A + W$ from $\mathscr{W}$ into $\mathbb{M}_n(\mathbb{C})$ and $T \to \|T\|$ from $\mathbb{M}_n(\mathbb{C})$ into $\mathbb{R}_+$. Thus we need to find subdifferentials of composition maps. For that we need a chain rule.

Proposition 1. *Let X, Y be any two Hilbert spaces. Let $g : Y \to \mathbb{R}$ be a convex function. Let $S : X \to Y$ be a bounded linear map and let $L : X \to Y$ be the affine map defined by $L(x) = S(x) + y_0$, for some $y_0 \in Y$. Then*

$$\partial(g \circ L)(x) = S^*\partial g\bigl(L(x)\bigr), \tag{7}$$

where S^* is the adjoint of S defined as

$$\bigl\langle S^*(y), x\bigr\rangle = \bigl\langle y, S(x)\bigr\rangle \quad \textit{for all } x \in X \textit{ and } y \in Y.$$

These elementary facts can be found in [11]. In this book the author deals with convex functions $f : \mathbb{R}^n \to \mathbb{R}$. The same proofs can be extended to functions $f : X \to \mathbb{R}$, where X is any Hilbert space. In our setting, g is the map $T \to \|T\|$. The subdifferential of this map has been calculated by Watson [14].

Proposition 2. *Let $A \in \mathbb{M}_n(\mathbb{C})$. Then*

$$\partial\|A\| = \operatorname{conv}\{uv^*:\ \|u\| = \|v\| = 1,\ Av = \|A\|u\}, \tag{8}$$

where conv D *denotes the convex hull of a set* D.

3. Proofs

Proof of Theorem 1. Suppose there exists a density matrix P such that $A^*AP = \|A\|^2 P$ and $AP \in \mathscr{W}^\perp$. Then for any $W \in \mathscr{W}$,

$$\begin{aligned}\|A+W\|^2 &= \big\|(A+W)^*(A+W)\big\| \\ &= \big\|A^*A + W^*A + A^*W + W^*W\big\|.\end{aligned}$$

Now for any $T \in \mathbb{M}_n(\mathbb{C})$,

$$\|T\| = \sup_{\|X\|_1=1} \big|\operatorname{tr}(TX)\big|, \tag{9}$$

where $\|\cdot\|_1$ denotes the trace norm. So,

$$\begin{aligned}\|A+W\|^2 &\geqslant \big|\operatorname{tr}\big(A^*AP + W^*AP + A^*WP + W^*WP\big)\big| \\ &\geqslant \operatorname{Re}\operatorname{tr}\big(A^*AP + W^*AP + A^*WP + W^*WP\big).\end{aligned} \tag{10}$$

Since $AP \in \mathscr{W}^\perp$, we have $\operatorname{Re}\operatorname{tr}(A^*WP) = \operatorname{Re}\operatorname{tr}(W^*AP) = 0$. The matrices W^*W and P are positive semidefinite, therefore $\operatorname{tr}(W^*WP) \geqslant 0$ and by our assumption, $\operatorname{tr}(A^*AP) = \|A\|^2$. Using these in (10) we get that $\|A+W\|^2 \geqslant \|A\|^2$.

Conversely, suppose

$$\|A+W\| \geqslant \|A\| \quad \text{for all } W \in \mathscr{W}. \tag{11}$$

Let $S : \mathscr{W} \to \mathbb{M}_n(\mathbb{C})$ be the inclusion map. Then $S^* : \mathbb{M}_n(\mathbb{C}) \to \mathscr{W}$ is the projection onto the subspace $\mathscr{W}$. Let $L : \mathscr{W} \to \mathbb{M}_n(\mathbb{C})$ be the map defined as

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

Let $g : \mathbb{M}_n(\mathbb{C}) \to \mathbb{R}$ be the map taking an $n \times n$ matrix W to $\|W\|$. Then (11) can be rewritten as

$$(g \circ L)(W) \geqslant (g \circ L)(0),$$

that is, $g \circ L$ is minimized at 0. Therefore $0 \in \partial(g \circ L)(0)$. Using Proposition 1, we get

$$0 \in S^*\partial\|A\|. \tag{12}$$

By Proposition 2,

$$S^*\partial\|A\| = \operatorname{conv}\{S^*\big(uv^*\big):\ \|u\| = \|v\| = 1,\ Av = \|A\|u\}. \tag{13}$$

From (12) and (13) it follows that there exist unit vectors $u_{(i)}, v_{(i)}$ and numbers t_i such that $0 \leqslant t_i \leqslant 1$, $\sum t_i = 1$, $Av_{(i)} = \|A\|u_{(i)}$ and

$$S^*\Big(\sum t_i u_{(i)} v_{(i)}^*\Big) = 0. \tag{14}$$

Let $P = \sum t_i v_{(i)} v_{(i)}^*$. Then $P \geqslant 0$ and $\operatorname{tr} P = 1$. Note that

$$\begin{aligned}AP &= \sum t_i A v_{(i)} v_{(i)}^* \\ &= \|A\| \sum t_i u_{(i)} v_{(i)}^*.\end{aligned}$$

So, from (14) we get $S^*(AP) = 0$, that is, $AP \in \mathscr{W}^\perp$. Since each $v_{(i)}$ is a right singular vector for the singular value $\|A\|$ of A, we have $A^*Av_{(i)} = \|A\|^2 v_{(i)}$. Using this we obtain

$$\begin{aligned} A^*AP &= \sum t_i A^*Av_{(i)}v_{(i)}^* \\ &= \sum t_i \|A\|^2 v_{(i)}v_{(i)}^* \\ &= \|A\|^2 P. \end{aligned} \tag{15}$$

Now let $m(A)=k$. We show that if P satisfies (15), then rank $P \leqslant k$. First note that A^*A and P commute and therefore can be diagonalized simultaneously. So we can assume A^*A and P in (15) to be diagonal matrices. By hypothesis, k of the diagonal entries of A^*A are equal to $\|A\|^2$. Let $A^*A = \begin{bmatrix} \|A\|^2 & & & & & \\ & \ddots & & & & \\ & & \|A\|^2 & & & \\ & & & s_{k+1}^2 & & \\ & & & & \ddots & \\ & & & & & s_n^2 \end{bmatrix}$, where $s_j < \|A\|$ for all $k+1 \leqslant j \leqslant n$. If $P = \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix}$, then from (15) we obtain

$$\left(s_j^2 - \|A\|^2\right)p_j = 0 \quad \text{for all } k+1 \leqslant j \leqslant n.$$

So $p_j = 0$ for all $k+1 \leqslant j \leqslant n$. Hence rank $P \leqslant k$. $\square$

Proof of Theorem 2. We first show that it is sufficient to prove the result when $\mathscr{B}$ is a subalgebra of block diagonal matrices in $\mathbb{M}_n(\mathbb{C})$. If $\mathscr{B}$ is any subalgebra of $\mathbb{M}_n(\mathbb{C})$, then there exist $n_1, n_2, \ldots, n_k$ with $\sum_i n_i = n$ such that $\mathscr{B}$ is $*$-isomorphic to $\bigoplus_i \mathbb{M}_{n_i}(\mathbb{C})$, the $*$-isomorphism $\varphi : \mathscr{B} \to \bigoplus_i \mathbb{M}_{n_i}(\mathbb{C})$ being $\varphi(X) = V^*XV$ for some unitary matrix $V \in \mathbb{M}_n(\mathbb{C})$ (see [9, p. 249], [10, p. 74]). By definition

$$\operatorname{dist}(A, \mathscr{B}) = \min_{W \in \mathscr{B}} \|A - W\|.$$

Let $\tilde{A}$ denote the matrix V^*AV. Since $\|\cdot\|$ is unitarily invariant, we get

$$\operatorname{dist}(A, \mathscr{B}) = \operatorname{dist}\left(\tilde{A}, \bigoplus_i \mathbb{M}_{n_i}(\mathbb{C})\right). \tag{16}$$

Next we show that for any density matrix P,

$$\begin{aligned} &\max\left\{\operatorname{tr}\left(A^*AP - \mathcal{C}_{\mathscr{B}}(AP)^*\mathcal{C}_{\mathscr{B}}(AP)\mathcal{C}_{\mathscr{B}}(P)^{-1}\right):\ P \geqslant 0,\ \operatorname{tr} P = 1\right\} \\ &\quad = \max\left\{\operatorname{tr}\left(\tilde{A}^*\tilde{A}\tilde{P} - \mathcal{C}(\tilde{A}\tilde{P})^*\mathcal{C}(\tilde{A}\tilde{P})\mathcal{C}(\tilde{P})^{-1}\right):\ \tilde{P} \geqslant 0,\ \operatorname{tr} \tilde{P} = 1\right\}, \end{aligned} \tag{17}$$

where $\mathcal{C}$ is the pinching map as defined in (4). Since

$$\operatorname{tr}(XY) = \operatorname{tr}(YX), \tag{18}$$

we have

$$\operatorname{tr}\left(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\right) = \operatorname{tr}\left(V^*A^*APV - V^*\mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}V\right).$$

Now note that for any $X \in \mathbb{M}_n(\mathbb{C})$, we have $V^*\mathcal{C}(X)V = \mathcal{C}(V^*XV)$. Therefore the above expression is the same as

$$\operatorname{tr}\left(\tilde{A}^*\tilde{A}\tilde{P} - \mathcal{C}(\tilde{A}\tilde{P})^*\mathcal{C}(\tilde{A}\tilde{P})\mathcal{C}(\tilde{P})^{-1}\right). \tag{19}$$

This gives (17). So it is enough to prove (5) when $\mathscr{B}$ is a subalgebra of block diagonal matrices. We first show that

$$\max\left\{\operatorname{tr}\left(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\right):\ P \geqslant 0,\ \operatorname{tr} P = 1\right\} \leqslant \operatorname{dist}(A, \mathscr{B})^2. \tag{20}$$

Let P be any density matrix. Then $\operatorname{tr}(A^*AP) \leqslant \|A\|^2$. Therefore

$$\operatorname{tr}\left(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\right) \leqslant \|A\|^2. \tag{21}$$

Let $B \in \mathscr{B}$. Applying the translation $A \mapsto A + B$ in (21), we get

$$\operatorname{tr}\big((A+B)^*(A+B)P - \mathcal{C}\big((A+B)P\big)^*\mathcal{C}\big((A+B)P\big)\mathcal{C}(P)^{-1}\big) \leqslant \|A+B\|^2. \tag{22}$$

We show that the expression on the left hand side is invariant under this translation. By expanding the expression on the left hand side of (22), we get that it is equal to

$$\begin{aligned}&\big(\operatorname{tr}\big(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\big)\big) + \big(\operatorname{tr}\big(B^*AP - \mathcal{C}(BP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\big)\big)\\ &\quad + \big(\operatorname{tr}\big(A^*BP - \mathcal{C}(AP)^*\mathcal{C}(BP)\mathcal{C}(P)^{-1}\big)\big) + \big(\operatorname{tr}\big(B^*BP - \mathcal{C}(BP)^*\mathcal{C}(BP)\mathcal{C}(P)^{-1}\big)\big).\end{aligned} \tag{23}$$

We show that all the terms in (23) are zero except the first one. We shall prove that the second term

$$\operatorname{tr}\big(B^*AP - \mathcal{C}(BP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\big) \tag{24}$$

in (23) is zero. The proof for the other two terms is similar.

By using (3) and (18), the expression in (24) is equal to

$$\operatorname{tr}\big(B^*\mathcal{C}(AP)\big(I - \mathcal{C}(P)^{-1}\mathcal{C}(P)\big)\big). \tag{25}$$

If $\mathcal{C}(P)$ is invertible, then this is clearly zero. So let $\mathcal{C}(P)$ be not invertible. This means that if $\mathcal{C}(P) = \begin{bmatrix} P_1 & & \\ & \ddots & \\ & & P_k \end{bmatrix}$, then there exists i, $1 \leqslant i \leqslant k$, such that P_i is not invertible. Let U denote the block diagonal unitary matrix

$$U = \begin{bmatrix} U_1 & & \\ & \ddots & \\ & & U_k \end{bmatrix}, \tag{26}$$

where $U_j = I$, if P_j is invertible and $U_j^* P_j U_j = \begin{bmatrix} \Lambda_j & \\ & O \end{bmatrix}$, if P_j is not invertible. (Here Λ_j is the diagonal matrix with eigenvalues of P_j as its diagonal entries.) Let X' denote the matrix U^*XU. Then from (3) and (18), we get that the expression in (25) is the same as

$$\operatorname{tr}\big(B'^*\mathcal{C}\big(A'P'\big)\big(I - \mathcal{C}\big(P'\big)^{-1}\mathcal{C}\big(P'\big)\big)\big). \tag{27}$$

Now $\mathcal{C}(P') = \begin{bmatrix} \Lambda_1 & & & & \\ & O & & & \\ & & \Lambda_2 & & \\ & & & O & \\ & & & & \ddots \end{bmatrix}$. Write A' and P' as $2k$-block matrices,

$$A' = \big(A'_{rs}\big)_{r,s=1,\dots,2k} \quad \text{and} \quad P' = \big(P'_{rs}\big)_{r,s=1,\dots,2k},$$

respectively such that whenever P_i is not invertible, we have $P'_{2i-1,2i-1} = \Lambda_i$ and $P'_{2i,2i} = O$.

For $1 \leqslant r \leqslant 2k$, the (r,r)-entry of $A'P'$ is $\sum_{s=1}^{2k} A'_{rs}P'_{sr}$. Suppose $P'_{rr} = O$. Since $P' \geqslant 0$, we have $P'_{rs} = P'_{sr} = O$ for all $s = 1, \dots, 2k$. Hence the (r,r)-entry of $A'P'$ is zero. So let $P'_{rr} \neq O$. Then the (r,r)-entry of $(I - \mathcal{C}(P')^{-1}\mathcal{C}(P'))$ is zero. Therefore the expression in (27) is zero, and hence the expression in (25) is zero. Therefore from (22), we obtain

$$\operatorname{tr}\big(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\big) \leqslant \big\|(A+B)\big\|^2,$$

for all $B \in \mathscr{B}$ and for all density matrices P. Eq. (20) now follows from here.

To show equality in (20), let $\operatorname{dist}(A, \mathscr{B}) = \|A_0\|$, where $A_0 = A - B_0$ for some $B_0 \in \mathscr{B}$. Then A_0 is orthogonal to $\mathscr{B}$. By Theorem 1 there exists a density matrix P such that

$$A_0^*A_0P = \|A_0\|^2 P \tag{28}$$

and

$$\mathcal{C}(A_0P) = 0, \quad \text{that is,} \quad \mathcal{C}(AP) = \mathcal{C}(B_0P). \tag{29}$$

From (28) we get that

$$\|A_0\|^2 = \operatorname{tr}\big((A-B_0)^*(A-B_0)P\big) = \operatorname{tr}\big(A^*AP\big) - \operatorname{tr}\big(B_0^*AP\big) - \operatorname{tr}\big(A^*B_0P\big) + \operatorname{tr}\big(B_0^*B_0P\big).$$

By using (3) and (18), we obtain

$$\|A_0\|^2 = \operatorname{tr}\big(A^*AP\big) - \operatorname{tr}\big(B_0^*\mathcal{C}(AP)\big) - \operatorname{tr}\big(B_0\mathcal{C}(AP)^*\big) + \operatorname{tr}\big(B_0^*\mathcal{C}(B_0P)\big). \tag{30}$$

Substituting (29) in (30) we get

$$\|A_0\|^2 = \operatorname{tr}\big(A^*AP\big) - \operatorname{tr}\big(B_0^*B_0P\big). \tag{31}$$

Now consider $\operatorname{tr}(\mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1})$. From (29) we see that this is the same as $\operatorname{tr}(B_0^*B_0\mathcal{C}(P)\mathcal{C}(P)^{-1}\mathcal{C}(P))$. If $\mathcal{C}(P)$ is invertible, then this is equal to $\operatorname{tr}(B_0^*B_0P)$. If $\mathcal{C}(P)$ is not invertible, then we define U as done in (26). From (3) and (18), we obtain

$$\operatorname{tr}\big(B_0^*B_0\mathcal{C}(P)\mathcal{C}(P)^{-1}\mathcal{C}(P)\big) = \operatorname{tr}\big(B_0'^*B_0'\mathcal{C}\big(P'\big)\mathcal{C}\big(P'\big)^{-1}\mathcal{C}\big(P'\big)\big).$$

By definition of U, this is equal to $\operatorname{tr}(B_0'^*B_0'\mathcal{C}(P'))$, which again by (3) and (18), is the same as $\operatorname{tr}(B_0^*B_0\mathcal{C}(P))$. Therefore from (31) we have

$$\operatorname{dist}(A,\mathscr{B})^2 = \|A_0\|^2 = \operatorname{tr}\big(A^*AP - \mathcal{C}(AP)^*\mathcal{C}(AP)\mathcal{C}(P)^{-1}\big). \qquad \square$$

4. Remarks

1. It is clear from the proof of Theorem 1 that the condition $A^*AP = \|A\|^2P$ can be replaced by the weaker condition $\operatorname{tr}(A^*AP) = \|A\|^2$ in the statement of Theorem 1.
2. As one would expect, the set $\{A\colon\ \|A+W\| \geqslant \|A\| \text{ for all } W \in \mathscr{W}\}$ need not be a subspace. As an example consider the subspace $\mathscr{W} = \mathbb{C}I$ of $\mathbb{M}_3(\mathbb{C})$. Let $A_1 = \begin{bmatrix} 0&1&0\\1&0&1\\0&1&0 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 0&0&1\\0&0&0\\1&0&0 \end{bmatrix}$. It can be checked from Theorem 1 that A_1, A_2 are orthogonal to $\mathscr{W}$. (Take $P = \begin{bmatrix} 0&0&0\\0&1&0\\0&0&0 \end{bmatrix}$ for A_1 and $P = \begin{bmatrix} 1&0&0\\0&0&0\\0&0&0 \end{bmatrix}$ for A_2, respectively.) Then $A_1 + A_2 = \begin{bmatrix} 0&1&1\\1&0&1\\1&1&0 \end{bmatrix}$, and $\|A_1 + A_2\| = 2$. But $\|A_1 + A_2 - \frac{1}{2}I\| = \frac{3}{2} < \|A_1 + A_2\|$. Hence $A_1 + A_2$ is not orthogonal to $\mathscr{W}$.
3. Let $\mathscr{W} = \{X\colon\ \operatorname{tr} X = 0\}$. Then $\mathscr{W}^\perp = \mathbb{C}I$. By Theorem 1, we know that if $A \in \mathscr{W}^\perp$ is such that $A^*A = \|A\|^2 \cdot I$, then A is orthogonal to $\mathscr{W}$. Therefore all the scalar matrices are orthogonal to $\mathscr{W}$. We show that if $A \notin \mathbb{C}I$, then there exists a matrix W with $\operatorname{tr} W = 0$ such that $\|A + W\| < \|A\|$. Let $\mathcal{D}(A)$ and $\mathcal{O}(A)$ denote the diagonal and off-diagonal parts of A, respectively. Then $\mathcal{O}(A) \in \mathscr{W}$, $A - \mathcal{O}(A) = \mathcal{D}(A)$ and $\|\mathcal{D}(A)\| \leqslant \|A\|$. So it is enough to find $W \in \mathscr{W}$ such that $\|\mathcal{D}(A) + W\| < \|\mathcal{D}(A)\|$. Let $\mathcal{D}(A) = \operatorname{diag}(a_1, \ldots, a_1, a_2, \ldots, a_2, \ldots, a_k, \ldots, a_k)$, where $k \geqslant 2$, each a_j $(1 \leqslant j \leqslant k)$ occurs on the diagonal n_j times and $n_1 + \cdots + n_k = n$. We can assume that $\|\mathcal{D}(A)\| = 1$. Take $W = \operatorname{diag}(\frac{a_2-a_1}{kn_1}, \ldots, \frac{a_2-a_1}{kn_1}, \frac{a_3-a_2}{kn_2}, \ldots, \frac{a_3-a_2}{kn_2}, \ldots, \frac{a_k-a_{k-1}}{kn_{k-1}}, \ldots, \frac{a_k-a_{k-1}}{kn_{k-1}}, \frac{a_1-a_k}{kn_k}, \ldots, \frac{a_1-a_k}{kn_k})$. Then $\operatorname{tr} W = 0$ and $\mathcal{D}(A) + W = \operatorname{diag}(\frac{(kn_1-1)a_1+a_2}{kn_1}, \ldots, \frac{(kn_1-1)a_1+a_2}{kn_1}, \frac{(kn_2-1)a_2+a_3}{kn_2}, \ldots, \frac{(kn_2-1)a_2+a_3}{kn_2}, \ldots, \frac{(kn_{k-1}-1)a_{k-1}+a_k}{kn_{k-1}}, \ldots, \frac{(kn_{k-1}-1)a_{k-1}+a_k}{kn_{k-1}}, \frac{(kn_k-1)a_k+a_1}{kn_k}, \ldots, \frac{(kn_k-1)a_k+a_1}{kn_k})$. It is easy to check that $\|\mathcal{D}(A) + W\| < 1$. Hence for this particular $\mathscr{W}$ we have that $\{A\colon\ \|A+W\| \geqslant \|A\| \text{ for all } W \in \mathscr{W}\} = \mathscr{W}^\perp = \mathbb{C}I$.
4. In Theorem 1, $m(A)$ is the best possible upper bound on rank P. Consider $\mathscr{W} = \{X\colon\ \operatorname{tr} X = 0\}$. From Remark 3, we get that if a matrix A is orthogonal to $\mathscr{W}$, then it has to be of the form $A = \lambda I$, for some $\lambda \in \mathbb{C}$. When $A \neq 0$ then $m(A) = n$. Let P be any density matrix satisfying $AP \in \mathscr{W}^\perp$. Then $AP = \mu I$, for some $\mu \in \mathbb{C}$, $\mu \neq 0$. If P also satisfies $A^*AP = \|A\|^2P$, then we get $P = \frac{\mu}{\lambda}I$. Hence rank $P = n = m(A)$.

5. For $n = 2$ and $\mathscr{B}$ any subalgebra of $\mathbb{M}_2(\mathbb{C})$, we can restrict maximum on the right hand side of (5) over rank one density matrices. By the same argument as in the proof of Theorem 2 it is sufficient to prove this for $\mathbb{D}_2(\mathbb{C})$, the subalgebra of diagonal matrices with complex entries. We show

$$\operatorname{dist}\big(A, \mathbb{D}_2(\mathbb{C})\big)^2 = \max_{\|x\|=1}\big(\|Ax\|^2 - \operatorname{tr}\Delta\big(Axx^*\big)^*\Delta\big(Axx^*\big)\Delta\big(xx^*\big)^{-1}\big), \tag{32}$$

where Δ is the projection onto $\mathbb{D}_2(\mathbb{C})$. From Theorem 2 we have

$$\max_{\|x\|=1}\big(\|Ax\|^2 - \operatorname{tr}\Delta\big(Axx^*\big)^*\Delta\big(Axx^*\big)\Delta\big(xx^*\big)^{-1}\big) \leqslant \operatorname{dist}\big(A, \mathbb{D}_2(\mathbb{C})\big)^2.$$

Note that

$$\operatorname{dist}\big(A, \mathbb{D}_2(\mathbb{C})\big) \leqslant \big\|\mathcal{O}(A)\big\|. \tag{33}$$

Let $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ and without loss of generality assume that $|b| \geqslant |c|$. Then $\|\mathcal{O}(A)\| = |b|$. For $x = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$,

$$\|Ax\|^2 - \operatorname{tr}\Delta\big(Axx^*\big)^*\Delta\big(Axx^*\big)\Delta\big(xx^*\big)^{-1} = \big\|\mathcal{O}(A)\big\|^2.$$

Combining this with (33), we obtain

$$\operatorname{dist}\big(A, \mathbb{D}_2(\mathbb{C})\big)^2 \leqslant \max_{\|x\|=1}\big(\|Ax\|^2 - \operatorname{tr}\Delta\big(Axx^*\big)^*\Delta\big(Axx^*\big)\Delta\big(xx^*\big)^{-1}\big).$$

6. For $n = 2$ and $\mathscr{B}$ any subalgebra of $\mathbb{M}_2(\mathbb{C})$, we note that

$$\big\{A\colon\ \|A + W\| \geqslant \|A\| \text{ for all } W \in \mathscr{B}\big\} = \mathscr{B}^{\perp}.$$

Again it is enough to show that

$$\big\{A\colon\ \|A + W\| \geqslant \|A\| \text{ for all } W \in \mathbb{D}_2(\mathbb{C})\big\} = \mathbb{D}_2(\mathbb{C})^{\perp}.$$

If A is an off-diagonal 2×2 matrix, that is, $A = \left[\begin{smallmatrix} 0 & b \\ c & 0 \end{smallmatrix}\right]$ then by Theorem 2.1 in [5], we obtain $\|A + W\| \geqslant \|A\|$ for all $W \in \mathbb{D}_2(\mathbb{C})$. Conversely let $A \in \mathbb{M}_2(\mathbb{C})$ be such that $\|A + W\| \geqslant \|A\|$ for all $W \in \mathbb{D}_2(\mathbb{C})$. Then by taking $W = -\mathcal{D}(A)$, we have $A + W = \mathcal{O}(A)$. Again by using Theorem 2.1 in [5], we obtain that $\|\mathcal{O}(A)\| = \|A\|$. So A is of the form $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, where $\|A\| = \max\{|b|, |c|\}$. Since norm of each row and each column is less than or equal to $\|A\|$, we get that $a = d = 0$. Hence $A \in \mathbb{D}_2(\mathbb{C})^{\perp}$.

Acknowledgements

I would like to thank Professor Rajendra Bhatia for several useful discussions and Professor Ajit Iqbal Singh for helpful comments in this paper.

References

[1] E. Andruchow, G. Larotonda, L. Recht, A. Varela, A characterization of minimal Hermitian matrices, Linear Algebra Appl. 436 (2012) 2366–2374.
[2] K.M.R. Audenaert, Variance bounds, with an application to norm bounds for commutators, Linear Algebra Appl. 432 (2010) 1126–1143.
[3] R. Bhatia, Matrix Analysis, Springer, 1997.
[4] R. Bhatia, Positive Definite Matrices, Princeton University Press, 2007.
[5] R. Bhatia, M.-D. Choi, C. Davis, Comparing a matrix to its off-diagonal part, Oper. Theory Adv. Appl. 40/41 (1989) 151–164.
[6] R. Bhatia, P. Šemrl, Orthogonality of matrices and some distance problems, Linear Algebra Appl. 287 (1999) 77–86.
[7] R. Bhatia, R. Sharma, Some inequalities for positive linear maps, Linear Algebra Appl. 436 (2012) 1562–1571.
[8] T. Bhattacharyya, P. Grover, Characterization of Birkhoff–James orthogonality, J. Math. Anal. Appl. 407 (2013) 350–358.
[9] J.B. Conway, A Course in Functional Analysis, Springer, 1996.
[10] K.R. Davidson, C^*-Algebras by Example, Hindustan Book Agency, 1996.
[11] J.B. Hiriart-Urruty, C. Lemarèchal, Fundamentals of Convex Analysis, Springer, 2000.
[12] J. Liesen, P. Tichý, On best approximations of polynomials in matrices in the matrix 2-norm, SIAM J. Matrix Anal. Appl. 31 (2009) 853–863.

[13] M.A. Rieffel, Standard deviation is a strongly Leibniz seminorm, arXiv:1208.4072v2 [math.OA].
[14] G.A. Watson, Characterization of the subdifferential of some matrix norms, Linear Algebra Appl. 170 (1992) 33–45.
[15] K. Zietak, Properties of linear approximations of matrices in the spectral norm, Linear Algebra Appl. 183 (1993) 41–60.
[16] K. Zietak, On approximation problems with zero-trace matrices, Linear Algebra Appl. 247 (1996) 169–183.