




# Towards a Resolution of the Spin Alignment Problem

Mohammad A. Alhejji<sup>1,2,3</sup> , Emanuel Knill<sup>1,4</sup>

- <sup>1</sup> National Institute of Standards and Technology, Boulder, CO 80305, USA
- <sup>2</sup> JILA, University of Colorado, 440 UCB, Boulder, CO 80309, USA
- <sup>3</sup> Department of Physics, University of Colorado, Boulder, CO, USA.  
E-mail: [mohammad.alhejji@colorado.edu](mailto:mohammad.alhejji@colorado.edu)
- <sup>4</sup> Center for Theory of Quantum Matter, University of Colorado, Boulder, CO 80309, USA

Received: 7 August 2023 / Accepted: 24 February 2024  
Published online: 29 April 2024 – © The Author, under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature. Parts of this work were authored by a US Federal Government author and are not under copyright protection in the US; foreign copyright protection may apply 2024

**Abstract:** Consider minimizing the entropy of a mixture of states by choosing each state subject to constraints. If the spectrum of each state is fixed, we expect that in order to reduce the entropy of the mixture, we should make the states less distinguishable in some sense. Here, we study a class of optimization problems that are inspired by this situation and shed light on the relevant notions of distinguishability. The motivation for our study is the recently introduced spin alignment conjecture. In the original version of the underlying problem, each state in the mixture is constrained to be a freely chosen state on a subset of  $n$  qubits tensored with a fixed state  $Q$  on each of the qubits in the complement. According to the conjecture, the entropy of the mixture is minimized by choosing the freely chosen state in each term to be a tensor product of projectors onto a fixed maximal eigenvector of  $Q$ , which maximally “aligns” the terms in the mixture. We generalize this problem in several ways. First, instead of minimizing entropy, we consider maximizing arbitrary unitarily invariant convex functions such as Fan norms and Schatten norms. To formalize and generalize the conjectured required alignment, we define *alignment* as a preorder on tuples of self-adjoint operators that is induced by majorization. We prove the generalized conjecture for Schatten norms of integer order, for the case where the freely chosen states are constrained to be classical, and for the case where only two states contribute to the mixture and  $Q$  is proportional to a projector. The last case fits into a more general situation where we give explicit conditions for maximal alignment. The spin alignment problem has a natural “dual” formulation, versions of which have further generalizations that we introduce.

## Contents

1. Introduction	2
2. Mathematical Preliminaries	4
3. Spin Alignment Problems and the Alignment Ordering	7
4. Results	10

4.1	First observations	10
4.2	Classical spin alignment	11
4.3	Schatten norms of integer order	13
4.4	The case where $\text{supp}(\mu) \leq 2$ and $Q$ is proportional to a projector	14
5.	"Dual" Problems to Spin Alignment and Their Generalizations	16
6.	Concluding Remarks	17

## 1. Introduction

Minimizing dispersion in a communication signal is often necessary in both the theory and practice of information processing. Informally, the term dispersion is used here so that the more dispersion in a signal the more information content it carries. How dispersion is quantified depends on context. For example, in a setting where sources are taken to generate signals in an independent and identically-distributed manner and processes are assumed to be memoryless, a useful measure of dispersion is the Shannon entropy for classical signals associated with probability distributions [1], or analogously the von Neumann entropy for quantum signals associated with quantum states [2]. By this point, myriads of measures of dispersion have been introduced and extensively studied [3–5]. A sensible requirement for a measure of dispersion is that it does not decrease under any process where labels designating outcomes are confused. Such a process is represented classically by a doubly-stochastic transition matrix; the quantum generalization of which is a mixed-unitary channel. Put in other words, we require measures of dispersion to be *Schur-concave*. In this work, we switch from concave to convex and primarily focus on convex Schur-convex functions. To be precise, we study how unitarily invariant convex functions behave under mixing of signals. The value of these functions at a mixture of signals depends on the spectra of the individual signals supported in the mixture as well as the alignment or overlap of the signals. It is our aim to shed light on the latter kind of dependence.

Maximizing unitarily invariant convex functions is less ubiquitous in quantum information than minimizing them, and is generally a more difficult task. This difficulty merits more analysis and research. Quantum channel minimum output entropy is an information-theoretic quantity that is defined by a minimization of a Schur-concave function. A reason it is of interest is the equivalence of the additivity (under tensor multiplication) conjectures of the minimum output entropy and the Holevo information [6]. The latter quantity is relevant because its regularization is a formula for the capacity for classical communication [7]. Despite an existence proof due to Hastings [8], an explicit example of a quantum channel with a finite-dimensional domain and a strictly super-additive minimum output entropy is yet to be found. In addition, such minimizations are relevant in broadcast channel scenarios with privacy concerns. Ideally, the entropy of the environment or the adversary is kept at a minimum. Depending on the channels in question, the optimal states may be mixtures, as in the case of platypus channels [9].

We are interested in situations where we can unambiguously say that a signal  $a$  contains less dispersion than a signal  $b$ . That is,  $f(b) \leq f(a)$  for all Schur-convex functions  $f$ . Luckily, we need not argue such an inequality for every such  $f$ . It suffices to check that  $a$  majorizes  $b$ . In this sense, of all the ways of measuring dispersion, majorization is the most stringent [10, 11]. Majorization theory, which is nearly a century old [12, 13], is the basis of Nielsen's celebrated theorem for pure bipartite state conversion using local operations and classical communication (LOCC) [14]. This is one example among many of majorization giving insights into information processing.

Here, we elucidate what we mean by alignment and how it is relevant to the behavior of dispersion under mixing. For a self-adjoint operator  $S$ , let  $\lambda(S)$  denote the tuple whose entries are the eigenvalues of  $S$  (with multiplicity) ordered non-increasingly.

**Definition 1.1** (*Perfect alignment*). Let  $S_1, \dots, S_\ell$  be self-adjoint operators on  $\mathbb{C}^d$ . They are said to be *perfectly aligned* if there exists an ordered orthonormal basis  $\{|\phi_i\rangle\}_{i=1}^d$  such that  $S_1 = \sum_{i=1}^d \lambda_i(S_1) |\phi_i\rangle\langle\phi_i|$ ,  $\dots$ ,  $S_\ell = \sum_{i=1}^d \lambda_i(S_\ell) |\phi_i\rangle\langle\phi_i|$ .

As an example, we note that self-adjoint projection operators  $P_1, \dots, P_\ell$  are perfectly aligned if and only if their supports are nested. This definition is motivated by Fan's majorization relation [15], which in the case of states  $\rho_1, \dots, \rho_\ell$  and a probability measure  $p = (p_i)_{i=1}^\ell$  reads

$$\lambda\left(\sum_{i=1}^{\ell} p_i \rho_i\right) \preceq \sum_{i=1}^{\ell} p_i \lambda(\rho_i). \quad (1.1)$$

Here  $b \preceq a$  means  $b$  is majorized by  $a$ . These inequalities are simultaneously saturated if and only if the states are perfectly aligned. Hence, to minimize the dispersion in a mixture of states subject to spectral constraints, we should choose the states to be perfectly aligned.

More generally, we may consider a set  $G$  whose elements are  $\ell$ -tuples of states  $(\rho_i)_{i=1}^\ell$  that satisfy constraints including that for each  $i \in [\ell]$ ,  $\lambda(\rho_i)$  is given. We may then ask if there is a tuple  $(\tilde{\rho}_i)_{i=1}^\ell \in G$  such that for all probability measures  $(p_i)_{i=1}^\ell$ ,  $\lambda(\sum_{i=1}^\ell p_i \tilde{\rho}_i)$  majorizes  $\lambda(\sum_{i=1}^\ell p_i \sigma_i)$  for all  $(\sigma_i)_{i=1}^\ell \in G$ . If  $G$  contains a tuple of perfectly aligned states, the answer is yes by way of the relation in Eq. 1.1. The following is an example of a situation where constraints preclude the existence of such a tuple.

*Example 1.2.* Let  $\tau$  be a state acting on  $\mathbb{C}^d$  of rank strictly less than  $d$ . Let  $|\alpha\rangle$  be a unit vector in  $\text{supp}(\tau) := \ker(\tau)^\perp$  and let  $|e\rangle$  be a unit vector in  $\ker(\tau)$ . Consider the set of pairs of states of the form  $(\tau, |v_\gamma\rangle\langle v_\gamma|)$ , where  $|v_\gamma\rangle := \sqrt{\gamma}|\alpha\rangle + \sqrt{1-\gamma}|e\rangle$  for some  $\gamma \in [0, 1]$ . If  $|\alpha\rangle$  is a maximal eigenvector of  $\tau$ , then  $\tau$  is perfectly aligned with  $|v_1\rangle\langle v_1|$ . Hence, we can see via Eq. 1.1 that for all  $p \in [0, 1]$ ,  $p\tau + (1-p)|v_1\rangle\langle v_1|$  majorizes  $p\tau + (1-p)|v_\gamma\rangle\langle v_\gamma|$  for all  $\gamma \in [0, 1]$ . If  $|\alpha\rangle$  is not a maximal eigenvector of  $\tau$ , then  $\tau$  is not perfectly aligned with  $|v_\gamma\rangle\langle v_\gamma|$  for any  $\gamma \in [0, 1]$ . However, it is natural to conjecture that the choice  $\gamma = 1$  minimizes dispersion in the mixture anyway, as that is where the two states are as overlapped or aligned as possible. We give a proof of this conjecture in App. C.

We use the spin alignment problem as a case study of these notions of alignment with the aim of refining Fan's majorization relation. Introduced by Leditzky et al. [9], the problem arose in the context of proving additivity of the coherent information of a class of quantum channels with peculiar communication properties. For a qudit state  $Q$  and a probability measure  $\mu$ , they considered  $n$ -qudit states of the form

$$\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c}. \quad (1.2)$$

Each term in the mixture corresponds to a subset  $I \subseteq [n]$  and is a tensor product of a state  $\rho_I$  of the qudits in  $I$  and a fixed state  $Q$  on each of the qudits in the complement  $I^c$ . Their task was to pick the tuple of states  $(\rho_I)_{I \subseteq [n]}$  so that the overall operator in Eq. 1.2 has minimum von Neumann entropy. It is not too difficult to see that an optimal choice has

to be one where states in the tuple are pure (see Lemma 3.2 below). However, depending on  $\mu$ , there may be no such choice where the summands in Eq. 1.2 are perfectly aligned. Leditzky et al. conjectured that the von Neumann entropy is minimized by choosing the tuple of states to be  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ , where  $|q_1\rangle$  is a maximal eigenvector of  $Q$ . With this tuple, the summands appear to overlap maximally. They showed that if this conjecture is true, then their channels have additive coherent information. For more information on these channels, see [9] and the companion paper [16].

The structure of this paper is as follows. In Sect. 2, we introduce notation and the relevant mathematical background. We formally define and motivate the class of spin alignment problems in Sect. 3. In Sect. 4, we provide our results, which include resolutions for various instances of spin alignment. In Sect. 5, we introduce a general class of optimization problems that is “dual” to spin alignment. In Sect. 6, we conclude and discuss open problems and follow-up lines of inquiry. Appendices App. A and App. B contain proofs of technical lemmas. Appendix App. C contains a resolution of the problem stated in Example 1.2.

## 2. Mathematical Preliminaries

In this work, we are concerned with finite-dimensional inner product spaces over  $\mathbb{C}$ . We denote such spaces by  $\mathcal{H}, \mathcal{K}$ , etc. The dimension of  $\mathcal{H}$  is denoted by  $d_{\mathcal{H}}$ . The set of operators on  $\mathcal{H}$  is denoted by  $\mathcal{B}(\mathcal{H})$ .  $\mathcal{S}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$  denotes the real vector space of self-adjoint operators on  $\mathcal{H}$ .  $\mathcal{P}(\mathcal{H}) \subset \mathcal{S}(\mathcal{H})$  denotes the convex cone of positive semi-definite operators, and  $\mathcal{D}(\mathcal{H}) \subset \mathcal{P}(\mathcal{H})$  denotes the convex set of states, that is the set of positive semi-definite operators with unit trace. For  $R \in \mathcal{B}(\mathcal{H})$ ,  $\sigma(R) \in \mathbb{R}^{d_{\mathcal{H}}}$  is the vector whose entries are the singular values of  $R$  ordered non-increasingly. Similarly, for  $S \in \mathcal{S}(\mathcal{H})$ ,  $\lambda(S) \in \mathbb{R}^{d_{\mathcal{H}}}$  is the vector whose entries are the eigenvalues of  $S$  ordered non-increasingly. We use the word projector to refer to a self-adjoint projection operator and the word qudit to refer to a physical system that is modeled on a complex inner product space of dimension  $d$ . For  $n \in \mathbb{N}$ , the set  $\{1, \dots, n\}$  is denoted by  $[n]$  and its power set is denoted by  $2^{[n]}$ . Given a measure  $\mu$  on a countable set, its support is denoted by  $\text{supp}(\mu)$ . If  $x \in \mathbb{R}^d$ , then  $x^{\downarrow}$  and  $x^{\uparrow}$  denote the vectors in  $\mathbb{R}^d$  with the entries of  $x$  ordered non-increasingly and non-decreasingly, respectively. Given  $\mathcal{H}$ , we distinguish an ordered orthonormal basis  $\{|q_i\rangle\}_{i=1}^{d_{\mathcal{H}}}$ . For  $S \in \mathcal{S}(\mathcal{H})$ ,  $S^{\downarrow} := \sum_{i=1}^{d_{\mathcal{H}}} \lambda_i(S) |q_i\rangle\langle q_i|$ . Real vectors  $v_1, \dots, v_s$  in  $\mathbb{R}^d$  are said to be similarly ordered if there exists a permutation matrix  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $v_1^{\downarrow} = Av_1, \dots, v_s^{\downarrow} = Av_s$ . Similar ordering for real vectors is analogous to perfect alignment for self-adjoint operators. Observe that for an arbitrary vector  $u \in \mathbb{R}^d$  and  $v \in \text{span}(\underbrace{(1, \dots, 1)}_{d \text{ times}})$ , the two are similarly ordered.

Analogously, an arbitrary self-adjoint operator and an arbitrary operator in the real span of the identity operator are perfectly aligned. This observation shows that neither similar ordering nor perfect alignment is transitive when the dimension of the underlying space is greater than 1.

Here is a brief review of majorization theory. Given  $x, y \in \mathbb{R}^d$ ,  $x$  is said to majorize  $y$  if

$$\sum_{i=1}^k x_i^{\downarrow} \geq \sum_{i=1}^k y_i^{\downarrow} \quad (2.1)$$

for each  $k \in [d - 1]$  and  $\sum_{i=1}^d x_i^\downarrow = \sum_{i=1}^d y_i^\downarrow$ . This statement is denoted with  $x \succeq y$ . Majorization is defined on  $\mathcal{S}(\mathcal{H})$  spectrally. For  $B, C \in \mathcal{S}(\mathcal{H})$ , if  $\lambda(B) \succeq \lambda(C)$ , then  $B$  is said to majorize  $C$ , and we denote that with  $B \succeq C$ . Majorization induces a preorder on  $\mathbb{R}^d$  and  $\mathcal{S}(\mathcal{H})$ , respectively. Given a set, a majorant, if it exists, is an element of the set that majorizes all elements in the set.

There are several useful characterizations of majorization, two of which are relevant to this work. The first is based on doubly-stochastic processing (see Theorem II.1.10 on page 33 in Ref. [10]):  $x$  majorizes  $y$  if and only if there exist a tuple of permutation matrices  $(P_i)_i$  and a probability measure  $(p_i)_i$  such that

$$y = \sum_i p_i P_i x. \quad (2.2)$$

In words,  $x$  majorizes  $y$  if and only if  $y$  belongs to the convex hull of the orbit of  $x$  under the action of the symmetric group. That is,  $y = Ax$  for some doubly-stochastic matrix  $A$ . Analogously,  $B \succeq C$  if and only if

$$C = \sum_i p_i U_i B U_i^* \quad (2.3)$$

for a tuple of unitary operators  $(U_i)_i$  and a probability measure  $(p_i)_i$ . Put differently,  $C$  is the value of a mixed-unitary quantum channel at  $B$ .

The second characterization is based on unjust transfers, which are related to the better-known  $T$ -transforms [11]. For distinct  $i, j \in [d]$ , and  $t \in [0, 1]$ , a  $T$ -transform (also known as a Robin Hood operation [17]) is a linear map with action

$$(x_1, \dots, x_i, \dots, x_j, \dots, x_d) \mapsto (x_1, \dots, (1-t)x_i + tx_j, \dots, (1-t)x_j + tx_i, \dots, x_d). \quad (2.4)$$

A  $T$ -transform is a convex combination of the identity and a transposition. The statement  $x \succeq y$  holds if and only if  $y$  can be obtained from  $x$  by a finite number of  $T$ -transforms (see Lemma B.1 on page 32 in Ref. [11]). A direct consequence of this is that  $x \succeq y$  implies  $x \oplus z \succeq y \oplus z$  for all  $x, y \in \mathbb{R}^{d_1}$  and  $z \in \mathbb{R}^{d_2}$ . A transfer on  $\mathbb{R}^d$  is an affine transformation determined by a pair of distinct indices  $i, j \in [d]$  and an amount  $\varepsilon \geq 0$ . It has the following action

$$(x_1, \dots, x_i, \dots, x_j, \dots, x_d) \mapsto (x_1, \dots, x_i + \varepsilon, \dots, x_j - \varepsilon, \dots, x_d). \quad (2.5)$$

Here,  $i$  is the receiving entry of the transfer and  $j$  is the giving entry. If  $x_i \geq x_j$ , then when applied to  $x$ , the transfer is said to be *unjust*.

**Proposition 2.1.** *Let  $x, y \in \mathbb{R}^d$ . Then,  $y = Tx$  for some  $T$ -transform  $T$  if and only if  $x^\downarrow$  can be obtained from  $y^\downarrow$  with an unjust transfer.*

*Proof.* Since both  $T$ -transforms and transfers affect at most a pair of coordinates, we may assume without loss of generality that  $d = 2$ . Denote  $x = (x_1, x_2)$  and  $y = (y_1, y_2)$ .

Suppose that there is a  $T$ -transform  $T$  such that  $y = Tx$ . Then, for some  $t \in [0, 1]$

$$(y_1, y_2) = ((1-t)x_1 + tx_2, (1-t)x_2 + tx_1). \quad (2.6)$$

Because  $y_1^\downarrow$  is a convex combination of  $x_1$  and  $x_2$ , we have  $x_1^\downarrow \geq y_1^\downarrow$ . Let  $\tilde{\varepsilon} := x_1^\downarrow - y_1^\downarrow$ , and observe that  $x_1^\downarrow = y_1^\downarrow + \tilde{\varepsilon}$  and  $x_2^\downarrow = x_1^\downarrow + x_2^\downarrow - x_1^\downarrow = y_1^\downarrow + y_2^\downarrow - x_1^\downarrow = y_2^\downarrow - \tilde{\varepsilon}$ , where we used the fact that  $x_1 + x_2 = y_1 + y_2$ .

Conversely, suppose there exists an unjust transfer of  $\varepsilon > 0$  (the case where  $\varepsilon = 0$  is trivial) that takes  $y^\downarrow$  to  $x^\downarrow$ . That is,  $x_1^\downarrow = y_1^\downarrow + \varepsilon$  and  $x_2^\downarrow = y_2^\downarrow - \varepsilon$ . Observe that for  $\tilde{t} := \frac{\varepsilon}{(y_1^\downarrow - y_2^\downarrow) + 2\varepsilon}$ ,

$$y_1^\downarrow = (1 - \tilde{t})x_1^\downarrow + \tilde{t}x_2^\downarrow, \quad y_2^\downarrow = (1 - \tilde{t})x_2^\downarrow + \tilde{t}x_1^\downarrow. \quad (2.7)$$

Hence, there exists a  $T$ -transform  $T'$  such that  $T'x^\downarrow = y^\downarrow$ . Without loss of generality, we may take  $x = x^\downarrow$ . If  $y \neq y^\downarrow$ , we may compose a transposition with  $T'$  to get another  $T$ -transform  $T$  that satisfies  $Tx = y$ .  $\square$

From this proposition, we conclude that  $x \succeq y$  if and only if it is possible to get to  $x$  from  $y$  with a finite number of transpositions and unjust transfers. In the particular case of  $x, y \in \mathbb{R}^2$ ,  $x \succeq y$  if and only if  $x$  can be obtained from  $y$  by an unjust transfer, or an unjust transfer and a transposition. This characterization may be extended to self-adjoint operators as follows. For  $B, C \in \mathcal{S}(\mathcal{H})$ ,  $B \succeq C$  if and only if there exists a finite number of transpositions and unjust transfers that take  $\lambda(C)$  to  $\lambda(B)$ .

Let  $f$  be a real-valued function such that  $\text{dom}(f)$  is a convex set contained in either  $\mathbb{R}^d$  or  $\mathcal{S}(\mathcal{H})$ . Suppose that  $\text{dom}(f)$  is closed under the action of the symmetric group if it is contained in  $\mathbb{R}^d$  and closed under the action of the unitary group if it is contained in  $\mathcal{S}(\mathcal{H})$ . We say that  $f$  is Schur-convex if for all  $a, b \in \text{dom}(f)$ ,  $a \succeq b$  implies  $f(a) \geq f(b)$ . If  $-f$  is Schur-convex, then  $f$  is said to be Schur-concave. Let  $F$  be a real-vector-valued function such that  $\text{dom}(F)$  is a convex set contained in either  $\mathbb{R}^d$  or  $\mathcal{S}(\mathcal{H})$ . Similarly, suppose that  $\text{dom}(F)$  is closed under the action of the symmetric group if it is contained in  $\mathbb{R}^d$  and closed under the action of the unitary group if it is contained in  $\mathcal{S}(\mathcal{H})$ .  $F$  is called strictly isotone if for all  $a, b \in \text{dom}(F)$  such that  $a \succeq b$ , it holds that  $F(a) \succeq F(b)$ .

An important tool in our analysis is Fan's maximum principle [15] for self-adjoint operators. It states that for  $T \in \mathcal{S}(\mathbb{C}^d)$ ,  $k \in [d]$ ,

$$\sum_{i=1}^k \lambda_i(T) = \max\{\text{tr}(TP) \mid P \text{ is a rank-}k \text{ projector}\}. \quad (2.8)$$

It follows directly from this principle that

$$\sum_{i=1}^s S_i \preceq \sum_{i=1}^s S_i^\downarrow \quad (2.9)$$

for any self-adjoint  $S_1, \dots, S_s$ . Moreover, the inequalities above are simultaneously satisfied with equality if and only if  $S_1, \dots, S_s$  are perfectly aligned.

We are interested in maximizing continuous functions over  $\mathcal{D}(\mathcal{H}^{\otimes n})$ , specifically those that are convex and unitarily invariant. Of particular interest are unitarily invariant norms. Generically denoted by  $||| \cdot |||$ , these are operator norms that satisfy  $||| \cdot ||| = |||U(\cdot)V|||$  for all unitaries  $U$  and  $V$ . For an operator  $R \in \mathcal{B}(\mathbb{C}^d)$ , its Schatten  $p$ -norm for  $p \in [1, \infty)$  is defined as

$$|||R|||_p := \text{tr}(|R|^p)^{\frac{1}{p}}, \quad (2.10)$$

where  $|R|$  denotes the absolute value  $\sqrt{R^*R}$ . Schatten norms satisfy  $\lim_{p \rightarrow \infty} ||| \cdot |||_p = ||| \cdot |||$ , where  $||| \cdot |||$  is the operator norm. Both Rényi entropies and Tsallis entropies [4, 5]

are monotonic functions of Schatten  $p$ -norms for  $p \in (1, \infty)$ . For  $R \in \mathcal{B}(\mathbb{C}^d)$ ,  $k \in [d]$ , the Fan norm of order  $k$  is the sum of the  $k$  largest singular values of  $R$ :

$$\|R\|_{(k)} := \sum_{i=1}^k \sigma_i(R). \quad (2.11)$$

For positive semi-definite  $R$ ,  $\|R\|_{(k)}$  is the sum of its  $k$  largest eigenvalues. Both Schatten norms and Fan norms are unitarily invariant norms. For any  $A, B \in \mathcal{B}(\mathbb{C}^d)$ ,  $\|A\|_{(k)} \leq \|B\|_{(k)}$  for all  $k \in [d]$  implies  $\|A\| \leq \|B\|$  for every unitarily invariant norm  $\|\cdot\|$ . This is Fan's dominance theorem (see Theorem IV.2.2 on page 93 in Ref. [10]).

For readability, we index factors in tensor products with subsets, each denoting the domain of the corresponding factor. For example, given  $O \in \mathcal{B}(\mathcal{H})$  and  $J \subseteq [n]$ ,  $O^{\otimes J}$  denotes the fully factorizable operator acting on  $|J|$  specific tensor copies of  $\mathcal{H}$ —specified to be corresponding to  $J$ —with each of its factors is equal to  $O$ . In general, we do not write the curly brackets of sets in subscripts.

### 3. Spin Alignment Problems and the Alignment Ordering

**Definition 3.1.** (*Spin alignment problem*). Let a local dimension  $d$ , a number of qudits  $n$ , a probability measure  $\mu$ , and a state  $Q = \sum_{i=1}^d \lambda_i(Q) |q_i\rangle\langle q_i| \in \mathcal{D}(\mathbb{C}^d)$  be given. If  $f$  is a continuous, unitarily invariant, convex function whose domain includes  $\mathcal{D}((\mathbb{C}^d)^{\otimes n})$ , then the spin alignment problem associated with  $(d, n, \mu, Q, f)$  is to maximize the objective function  $f(\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c})$  over all state tuples  $(\rho_I)_{I \subseteq [n]}$ . The operator argument

$$\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} \quad (3.1)$$

is called an *alignment operator*.

Since the objective function is continuous and the set of alignment operators is compact, an optimal point exists. For concave functions, such as the von Neumann entropy, the associated spin alignment problem is a minimization problem. The orthonormal basis  $(|q_i\rangle)_{i=1}^d$  and bases arising from taking tensor powers of it are called *computational* bases. Operators that are diagonal in these bases are called *classical*. We refer to assertions that the state tuple  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$  is optimal as spin alignment conjectures. We mostly focus on the dependence of the problems on the state  $Q$  and the objective function  $f$  and implicitly take the remaining parameters to be arbitrary. For example, when we speak of the spin alignment problem associated with a state  $Q$ , we mean the collection of all spin alignment problems with state  $Q$ .

We conjecture that the spectrum of an alignment operator with state tuple  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$  majorizes the spectra of alignment operators arising with other state tuples for the same  $Q$  and  $\mu$ . That is, it has the largest possible value of the order- $k$  Fan norm for each  $k \in [d^n]$ . By the doubly-stochastic processing characterization of majorization, if this conjecture is true, then every spin alignment conjecture is true. We call this the *strong* spin alignment conjecture. In cases where  $Q$  is pure, the conjecture follows directly from Eq. 2.9. So, we focus on the cases where  $Q$  is mixed.



Spin alignment problems can be extended to cases where the operators  $Q$  are not the same in each term in the sum in Eq. 3.1. For example, spin alignment conjectures can be considered in the more general case where the alignment operators are of the form

$$\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q_{I^c}, \quad (3.2)$$

where for  $I^c \subseteq [n]$ ,  $Q_{I^c} = \bigotimes_{j \in I^c} Q_{I^c, j}$  such that for each  $j \in [n]$ , the operators in the tuple  $(Q_{I^c, j})_{I^c \subseteq [n]}$  are perfectly aligned. In this work, we consider only alignment operators as in Eq. 3.1, though many of our results apply for these more general alignment operators.

**Lemma 3.2.** *For any spin alignment problem, it suffices to restrict to state tuples of pure states.*

*Proof.* Notice that the alignment operator is a linear function of the state tuple, so convex combinations of state tuples correspond to convex combinations of alignment operators. By convexity of the objective function, there exists an extremal point that is optimal. The extremal points of state tuples are pure-state tuples.  $\square$

**Remark 3.3.** If the spin alignment conjecture for a function  $g$  is true, then it also holds that  $f = h \circ g$  attains its maximum over alignment operators at  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$  whenever  $h$  is monotonically increasing. If  $h$  is monotonically decreasing, then  $f$  attains a minimum there. Examples of this include Schatten norms and the corresponding Rényi entropies.

In addition to showing additivity of coherent information for platypus channels, proving the strong spin alignment conjecture would formalize intuitions we have about how dispersion behaves under mixing. Namely, to minimize dispersion in a mixture, the optimal procedure ought to be one where each signal has the lowest dispersion possible and the signals collectively overlap as much as possible. This can be considered a refinement of Fan's majorization relation Eq. 1.1 because in the spin alignment problem, while each summand has minimum dispersion by a pure state choice, we are not always free to choose the bases so that they are perfectly aligned.

When operators cannot be chosen to be perfectly aligned to achieve minimum dispersion, we need a way of comparing the alignment of different tuples of operators. The next definition provides such a way while restricting the tuples to have matching spectra.

**Definition 3.4 (Alignment).** Let  $\mathcal{T} = (T_1, T_2, \dots, T_\ell)$  and  $\mathcal{R} = (R_1, R_2, \dots, R_\ell)$  be two tuples of self-adjoint operators such that for each  $i \in [\ell]$ ,  $\lambda(T_i) = \lambda(R_i)$ . Then,  $\mathcal{T}$  is *more aligned* than  $\mathcal{R}$  if for all probability measures  $(p_i)_{i=1}^\ell$ , the following majorization relation holds

$$\sum_{i=1}^{\ell} p_i T_i \succeq \sum_{i=1}^{\ell} p_i R_i. \quad (3.3)$$

Alignment can be used to construct a preorder on any set of  $\ell$ -tuples of operators with matching spectra. We are interested in cases where there exists a maximal element according to this ordering. The basic example is one where the set in question contains a tuple of perfectly aligned operators. The strong spin alignment conjecture may be phrased as follows.



**Conjecture 3.5.** *The tuple  $(|q_1\rangle\langle q_1|^{\otimes I} \otimes Q^{\otimes I^c})_{I \subseteq [n]}$  is more aligned than every state tuple of the form  $(|\psi_I\rangle\langle \psi_I|_I \otimes Q^{\otimes I^c})_{I \subseteq [n]}$ .*

A key reason to consider alignment as opposed to perfect alignment is that the latter does not behave well with respect to tensor multiplication. Specifically, the coordinate-wise tensor product of two perfectly aligned tuples is not necessarily perfectly aligned. Even worse, the perfect alignment of  $A$  and  $B$  does not imply the perfect alignment of  $C \otimes A$  and  $C \otimes B$  for self-adjoint  $A$ ,  $B$ , and  $C$ .

*Example 3.6.* Consider  $D = |q_1\rangle\langle q_1| + \frac{1}{2}|q_2\rangle\langle q_2| + \frac{1}{3}|q_3\rangle\langle q_3|$ . It is clear that  $D$  and  $|q_1\rangle\langle q_1|$  are perfectly aligned. However,  $D \otimes D$  and  $D \otimes |q_1\rangle\langle q_1|$  are not perfectly aligned.

In contrast,  $(C \otimes A^\downarrow, C \otimes B^\downarrow)$  is more aligned than  $(C \otimes A, C \otimes B)$ . This is a simple consequence of Eq. 2.9 and the following lemma, which appears in [18]. We include a proof of the lemma for completeness.

**Lemma 3.7.** *Let  $A, B \in \mathcal{S}(\mathcal{H})$  and  $C \in \mathcal{S}(\mathcal{K})$ . If  $A \succeq B$ , then  $C \otimes A \succeq C \otimes B$ .*

*Proof.* This follows from the doubly stochastic characterization of majorization.  $A \succeq B$  is equivalent to the existence of a mixed-unitary quantum channel  $\mathcal{N}$  such that  $B = \mathcal{N}(A)$ . It is clear that  $\text{id} \otimes \mathcal{N}$  is mixed-unitary and  $\text{id} \otimes \mathcal{N}(C \otimes A) = C \otimes B$ .  $\square$

We note that the converse of this lemma does not hold. Aside from trivial counterexamples where  $C = 0$ , others arise in the context of catalyst-enabled quantum pure-state interconversion with LOCC [19].

If we restrict the factors on the left to be positive semi-definite, we can prove the following statement.

**Proposition 3.8.** *Let  $C_1, \dots, C_\ell \in \mathcal{P}(\mathcal{K})$  be simultaneously diagonalizable. For every  $A_1, \dots, A_\ell \in \mathcal{S}(\mathcal{H})$ ,  $(C_i \otimes A_i^\downarrow)_{i=1}^\ell$  is more aligned than  $(C_i \otimes A_i)_{i=1}^\ell$ .*

*Proof.* Let  $p = (p_i)_{i=1}^\ell$  be an arbitrary probability measure. Let  $\{|j\rangle\}_{j=1}^{d_{\mathcal{K}}}$  be an orthonormal basis for  $\mathcal{K}$  where  $C_1, \dots, C_\ell$  are simultaneously diagonal. For each  $i \in [\ell]$ , let  $\sum_{j=1}^{d_{\mathcal{K}}} c_{j|i} |j\rangle\langle j|$  be the corresponding spectral decomposition of  $C_i$ . Observe that

$$\sum_{i=1}^\ell p_i C_i \otimes A_i^\downarrow = \sum_{i=1}^\ell p_i \sum_{j=1}^{d_{\mathcal{K}}} c_{j|i} |j\rangle\langle j| \otimes A_i^\downarrow = \sum_{j=1}^{d_{\mathcal{K}}} |j\rangle\langle j| \otimes \left( \sum_{i=1}^\ell p_i c_{j|i} A_i^\downarrow \right). \quad (3.4)$$

For each  $j \in [d_{\mathcal{K}}]$ , we know from Eq. 2.9 that  $\sum_{i=1}^\ell p_i c_{j|i} A_i^\downarrow \succeq \sum_{i=1}^\ell p_i c_{j|i} A_i$ . Hence, there exists a mixed-unitary channel  $\mathcal{N}_j$  that takes the left-hand side to the right-hand side. Let  $\mathcal{N}$  be the global channel that conditional on  $j$  applies  $\mathcal{N}_j$  for each  $j \in [d_{\mathcal{K}}]$ . It is a mixed-unitary quantum channel and it takes  $\sum_{i=1}^\ell p_i C_i \otimes A_i^\downarrow$  to  $\sum_{i=1}^\ell p_i C_i \otimes A_i$ .  $\square$

The strong spin alignment conjecture subsumes a conjecture about maximizing unitarily invariant norms at the output of a tensor product of depolarizing channels. For  $q \in [\frac{-1}{d^2-1}, 1]$ , the qudit depolarizing channel  $\Delta_{d,q}$  has the action:

$$\Delta_{d,q}(\cdot) := q \text{id}(\cdot) + (1-q) \frac{\mathbb{1}}{d} \text{tr}(\cdot). \quad (3.5)$$

In [20], to show additivity of the Holevo information of depolarizing channels, King proves that for any  $p \in [1, \infty)$ , the Schatten  $p$ -norm of the output of a tensor product of depolarizing channels is maximized at a pure product input. It is reasonable to extrapolate and suspect that this holds for all unitarily invariant norms. The basic idea is that any entanglement in the input manifests as excess dispersion at the output. For  $i \in [n]$ , let  $q_i \in [0, 1]$  and notice that

$$\bigotimes_{i=1}^n \Delta_{d, q_i}(\rho_{[n]}) = \sum_{I \subseteq [n]} v_I \rho_I \otimes \left(\frac{\mathbb{1}}{d}\right)^{\otimes I^c}, \quad (3.6)$$

where  $v_I = \prod_{i \in I} q_i \prod_{j \in I^c} (1 - q_j)$  and  $\rho_I$  denotes the marginal state on  $I$  of  $\rho_{[n]}$ . This is an alignment operator with  $Q = \frac{\mathbb{1}}{d}$  arising from a state tuple satisfying an instance of the quantum marginal problem [21]. Since the conjectured optimal point  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$  satisfies an instance of the quantum marginal problem, it follows that a proof of the strong spin alignment conjecture for  $Q = \frac{\mathbb{1}}{d}$  implies a resolution for this problem. This problem is analogous to the now-solved problem of maximizing unitarily invariant output norms for single-mode Gaussian channels. Specifically, it was shown in [22] that coherent states are optimal.

## 4. Results

**4.1. First observations.** First, we show that the spin alignment conjecture is true for the case where the objective function is  $\lambda_1$ .

**Proposition 4.1.**  $\lambda_1(\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c})$  is maximized at  $(\rho_I)_{I \subseteq [n]} = (|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ .

*Proof.* This follows from Eq. 2.8. Specifically,

$$\lambda_1\left(\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c}\right) \leq \sum_{I \subseteq [n]} \mu_I \lambda_1(\rho_I \otimes Q^{\otimes I^c}) \leq \sum_{I \subseteq [n]} \mu_I \lambda_1(Q^{\otimes I^c}). \quad (4.1)$$

At  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ , both inequalities are satisfied with equality.  $\square$

This proves one of  $d^n - 1$  conditions in Eq. 2.1 necessary to prove the strong spin alignment conjecture. The difficulty in proving the remaining ones lies in the fact that summands of alignment operators with pure-state tuples cannot be made perfectly aligned in general.

As explained in the previous section, the strong spin alignment conjecture holds for pure  $Q$ . Next, for  $Q$  of rank 2 or more, we prove a reduction to cases where  $\lambda_1(Q) = \lambda_2(Q)$ .

**Proposition 4.2.** Suppose  $Q = \sum_{i=1}^d \lambda_i(Q) |q_i\rangle\langle q_i| \in \mathcal{D}(\mathcal{H})$  has rank  $\geq 2$ . Define

$$\tilde{Q} = \frac{1}{1 - (\lambda_1(Q) - \lambda_2(Q))} \left( \lambda_2(Q) (|q_1\rangle\langle q_1| + |q_2\rangle\langle q_2|) + \sum_{i=3}^d \lambda_i(Q) |q_i\rangle\langle q_i| \right). \quad (4.2)$$

If for all  $\tilde{\mu}$ ,  $f(\sum_{I \subseteq [n]} \tilde{\mu}_I \rho_I \otimes \tilde{Q}^{\otimes I^c})$  is maximized at  $(\rho_I)_{I \subseteq [n]} = (|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ , then for all  $\mu$ ,  $f(\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c})$  is maximized there as well.

*Proof.* Denote  $\varepsilon := \lambda_1(Q) - \lambda_2(Q)$ . Observe that  $Q = (1 - \varepsilon)\tilde{Q} + \varepsilon|q_1\rangle\langle q_1|$ . That is,  $Q$  lies on the line segment connecting  $\tilde{Q}$  and  $|q_1\rangle\langle q_1|$ . For any probability measure  $\mu$ , we can express

$$\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} = \sum_{K \subseteq [n]} \tilde{\mu}_K \tilde{\rho}_K \otimes \tilde{Q}^{\otimes K^c}, \quad (4.3)$$

where each factor of  $Q$  on the left-hand side was replaced by  $(1 - \varepsilon)\tilde{Q} + \varepsilon|q_1\rangle\langle q_1|$  and  $\mu_I \rho_I \otimes \varepsilon^{|J|}(1 - \varepsilon)^{|(I \cup J)^c|}|q_1\rangle\langle q_1|^{\otimes J}$  for  $I \cup J = K$  were summed to obtain  $\tilde{\mu}_K \tilde{\rho}_K$ . By assumption,

$$f\left(\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c}\right) = f\left(\sum_{K \subseteq [n]} \tilde{\mu}_K \tilde{\rho}_K \otimes \tilde{Q}^{\otimes K^c}\right) \quad (4.4)$$

$$\begin{aligned} &\leq f\left(\sum_{K \subseteq [n]} \tilde{\mu}_K |q_1\rangle\langle q_1|^{\otimes K} \otimes \tilde{Q}^{\otimes K^c}\right) \\ &= f\left(\sum_{I \subseteq [n]} \mu_I |q_1\rangle\langle q_1|^{\otimes I} \otimes Q^{\otimes I^c}\right), \end{aligned} \quad (4.5)$$

where we reversed the process used to obtain the right-hand side of Eq. 4.3 for  $(\rho_I)_{I \subseteq [n]} = (|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$  to arrive at the last equality.  $\square$

*Remark 4.3.* For  $d = 2$ , this reduction implies that  $Q = \frac{1}{2}$  may be assumed without loss of generality. One can therefore apply the argument in Sect. VI.B of [9] to show that the strong spin alignment conjecture holds for  $n = d = 2$ .

The last result of this subsection relates to problem instances where the union of the elements of the support of the probability measure  $\mu$  is not  $[n]$ . That is, those instances where  $\mu$  is such that there exists nonempty  $K^c \subseteq [n]$  such that  $\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} = (\sum_{J \subseteq [n] \setminus K^c} \mu_J \rho_J \otimes Q^{\otimes J^c}) \otimes Q^{\otimes K^c}$ . If the objective function is multiplicative under tensor multiplication, such as Schatten norms [23], then we may without loss of generality trace over the qudits in  $K^c$  and consider the optimization problem on the rest. Not all Fan norms are multiplicative under tensor multiplication. However, to prove the strong spin alignment conjecture, it may be assumed without loss of generality that the union of the support of  $\mu$  is  $[n]$ .

**Proposition 4.4.** *For every  $n_1, n_2 \in \mathbb{N}$ , if the strong spin alignment conjecture holds for  $\sum_{I \subseteq [n_1]} \mu_I \rho_I \otimes Q^{\otimes I^c}$ , then it holds for  $\sum_{I \subseteq [n_1]} \mu_I \rho_I \otimes Q^{\otimes I^c} \otimes Q^{\otimes n_2}$ .*

*Proof.* This is a direct consequence of Lemma 3.7.  $\square$

**4.2. Classical spin alignment.** We prove a classical version of the strong spin alignment conjecture. That is, we prove it with the extra assumption that  $(\rho_I)_{I \subseteq [n]}$  is such that  $\rho_I$  is diagonal in the computational basis for each  $I \subseteq [n]$ .

**Proposition 4.5.** *For every probability measure  $\mu$  and tuple of classical states  $(\rho_I)_{I \subseteq [n]}$ ,*

$$\sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} \preceq \sum_{I \subseteq [n]} \mu_I |q_1\rangle\langle q_1|^{\otimes I} \otimes Q^{\otimes I^c}, \quad (4.6)$$

*Proof.* By the same argument as in Lemma 3.2, without loss of generality, the state tuple  $(\rho_I)_{I \subseteq [n]}$  may be taken to consist of pure classical states. For  $I \subseteq [n]$ , we denote  $\rho_I = |t_I\rangle\langle t_I|$ , where  $t_I$  is a string of length  $|I|$  over the alphabet  $[d]$ . For example,  $|123\rangle\langle 123| := |q_1\rangle\langle q_1| \otimes |q_2\rangle\langle q_2| \otimes |q_3\rangle\langle q_3|$ .

The idea of the proof is to start with an arbitrary classical assignment  $(t_I)_{I \subseteq [n]}$  and flip to 1 sequentially to improve the assignment making use of the transitivity of the majorization ordering. We start by considering the first qudit. We write the alignment operator in the form

$$\sum_{I \subseteq [n]} \mu_I |t_I\rangle\langle t_I| \otimes Q^{\otimes I^c} = Q \otimes B + |1\rangle\langle 1| \otimes C + \sum_{j=2}^d |j\rangle\langle j| \otimes D_j, \quad (4.7)$$

where  $B$ ,  $C$ , and  $D_j$  can be expressed as

$$B = \sum_{I \subseteq [n] \setminus \{1\}} \nu_I |t'_I\rangle\langle t'_I| \otimes Q^{I^c}, \quad (4.8)$$

$$C = \sum_{I \subseteq [n] \setminus \{1\}} w_I |t''_I\rangle\langle t''_I| \otimes Q^{I^c}, \quad (4.9)$$

$$D_j = \sum_{I \subseteq [n] \setminus \{1\}} \kappa_{j,I} |t_{j,I}\rangle\langle t_{j,I}| \otimes Q^{I^c}, \quad (4.10)$$

where the summand  $Q \otimes B$  accounts for all  $t_I$  with  $1 \notin I$ ,  $|1\rangle\langle 1| \otimes C$  accounts for all  $t_I$  with  $1 \in I$  and first letter 1 (the “good”  $t_I$ ) and the  $D_j$  account for the rest. The measures  $\nu$ ,  $w$ , and  $\kappa_j$  are positive, and the operators  $B$ ,  $C$ , and  $D_j$  are positive semi-definite and classical. We consider each  $j \in [d] \setminus \{1\}$  in sequence and flip the first entries of those  $t_I$  contributing to  $D_j$  from  $j$  to 1. This implements the following change in the classical assignment  $(t_I)_{I \subseteq [n]}$ : For every  $t_I$  of the form  $t_I = |j\rangle\langle j| \otimes |t_{j,I}\rangle\langle t_{j,I}|$  contributing to the summand  $|j\rangle\langle j| \otimes D_j$ , replace it with  $|1\rangle\langle 1| \otimes |t_{j,I}\rangle\langle t_{j,I}|$ . The representation in Eq. 4.7 changes according to the assignments  $D_1 \leftarrow D_1 + D_j$  and  $D_j \leftarrow 0$ , so that the new alignment operator is compatible with  $Q$  and  $\mu$ . For a string  $r$  of length  $n-1$  over  $[d]$ , let  $b_r$ ,  $c_r$ , and  $d_{j,r}$  be the eigenvalues for the eigenstate  $|r\rangle$  of  $B$ ,  $C$ , and  $D_j$  before the change. The change in the spectrum can be implemented by composing transfers of the form:  $d_{j,r}$  is taken from the eigenvalue of  $|j\rangle \otimes |r\rangle$  and given to the eigenvalue of  $|1\rangle \otimes |r\rangle$  for each string  $r$ . These transfers may be implemented in parallel. For a string  $r$ , those two entries in the spectrum before the change are

$$(q_1 b_r + c_r, q_j b_r + d_{j,r}), \quad (4.11)$$

while after, they are

$$(q_1 b_r + c_r + d_{j,r}, q_j b_r). \quad (4.12)$$

To see that the spectrum after the transfer majorizes the one before it, observe that

$$q_1 b_r + c_r + d_{j,r} \geq \max(q_1 b_r + c_r, q_j b_r + d_{j,r}). \quad (4.13)$$

That is, the larger value of the pair did not decrease after the transfer. Hence, the transfer is either unjust or its composition with a transposition is unjust. By transitivity of majorization, the same process may now be done to the remaining qudits.  $\square$

**4.3. Schatten norms of integer order.** We prove the spin alignment conjecture for Schatten norms of integer order. By extension, this implies that any monotonic function of these norms, such as the corresponding Rényi entropies, is optimized at the conjectured optimal state tuple  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ . The proof relies on the fact that overlaps between summands in an alignment operator are simultaneously maximized at the conjectured optimal state tuple. Here, overlap means the absolute value of the trace of any product of these summands. In fact, the overlap lemma Lemma A.1 in App. A establishes the stronger statement that every unitarily invariant norm of such a product is maximized at  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ .

**Corollary 4.6.** *Let  $I_1, \dots, I_\ell$  be a family of subsets of  $[n]$ . For any tuple  $(R_{I_i})_{i=1}^\ell$  of operators each with trace norm at most 1, the following inequality holds*

$$\left| \text{tr} \left( \prod_{i=1}^\ell R_{I_i} \otimes Q^{\otimes I_i^c} \right) \right| \leq \text{tr} \left( \prod_{i=1}^\ell |q_1\rangle\langle q_1|^{\otimes I_i} \otimes Q^{\otimes I_i^c} \right). \quad (4.14)$$

*Proof.* The result follows from an application of Lemma A.1 in the following chain of inequalities:

$$\begin{aligned} \left| \text{tr} \left( \prod_{i=1}^\ell R_{I_i} \otimes Q^{\otimes I_i^c} \right) \right| &\leq \left| \text{tr} \left( \prod_{i=1}^\ell R_{I_i} \otimes Q^{\otimes I_i^c} \right) \right| \\ &\leq \left| \text{tr} \left( \prod_{i=1}^\ell |q_1\rangle\langle q_1|^{\otimes I_i} \otimes Q^{\otimes I_i^c} \right) \right| \quad \text{by Lemma A.1} \\ &= \text{tr} \left( \prod_{i=1}^\ell |q_1\rangle\langle q_1|^{\otimes I_i} \otimes Q^{\otimes I_i^c} \right). \end{aligned} \quad (4.15)$$

□

**Theorem 4.7.** *For arbitrary  $m \in \mathbb{N}$ , the following inequality holds*

$$\left\| \sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} \right\|_m \leq \left\| \sum_{I \subseteq [n]} \mu_I |q_1\rangle\langle q_1|^{\otimes I} \otimes Q^{\otimes I^c} \right\|_m \quad (4.16)$$

*Proof.* Notice that  $\left\| \sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} \right\|_m^m$  is a linear combination of traces of products of operators of the form  $\rho_I \otimes Q^{\otimes I^c}$  for some  $I \subseteq [n]$ . By Corollary 4.6, the absolute value of each term is maximized at  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ . By the triangle inequality for the absolute value and the fact that the terms in the combination have equal phase at  $(|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ ,

$$\left\| \sum_{I \subseteq [n]} \mu_I \rho_I \otimes Q^{\otimes I^c} \right\|_m^m \leq \left\| \sum_{I \subseteq [n]} \mu_I |q_1\rangle\langle q_1|^{\otimes I} \otimes Q^{\otimes I^c} \right\|_m^m. \quad (4.17)$$

The claim follows from the fact that  $m$ th root function is monotonically non-decreasing on  $[0, \infty)$ . □

We note here that generalizing this result to Schatten norms of arbitrary order would imply the original spin alignment conjecture as stated in [9]. However, even that would not be sufficient for proving Conjecture 3.5 as can be seen from the equivalence of catalytic majorization and an infinite set of inequalities for (generalized) Rényi entropies [24].

**4.4. The case where  $\text{supp}(\mu) \leq 2$  and  $Q$  is proportional to a projector.** We now prove the strong spin alignment conjecture for alignment operators of the form

$$(1-t)|\psi_1\rangle\langle\psi_1|_{I_1} \otimes Q^{\otimes I_1^c} + t|\psi_2\rangle\langle\psi_2|_{I_2} \otimes Q^{\otimes I_2^c}, \quad (4.18)$$

where  $t \in [0, 1]$ ,  $I_1, I_2 \subseteq [n]$ , and  $Q$  is proportional to a projector. Since we have established that we may take  $I_1 \cup I_2 = [n]$  without loss of generality (see Lemma 4.4), the cases where  $I_1 \subseteq I_2$  or  $I_1 \supseteq I_2$  are simply resolved by Eq. 2.9. To address the other cases, we consider the more general problem of the alignment of two projectors subject to a certain overlap constraint.

Let  $\mathcal{K}$  be given. For each tuple  $(r_1, r_2, c)$ , where  $r_1, r_2 \in [d_{\mathcal{K}}] \cup \{0\}$  and  $c \geq 0$ , let  $G_{r_1, r_2, c}$  denote the set of pairs of projectors  $(P_1, P_2)$  satisfying the conditions:  $\text{rank}(P_1) = r_1$ ,  $\text{rank}(P_2) = r_2$ , and  $\text{tr}(|P_1 P_2|) \leq c$ .

**Lemma 4.8.**  $G_{r_1, r_2, c}$  is nonempty if and only if  $r_1 + r_2 - d_{\mathcal{K}} \leq \lfloor c \rfloor$ .

*Proof.* Suppose  $r_1 + r_2 - d_{\mathcal{K}} > \lfloor c \rfloor$ . If  $P_1$  and  $P_2$  are projectors satisfying  $\text{rank}(P_1) = r_1$  and  $\text{rank}(P_2) = r_2$ , then from basic linear algebra

$$\dim(\text{supp}(P_1) \cap \text{supp}(P_2)) \geq r_1 + r_2 - d_{\mathcal{K}} > \lfloor c \rfloor. \quad (4.19)$$

Hence,  $P_1 P_2$  acts as the identity on a subspace of dimension at least  $\lfloor c \rfloor + 1$  and so  $\text{tr}(|P_1 P_2|) > c$ . This implies  $G_{r_1, r_2, c}$  is empty.

If on the other hand  $r_1 + r_2 - d_{\mathcal{K}} \leq \lfloor c \rfloor$ , let  $P$  be a projector of rank equal to  $\max(r_1 + r_2 - d_{\mathcal{K}}, 0)$ . Consider the two projectors  $P_1 = P + R_1$  and  $P_2 = P + R_2$ , where  $R_1, R_2$  are projectors satisfying  $R_1 \perp P$ ,  $R_2 \perp P$ ,  $R_1 \perp R_2$ , and  $\text{rank}(R_1) = r_1 - \text{rank}(P)$ ,  $\text{rank}(R_2) = r_2 - \text{rank}(P)$ . Then  $P_1 P_2 = P$ ,  $\text{tr}(|P_1 P_2|) = \text{tr}(P) \leq \lfloor c \rfloor$ , and so  $(P_1, P_2) \in G_{r_1, r_2, c}$ .  $\square$

From now on, we take for granted that  $r_1 + r_2 - d_{\mathcal{K}} \leq \lfloor c \rfloor$ . Moreover, since  $\text{tr}(|P_1 P_2|) \leq \min(\text{tr}(P_1), \text{tr}(P_2))$  for any two projectors  $P_1$  and  $P_2$ , we take  $c \leq \min(r_1, r_2)$ . Otherwise, the constraint on the product may as well not be considered.

The crucial ingredient of our analysis is a characterization of the angle between two subspaces due to Camille Jordan [25]. This characterization, commonly referred to as Jordan's lemma, allows us to reason about the relationship between the eigenvalues of a sum of two projectors and the singular values of their product.

**Theorem 4.9.** *There exists  $(P_1^{\text{opt}}, P_2^{\text{opt}}) \in G_{r_1, r_2, c}$  that is maximal in the alignment ordering on  $G_{r_1, r_2, c}$ , and it satisfies  $\text{tr}(|P_1^{\text{opt}} P_2^{\text{opt}}|) = c$ . Moreover, if  $c \in \mathbb{N}$ , then  $[P_1^{\text{opt}}, P_2^{\text{opt}}] = 0$ .*

*Proof.* It suffices to show that there exists  $(P_1^{\text{opt}}, P_2^{\text{opt}}) \in G_{r_1, r_2, c}$  such that

$$s_1 P_1^{\text{opt}} + s_2 P_2^{\text{opt}} \geq s_1 P_1 + s_2 P_2 \quad (4.20)$$

holds for all  $s_1, s_2 \geq 0$  and  $(P_1, P_2) \in G_{r_1, r_2, c}$ . We show existence by explicit construction. Let arbitrary  $s_1, s_2 \geq 0$  be given and consider an arbitrary pair  $(P_1, P_2) \in G_{r_1, r_2, c}$ . We modify the pair such that it remains in  $G_{r_1, r_2, c}$  and the corresponding Fan norms do not decrease until we reach an optimal pair.

We can decompose  $\mathcal{K}$  into minimal subspaces invariant under both  $P_1$  and  $P_2$ . According to Jordan's lemma, these subspaces are either one- or two-dimensional. Thus, we may write  $\mathcal{K} = \mathcal{K}^{(1)} \oplus \mathcal{K}^{(2)}$ , where  $\mathcal{K}^{(1)} = \bigoplus_{i=1}^{m_1} \mathcal{K}_{i_1}^{(1)}$  is the direct sum of invariant subspaces

of dimension 1, and  $\mathcal{K}^{(2)} = \bigoplus_{i_2=1}^{m_2} \mathcal{K}_{i_2}^{(2)}$  is the direct sum of invariant subspaces of dimension 2. Moreover, for each  $i_2 \in [m_2]$ ,  $P_1|_{\mathcal{K}_{i_2}}$  and  $P_2|_{\mathcal{K}_{i_2}}$  are projectors of rank 1. Denote the overlap  $\text{tr}(|P_1 P_2|) = \tilde{c}_1 + \tilde{c}_2$  where  $\tilde{c}_1 = \text{tr}(|P_1 P_2|_{\mathcal{K}^{(1)}})$  and  $\tilde{c}_2 = \text{tr}(|P_1 P_2|_{\mathcal{K}^{(2)}})$ . Observe that  $\tilde{c}_1$  is an integer as it equals the dimension of  $\text{supp}(P_1|_{\mathcal{K}^{(1)}}) \cap \text{supp}(P_2|_{\mathcal{K}^{(1)}})$ .

First, we reduce to the case where  $m_2 \in \{0, 1\}$ . For each  $i_2 \in [m_2]$ , we notate

$$P_1|_{\mathcal{K}_{i_2}^{(2)}} =: |\alpha_{i_2}\rangle\langle\alpha_{i_2}|, P_2|_{\mathcal{K}_{i_2}^{(2)}} =: |\beta_{i_2}\rangle\langle\beta_{i_2}|. \quad (4.21)$$

Notice that  $\tilde{c}_2 = \sum_{i_2=1}^{m_2} |\langle\alpha_{i_2}|\beta_{i_2}\rangle|$ . By Lemma B.2, the vector  $(|\langle\alpha_{i_2}|\beta_{i_2}\rangle|)_{i_2=1}^{m_2}$  is majorized by the vector  $(\underbrace{1, \dots, 1}_{\lceil \tilde{c}_2 \rceil \text{ times}}, \tilde{c}_2 - \lceil \tilde{c}_2 \rceil, 0, \dots, 0)$ . For each  $i_2 \in [m_2]$ , there are

$|\alpha'_{i_2}\rangle, |\beta'_{i_2}\rangle \in \mathcal{K}_{i_2}^{(2)}$ , such that

$$(|\langle\alpha'_{i_2}|\beta'_{i_2}\rangle|)_{i_2=1}^{m_2} = (\underbrace{1, \dots, 1}_{\lceil \tilde{c}_2 \rceil \text{ times}}, \tilde{c}_2 - \lceil \tilde{c}_2 \rceil, 0, \dots, 0). \quad (4.22)$$

Specifically, for each 1 on the right-hand side we choose the two vectors to be equal and for each 0 we choose them to be orthogonal. For the remaining entry, we pick the two such that they have overlap  $\tilde{c}_2 - \lceil \tilde{c}_2 \rceil$ . By Lemma B.1, the eigenvalues of  $(s_1 P_1 + s_2 P_2)|_{\mathcal{K}^{(2)}}$  are a strictly isotone function of the nonzero singular values of  $P_1 P_2|_{\mathcal{K}^{(2)}}$ , and so we may make the reassignments

$$|\alpha_{i_2}\rangle \leftarrow |\alpha'_{i_2}\rangle, |\beta_{i_2}\rangle \leftarrow |\beta'_{i_2}\rangle \quad (4.23)$$

for each  $i_2 \in [m_2]$ . Save for at most one, this replaces each  $\mathcal{K}_{i_2}^{(2)}$  with two one-dimensional invariant subspaces. That is, we may assume without loss of generality that  $m_2 \in \{0, 1\}$  and that  $\tilde{c}_2 < 1$ .

Second, we argue that  $\tilde{c}_1$  may be taken to be at least  $\lceil c \rceil - 1$  without loss of generality. Suppose that  $\tilde{c}_1 < \lceil c \rceil - 1$ . Recall that the tuple of eigenvalues  $\lambda((s_1 P_1 + s_2 P_2)|_{\mathcal{K}^{(1)}})$  contains  $s_1 + s_2$  with multiplicity  $\tilde{c}_1$  along with the smaller eigenvalues  $s_1, s_2$  and 0. The number of one-dimensional invariant subspaces with eigenvalue  $s_1$  is at least  $(r_1 - 1) - \tilde{c}_1$ . Similarly, The number of one-dimensional invariant subspaces with eigenvalue  $s_2$  is at least  $(r_2 - 1) - \tilde{c}_1$ . The reason 1 is subtracted is to allow for the case where  $m_2 = 1$ . Since  $\lceil c \rceil \leq \min(r_1, r_2)$ , we may pair up  $(\lceil c \rceil - 1) - \tilde{c}_1$  many of these two kinds of invariant one-dimensional subspaces to get  $(\lceil c \rceil - 1) - \tilde{c}_1$  two-dimensional subspaces. We adjust the projectors so that they are equal when restricted to these subspaces. For each such subspace, this causes a transfer on the spectrum of  $s_1 P_1 + s_2 P_2$  of the form:

$$s_1 \leftarrow 0, s_2 \leftarrow s_1 + s_2. \quad (4.24)$$

Since this is either an unjust transfer or an unjust transfer composed with a transposition, the Fan norms of  $s_1 P_1 + s_2 P_2$  do not decrease afterwards. Hence, we may, without loss of generality, take  $\tilde{c}_1 \geq \lceil c \rceil - 1$ .

Now, suppose  $c$  is an integer. If  $m_2 = 1$ , then we may adjust the action of the projectors on the sole invariant two-dimensional subspace so that they are equal when restricted to it. Then,  $\tilde{c}_1$  becomes equal to  $c$ . If  $m_2 = 0$ , then either  $\tilde{c}_1 = c$  or there exists two invariant one-dimensional subspaces where the two projectors are nonzero and orthogonal. In the latter case, we may adjust the actions of the projectors on the direct sum of these two subspaces so that the two are equal there. In this case also  $\tilde{c}_1$



becomes equal to  $c$ . In both cases, the arising pair  $(P_1^{\text{opt}}, P_2^{\text{opt}})$  is a maximal pair in the alignment preorder. Notice that  $\text{tr}(|P_1^{\text{opt}} P_2^{\text{opt}}|) = c$  and  $[P_1^{\text{opt}}, P_2^{\text{opt}}] = 0$ .

On the other hand, consider the case where  $c$  is not an integer. Then,  $\lceil c \rceil - 1 = \lfloor c \rfloor$ . If  $m_2 = 1$ , then we adjust the overlap in the two-dimensional invariant subspace until  $\tilde{c}_2 = c - \lfloor c \rfloor$ . That is, we make the projectors align there as much as possible. From Eq. B.3, this is either an unjust transfer or an unjust transfer composed with a transposition on the spectrum of  $s_1 P_1 + s_2 P_2$ . If  $m_2 = 0$ , then consider again two invariant one-dimensional subspaces where the two projectors are nonzero and orthogonal. Such a pair exists because  $\tilde{c}_1 = \lfloor c \rfloor < c \leq \min(r_1, r_2)$ . Then, we adjust the action of the two projectors on the direct sum of these two subspaces so that  $m_2 = 1$  and  $\tilde{c}_2 = c - \lfloor c \rfloor$ . Again, the pair arising after these adjustments  $(P_1^{\text{opt}}, P_2^{\text{opt}})$  is a desired maximal pair.  $\square$

By Thm. 4.9, if  $c \in \mathbb{N}$ , then any pair of commuting projectors in  $G_{r_1, r_2, c}$  whose product has rank  $c$  is optimal. By Lemma A.1, the projectors in the alignment operator in Eq. 4.18 have overlap at most  $(\text{rank}(Q))^{(|I_1 \cup I_2|)^c}$ .

**Corollary 4.10.** *For  $t \in [0, 1]$ ,  $I_1, I_2 \subseteq [n]$  and  $Q$  proportional to a projector,*

$$(1-t)|q_1\rangle\langle q_1|^{\otimes I_1} \otimes Q^{\otimes I_1^c} + t|q_1\rangle\langle q_1|^{\otimes I_2} \otimes Q^{\otimes I_2^c} \geq (1-t)|\psi_1\rangle\langle\psi_1|_{I_1} \otimes Q^{\otimes I_1^c} \\ + t|\psi_2\rangle\langle\psi_2|_{I_2} \otimes Q^{\otimes I_2^c}.$$

## 5. "Dual" Problems to Spin Alignment and Their Generalizations

We introduce a general class of conjectures that articulates the intuition that under a fixed global spectrum constraint, the least locally dispersed states are classical. The inspiration came from considering spin alignment problems with  $Q \propto \mathbb{1}$  and using Fan's maximum principle Eq. 2.8. Specifically, for a positive measure  $\nu$  and a state tuple  $(|\psi_I\rangle\langle\psi_I|)_{I \subseteq [n]}$ , we consider

$$\max_{P: \text{rank}(P)=k} \text{tr}(P_{[n]} \sum_{I \subseteq [n]} \nu_I |\psi_I\rangle\langle\psi_I| \otimes \mathbb{1}^{\otimes I^c}) = \max_{P: \text{rank}(P)=k} \sum_{I \subseteq [n]} \nu_I \text{tr}(P_I |\psi_I\rangle\langle\psi_I|) \quad (5.1)$$

$$\leq \max_{P: \text{rank}(P)=k} \sum_{I \subseteq [n]} \nu_I \lambda_1(P_I). \quad (5.2)$$

When  $(|\psi_I\rangle\langle\psi_I|)_{I \subseteq [n]} = (|q_1\rangle\langle q_1|^{\otimes I})_{I \subseteq [n]}$ , the inequality is satisfied with equality. This maximization problem may be generalized by letting the optimization be over quantum states with a prescribed global spectrum, and by replacing  $\lambda_1$  with any continuous, convex, unitarily invariant function. Moreover, since the objective function is convex, we may take the convex closure of the feasible set without loss of generality. We formalize this below.

Let complex inner product spaces  $\mathcal{H}_1, \dots, \mathcal{H}_n$  be given and denote  $\mathcal{H}_I = \bigotimes_{i \in I} \mathcal{H}_i$  for  $I \subseteq [n]$ . For a probability measure  $p \in \mathbb{R}^{\prod_{i \in [n]} d_{\mathcal{H}_i}}$ , we consider the maximization problem:

$$\max_{\tau_{[n]} \in \mathcal{D}(\mathcal{H}_{[n]})} \sum_{I \subseteq [n]} f_I(\tau_I), \quad (5.3)$$

$$\text{subject to: } \lambda(\tau_{[n]}) \leq p, \quad (5.4)$$

where for all  $I \subseteq [n]$ ,  $f_I$  is continuous, convex and unitarily invariant on a domain containing  $\mathcal{D}(\mathcal{H}_I)$ . We conjecture that there exists a classical optimal point. This is self-evident in the case where  $p$  has only one nonzero value, as all the functions in the sum can be maximized simultaneously. More interesting are the cases where the optimization is over strictly mixed states. An example of such a problem in the quantum information literature appears in Ref. [26]. Namely, Eq. 39 there gives an upper bound on the fidelity of entanglement generating pure codes in terms of a sum as above where for each  $I \subseteq [n]$ ,  $f_I(\cdot) \propto \|\cdot\|_\alpha^\alpha$  for a fixed  $\alpha \in [1, 2]$ .

Below, we show that this “dual” conjecture holds in the special case where non-constant functions in  $(f_I)_{I \subseteq [n]}$  correspond to members of a partition of  $[n]$ .

**Proposition 5.1.** *Let  $\mathcal{W} \subseteq 2^{[n]}$  be a partition of  $[n]$ . The problem*

$$\max_{\tau_{[n]} \in \mathcal{D}(\mathcal{H}_{[n]})} \sum_{I \in \mathcal{W}} f_I(\tau_I), \quad (5.5)$$

$$\text{subject to: } \lambda(\tau_{[n]}) \leq p. \quad (5.6)$$

*has a classical solution.*

*Proof.* Without loss of generality, let  $\tau_{[n]} \in \mathcal{D}(\mathcal{H}_{[n]})$  have spectrum  $p$ . For each  $I \in \mathcal{W}$ , if necessary, a unitary quantum channel  $\mathcal{U}_I \otimes \text{id}_{I^c}$  may be applied so that  $\tau_I$  is classical. This does not affect the objective function. For each  $I \in \mathcal{W}$ , let  $\mathcal{N}_{\tau, I}$  denote the quantum channel that fully decoheres in a basis of  $\tau_I$ . That is, it has the pinching action  $\mathcal{N}_{\tau, I}(\cdot) = \text{diag}(\cdot)$ . It is clear that  $\tau'_{[n]} = \bigotimes_{I \in \mathcal{W}} \mathcal{N}_{\tau, I}(\tau_{[n]})$  is classical, and for each  $I \in \mathcal{W}$ ,  $\tau'_I = \tau_I$ . Since pinching is mixed-unitary (see chapter 4 of Ref. [27]), then  $\lambda(\tau'_{[n]}) \leq p$ . We may optimize further by noticing that  $\tau'_{[n]}$  is a convex combination of classical states each with spectrum equal to  $p$ . One of them has a higher objective value.  $\square$

## 6. Concluding Remarks

We generalized and systematically studied spin alignment problems. We gave non-trivial examples of cases where tuples of spectrally constrained self-adjoint operators are more aligned than others. These can be considered refinements of Ky Fan’s relation Eq. 1.1. We conclude with a non-exhaustive list of problems in this area that we hope will help guide future research.

1. (*Schatten norms of non-integer order*) We used Corollary 4.6 to resolve the spin alignment problem for Schatten norms of integer order. It is of interest to know if a similar approach can be used in cases where the order of the norm is not an integer. Suppose that  $A_0, A_1, B_0, B_1 \geq 0$  satisfy  $\lambda(A_0) = \lambda(B_0)$ ,  $\lambda(A_1) = \lambda(B_1)$ . For a string  $s \in \{0, 1\}^*$ , define the ordered products  $\Pi_s(A_0, A_1) := \Pi_{i \in [|s|]} A_{s(i)}$  and  $\Pi_s(B_0, B_1) := \Pi_{i \in [|s|]} B_{s(i)}$ . If for all strings  $s$ ,  $\text{tr}(\Pi_s(A_0, A_1)) \geq |\text{tr}(\Pi_s(B_0, B_1))|$ , does it hold that  $\|A_0 + A_1\|_p \geq \|B_0 + B_1\|_p$  for all  $p \in [1, \infty)$ ? If so, then the overlap lemma may be used to prove the spin alignment conjecture for all Schatten norms, and by extension for the von Neumann entropy as well.
2. (*Algebraic approximations of Fan norms*) The Schatten norms of integer order may be said to be algebraic norms, in that they can be defined using polynomials of the absolute value of the operator in question. Since  $\lim_{m \rightarrow \infty} \|\cdot\|_m = \|\cdot\|$ , they approximate the Fan norm of order 1 arbitrarily well. Do there exist algebraic operator norms that approximate other Fan norms? If so, is there any interesting structure to the corresponding polynomials, especially with regard to spin alignment?

3. (*Separable refinement of Ky Fan's relation*) Given  $A_0, A_1, B_0, B_1 \geq 0$ , does the following majorization relation

$$\lambda(A_0 \otimes B_0 + A_1 \otimes B_1) \preceq \lambda(A_0^\downarrow \otimes B_0^\downarrow + A_1^\downarrow \otimes B_1^\downarrow) \quad (6.1)$$

hold? In the special case where  $\text{rank}(B_0) = \text{rank}(B_1) = 1$ , we know the answer to be yes by way of a characterization of separable states due to Nielsen and Kempe [28]. Specifically, they showed that separable states contain more dispersion globally than locally. If the relation Eq. 6.1 holds more generally, then it may be used to prove a separable version of the strong spin alignment conjecture.

**Acknowledgement** The authors thank Eli Halperin, Graeme Smith, and Mark Wilde for fruitful discussions regarding the manuscript, and Felix Leditzky and Scott Glancy for comments that helped improve the presentation of the material. The authors also thank the anonymous referee for their careful reading of the manuscript and useful comments. At the time this work was performed, M. A. Alhejji was supported as an Associate in the Professional Research Experience Program (PREP) operated jointly by NIST and the University of Colorado Boulder. This is a contribution of the National Institute of Standards and Technology, not subject to U.S. copyright.

**Data availability** The manuscript has no associated data.

#### Declarations

**Conflict of interests** We declare no conflict of interests.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Appendix A: The overlap Lemma

**Lemma A.1.** (The overlap lemma) *Let  $I_1, \dots, I_\ell$  be a family of subsets of  $[n]$ . For  $i \in [\ell]$ , let  $Q_{I_i^c} = \bigotimes_{j \in I_i^c} Q_{I_i^c, j}$ , where for each  $j \in I_i^c$ ,  $Q_{I_i^c, j} \in \mathcal{B}(\mathcal{H}_j)$  has largest singular value  $\alpha_{i,j} > 0$  and is of the form  $Q_{I_i^c, j} = \alpha_{i,j} |q_1\rangle\langle q_1| \oplus W_{i,j}$ . For any unitarily invariant norm  $||| \cdot |||$ , the maximization problem*

$$\max_{(R_{I_i})_{i=1}^\ell} ||| \prod_{i=1}^\ell R_{I_i} \otimes Q_{I_i^c} |||, \quad (A.1)$$

where the variable tuple  $(R_{I_i})_{i=1}^\ell$  is of operators each with trace norm at most 1, has  $(|q_1\rangle\langle q_1|^{\otimes I_i})_{i=1}^\ell$  as a solution.

We argue by induction on the number of sets  $\ell$ . For that, we need the following statement.

**Lemma A.2.** *Let  $\mathcal{K}_1, \mathcal{K}_2$ , and  $\mathcal{K}_3$  be given complex inner product spaces. Let  $A_1 \in \mathcal{B}(\mathcal{K}_1)$  and  $A_3 \in \mathcal{B}(\mathcal{K}_3)$  be such that  $\|A_1\| \leq 1$  and  $\|A_3\| \leq 1$ . For all  $T_{2,3} \in \mathcal{B}(\mathcal{K}_2 \otimes \mathcal{K}_3)$  and  $S_{1,2} \in \mathcal{B}(\mathcal{K}_1 \otimes \mathcal{K}_2)$  of trace norm at most 1, it holds that  $\|(A_1 \otimes T_{2,3})(S_{1,2} \otimes A_3)\|_1 \leq 1$ .*

*Proof.* Because  $\|\cdot\|_1$  is homogeneous and convex, it suffices to prove the inequality holds in the cases where  $S_{1,2} = |s\rangle\langle\tilde{s}|_{1,2}$  and  $T_{2,3} = |t\rangle\langle\tilde{t}|_{2,3}$  for arbitrary unit vectors  $|s\rangle, |\tilde{s}\rangle \in \mathcal{K}_1 \otimes \mathcal{K}_2$  and  $|t\rangle, |\tilde{t}\rangle \in \mathcal{K}_2 \otimes \mathcal{K}_3$ . Since  $\|A_1\| \leq 1$  and  $\|A_3\| \leq 1$ , the norms of  $|v\rangle_{1,2} = A_1 \otimes \mathbb{1}_2 |s\rangle_{1,2}$  and  $|u\rangle_{2,3} = \mathbb{1}_2 \otimes A_3^* |\tilde{t}\rangle_{2,3}$  are bounded from above by 1. Hence, it suffices to prove that

$$\|(\mathbb{1}_1 \otimes |t\rangle\langle u|_{2,3})(|v\rangle\langle\tilde{s}|_{1,2} \otimes \mathbb{1}_3)\|_1 \leq 1 \quad (\text{A.2})$$

for arbitrary unit vectors  $|v\rangle, |\tilde{s}\rangle \in \mathcal{K}_1 \otimes \mathcal{K}_2$  and  $|t\rangle, |u\rangle \in \mathcal{K}_2 \otimes \mathcal{K}_3$ .

Consider the inner factor operator  $\langle u|_{2,3}|v\rangle_{1,2} : \mathcal{K}_3 \rightarrow \mathcal{K}_1$ . To see that its trace norm is at most 1, let  $|v\rangle_{1,2} = \sum_i \sqrt{\lambda_i} |\alpha_i\rangle_1 \otimes |\beta_i\rangle_2$  and  $|u\rangle_{2,3} = \sum_j \sqrt{\mu_j} |\gamma_j\rangle_2 \otimes |\delta_j\rangle_3$  be Schmidt decompositions. Then,

$$\langle u|_{2,3}|v\rangle_{1,2} = \sum_{i,j} \sqrt{\mu_j} \sqrt{\lambda_i} \langle \gamma_j | \beta_i \rangle \langle \alpha_i \rangle_1 \langle \delta_j \rangle_3 \quad (\text{A.3})$$

and

$$\|\langle u|_{2,3}|v\rangle_{1,2}\|^2 = \sum_{i,j,j'} \sqrt{\mu_j \mu_{j'}} \lambda_i \langle \gamma_j | \beta_i \rangle \langle \beta_i | \gamma_{j'} \rangle \langle \delta_{j'} \rangle_3 \langle \delta_j \rangle_3. \quad (\text{A.4})$$

Denote  $b_j = \langle \gamma_j | (\sum_i \lambda_i |\beta_i\rangle\langle\beta_i|) | \gamma_j \rangle$  and notice that  $\sum_j b_j \leq 1$ . It follows from the Schur-concavity of  $\text{tr}(\sqrt{\cdot})$  and the Schur-Horn theorem that

$$\|\langle u|_{2,3}|v\rangle_{1,2}\|_1 = \text{tr}(\sqrt{|\langle u|_{2,3}|v\rangle_{1,2}|^2}) \quad (\text{A.5})$$

$$\leq \text{tr}(\sqrt{\text{diag}(|\langle u|_{2,3}|v\rangle_{1,2}|^2)}) \quad (\text{A.6})$$

$$= \sum_j \sqrt{\mu_j b_j} \leq 1. \quad (\text{A.7})$$

Finally,  $\|\langle u|_{2,3}|v\rangle_{1,2} \otimes |t\rangle_{2,3} \langle \tilde{s}|_{1,2}\|_1 = \|\langle u|_{2,3}|v\rangle_{1,2}\|_1 \| |t\rangle_{2,3} \langle \tilde{s}|_{1,2} \|_1 \leq 1$ .  $\square$

*Proof of Lemma A.1.* Since  $\|\cdot\|$  is homogeneous, it may be assumed without loss of generality for each  $i \in [\ell]$  that  $\|Q_{I_i^c, j}\| = 1$  for all  $j \in I_i^c$ . We proceed via induction on  $\ell$ . If  $\ell = 1$ , the statement of the lemma follows from the facts that  $\| |q_1\rangle\langle q_1|^{\otimes I_1} \|_1 = 1$  and that  $\|\cdot\|$  is convex and unitarily invariant. Suppose that the statement holds in cases where the family of sets has  $\ell > 1$  elements. Consider  $\prod_{i=1}^{\ell+1} R_{I_i} \otimes Q_{I_i^c}$  and notice that

$$R_{I_1} \otimes Q_{I_1^c} R_{I_2} \otimes Q_{I_2^c} = \underbrace{Q_{I_1^c \cap I_2} R_{I_1} R_{I_2} Q_{I_2^c \cap I_1}}_{\tilde{R}_{I_1 \cup I_2}} \otimes \underbrace{Q_{I_1^c \setminus I_2} Q_{I_2^c \setminus I_1}}_{\tilde{Q}_{(I_1 \cup I_2)^c}}. \quad (\text{A.8})$$

The operator  $\tilde{Q}_{(I_1 \cup I_2)^c}$  is completely factorizable on  $(I_1 \cup I_2)^c$  and has maximal singular value 1. Moreover, each of its factors can be written as  $|q_1\rangle\langle q_1| \oplus W$  for some operator

W. By Lemma A.2,  $\|\tilde{R}_{I_1 \cup I_2}\|_1 \leq 1$  and so we may estimate

$$\| \prod_{i=1}^{\ell+1} R_{I_i} \otimes Q_{I_i^c} \| = \| \tilde{R}_{I_1 \cup I_2} \otimes \tilde{Q}_{(I_1 \cup I_2)^c} \prod_{i=3}^{\ell+1} R_{I_i} \otimes Q_{I_i^c} \| \quad (\text{A.9})$$

$$\leq \| |q_1\rangle\langle q_1|^{\otimes I_1 \cup I_2} \otimes \tilde{Q}_{(I_1 \cup I_2)^c} \prod_{i=3}^{\ell+1} |q_1\rangle\langle q_1|^{\otimes I_i} \otimes Q_{I_i^c} \| \quad (\text{A.10})$$

$$= \| \prod_{i=1}^{\ell+1} |q_1\rangle\langle q_1|^{\otimes I_i} \otimes Q_{I_i^c} \|, \quad (\text{A.11})$$

where the inequality is by the induction hypothesis.  $\square$

## Appendix B: On the Relationship Between the Sum of Two Projectors and Their Product

In this appendix, we prove lemmas necessary to elucidate the relationship between a non-negative linear combination  $s_1 P_1 + s_2 P_2$  of two projectors  $P_1, P_2 \in \mathcal{S}(\mathcal{K})$  and their product  $P_1 P_2$ . By [25],  $\mathcal{K}$  may be decomposed into a direct sum of subspaces, each of dimension at most 2, that are invariant under the action of both  $P_1$  and  $P_2$ . Moreover, when restricted to each invariant subspace, the two projectors have rank at most 1. So, we may write  $\mathcal{K}$  as an orthogonal direct sum of one- and two-dimensional minimal invariant subspaces

$$\mathcal{K} = \bigoplus_{i_1=1}^{m_1} \mathcal{K}_{i_1}^{(1)} \oplus \bigoplus_{i_2=1}^{m_2} \mathcal{K}_{i_2}^{(2)}, \quad (\text{B.1})$$

where the  $\mathcal{K}_{i_1}^{(1)}$  are one-dimensional and the  $\mathcal{K}_{i_2}^{(2)}$  are two dimensional. By minimal, we mean that the subspaces contain no proper nonzero invariant subspace. For each  $i_2 \in [m_2]$ , we notate

$$P_1|_{\mathcal{K}_{i_2}^{(2)}} =: |\alpha_{i_2}\rangle\langle\alpha_{i_2}|, \quad P_2|_{\mathcal{K}_{i_2}^{(2)}} =: |\beta_{i_2}\rangle\langle\beta_{i_2}|. \quad (\text{B.2})$$

$P_1$  and  $P_2$  commute if and only if  $m_2 = 0$ . The difficulty in reasoning about the eigenvalues of a linear combination of two projectors lies in these subspaces where they do not commute.

When restricted to  $\bigoplus_{i_2=1}^{m_2} \mathcal{K}_{i_2}^{(2)}$ , the nonzero singular values of  $P_1 P_2$  are  $(|\langle\alpha_{i_2}|\beta_{i_2}\rangle|)_{i_2=1}^{m_2}$ . The eigenvalues of the restriction  $(s_1 P_1 + s_2 P_2)|_{\mathcal{K}_{i_2}^{(2)}}$  may be computed as

$$\frac{1}{2}((s_1 + s_2) \pm \sqrt{(s_1 - s_2)^2 + 4s_1 s_2 |\langle\alpha_{i_2}|\beta_{i_2}\rangle|^2}). \quad (\text{B.3})$$

Hence, the eigenvalues of  $s_1 P_1 + s_2 P_2$  are a function of the singular values of the product  $P_1 P_2$ . We show next that it is in fact a *strictly isotone* function (see page 41 of Ref. [10]). A strictly isotone function  $G$  is one that preserves the majorization ordering in the sense that if  $v \succeq w$ , then  $G(v) \succeq G(w)$ . That is, the less dispersed the singular values of  $P_1 P_2$ , the less dispersed the eigenvalues of  $s_1 P_1 + s_2 P_2$ . This is a consequence of the fact that for  $a, b \geq 0$ , the map  $x \mapsto \sqrt{a + bx^2}$  is convex on  $\mathbb{R}_{\geq 0}$ .

**Lemma B.1.** Let  $A \subseteq \mathbb{R}$  be convex and  $g : A \rightarrow \mathbb{R}_{\geq 0}$  be a convex function. For  $t \in \mathbb{R}$ , define the mapping  $G : A^m \rightarrow \mathbb{R}^{2m}$  with action

$$(v_1, \dots, v_m) \mapsto (t + g(v_1), \dots, t + g(v_m)) \oplus (t - g(v_1), \dots, t - g(v_m)). \quad (\text{B.4})$$

Then,  $G$  is strictly isotone.

*Proof.* Let  $v, w \in A^m$  be such that  $v \succeq w$ . Observe that for  $k \in [m]$ ,

$$\sum_{j=1}^k G(v)_j^\downarrow = kt + \sum_{j=1}^k g(v)_j^\downarrow \geq kt + \sum_{j=1}^k g(w)_j^\downarrow = \sum_{j=1}^k G(w)_j^\downarrow, \quad (\text{B.5})$$

where the inequality follows from the convexity of  $g$  and the doubly-stochastic characterization of majorization (see, for example, Theorem II.3.3 on page 41 of Ref. [10]). If  $k > m$ , then

$$\sum_{j=1}^k G(v)_j^\downarrow = kt + \sum_{j=1}^m g(v)_j^\downarrow - \sum_{j=1}^{k-m} g(v)_j^\uparrow = kt + \sum_{j=1}^{2m-k} g(v)_j^\downarrow \quad (\text{B.6})$$

$$\geq kt + \sum_{j=1}^{2m-k} g(w)_j^\downarrow = \sum_{j=1}^k G(w)_j^\downarrow. \quad (\text{B.7})$$

Since  $\sum_{j=1}^{2m} G(\cdot)_j = 2mt$ ,  $G(v) \succeq G(w)$ .  $\square$

The following lemma exhibits the maximal elements in the majorization ordering in a superset of the possible tuples of nonzero singular values of  $P_1 P_2 | \oplus_{i_2=1}^{m_2} \mathcal{K}_{i_2}^{(2)}$ .

**Lemma B.2.** Given  $m \in \mathbb{N}$ ,  $e \geq 0$ , define the set

$$S_e := \{x \in \mathbb{R}^m \mid \forall i \in [m], x_i \in [0, 1], \sum_{i=1}^m x_i = e\}. \quad (\text{B.8})$$

The element  $(\underbrace{1, \dots, 1}_{[e] \text{ times}}, e - [e], 0, \dots, 0)$  is a majorant of  $S_e$ .

*Proof.* Let  $x \in S_e$  be arbitrary. For  $k \in [m]$ ,  $k \leq [e]$ , observe that  $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k 1 = k$ . And for  $k > [e]$ ,  $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^m x_i^\downarrow = e$ .  $\square$

## Appendix C: Proof of the Conjecture in Example 1.2

Let  $\sum_{i=1}^{d-1} \lambda_i(\tau) |\tau_i\rangle \langle \tau_i|$  be a spectral decomposition of  $\tau$ . Let arbitrary  $p \in [0, 1]$  be given and consider the convex mixture  $\tau_\gamma := p\tau + (1-p)|v_\gamma\rangle \langle v_\gamma|$ . We wish to prove that  $\tau_\gamma \preceq \tau_1$ . Define

$$w := (\sqrt{p(1-p)\lambda_1(\tau)} \langle \tau_1 | \alpha \rangle, \dots, \sqrt{p(1-p)\lambda_{d-1}(\tau)} \langle \tau_{d-1} | \alpha \rangle)^T. \quad (\text{C.1})$$

The Gram matrix for the  $d$  vectors  $\sqrt{p\lambda_1}|\tau_1\rangle, \dots, \sqrt{p\lambda_{d-1}}|\tau_{d-1}\rangle, \sqrt{1-p}|v_\gamma\rangle$  is

$$M_\gamma := \left( \begin{array}{ccc|c} p\lambda_1(\tau) & \cdots & 0 & \sqrt{\gamma}w \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & p\lambda_{d-1}(\tau) & \\ \hline & & \sqrt{\gamma}w^* & (1-p) \end{array} \right) \quad (\text{C.2})$$

It is not difficult to show that (see [29] for example), up to zeros, the spectrum of  $M_\gamma$  is equal to the spectrum of  $\tau_\gamma$ . Hence, it suffices to show that  $M_\gamma \leq M_1$  for all  $\gamma \in [0, 1]$ . This statement is proven in the next paragraph.

Let  $|e_1\rangle, \dots, |e_d\rangle$  denote the orthonormal basis (ordered in the obvious way) used to write down the Gram matrices in Eq. C.2. Consider the unitary

$$V := \left( \sum_{i=1}^{d-1} |e_i\rangle\langle e_i| - |e_d\rangle\langle e_d| \right). \quad (\text{C.3})$$

Supposing it occurs with a probability  $q \in [0, 1]$ , we have

$$(1-q)M_1 + qVM_1V^* = \left( \begin{array}{ccc|c} p\lambda_1(\tau) & \cdots & 0 & (1-2q)w \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & p\lambda_{d-1}(\tau) & \\ \hline & & (1-2q)w^* & (1-p) \end{array} \right) \quad (\text{C.4})$$

If  $q = \frac{1-\sqrt{\gamma}}{2}$ , then  $(1-q)M_1 + qVM_1V^* = M_\gamma$ . Hence, for all  $\gamma \in [0, 1]$ , there exists a mixed-unitary channel that takes  $M_1$  to  $M_\gamma$ .

## References

- Shannon, C.E.: A mathematical theory of communication, The. Bell Syst. Tech. J. **27**, 379–423 (1948)
- Schumacher, B.: Quantum coding, Phys. Rev. A **51**, 2738–2747 (1995)
- Chehade, S.S., Vershynina, A.: Quantum entropies. Scholarpedia **14**, 53131 (2019)
- Rényi, A.: On measures of information and entropy. In: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability (Berkeley and Los Angeles, USA), Lecture Notes in Computer Science, vol. I, University of California Press, pp. 547–561 (1960)
- Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. J. Stat. Phys. **52**, 479–487 (1988)
- Shor, P.W.: Equivalence of additivity questions in quantum information theory. Commun. Math. Phys. **246**(3), 453–472 (2004)
- Schumacher, B., Westmoreland, M.D.: Sending classical information via noisy quantum channels. Phys. Rev. A **56**, 131–138 (1997)
- Hastings, M.B.: Superadditivity of communication capacity using entangled inputs. Nat. Phys. **5**(4), 255–257 (2009)
- Leditzky, F., Leung, D., Siddhu, V., Smith, G., Smolin, J.A.: The platypus of the quantum channel zoo. IEEE Trans. Inf. Theory **69**(6), 3825–3849 (2023)
- Bhatia, R.: Matrix Analysis, vol. 169. Springer, Berlin (1997)
- Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: Theory of Majorization and Its Applications, vol. 143, 2nd edn. Springer, Berlin (2011)
- Hardy, G.H., Littlewood, J.E., Pólya, G.: Some simple inequalities satisfied by convex functions. Messenger Math. **58**, 145–152 (1929)
- Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities, Cambridge University Press (1934)
- Nielsen, M.A.: Conditions for a class of entanglement transformations. Phys. Rev. Lett. **83**, 436–439 (1999)
- Fan, Ky.: On a theorem of Weyl concerning eigenvalues of linear transformations i\*. In: Proceedings of the National Academy of Sciences **35**(11), 652–655 (1949)



16. Leditzky, F., Leung, D., Siddhu, V., Smith, G., Smolin, J.A.: Generic nonadditivity of quantum capacity in simple channels. *Phys. Rev. Lett.* **130**, 200801 (2023)
17. Barry, C.: Arnold, inequality and majorization: robin hood in unexpected places. *Calcutta Statist. Assoc. Bull.* **63**(1–4), 71–80 (2011)
18. James, V.: Bondar, Schur majorization inequalities for symmetrized sums with applications to tensor products. *Linear Algebra Appl.* **360**, 1–13 (2003)
19. Jonathan, D., Plenio, M.B.: Entanglement-assisted local manipulation of pure quantum states. *Phys. Rev. Lett.* **83**, 3566–3569 (1999)
20. King, C.: The capacity of the quantum depolarizing channel. *IEEE Trans. Inf. Theory* **49**(1), 221–229 (2003)
21. Schilling, C.: The quantum marginal problem, *Mathematical Results in Quantum Mechanics*, WORLD SCIENTIFIC, November (2014)
22. Mari, A., Giovannetti, V., Holevo, A.S.: Quantum state majorization at the output of bosonic gaussian channels. *Nat. Commun.* **5**(1), 3826 (2014)
23. Aubrun, G., Nechita, I.: The multiplicative property characterizes  $\ell_p$  and  $l_p$  norms. *Confluen. Math.* **03**(04), 637–647 (2011)
24. Klimesh, M.: Inequalities that collectively completely characterize the catalytic majorization relation (2007)
25. Jordan, C.: Essai sur la géométrie à  $n$  dimensions. *Bull. Soc. Math. France* **3**, 103–174 (fr) (1875)
26. Wilde, M.M., Winter, A.: Strong converse for the quantum capacity of the erasure channel for almost all codes. In: 9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2014), Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany, pp. 52–66 (en) (2014)
27. Watrous, J.: *The Theory of Quantum Information*. Cambridge University Press, Cambridge (2018)
28. Nielsen, M.A., Kempe, J.: Separable states are more disordered globally than locally. *Phys. Rev. Lett.* **86**, 5184–5187 (2001)
29. Fannes, M., De Melo, F., Roga, W., Życzkowski, K.: Matrices of fidelities for ensembles of quantum states and the holevo quantity. *Quantum Info. Comput.* **12**(5–6), 472–489 (2012)

Communicated by G. Chiribella