



BIRKHOFF–JAMES ORTHOGONALITY IN THE TRACE NORM, WITH APPLICATIONS TO QUANTUM RESOURCE THEORIES*

NATHANIEL JOHNSTON^{§†}, SHIRIN MOEIN^{†§¶‡}, RAJESH PEREIRA[§], AND SARAH PLOSKER^{§¶}

Abstract. Numerous results are presented that characterize when a complex Hermitian matrix is Birkhoff–James orthogonal, in the trace norm, to a (Hermitian) positive semidefinite matrix or set of positive semidefinite matrices. For example, a simple-to-test criterion that determines which Hermitian matrices are Birkhoff–James orthogonal, in the trace norm, to the set of all positive semidefinite diagonal matrices is developed. Applications in the theory of quantum resources are explored. For example, the quantum states that have modified trace distance of coherence equal to 1 (the maximal possible value) are characterized, and a connection between the modified trace distance of 2-entanglement and the NPPT bound entanglement problem is established.

Key words. Trace norm, Birkhoff–James orthogonality, Quantum coherence, Quantum entanglement, Quantum resource theory.

AMS subject classifications. 15A03, 15A60, 15B57.

1. Introduction. Birkhoff–James orthogonality was introduced in [20] to provide a definition of orthogonality in normed vector spaces that extends the usual one from inner product spaces. Necessary and sufficient conditions for Birkhoff–James orthogonality of matrices in the operator norm were given in [5], and a study of Birkhoff–James orthogonality of matrices in the Schatten p -norms as well as operator norm was performed in [25]. Birkhoff–James orthogonality of a given Hermitian matrix to every member of the subspace of real diagonal matrices, under the operator norm, was studied in [2].

In this work, we characterize when a given complex Hermitian matrix is Birkhoff–James orthogonal in the trace norm to a given (Hermitian) positive semidefinite matrix. We then explore numerous consequences of this main result, such as a simple-to-check characterization of when a given Hermitian matrix is Birkhoff–James orthogonal in the trace norm to every positive semidefinite diagonal matrix.

Our result has several natural applications in quantum information theory, where (mixed) quantum states, also known as density matrices, are positive semidefinite trace-one matrices. When working with a quantum resource theory, it is natural to ask how far a given quantum state is from a particular convex set of states of interest [29], and the natural norm to use to measure distance is the trace norm [36, Chapter 3]. For example, in the resource theory of quantum coherence [3], which is of particular interest in quantum optics, quantum biology, and quantum thermodynamics [11, 32], the aforementioned convex set of states is exactly those that are diagonal. In this case, our results concerning Birkhoff–James orthogonality to all positive semidefinite diagonal matrices provide a characterization of states that are a trace distance of 1 (the largest

*Received by the editors on May 24, 2022. Accepted for publication on November 8, 2022. Handling Editor: Panagiotis Psarrakos. Corresponding Author: Nathaniel Johnston

[†]Department of Mathematics and Computer Science, Mount Allison University, Sackville, NB, Canada E4L 1E4 njohnston@mta.ca.

[‡]Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran 84156-83111 smoein@mta.ca.

[§]Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada N1G 2W1 pereirar@uoguelph.ca.

[¶]Department of Mathematics and Computer Science, Brandon University, Brandon, MB, Canada R7A 6A9 ploskers@brandonu.ca.



such distance possible) from that set, thus extending a result from [23] from pure states to mixed states.

We also extend our results to several other quantum resource theories, including those of entanglement and k -coherence. In particular, when applying our result to the resource theory of 2-entanglement (i.e., Schmidt number 2 [33]), we show that the long-standing NPPT bound entanglement problem [18] has a natural rephrasing in terms of Birkhoff–James orthogonality.

1.1. Notation and terminology. We now introduce the mathematical preliminaries that we need to discuss and prove our results concerning Birkhoff–James orthogonality. We defer a brief introduction to quantum information theory to Section 4, when we need it.

Let M_n be the set of all $n \times n$ matrices with complex entries and let M_n^+ be the subset of them that are (Hermitian) positive semidefinite. We use bold lower case letters such as \mathbf{v} and \mathbf{w} to denote vectors in \mathbb{C}^n , with the entries of $\mathbf{v} \in \mathbb{C}^n$ denoted by v_1, v_2, \dots, v_n . We denote the standard basis vectors by $\{\mathbf{e}_j\}_{j=1}^n$, and we use $A_{i,j}$ to denote the (i, j) -entry of a matrix $A \in M_n$. The eigenvalues of A are denoted by λ_j or by μ_j , while the singular values of A —the non-negative square roots of the eigenvalues of AA^* —are denoted by $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. If A and B are Hermitian matrices, we use the notation $A \geq B$ (or $B \leq A$) to denote $A - B$ being positive semidefinite.

The trace norm of A is $\|A\|_{\text{tr}} \stackrel{\text{def}}{=} \sum_{j=1}^n \sigma_j$, and the operator norm of A is $\|A\| \stackrel{\text{def}}{=} \sigma_1$. If A is Hermitian (or even just normal) then its singular values are the absolute values of its eigenvalues, so $\|A\|_{\text{tr}} = \sum_{j=1}^n |\lambda_j|$. When discussing norms in general, we use the notation $\|\cdot\|$, to avoid confusion with the operator norm. Given a Hermitian matrix A , we denote the eigenprojection matrices, projecting onto the direct sum of the eigenspaces corresponding to the positive, zero, and negative eigenvalues, respectively, by P^+ , P_0 , and P^- , and we note that $P^+ + P_0 + P^- = I$. For simplicity of terminology, we refer to these as the orthogonal projections onto the positive, zero, and negative eigenspaces of A from now on.

As a generalization of orthogonality of vectors in a Hilbert space, we have the following notion of orthogonality in Banach spaces [20]:

DEFINITION 1.1. *Suppose $(\mathcal{X}, \|\cdot\|)$ is a Banach space over \mathbb{R} , and $A, B \in \mathcal{X}$. We say that A is Birkhoff–James orthogonal to B if*

$$\|A\| \leq \|A + \lambda B\| \quad \text{for all } \lambda \in \mathbb{R}.$$

We note that this notion of orthogonality is homogeneous and additive, but not symmetric: if A is Birkhoff–James orthogonal to B , then it is not necessarily the case that B is Birkhoff–James orthogonal to A .

Here, we are interested in the case where a Hermitian matrix A is Birkhoff–James orthogonal under the trace norm $\|\cdot\|_{\text{tr}}$ to a positive semidefinite matrix B . For this reason, from now on we simply use the terminology *Birkhoff–James orthogonal* to signify Birkhoff–James orthogonal *with respect to the trace norm*, and we note that A being Birkhoff–James orthogonal to B means exactly that

$$\|A\|_{\text{tr}} \leq \|A + \lambda B\|_{\text{tr}} \quad \text{for all } \lambda \in \mathbb{R}.$$

Definition 1.1 of Birkhoff–James orthogonality works just fine if the field \mathbb{R} is replaced by \mathbb{C} . For now (and only now), we briefly consider Birkhoff–James orthogonality in this sense and recall the following result:

THEOREM 1.2. [25, Theorem 3.3] Let $A, B \in M_n$. Then, the following are equivalent.

- (a) A is Birkhoff–James orthogonal to B (in the trace norm).
- (b) There exists a matrix $M \in M_n$ with $\|M\| \leq 1$ such that $\text{Tr}(AM^*) = \|A\|_{\text{tr}}$ and $\text{Tr}(BM^*) = 0$.

While this result in this form is stated for the complex vector space M_n , it applies straightforwardly to the real vector space of Hermitian matrices as well. Indeed, if A is Hermitian and $\text{Tr}(AM^*)$ is real, then

$$\text{Tr}(AM^*) = \text{Tr}(AM) = \text{Tr}\left(A\left(\frac{M + M^*}{2}\right)\right),$$

and hence M can be chosen to be Hermitian in this case. We thus only consider Birkhoff–James orthogonality in the *real* vector space of Hermitian matrices from now on.

It is perhaps worth noting that Theorem 1.2 can be proved in a different way than was done in [25]: via the subgradient [37, Theorem 2], which is a useful tool in convex optimization. Indeed, the same is true of many of our upcoming results, but we avoid using the subgradient in this paper, opting instead for more elementary proofs.

1.2. Arrangement of the paper. In Section 2, we present and prove our main result, which characterizes when a Hermitian matrix is Birkhoff–James orthogonal to a positive semidefinite one. In Section 3, we apply our main result to the special case of Birkhoff–James orthogonality to diagonal matrices. In particular, we characterize which Hermitian matrices are Birkhoff–James orthogonal to all positive semidefinite diagonal matrices, and we briefly consider the problem of Birkhoff–James orthogonality with the set of *all* (not necessarily positive semidefinite) diagonal matrices.

In Section 4, we introduce quantum resource theories and show how our results characterize which quantum states are “most resourceful” in the sense of the modified trace distance of that resource. Furthermore, in Section 4.4 we show that Birkhoff–James orthogonality of certain matrices is equivalent to a long-standing conjecture from quantum information theory, which states that a certain quantum state is “bound entangled.” Finally, we close in Section 5 with some conclusions and open questions related to our work.

2. Main result. We now explore the question of when a Hermitian matrix is Birkhoff–James orthogonal to a given positive semidefinite matrix. First, however, we need the following well-known inequality that bounds the trace of a product of Hermitian matrices (see [4, Problem III.6.14], for example).

LEMMA 2.1. Suppose $H, M \in M_n$ are two Hermitian matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$, respectively. Then,

$$\text{Tr}(HM) \leq \sum_{j=1}^n \lambda_j \mu_j,$$

with equality if and only if there exists an orthonormal basis $\{\mathbf{v}_j\}_{j=1}^n \subset \mathbb{C}^n$ such that, for all j , \mathbf{v}_j is an eigenvector of both H and M corresponding to λ_j and μ_j , respectively.

We also need one more lemma before we will be able to state and prove our main result. In this lemma, and the rest of the paper, we use the term *strictly positive eigenspace* to mean the span of the eigenspaces corresponding to strictly positive eigenvalues of a matrix (and similarly for *strictly negative eigenspace*).

LEMMA 2.2. *Suppose $H, M \in M_n$ are two Hermitian matrices, and let P^+ and P^- denote the orthogonal projections onto the strictly positive and strictly negative eigenspaces of H , respectively. Then, $2P^+ - I \leq M \leq I - 2P^-$ if and only if $\|M\| \leq 1$ and $\text{Tr}(HM) = \|H\|_{\text{tr}}$.*

Proof. Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ denote the eigenvalues of H and M , respectively. Then, $\|M\| \leq 1$ if and only if $-1 \leq \mu_j \leq 1$ for all j . If this condition holds then Lemma 2.1 tells us that

$$\text{Tr}(HM) \leq \sum_{j=1}^n \lambda_j \mu_j \leq \sum_{j=1}^n |\lambda_j| = \|H\|_{\text{tr}}.$$

We thus have $\text{Tr}(HM) = \|H\|_{\text{tr}}$ if and only if both of the previous two inequalities are equalities. The first inequality is an equality if and only if there exists an orthonormal basis $\{\mathbf{v}_j\}_{j=1}^n \subset \mathbb{C}^n$ such that, for all j , \mathbf{v}_j is an eigenvector of both H and M corresponding to λ_j and μ_j , respectively. The second inequality is an equality if and only if $\mu_j = 1$ whenever $\lambda_j > 0$ and $\mu_j = -1$ whenever $\lambda_j < 0$. These two conditions together are equivalent to $2P^+ - I \leq M \leq I - 2P^-$, as claimed. \square

We now have enough machinery to state and prove our main result:

THEOREM 2.3. *Suppose $H, B \in M_n$ are Hermitian and B is positive semidefinite. Let P^+ and P^- denote the orthogonal projections onto the strictly positive and strictly negative eigenspaces of H , respectively. Then, H is Birkhoff–James orthogonal to B in the trace norm if and only if both*

$$\text{Tr}(BP^+) \leq \frac{1}{2} \text{Tr}(B) \quad \text{and} \quad \text{Tr}(BP^-) \leq \frac{1}{2} \text{Tr}(B).$$

Proof. Theorem 1.2 and Lemma 2.2 together imply that H is Birkhoff–James orthogonal to B in the trace norm if and only if there exists a Hermitian matrix M with $2P^+ - I \leq M \leq I - 2P^-$ and $\text{Tr}(BM) = 0$. Since B is positive semidefinite, we have $\text{Tr}(B(2P^+ - I)) \leq \text{Tr}(BM) \leq \text{Tr}(B(I - 2P^-))$. These inequalities can be rewritten as $\text{Tr}(B(I - 2P^+)) \geq 0$ and $\text{Tr}(B(I - 2P^-)) \geq 0$, which is equivalent to $\text{Tr}(BP^+) \leq \frac{1}{2} \text{Tr}(B)$ and $\text{Tr}(BP^-) \leq \frac{1}{2} \text{Tr}(B)$ by the linearity of the trace.

Conversely, suppose we have both $\text{Tr}(BP^+) \leq \frac{1}{2} \text{Tr}(B)$ and $\text{Tr}(BP^-) \leq \frac{1}{2} \text{Tr}(B)$. If either $\text{Tr}(B(I - 2P^+)) = 0$ or $\text{Tr}(B(I - 2P^-)) = 0$, we choose $M = 2P^+ - I$ or $M = I - 2P^-$, respectively. On the other hand, if we have $\alpha := \text{Tr}(B(I - 2P^+)) > 0$ and $\beta := \text{Tr}(B(I - 2P^-)) > 0$, we choose

$$M = \frac{1}{\alpha + \beta} (\beta(2P^+ - I) + \alpha(I - 2P^-)).$$

In either case, we then have $2P^+ - I \leq M \leq I - 2P^-$ and $\text{Tr}(BM) = 0$, which means that H is Birkhoff–James orthogonal to B in the trace norm. \square

3. Birkhoff–James orthogonality to diagonal matrices. We now consider the problem of determining which Hermitian matrices are Birkhoff–James orthogonal to every diagonal matrix. This problem was considered for the operator norm in [2] and was solved in small dimensions there, whereas we consider the trace norm version of it. We start with a result that determines when a Hermitian matrix is Birkhoff–James orthogonal in the trace norm to all *positive semidefinite* diagonal matrices:

THEOREM 3.1. *Suppose $H \in M_n$ is Hermitian, and let P^+ and P^- be the orthogonal projections onto its strictly positive and negative eigenspaces, respectively. Then, the following are equivalent:*

- (a) H is Birkhoff–James orthogonal in the trace norm to every positive semidefinite diagonal matrix.
- (b) H is Birkhoff–James orthogonal in the trace norm to $\mathbf{e}_j \mathbf{e}_j^*$ for all $1 \leq j \leq n$.
- (c) $P_{j,j}^+ \leq 1/2$ and $P_{j,j}^- \leq 1/2$ for all $1 \leq j \leq n$.

Proof. The equivalence of (b) and (c) follows from choosing $B = \mathbf{e}_j \mathbf{e}_j^*$ in Theorem 2.3. The equivalence of (a) and (c) similarly follows from choosing B to be an arbitrary diagonal positive semidefinite matrix scaled so that $\text{Tr}(B) = 1$ in Theorem 2.3. \square

We now start looking at the more difficult problem of characterizing which Hermitian matrices are Birkhoff–James orthogonal to *all* (not necessarily positive semidefinite) diagonal matrices. Theorem 3.1 provides a necessary condition, but it is not a sufficient one when $n \geq 3$, as demonstrated by the following example:

EXAMPLE 3.2. *The Hermitian matrix*

$$H = \begin{bmatrix} -1 & 5 & 2 \\ 5 & -1 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

has projections onto its negative and positive eigenspaces

$$P^- = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P^+ = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

respectively. It follows that $P_{j,j}^- \leq 1/2$ and $P_{j,j}^+ \leq 1/2$ for all $1 \leq j \leq 3$, so Theorem 3.1 tells us that H is Birkhoff–James orthogonal to every positive semidefinite diagonal matrix.

However, H is not Birkhoff–James orthogonal to every diagonal matrix. To see this, we can compute $\|H\|_{\text{tr}} = 12$, but if $D = \text{diag}(6, 6, -6/5)$ then $\|H + D\|_{\text{tr}} = 54/5 < 12$ (and in fact, semidefinite programming can be used to show that this D is optimal, so $H + D$ is Birkhoff–James orthogonal to every diagonal matrix).

On the other hand, if $n = 2$ then the condition of Theorem 3.1 is both necessary and sufficient for Birkhoff–James orthogonality to every diagonal matrix, as we will see shortly. Our starting point towards proving this fact is the following simple corollary, which solves this problem for invertible Hermitian matrices:

COROLLARY 3.3. *Suppose $H \in M_n$ is Hermitian and invertible, and let P^+ and P^- be the orthogonal projections onto its positive and negative eigenspaces, respectively. The following are equivalent:*

- (a) H is Birkhoff–James orthogonal in the trace norm to $\mathbf{e}_j \mathbf{e}_j^*$ for all $1 \leq j \leq n$.
- (b) H is Birkhoff–James orthogonal in the trace norm to every diagonal matrix.
- (c) $P^+ = (I + U)/2$ and $P^- = (I - U)/2$ for some Hermitian unitary U with zeros on its diagonal.
- (d) $P_{j,j}^+ = P_{j,j}^- = 1/2$ for all $1 \leq j \leq n$.

Proof. To see that (a) and (d) are equivalent, recall from Theorem 3.1 that H is Birkhoff–James orthogonal to each $\mathbf{e}_j \mathbf{e}_j^*$ if and only if $P_{j,j}^+ \leq 1/2$ and $P_{j,j}^- \leq 1/2$ for all $1 \leq j \leq n$. Since H is invertible, $P^+ + P^- = I$, so $P_{j,j}^+ + P_{j,j}^- = 1$ for all j . It follows that $P_{j,j}^+ \leq 1/2$ and $P_{j,j}^- \leq 1/2$ is equivalent to $P_{j,j}^+ = P_{j,j}^- = 1/2$, so (a) is equivalent to (d).

The fact that (c) implies (d) is trivial. To see that (d) implies (c), notice that (d) implies we can write $P^+ = I/2 + X$ for some matrix X with diagonal entries equal to 0. Since P^+ is Hermitian, so is X , and

since P^+ is a projection, we have $(I/2 + X)^2 = I/2 + X$, so $X^2 = I/4$, so $U := X/2$ is unitary. Then, $P^+ = (I + U)/2$, and since $P^+ + P^- = I$, we must have $P^- = (I - U)/2$.

The fact that (b) implies (a) is trivial. To complete the proof, we show that (d) implies (b) as follows. If (d) holds and B is diagonal, then it must be the case that $\text{Tr}(BP^+) = (1/2)\text{Tr}(B)$ and $\text{Tr}(BP^-) = (1/2)\text{Tr}(B)$. It follows from Theorem 2.3 that H is Birkhoff–James orthogonal to B in the trace norm, completing the proof. \square

THEOREM 3.4. *A Hermitian matrix $H \in M_2$ is Birkhoff–James orthogonal to every diagonal matrix if and only if it is Birkhoff–James orthogonal to $\mathbf{e}_1\mathbf{e}_1^*$ and $\mathbf{e}_2\mathbf{e}_2^*$.*

Proof. The “only if” direction is trivial, and Corollary 3.3 gives the result when H is invertible, so we only prove the “if” direction in the case when $\text{rank}(H) = 1$.

In this case, H has the form $H = \pm\mathbf{v}\mathbf{v}^*$ for some $\mathbf{v} \in \mathbb{C}^2$ (without loss of generality, we will assume that $H = \mathbf{v}\mathbf{v}^*$). Then, $P^+ = \mathbf{v}\mathbf{v}^*/\|\mathbf{v}\|^2$, and we see that if $P_{j,j}^+ \leq 1/2$ for $j = 1, 2$, then in fact we must have $P_{j,j}^+ = 1/2$ for $j = 1, 2$. It follows that

$$H = \frac{\|\mathbf{v}\|^2}{2} \begin{bmatrix} 1 & e^{i\theta} \\ e^{-i\theta} & 1 \end{bmatrix},$$

for some $\theta \in \mathbb{R}$. We can then choose $M = e^{i\theta}\mathbf{e}_1\mathbf{e}_2^* + e^{-i\theta}\mathbf{e}_2\mathbf{e}_1^*$ in Theorem 1.2 to see that H is Birkhoff–James orthogonal to every diagonal matrix. \square

We now present several corollaries of Theorem 3.1 that place restrictions on which Hermitian matrices are Birkhoff–James orthogonal to the diagonal matrices $\mathbf{e}_j\mathbf{e}_j^*$ ($1 \leq j \leq n$). To start, we note that these conditions bound the inertia of such matrices quite strongly.

COROLLARY 3.5. *Suppose $H \in M_n$ is Hermitian and Birkhoff–James orthogonal to $\mathbf{e}_j\mathbf{e}_j^*$ for each $1 \leq j \leq n$, and let μ_+ and μ_- denote how many eigenvalues it has that are positive and negative, respectively. Then,*

$$\mu_- \leq n/2 \quad \text{and} \quad \mu_+ \leq n/2.$$

Proof. Theorem 3.1 tells us that the positive and negative eigenprojections P^+ and P^- satisfy $P_{j,j}^+ \leq 1/2$ and $P_{j,j}^- \leq 1/2$ for all $1 \leq j \leq n$. Taking the trace, we obtain $\sum_{j=1}^n P_{j,j}^+ = \mu_+ \leq n/2$, and similarly for μ_- . \square

COROLLARY 3.6. *Suppose $H \in M_n$ is Hermitian and n is odd. If H is Birkhoff–James orthogonal to $\mathbf{e}_j\mathbf{e}_j^*$ for each $1 \leq j \leq n$, then it must have 0 as an eigenvalue.*

Proof. Let μ_+ , μ_- , and μ_0 denote how many eigenvalues H has that are positive, negative, and 0, respectively. Suppose (for the sake of establishing a contradiction) that $\mu_0 = 0$. Then, $\mu_+ + \mu_0 + \mu_- = n$ implies that $\mu_+ + \mu_- = n$. However, Corollary 3.5 implies, since n is odd, that $\mu_+ \leq (n - 1)/2$ and similarly for μ_- , which yields $\mu_+ + \mu_- \leq n - 1 < n$, a contradiction. \square

The above Corollary 3.6 really is specific to matrices of odd size: in the even-dimensional case, matrices that are Birkhoff–James orthogonal to every diagonal matrix do not necessarily have 0 as an eigenvalue. For example, Theorem 3.4 tells us that the invertible matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

is Birkhoff–James orthogonal to every diagonal matrix.

We now present a *necessary and sufficient* condition for a Hermitian matrix to be Birkhoff–James orthogonal to every diagonal matrix (in contrast with Theorem 3.1, which provides a necessary condition).

COROLLARY 3.7. *Suppose $H \in M_n$ is Hermitian and let P^+ , P^- , P_0 denote the orthogonal projections onto the strictly positive eigenspaces, strictly negative eigenspaces, and null space of H , respectively. Then, H is Birkhoff–James orthogonal to every diagonal matrix if and only if there exists $X \in M_n$ with X Hermitian and $-I \leq X \leq I$ such that the matrix $M = P^+ - P^- + P_0XP_0$ has all of its main diagonal entries equal to zero.*

Proof. We note that if $-I \leq X \leq I$, then $2P^+ - I = P^+ - P^- - P_0 \leq P^+ - P^- + P_0XP_0 \leq P^+ - P^- + P_0 = I - 2P^-$. Hence by Lemma 2.2, the existence of an $X \in M_n$ with X Hermitian and $-I \leq X \leq I$ and $M = P^+ - P^- + P_0XP_0$ has all of its main diagonal entries equal to zero is equivalent to the existence of a Hermitian M having all of its main diagonal entries equal to zero and satisfying $\|M\| = 1$ and $\text{Tr}(HM) = \|H\|_{\text{tr}}$ which by Theorem 1.2 is equivalent to H being Birkhoff–James orthogonal to every diagonal matrix. \square

We note that any choice of Hermitian U in any polar decomposition of H corresponds to $U = P^+ - P^- + P_0XP_0$ where X is chosen to be Hermitian and unitary; further, if H is invertible then $P_0 = 0$ and $P^+ - P^-$ is the unitary in the unique polar decomposition of H . The following theorem is a consequence of these observations together with Corollary 3.7.

THEOREM 3.8. *Suppose $H \in M_n$ is a Hermitian matrix. If there is a polar decomposition $H = UP$ in which the unitary U is Hermitian and has all of its diagonal entries equal to 0, then H is Birkhoff–James orthogonal to every diagonal matrix. If H is invertible then the converse also holds.*

After some work, it can be seen that Theorem 3.8 is essentially equivalent to a special case of [12, Theorem 1.1]. It is perhaps worth recalling that if H is invertible then its polar decomposition is unique, so the necessary and sufficient condition of Theorem 3.8 is easy to check in this case.

3.1. When the given matrix is positive semidefinite. In addition to the 2-dimensional case that we saw in Theorem 3.4, there are some other extra conditions that we can add to H that ensure that the (easy-to-check) properties of Theorem 3.1 are equivalent to H being Birkhoff–James orthogonal to every diagonal matrix. In particular, positive semidefiniteness of H lets us increase the dimension a bit:

THEOREM 3.9. *Suppose $H \in M_n^+$ is positive semidefinite and at least one of the following conditions hold: (a) $n \leq 4$, (b) $\text{rank}(H) \in \{1, n/2, n\}$, or (c) the orthogonal projection onto $\text{range}(H)$ has constant diagonal. Then, H is Birkhoff–James orthogonal to every diagonal matrix if and only if it is Birkhoff–James orthogonal to $\mathbf{e}_j\mathbf{e}_j^*$ for all $1 \leq j \leq n$.*

Proof. Since the “only if” implication of this theorem is trivial, we only prove the “if” direction.

We start with the rank-1 case of part (b). Notice that in this case, we can write $H = \mathbf{v}\mathbf{v}^*$. By the same argument used in the proof of [23, Theorem 2], there exists a matrix $M = (\mathbf{v}\mathbf{v}^* - \mathbf{w}\mathbf{w}^*)/\|\mathbf{v}\|^2$ with all of its diagonal entries equal to 0, $|w_j| = |v_j|$ for all $1 \leq j \leq n$, and $\mathbf{v}^*\mathbf{w} = 0$ (this matrix M was called Y in the proof of [23, Theorem 2]). For an arbitrary diagonal matrix D , we thus have $\text{Tr}(DM^*) = 0$, $\|M\| \leq 1$, and $\text{Tr}(HM^*) = \mathbf{v}^*M\mathbf{v} = 1 = \|H\|_{\text{tr}}$, so Theorem 1.2 tells us that H is Birkhoff–James orthogonal to D .

To see that part (c) of the theorem holds, let c be the constant diagonal value (i.e., if P is the orthogonal projection onto $\text{range}(H)$ then $c = P_{j,j}$ for all j). If we let $M = P/c - I$, then by Lemma 2.2 we have $\|M\| \leq 1$ and $\text{Tr}(HM) = \|H\|_{\text{tr}}$. Since all diagonal entries of M are zero, $\text{Tr}(MD) = 0$ for all

diagonal matrices. It then follows from Theorem 1.2 that H is Birkhoff–James orthogonal to every diagonal matrix.

The rank $n/2$ case of part (b) then holds because, in this case we must have $\text{Tr}(P) = \text{rank}(H) = n/2$. When H is Birkhoff–James orthogonal to $\mathbf{e}_j \mathbf{e}_j^*$ for all $1 \leq j \leq n$, we must have all diagonal entries of P exactly equal to $1/2$ to also satisfy both the trace condition and condition (c) of Theorem 3.1. Part (c) of this theorem (which we already proved) then implies the result.

The rank n case of part (b) follows immediately from Corollary 3.3.

Finally, to see that part (a) of the theorem holds, notice that if $n = 2$ then Theorem 3.4 gives the result. If $n = 3$, then Corollary 3.5 tells us that $\text{rank}(H) = \mu_+ \leq n/2 = 3/2$, so $\text{rank}(H) = 1$, and we already proved the rank-1 case. If $n = 4$ then Corollary 3.5 tells us that $\text{rank}(H) = \mu_+ \leq n/2 = 2$. Regardless of whether $\text{rank}(H) = 1$ or $\text{rank}(H) = 2 = n/2$, case(b) of this theorem shows that we are already done. \square

We note that Theorem 3.9 is in a sense tight: if all three of the conditions (a)–(c) fail, then the conclusion of the theorem needs not hold, as demonstrated by the next example that has $n = 5$ and $1 < \text{rank}(H) = 2 < n/2$.

EXAMPLE 3.10. Let $\alpha = \sqrt{(\sqrt{10} - 1)/3}$ and let $P \in M_5^+$ be the orthogonal projection onto the subspace $\text{span}((1, 1, 1, 0, 0), (0, 0, 1, \alpha, 1/\alpha))$.

Then the following claims are all straightforward to verify:

- (1) P is positive semidefinite with rank 2;
- (2) $P_{j,j} \leq 1/2$ for all j (with $P_{5,5} = 1/2$ exactly), so Theorem 3.1 tells us that it is Birkhoff–James orthogonal to every positive semidefinite diagonal matrix. However;
- (3) If $D = \text{diag}(0, 0, 3, -1, 3)$ then $\|P - D/40\|_{\text{tr}} \approx 1.99441235 < 2 = \|P\|_{\text{tr}}$, so P is not Birkhoff–James orthogonal to D .

4. Applications to quantum resource theories. In quantum information theory, a *mixed quantum state* or *density matrix* is a positive semidefinite trace-one matrix, typically denoted by $\rho \in M_n$ or $\sigma \in M_n$. A *pure quantum state* is a unit vector in \mathbb{C}^n , which we denote using “bra-ket” notation: $|v\rangle \in \mathbb{C}^n$ is a unit column vector, while $\langle v| \stackrel{\text{def}}{=} |v\rangle^*$ is the corresponding (dual) row vector. From now on, whenever we use lowercase Greek letters like ρ , we are assuming that it is a quantum state normalized to have $\text{Tr}(\rho) = 1$, and similarly if we use $|v\rangle$, then we are assuming it is a *unit* vector. For un-normalized matrices and vectors, we denote them like $A \in M_n^+$ and $\mathbf{v} \in \mathbb{C}^n$, just like in the earlier sections of this paper.

A quantum resource theory [8] consists of two things: (1) a closed convex cone of positive semidefinite matrices $C \subseteq M_n^+$, and (2) a set of linear maps that send C back to itself (for our purposes, only (1) is relevant). The quantum states $\rho \in C$ (i.e., the members of C with trace 1) are called “free states” and are thought of as the states that are “useless” for some given task in quantum information theory or quantum computation.

The most well-known resource theory is that of *entanglement*, where $C = C_{\text{ent}}$ is the set of *separable* quantum states [13, 19, 38]:

$$C_{\text{ent}} \stackrel{\text{def}}{=} \left\{ \sum_j X_j \otimes Y_j : X_j \in M_m^+, Y_j \in M_n^+ \text{ for all } j \right\} \subseteq (M_m \otimes M_n)^+.$$

However, numerous other resource theories are studied as well [29]. For example, another widely used resource theory is that of *coherence* [1, 3, 11, 32], in which the set of free states $C = C_{\text{coh}}$ consists of exactly those that are diagonal (in the standard basis):

$$C_{\text{coh}} \stackrel{\text{def}}{=} \left\{ \sum_j x_j \mathbf{e}_j \mathbf{e}_j^* : 0 \leq x_j \in \mathbb{R} \text{ for all } j \right\} \subseteq M_n^+.$$

In this context, the members of C_{coh} are called *incoherent*.

Given a closed convex cone C , we can define the following *modified trace distance of C* , which simply measures the distance from a given quantum state to C :

$$(4.1) \quad D_C(\rho) \stackrel{\text{def}}{=} \min \{ \|\rho - X\|_{\text{tr}} : X \in C \}.$$

This quantity is a “proper” (i.e., physically relevant) measure of the resource specified by C in the sense of [29]. When $C = C_{\text{coh}}$, it is the *modified trace distance of coherence* [39], and when $C = C_{\text{ent}}$, it is the *modified trace distance of entanglement* [30]. For notational convenience, we omit the central level of subscripting when denoting this distance. For example, we denote the modified trace distance of coherence simply by D_{coh} , rather than $D_{C_{\text{coh}}}$.

As a bit of a historical note, we comment that an analogous (non-modified) *trace distance of C* was originally defined in the $C = C_{\text{coh}}$ case in [28] via the minimization

$$\min \{ \|\rho - X\|_{\text{tr}} : X \in C, \text{Tr}(X) = 1 \}.$$

However, this quantity was subsequently shown to *not* be a “proper” measure of coherence (i.e., it lacked certain physically desirable properties that a quantification of a quantum resource should satisfy). The “modified” trace distance of C was then introduced to fix this problem [39].

By applying our theorems concerning Birkhoff–James orthogonality (in particular, Theorem 2.3) in this setting, we get the following characterization of which mixed quantum states have $D_C(\rho) = 1$. These states are of interest because they are exactly those that are “most resourceful” or “most useful”: it is clear from the definition of Equation (4.1) that it is always the case that $D_C(\rho) \leq 1$.

THEOREM 4.1. *Suppose $C \subseteq M_n^+$ is a closed convex cone, $\rho \in M_n^+$ is a mixed quantum state, and P is the orthogonal projection onto $\text{range}(\rho)$. The following are equivalent:*

- (a) $D_C(\rho) = 1$.
- (b) ρ is Birkhoff–James orthogonal in the trace norm to every member of C .
- (c) $\text{Tr}(P\sigma) \leq 1/2$ for every mixed quantum state $\sigma \in C$.

Proof. To see that (a) is equivalent to (b), note (by definition) that ρ is Birkhoff–James orthogonal in the trace norm to every member of C if and only if

$$1 = \|\rho\|_{\text{tr}} \leq \|\rho + \lambda X\|_{\text{tr}}$$

for all $\lambda \in \mathbb{R}$ and $X \in C$. However, we can safely ignore positive values of λ since adding the positive semidefinite matrix λX to the positive semidefinite matrix ρ can only increase its eigenvalues (and thus its

trace norm). We can also safely ignore all negative values of λ other than $\lambda = -1$, since C is a cone so we can absorb $|\lambda|$ into X .

It follows that (b) is equivalent to $\|\rho - X\|_{\text{tr}} \geq 1$ for all $X \in C$, which (by Equation (4.1)) is equivalent to $D_C(\rho) \geq 1$. Since $D_C(\rho) \leq 1$ always holds, this establishes the claim that (a) and (b) are equivalent.

To see that (b) and (c) are equivalent, we recall from Theorem 2.3 that ρ is Birkhoff–James orthogonal to every $X \in C$ if and only if $\text{Tr}(XP) \leq \text{Tr}(X)/2$. Since this inequality is scale-invariant, without loss of generality we may set $\sigma = X/\text{Tr}(X)$ to see that (b) holds if and only if $\text{Tr}(P\sigma) \leq 1/2$ for every mixed quantum state $\sigma \in C$. \square

Since C is a convex cone, condition (c) of Theorem 4.1 can be weakened slightly to say that $\text{Tr}(P\sigma) \leq 1/2$ for every mixed quantum state σ that lies on an extreme ray of C . For all of the cones C that we will explore (i.e., the cones defining k -coherence or k -entanglement), these extreme quantum states are exactly the pure states (i.e., the rank-1 matrices). So for our purposes, the conditions of Theorem 4.1 are furthermore equivalent to $\langle v|P|v \rangle \leq 1/2$ for every pure quantum state $|v\rangle\langle v| \in C$.

4.1. The modified trace distance of coherence. The modified trace distance of coherence was shown to be of limited use for pure states; in that it reaches its maximum value (i.e., $D_{\text{coh}}(|v\rangle\langle v|) = 1$) on nearly all pure states (and therefore does not differentiate between these states) in [23]. We now provide the natural generalization of this result to the case of arbitrary-rank (i.e., not necessarily pure) quantum states:

THEOREM 4.2. *Suppose $\rho \in M_n^+$ is a mixed quantum state and P is the orthogonal projection onto $\text{range}(\rho)$. Then $D_{\text{coh}}(\rho) = 1$ if and only if $P_{j,j} \leq 1/2$ for all $1 \leq j \leq n$.*

Proof. This follows immediately from the equivalence of conditions (a) and (c) in Theorem 4.1: the only pure quantum states $|v\rangle\langle v| \in C$ come from choosing $|v\rangle = \mathbf{e}_j$ for some $1 \leq j \leq n$, so $D_{\text{coh}}(\rho) = 1$ if and only if $P_{j,j} = \mathbf{e}_j^* P \mathbf{e}_j \leq 1/2$ for all $1 \leq j \leq n$. \square

Indeed, the above theorem is a direct generalization of [23, Theorem 2] from the rank-1 case to the general case (in the rank-1 case $\rho = |v\rangle\langle v|$, notice that $P_{j,j} = |v_j|^2$).

In addition to the condition of Theorem 4.2 being easy to check, it has the interesting consequence of bounding the possible rank of any state with $D_{\text{coh}}(\rho) = 1$. Not only are low-rank states usually maximally coherent (in the sense of D_{coh}), but high-rank states are *never* maximally coherent:

COROLLARY 4.3. *Suppose $\rho \in M_n^+$ is a mixed quantum state with $D_{\text{coh}}(\rho) = 1$. Then, $\text{rank}(\rho) \leq n/2$.*

Proof. If $D_{\text{coh}}(\rho) = 1$, then Theorem 4.2 tells us that the projection P onto $\text{range}(\rho)$ satisfies $P_{j,j} \leq 1/2$ for all j , so $\text{Tr}(P) \leq n/2$. Since $\text{Tr}(P) = \text{rank}(P) = \text{rank}(\rho)$, the result follows. \square

In particular, the above corollary tells us that all quantum states with maximal modified trace distance of coherence in the $n = 2$ and $n = 3$ cases are necessarily pure (i.e., rank-1). This is no longer true when $n \geq 4$, however.

4.2. k -Coherence. As a generalization of coherence, for an integer $1 \leq k \leq n$, the resource theory of k -coherence [31] makes use of the following set $C = C_{k\text{-coh}}$ as its free states:

$$C_{k\text{-coh}} \stackrel{\text{def}}{=} \left\{ \sum_j \mathbf{v}_j \mathbf{v}_j^* : \text{each } \mathbf{v}_j \in \mathbb{C}^n \text{ has at most } k \text{ non-zero entries} \right\} \subseteq M_n^+.$$

In the $k = 1$ case, this reduces down to exactly the set of incoherent states (i.e., $C_{1\text{-coh}} = C_{\text{coh}}$), whereas the $k = n$ case simply reduces to the set of *all* mixed states (i.e., $C_{n\text{-coh}} = M_n^+$). For intermediate values of k , these sets satisfy the obvious chain of inclusions

$$C_{\text{coh}} = C_{1\text{-coh}} \subsetneq C_{2\text{-coh}} \subsetneq C_{3\text{-coh}} \subsetneq \cdots \subsetneq C_{(n-1)\text{-coh}} \subsetneq C_{n\text{-coh}} = M_n^+.$$

By using the fact that the pure states in $C_{k\text{-coh}}$ are the unit vectors with at most k non-zero entries, we immediately get a characterization of the mixed states that maximize the modified trace distance of k -coherence:

THEOREM 4.4. *Suppose $\rho \in M_n^+$ is a mixed quantum state and P is the orthogonal projection onto $\text{range}(\rho)$. Then, $D_{k\text{-coh}}(\rho) = 1$ if and only if the operator norm of every $k \times k$ principal submatrix of P is at most $1/2$.*

Proof. This follows immediately from the equivalence of conditions (a) and (c) in Theorem 4.1: the pure quantum states $|v\rangle\langle v| \in C$ come from choosing the unit vectors $|v\rangle \in \mathbb{C}^n$ with at most k non-zero entries, so $D_{\text{coh}}(\rho) = 1$ if and only if $\langle v|P|v\rangle \leq 1/2$ for all such vectors. By focusing on all the $|v\rangle$ that have their non-zero entries in particular locations, we see that this is equivalent to the principal submatrix of P corresponding to those locations having operator norm at most $1/2$. \square

Indeed, the above theorem reduces to Theorem 4.2 in the special case when $k = 1$, since the 1×1 principal submatrices of P are exactly its diagonal entries.

The following concept will be useful in what follows. Let $\|A\|_{(k)}$ denote the maximal operator norm of a $k \times k$ principal submatrix of A (for example, if $k = 1$ then $\|A\|_{(k)}$ is the maximal diagonal entry of A). For $k > 1$, this is a norm (but that fact won't be important for the coming argument). We will present a useful characterization of $\|P\|_{(k)}$ when P is a projection. First, we need the following lemma.

LEMMA 4.5. *If $P, Q \in M_n^+$ are orthogonal projections then $\|PQP\| = \|QPQ\|$.*

Proof. Since PQP and QPQ are both positive semidefinite matrices, their operator norms are equal to their largest eigenvalues. We note that it is well-known that if A and B are two square matrices, then AB and BA have the same eigenvalues. Our result now follows from the fact that $PQP = PQ^2P = (PQ)(QP)$ and $QPQ = QP^2Q = (QP)(PQ)$. \square

While this fact will not be used in the paper, we note that $\|PQP\|$ has geometric significance being the square of the cosine of the first canonical angle between the range subspaces of P and Q (see [26, 24] for applications of canonical angles to quantum information).

LEMMA 4.6. *Let $P \in M_n$ be a projection and let $1 \leq k \leq n$. Then, we have the equality*

$$\|P\|_{(k)} = \max \left\{ \sum_{j=1}^k (|v_j|^\downarrow)^2 : |v\rangle \in \text{range}(P) \right\},$$

where $|v_j|^\downarrow$ is the j -th largest modulus of an entry of $|v\rangle$.

Proof. Let i_1, i_2, \dots, i_k be the rows of the $k \times k$ principal submatrix of P with the largest operator norm. Let Q be the projection onto $\text{span}\{\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_k}\}$. Then, $\|P\|_{(k)} = \|QPQ\| = \|PQP\|$. It is well-known that the operator norm of a Hermitian matrix is equal to its numerical radius (i.e., $\|H\| = \max_{\|x\|=1} |\langle x|H|x\rangle|$), and we apply this fact to PQP . Let $|v\rangle$ be the unit vector which maximizes $\langle x|PQP|x\rangle$. If $|v\rangle \notin \text{range}(P)$,

we can replace $|x\rangle = |v\rangle$ with $|x\rangle = P|v\rangle/\|P|v\rangle\|$ giving us an even larger value for $\langle x|PQP|x\rangle$; hence, $|v\rangle \in \text{range}(P)$. Therefore $\|P\|_{(k)} = \langle v|PQP|v\rangle = \langle v|Q|v\rangle = \sum_{j=1}^k |v_{i_j}|^2 = \sum_{j=1}^k (|v_j|^\downarrow)^2$. The final equality follows from the fact that if $\{v_{i_j}\}_{j=1}^k$ were not the k largest entries of $|v\rangle$ in modulus, this would contradict our assumption that i_1, i_2, \dots, i_k are the rows of the $k \times k$ principal submatrix of P with the largest operator norm. \square

We use the above lemmas to get a rank bound that generalizes Corollary 4.3 to this setting of k -coherence:

COROLLARY 4.7. *Suppose $\rho \in M_n^+$ is a mixed quantum state with $D_{k\text{-coh}}(\rho) = 1$. Then*

$$\text{rank}(\rho) \leq \frac{n(n+1-2k)}{2(n-k)}.$$

Proof. Theorem 4.4 tells us that if $D_{k\text{-coh}}(\rho) = 1$ then $\|P\|_{(k)} \leq 1/2$. On the other hand, we also have some lower bounds on $\|P\|_{(k)}$: it must be the case that $\|P\|_{(1)} \geq \text{Tr}(P)/n$, since otherwise the sum of the diagonal entries of P would be strictly less than $n(\text{Tr}(P)/n) = \text{Tr}(P)$, which is a contradiction.

For larger values of k , we claim that $\|P\|_{(k)} \geq \|P\|_{(1)} + (1 - \|P\|_{(1)})(k-1)/(n-1)$. If we let $|w\rangle$ be a particular $|v\rangle$ attaining the supremum for $\|P\|_{(1)}$ given in Lemma 4.6, then we have

$$\sum_{j=2}^n (|w_j|^\downarrow)^2 = \sum_{j=1}^n (|w_j|^\downarrow)^2 - \|P\|_{(1)} = 1 - \|P\|_{(1)}.$$

Since the $|w_j|^\downarrow$'s are sorted in non-increasing order, this implies

$$\sum_{j=2}^k (|w_j|^\downarrow)^2 \geq \frac{k-1}{n-1}(1 - \|P\|_{(1)}).$$

Finally, this gives us

$$(4.2) \quad \|P\|_{(k)} \geq \sum_{j=1}^k (|w_j|^\downarrow)^2 = \|P\|_{(1)} + \sum_{j=2}^k (|w_j|^\downarrow)^2 \geq \|P\|_{(1)} + \frac{k-1}{n-1}(1 - \|P\|_{(1)}).$$

If we substitute the inequality $\|P\|_{(1)} \geq \text{Tr}(P)/n = \text{rank}(\rho)/n$ into Inequality (4.2), and then use the fact that $1/2 \geq \|P\|_{(k)}$ and solve for $\text{rank}(\rho)$, we get exactly the inequality given in the statement of the corollary. \square

In the special case when $k = 1$, Corollary 4.7 says that $D_{\text{coh}}(\rho) = 1$ implies $\text{rank}(\rho) \leq n/2$, agreeing with Corollary 4.3. At the other extreme, if $k > n/2$ then Corollary 4.7 says that $D_{k\text{-coh}}(\rho) = 1$ implies $\text{rank}(\rho) \leq 0$, which of course is impossible. It follows that if $D_{k\text{-coh}}(\rho) = 1$ then $k \leq n/2$. Furthermore, in the $k = n/2$ case if $D_{k\text{-coh}}(\rho) = 1$ then $\text{rank}(\rho) \leq 1$, so ρ must be pure.

We end this subsection by making a (somewhat tangential) note that in the numerical linear algebra literature, there is a concept called the factor width of a matrix first defined in [6]. That paper deals entirely with real matrices, so we give the natural extension of the definition to complex matrices.

DEFINITION 4.8. [6, Definition 1] *The factor width of a complex positive semidefinite matrix A is the smallest integer k for which there exists a complex (rectangular) matrix V where $A = VV^*$ and each column of V contains at most k non-zero entries.*

It is easy to see that a density matrix is k -coherent if and only if it has factor width at most k . In [6], there is a characterization of matrices which have factor rank at most two. We first need the following definitions:

DEFINITION 4.9. [34] Let $A \in M_n$, then its comparison matrix (denoted $M(A)$) is an n by n matrix with $M(A)_{i,i} = |A_{i,i}|$ for all i and $M(A)_{i,j} = -|A_{i,j}|$ whenever $i \neq j$.

DEFINITION 4.10. A square matrix is said to be an H -matrix if its comparison matrix is positive semidefinite.

We can now restate the main result of [6] in terms of 2-coherent states.

THEOREM 4.11. [6, Theorem 9] A density matrix is 2-coherent if and only if it is an H -matrix.

4.3. Entanglement and k -Entanglement. When we apply Theorem 4.1 to the resource theory of entanglement (i.e., when we take $C = C_{\text{ent}}$), we see that the modified trace distance of entanglement satisfies $D_{\text{ent}}(\rho) = 1$ if and only if $\langle v|P|v \rangle \leq 1/2$ for every pure product state $|v\rangle = |x\rangle \otimes |y\rangle \in \mathbb{C}^m \otimes \mathbb{C}^n$, where P is the orthogonal projection onto $\text{range}(\rho)$. In the terminology of [21, 22], this is equivalent to the statement that $\|P\|_{S(1)} \leq 1/2$, where $\|\cdot\|_{S(1)}$ is the norm defined by

$$(4.3) \quad \|X\|_{S(1)} \stackrel{\text{def}}{=} \max \{ \langle x| \otimes \langle y| X (|x\rangle \otimes |y\rangle) : |x\rangle \in \mathbb{C}^m, |y\rangle \in \mathbb{C}^n \}.$$

A natural generalization of entanglement is the resource theory of k -entanglement, in which the set $C = C_{k\text{-ent}}$ of free states consists of states with Schmidt number [33] at most k (where $1 \leq k \leq \min\{m, n\}$ is an integer):

$$C_{k\text{-ent}} \stackrel{\text{def}}{=} \left\{ \sum_j \mathbf{v}_j \mathbf{v}_j^* : \text{there exist } \{\mathbf{x}_{i,j}\} \subset \mathbb{C}^m, \{\mathbf{y}_{i,j}\} \subset \mathbb{C}^n \text{ such that } \mathbf{v}_j = \sum_{i=1}^k \mathbf{x}_{i,j} \otimes \mathbf{y}_{i,j} \text{ for all } j \right\}.$$

By the spectral decomposition theorem, if $k = 1$ then we have $C_{1\text{-ent}} = C_{\text{ent}}$, whereas if $k = \min\{m, n\}$ then we have $C_{\min\{m,n\}\text{-ent}} = (M_m \otimes M_n)^+$. For intermediate values of k , these sets satisfy a chain of inclusions that is analogous to that of k -coherence:

$$C_{\text{ent}} = C_{1\text{-ent}} \subsetneq C_{2\text{-ent}} \subsetneq C_{3\text{-ent}} \subsetneq \cdots \subsetneq C_{(\min\{m,n\}-1)\text{-ent}} \subsetneq C_{\min\{m,n\}\text{-ent}} = (M_m \otimes M_n)^+.$$

In this setting, the natural generalization of the norm defined in Equation (4.3) is

$$\|X\|_{S(k)} \stackrel{\text{def}}{=} \max \left\{ \langle v|X|v \rangle : \text{there exist } \{\mathbf{x}_i\} \subset \mathbb{C}^m, \{\mathbf{y}_i\} \subset \mathbb{C}^n \text{ such that } |v\rangle = \sum_{i=1}^k \mathbf{x}_i \otimes \mathbf{y}_i \right\}.$$

This quantity is also a norm, and an almost identical argument to the one from the start of this section demonstrates the following theorem:

THEOREM 4.12. Suppose $\rho \in (M_m \otimes M_n)^+$ is a mixed quantum state and P is the orthogonal projection onto $\text{range}(\rho)$. Then $D_{k\text{-ent}}(\rho) = 1$ if and only if $\|P\|_{S(k)} \leq 1/2$.

Unlike the case of coherence, the condition of Theorem 4.12 is difficult to check even in the $k = 1$ case. Indeed, computation of $\|\cdot\|_{S(1)}$ is NP-hard [10, 14] (which is expected, since computation of D_{ent} is also NP-hard). Despite this, numerous bounds on $\|P\|_{S(k)}$ are known, which let us derive rank bounds like the following:

COROLLARY 4.13. *Suppose $\rho \in (M_m \otimes M_n)^+$ is a mixed quantum state with $D_{k\text{-ent}}(\rho) = 1$. Then*

$$\begin{aligned} \text{rank}(\rho) &\leq \frac{mn(\min\{m, n\} - 2k + 1)}{2(\min\{m, n\} - k)} \quad \text{and} \\ \text{rank}(\rho) &\leq \frac{(n + m + 2 - 4k)^2 - (n - m)^2}{4}. \end{aligned}$$

Proof. We use [21, Theorem 4.15], which says that every orthogonal projection P satisfies the pair of inequalities

$$\begin{aligned} \|P\|_{S(k)} &\geq \frac{(k - 1)mn + (\min\{m, n\} - k)\text{rank}(P)}{mn(\min\{m, n\} - 1)} \quad \text{and} \\ \|P\|_{S(k)} &\geq \min \left\{ 1, \frac{k}{\left\lfloor \frac{1}{2}(n + m - \sqrt{(n - m)^2 + 4\text{rank}(P) - 4}) \right\rfloor} \right\}. \end{aligned}$$

If $D_{k\text{-ent}}(\rho) = 1$, then Theorem 4.12 tells us that $1/2 \geq \|P\|_{S(k)}$, which gives the claimed result after a bit of routine (but somewhat ugly) algebra. \square

In particular, the first inequality of Corollary 4.13 tells us that if $k > \min\{m, n\}/2$ then $\text{rank}(\rho) \leq 0$, which is impossible. It follows that there only exist quantum states ρ with $D_{k\text{-ent}}(\rho) = 1$ when $k \leq \min\{m, n\}/2$. Furthermore, in the extreme case when $k = \min\{m, n\}/2$, the second inequality of Corollary 4.13 says that any state with $D_{k\text{-ent}}(\rho) = 1$ must have $\text{rank}(\rho) \leq |n - m + 1|$ (so if we also have $m = n$ then ρ must be pure). At the other extreme, if $k = 1$ then the first inequality of Corollary 4.13 is the stronger one, and it says that if $D_{\text{ent}}(\rho) = 1$ then $\text{rank}(\rho) \leq mn/2$.

4.4. An application to the NPPT bound entanglement problem. Given a quantum state $\rho \in (M_m \otimes M_n)^+$, one often wants to know if it can be *distilled*: transformed into (an arbitrarily good approximation of) the *pure maximally entangled state*

$$|\phi_+\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{\min\{m, n\}}} \left(\sum_{j=1}^{\min\{m, n\}} \mathbf{e}_j \otimes \mathbf{e}_j \right),$$

via local operations [16]. In some cases where ρ is undistillable, multiple tensor copies of it are distillable [35], so we say that ρ is *r-copy (un)distillable* if $\rho^{\otimes r} \in (M_m^{\otimes r} \otimes M_n^{\otimes r})^+$ is (un)distillable.

It is straightforward to show that if ρ is separable (i.e., if $\rho \in C_{\text{ent}}$) then it is *r-copy undistillable* for all $r \geq 1$. More surprisingly, it has also been shown that there is a class of entangled (i.e., not separable) states called *PPT states* that are *r-copy undistillable* for all $r \geq 1$ [17]. Entangled states with this undistillability property are called *bound entangled*.

The *NPPT bound entanglement* problem asks whether or not there are any other (i.e., non-PPT) states that are *r-copy undistillable* for all $r \geq 1$, and it is one of the central open questions in the theory of quantum entanglement [18]. It has been shown that it suffices to answer this problem for a single-parameter family of states called *Werner states* [15], and it has furthermore been conjectured that the specific Werner state

$$(4.4) \quad \rho_{2/n} \stackrel{\text{def}}{=} \left(\frac{1}{n^2 - 2} \right) I - \frac{2}{n(n^2 - 2)} \left(\sum_{i,j=1}^n \mathbf{e}_i \mathbf{e}_j^* \otimes \mathbf{e}_j \mathbf{e}_i^* \right) \in (M_n \otimes M_n)^+,$$

is bound entangled for all $n \geq 4$ [9]. Indeed, it is straightforward to check that $\rho_{2/n}$ is not PPT and that it is (1-copy) undistillable. However, even proving the simplest case of 2-copy undistillability of this state when $n = 4$ remains elusive, despite significant effort [7, 27].

It was shown in [21] that $\rho_{2/n}$ is r -copy undistillable if and only if the orthogonal projections defined recursively by

$$(4.5) \quad \begin{aligned} P_1 &\stackrel{\text{def}}{=} |\phi_+\rangle\langle\phi_+| \in (M_n \otimes M_n)^+ \quad \text{and} \\ P_r &\stackrel{\text{def}}{=} (I - P_1) \otimes P_{r-1} + P_1 \otimes (I - P_{r-1}) \in (M_n^{\otimes r} \otimes M_n^{\otimes r})^+ \quad \forall r \geq 2, \end{aligned}$$

satisfy $\|P_r\|_{S(2)} \leq 1/2$. Here, it is understood that the partitioning of the $2r$ tensor factors used to compute $\|P_r\|_{S(2)}$ is the one in which the first factors of P_1 always “stay together,” as do the second factors of P_1 . That is, if we label the two tensor factors of P_1 as $P_1 \in A \otimes B$ then we have $P_r \in (A^{\otimes r}) \otimes (B^{\otimes r})$, with the $S(2)$ -norm being computed across the central tensor cut, *not* $P_r \in (A \otimes B) \otimes \cdots \otimes (A \otimes B)$.

This connection with the $S(2)$ -norm leads to the following characterization of the bound entanglement problem in terms of Birkhoff–James orthogonality and the modified trace distance of 2-entanglement:

THEOREM 4.14. *Let $\rho_{2/n} \in (M_n \otimes M_n)^+$ be the Werner state defined in Equation (4.4) and let P_r be the orthogonal projection defined in Equation (4.5). The following are equivalent:*

- (a) $\rho_{2/n}$ is r -copy undistillable.
- (b) P_r is Birkhoff–James orthogonal to every member of $C_{2\text{-ent}}$.
- (c) There exists a quantum state $\sigma \in (M_n^{\otimes r} \otimes M_n^{\otimes r})^+$ with $\text{range}(\sigma) = \text{range}(P_r)$ and $D_{2\text{-ent}}(\sigma) = 1$.

Proof. By Theorem 4.12, we know that condition (c) happens if and only if $\|P_r\|_{S(2)} \leq 1/2$, and we already discussed the fact that $\|P_r\|_{S(2)} \leq 1/2$ is equivalent to $\rho_{2/n}$ being r -copy undistillable [21]. This shows that (c) is equivalent to (a).

The fact that $\|P_r\|_{S(2)} \leq 1/2$ is also equivalent to (b) (and thus (a) and (c) are equivalent to (b)) follows from the equivalence of conditions (b) and (c) of Theorem 4.1 when $C = C_{2\text{-ent}}$. \square

5. Conclusions and outlook. The notion of Birkhoff–James orthogonality is an extension of the usual notion of orthogonality. We considered Birkhoff–James orthogonality in the setting of $n \times n$ complex-valued matrices together with the trace norm. Our main result (Theorem 2.3) characterized when a Hermitian matrix H is Birkhoff–James orthogonal in the trace norm to a given positive semidefinite matrix B , via trace conditions involving the product of B and the positive and negative eigenprojection matrices of H . These eigenprojection matrices are key to our study, and we used them to further characterize when H is Birkhoff–James orthogonal in the trace norm to the set of all positive semidefinite diagonal matrices and when H is Birkhoff–James orthogonal in the trace norm to the set of all (not just positive semidefinite) diagonal matrices in certain special cases (e.g., when H has a certain rank, when H is positive semidefinite, and/or the dimension is small).

Our results have direct applications to quantum resource theories; we showed that the modified trace distance between any quantum state (positive semidefinite, trace-one matrix) ρ and a given closed convex cone C of positive $n \times n$ matrices equals 1 (the maximum possible value) precisely when ρ is Birkhoff–James orthogonal in the trace norm to every member of C . This allowed us to generalize a result from [23], which characterized when the modified trace distance of coherence of a given pure quantum state is maximal, in numerous ways: we now have a version of that result for non-pure states (i.e., matrices of higher rank), we

now have a version of that result that applies to resource theories other than just quantum coherence, and we now have a bound on how large the rank of such states can be.

Our results, when applied to the resource theory of 2-entanglement (i.e., Schmidt number 2), provide an intriguing connection to the NPPT bound entanglement problem. While we do not solve that problem (it has been open for over a decade and is notoriously difficult), we showed that it can be phrased in terms of Birkhoff–James orthogonality, thus opening up a wide variety of new tools that can be used to explore it.

Our results highlight the clear utility of Birkhoff–James orthogonality. In terms of linear algebraic considerations, one open problem would be to extend the work from Hermitian to normal matrices, which we believe is possible but we avoided the added complexity of the problem as we were interested in the applications to quantum resource theory, which involves positive semidefinite matrices. It would also be interesting to exactly characterize when a given Hermitian matrix H is Birkhoff–James orthogonal to every diagonal matrix, but our Examples 3.2 and 3.10 suggest that such a characterization might be quite delicate and difficult to pin down.

Acknowledgment. N.J. was supported by NSERC Discovery Grant numbers RGPIN-2016-04003 and RGPIN-2022-04098. R.P. was supported by NSERC Discovery Grant number 400550. S.P. was supported by NSERC Discovery Grant number 1174582, the Canada Foundation for Innovation (CFI) grant number 35711, and the Canada Research Chairs (CRC) Program grant number 231250.

REFERENCES

- [1] J. Aberg, *Quantifying Superposition*. E-print: [arXiv:quant-ph/0612146](https://arxiv.org/abs/2006.06121), 2006.
- [2] E. Andruchow, G. Larotonda, L. Recht, and A. Varella. A Characterization of Minimal Hermitian Matrices. *Linear Algebra Appl.*, 436:2366–2374, 2012. <https://doi.org/10.1016/j.laa.2011.08.042>, <https://www.sciencedirect.com/science/article/pii/S0024379511006264>.
- [3] T. Baumgratz, M. Cramer, and M.B. Plenio. Quantifying coherence. *Phys. Rev. Lett.*, 113:140401, 2014.
- [4] R. Bhatia. *Matrix Analysis*. Springer, 1997.
- [5] R. Bhatia and P. Šemrl. Orthogonality of matrices and some distance problems. *Linear Algebra Appl.*, 287:77–85, 1999.
- [6] E. Boman, D. Chen, O. Parekh, and S. Toledo. On factor width and symmetric H-matrices. *Linear Algebra Appl.*, 405:239–248, 2005.
- [7] L. Chen, H. He, X. Shi, and L.-J. Zhao. Proving the distillability problem of two-copy 4×4 Werner states for monomial matrices. *Quan. Inform. Process.*, 20:157, 2021.
- [8] E. Chitambar and G. Gour. Quantum resource theories. *Rev. Modern Phys.*, 91:025001, 2019.
- [9] D.P. DiVincenzo, P.W. Shor, J.A. Smolin, B.M. Terhal, and A.V. Thapliyal. Evidence for bound entangled states with negative partial transpose. *Phys. Rev. A*, 61:062312, 2000.
- [10] S. Gharibian. Strong NP-hardness of the quantum separability problem. *Quan. Inform. Comput.*, 10:343–360, 2010.
- [11] R.J. Glauber. Coherent and incoherent states of the radiation field. *Phys. Rev.*, 131:2766, 1963.
- [12] P. Grover. Orthogonality of matrices in the KY fan k-norms. *Linear Multilinear Algebra*, 65:496–509, 2017.
- [13] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474:1–75, 2009.
- [14] L. Gurvits. Classical deterministic complexity of Edmonds’ problem and quantum entanglement. in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, 2003, 10–19.
- [15] M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206–4216, 1999.
- [16] M. Horodecki, P. Horodecki, and R. Horodecki. Inseparable two spin-1/2 density matrices can be distilled to a singlet form. *Phys. Rev. Lett.*, 78:574–577, 1997.
- [17] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: Is there a ‘bound’ entanglement in nature? *Phys. Rev. Lett.*, 80:5239–5242, 1998.
- [18] P. Horodecki, L. Rudnicki, and K. Życzkowski. Five open problems in quantum information. E-print: [arXiv:2002.03233](https://arxiv.org/abs/2002.03233) [quant-ph], 2020. <https://arxiv.org/abs/2002.03233>.

- [19] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.
- [20] R.C. James. Orthogonality and linear functionals in normed linear spaces. *Trans. Amer. Math. Soc.*, 61:265–292, 1947.
- [21] N. Johnston and D.W. Kribs. A family of norms with applications in quantum information theory. *J. Math. Phys.*, 51:082202, 2010.
- [22] N. Johnston and D.W. Kribs. A family of norms with applications in quantum information theory II. *Quan. Inform. Comput.*, 11:104–123, 2011.
- [23] N. Johnston, C.-K. Li, and S. Plosker. The modified trace distance of coherence is constant on most pure states. *J. Phys. A: Math. Theor.*, 51:414010, 2018.
- [24] D.W. Kribs, D. Mammarella, and R. Pereira. Isoclinic subspaces and quantum error correction. *Oper. Matrices*, 15:571–580, 2021.
- [25] C.-K. Li and H. Schneider. Orthogonality of matrices. *Linear Algebra Appl.*, 347:115–122, 2002.
- [26] J. Li, R. Pereira, and S. Plosker. Some geometric interpretations of quantum fidelity. *Linear Algebra Appl.*, 487:158–171, 2015.
- [27] L. Pankowski, M. Piani, M. Horodecki, and P. Horodecki. A few steps more towards NPT bound entanglement. *IEEE Trans. Inform. Theory*, 56:4085–4100, 2010.
- [28] S. Rana, P. Parashar, and M. Lewenstein. Trace-distance measure of coherence. *Phys. Rev. A*, 93:012110, 2016.
- [29] B. Regula. Convex geometry of quantum resource quantification. *J. Phys. A: Math. Theor.*, 51:045303, 2018.
- [30] B. Regula, K. Fang, X. Wang, and M. Gu. One-shot entanglement distillation beyond local operations and classical communication. *New J. Phys.*, 21:103017, 2019. <https://doi.org/10.1088/1367-2630/ab4732>, <https://doi.org/10.1088/1367-2630/ab4732>.
- [31] M. Ringbauer, T.R. Bromley, M. Cianciaruso, L. Lami, W.Y.S. Lau, G. Adesso, A.G. White, A. Fedrizzi, and M. Piani. Certification and quantification of multilevel quantum coherence. *Phys. Rev. X*, 8:041007, 2018. <https://doi.org/10.1103/PhysRevX.8.041007>, <https://link.aps.org/doi/10.1103/PhysRevX.8.041007>.
- [32] E.C.G. Sudarshan. Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams. *Phys. Rev. Lett.*, 10:277, 1963.
- [33] B.M. Terhal and P. Horodecki. Schmidt number for density matrices. *Phys. Rev. A*, 61:040301(R), 2000.
- [34] R.S. Varga. On recurring theorems on diagonal dominance. *Linear Algebra Appl.*, 13:1–9, 1976.
- [35] J. Watrous. Many copies may be required for entanglement distillation. *Phys. Rev. Lett.*, 93:010502, 2004. <https://doi.org/10.1103/PhysRevLett.93.010502>, <https://link.aps.org/doi/10.1103/PhysRevLett.93.010502>.
- [36] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.
- [37] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.*, 170:33–45, 1992.
- [38] R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989.
- [39] X.-D. Yu, D.-J. Zhang, G.F. Xu, and D.M. Tong. Alternative framework for quantifying coherence. *Phys. Rev. A*, 94:060302(R), 2016.