Extrema in majorization order with applications to entropic continuity bounds

Eric P. Hanson

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences
University of Cambridge, Cambridge CB3 0WA, UK

Département de Mathématiques d'Orsay
Univ. Paris-Sud, Orsay 91405, France

January 16, 2018

Majorization is a pre-order of vectors, giving a sense in which one vector can be said to be more disordered than another. Two given vectors, however, may be incomparable. This concept has been extended to quantum mechanical states, to particular success in the theory of entanglement manipulation. Here, we investigate the majorization pre-order over the ε-ball of quantum states defined by the trace distance. We find, suprisingly, that the ε-ball admits a maximal and minimal state in majorization order, which are comparable to every other state in the set. This property depends in particular on the geometry described by the trace distance. We apply these minimal and maximal states to find explicit formulas for "smooothed" single-partite entropies, briefly consider implications for state estimation via the MaxEnt principle, and establish local continuity bounds for a broad class of functions. Next, we investigate further properties of the minimal state in majorization order, and in particular of the map taking the center of the ε-ball to the minimal state in this ball. We find that this map satisfies a so-called *semigroup* property, which allows us to derive continuity bounds which are uniform over the set of states from our local results. The construction of the minimal state is motivated by a more general result we derive via Fermat's rule of convex optimization, providing a first step to extending our results to functions of bipartite states, such as the conditional entropy.

Contributions Section 2, Appendix A, and the first part of Section 4 contain review material. Appendix B reviews the proof of the Audenaert-Fannes bound, as presented by [Win16], but also contains original research: a direct derivation of the necessity of the equality conditions for the bound, as well as the (rather immediate) application of Theorem 3.1 to formulate the local continuity bound. Section 3 summarizes the original contributions of this work, which are proven in Sections 4 to 6 and appendix B.

This research was conducted under the guidance of my PhD supervisor, Nilanjana Datta, to whom I'm indebted. Additionally, the introduction, statements of results, and some of the organization and wording of the proofs were done in collaboration with her, with language taken from our preprints [HD17a] and [HD17b], the first of which has been submitted to the *Journal of Mathematical Physics*. Mark Wilde suggested we study the continuity of the quantum conditional entropy, which led us to the present work. Andreas Winter suggested using our local results for local continuity bounds. We would also like to thank Koenraad Audenaert, Barbara Kraus, Sam Power, Cambyse Rouzé and Stephanie Wehner for helpful discussions. We would particularly like to thank Hamza Fawzi for insightful discussions about convex optimization.

Contents

1. Introduction

The concept of majorization, introduced in the early 20th century by Littlewood, Hardy and Polya [HLP29] (see also [Sch23] and [Hor54]) has turned out to be ubiquitous in a variety of widely diverse disciplines, ranging from various branches of mathematics to economics, to areas of physics such as entanglement theory and quantum thermodynamics. Majorization is a pre-ordering of some set such as the space $\mathbb{R}^n$ of real n-dimensional vectors. It can be interpreted as an ordering of mixedness or disorder. This is because two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ satisfy the majorization relation $\mathbf{x} \prec \mathbf{y}$ (read $\mathbf{x}$ is majorized by $\mathbf{y}$) if and only if $\mathbf{x}$ is in the convex hull of all vectors obtained by permuting the entries of $\mathbf{y}$. Hence, $\mathbf{x}$ is a random mixture of reshufflings of the entries of $\mathbf{y}$, and in this sense is more mixed (or disordered) than $\mathbf{y}$. Note that in general two arbitrary vectors need not be comparable in the sense of majorization, i.e. one could have $\mathbf{x} \not\prec \mathbf{y}$ and $\mathbf{y} \not\prec \mathbf{x}$. There are two important classes of functions which are, respectively, order-preserving and order-reversing in the majorization sense. These are the classes of Schur convex and Schur concave functions. For a Schur convex (resp. Schur concave) function $f(\cdot)$, $\mathbf{x} \prec \mathbf{y} \implies f(\mathbf{x}) \leq f(\mathbf{y})$ (resp. $f(\mathbf{x}) \geq f(\mathbf{y})$). Majorization coupled with Schur convexity (or Schur concavity) yields a plethora of useful mathematical inequalities. The theory of majorization is now recognized as a powerful mathematical tool which has found applications in a number of different fields including linear algebra, geometry, numerical analysis, operational research and quantum information theory [Mar11].

In the context of quantum information theory, we are interested in quantum states on a finite dimensional Hilbert space, namely trace-1 positive semi-definite operators on $\mathbb{C}^d$ for some dimension $d \in \mathbb{N}$. In this context, quantum states can be considered as a noncommutative analog of a probability distribution on a set of d elements. In fact, the eigenvalues of a quantum state are such a distribution. However, physical manipulations of quantum states can change their spectral projections as well as their eigenvalues, yielding the mathematical richness of quantum mechanics.

Majorization has proven to be an important tool in quantum information theory. We say that two quantum states ρ and ρ satisfy the majorization relation $\rho \prec \sigma$ if $\lambda(\rho) \prec \lambda(\sigma)$, where $\lambda(\rho)$ denotes the vector of eigenvalues of ρ. Since the von Neumann entropy, $S(\rho)$, is a Schur concave function, $\rho \prec \sigma$ implies $S(\rho) \geq S(\sigma)$. Moreover, since S increases with the mixedness of the state, the majorization relation $\rho \prec \sigma$ can be viewed as an ordering of mixedness, as in the case of vectors in $\mathbb{R}^n$.

An important class of problems in quantum information theory concerns the determination of optimal rates of information-processing tasks such as storage and transmission of information, and entanglement manipulation. These optimal rates can be viewed as the *operational quantities* in quantum information theory. They include the data compression limit of a quantum information source, the various capacities of a quantum channel, and the entanglement cost and distillable entanglement of a bipartite state. The aim in quantum information theory is to express these operational quantities in terms of suitable *entropic quantities*. Examples of the latter include the von Neumann entropy, Rényi entropies, conditional entropies, coherent information, and mutual information.

Deriving entropic characterizations of optimal rates of information theoretic tasks is one of the central problems of quantum information theory. In order to determine whether a given $R \in \mathbb{R}$ is an achievable rate for a particular task, it is sufficient to require that the final state of the protocol is approximately equal to the desired target state. For example, consider Schumacher compression of a quantum information source with source state ρ. Let $\mathcal{E}_n$ be an encoding map which compresses $\rho^{\otimes n}$ to a state in a Hilbert space of dimension $\sim 2^{nR}$, for some $R \in \mathbb{R}$, and let $\mathcal{D}_n$ be the corresponding decompression map. Then R is an achievable rate if for n large enough, the final state $(\mathcal{D}_n \circ \mathcal{E}_n)(\rho^{\otimes n})$ of the protocol is *approximately equal* to $\rho^{\otimes n}$. The notion of approximate equality is quantified in terms of a suitable metric on the space of quantum states. The trace distance is one such metric. For this choice of the metric, the set of states which are approximately equal to a desired state ρ are those which are at a trace distance of at most ε from it, for some threshold value $\varepsilon \in (0,1)$. These set of states define an ε-ball around ρ in the space of states $\mathcal{D}(\mathcal{H})$ on the Hilbert space $\mathcal{H}$:

$$B_\varepsilon(\rho) = \{\omega \in \mathcal{D}(\mathcal{H}) : \frac{1}{2}\|\omega - \rho\|_1 \leq \varepsilon\}.$$

This ε-ball also arises in the study of continuity properties of entropic functions, like the von Neumann entropy, which play a pivotal role in quantum information theory.

The geometry of the space of states depends on the chosen metric. Here we study the geometry of $B_\varepsilon(\rho)$, and properties of states in it, from the perspective of majorization theory. We prove that there are two states $\rho^*_\varepsilon, \rho_{*,\varepsilon} \in B_\varepsilon(\rho)$, satisfying the majorization relation

$$\rho^*_\varepsilon \prec \omega \prec \rho_{*,\varepsilon}, \qquad \forall\, \omega \in B_\varepsilon(\rho). \tag{1}$$

The state ρ^*_ε is unique as an element of $B_\varepsilon(\rho)$ satisfying (1), while $\rho_{*,\varepsilon}$ is unique up to unitary equivalence. Both states commute with ρ and can be constructed explicitly. These states were also independently found by Horodecki, Oppenheim, and Sparaciari [HOS17], and considered in the context of thermal majorization [Mee16; MNW17]. A priori there is no reason to expect that states in $B_\varepsilon(\rho)$ are comparable in majorization order. Given any two states ρ and σ on a Hilbert space $\mathcal{H}$, one may have both $\rho \not\prec \sigma$ and $\sigma \not\prec \rho$. Hence, one does not know a priori, that there is any state in the trace-ball around ρ which is comparable to every other state, let alone states which are, respectively, maximum and minimum in majorization order, as shown by (1). This is what makes our result novel, and it appears to be a special feature of the geometry of $B_\varepsilon(\rho)$[1].

Tools from the field of convex optimization theory often play a key role in the analysis of the above-mentioned quantities. This is because the relevant operational quantity of an information-processing task is often expressed as a convex optimization problem: one in which an entropic quantity is optimized over a suitable convex set.

The construction of the maximum entropy state can also be motivated by a result from convex optimization theory. Maximizing the entropy over states in $B_\varepsilon(\rho)$ can be transcribed into an optimization problem over all states involving a non-differentiable objective function. To solve the problem we employ the notion of subgradients (a generalization of the concept of a gradient) and Fermat's Rule of convex optimization (see e.g. [BC11]). The latter provides a simple characterization of the minimum of a convex function in terms of the zeroes of its subgradient. In fact, we first derive the following, more general, result which might be of independent interest. We consider a particular class of functions $\mathcal{F}$, and derive a necessary and sufficient condition under which a state in $B_\varepsilon(\rho)$ maximizes any function in this class. The von Neumann entropy, the conditional entropy, and the α-Rényi entropy (for $\alpha \in (0,1)$) are examples of functions in $\mathcal{F}$. The precise mathematical definition of the class $\mathcal{F}$ is given in Section 6.2. For the case of Schur concave functions, these conditions provide strong restrictions on the set of possible states ρ^*_ε which could satisfy (1), which leads to the construction of the maximum entropy state, ρ^*_ε. This state is unique and is shown to satisfy a *semigroup property*: $\rho^*_{\varepsilon_1+\varepsilon_2} = \sigma^*_{\varepsilon_2}$, where $\sigma = \rho^*_{\varepsilon_1}$.

[1] Whether any metric, other than the trace distance, induces ε-balls admitting extrema in the majorization order, is an interesting open question.

We also present several applications of the result (1): explicit formulas for ε-smoothed Schur concave and convex functions, local continuity bounds for such functions, and a universal construction of uniform continuity bounds for a family of entropies which generalizes the Audenaert-Fannes bound. Given the important role majorization plays in quantum and classical information theory, it can be expected that (1) will find further information-theoretic applications in the future.

Given a Schur concave function H on the set of quantum states on a Hilbert space $\mathcal{H}$, we define the following two *ε-smoothed functions*;

$$\overline{H}^{(\varepsilon)}(\rho) = \max_{\rho \in B_\varepsilon(\rho)} H(\rho) : \qquad \underline{H}^{(\varepsilon)}(\rho) = \min_{\rho \in B_\varepsilon(\rho)} H(\rho).$$

Such quantities arise naturally in the study of information-processing tasks in the framework of one-shot information theory [Tom16; Ren05]. The majorization relation (1) yields an explicit formula for these quantities, namely $\overline{H}^{(\varepsilon)}(\rho) = H(\rho_\varepsilon^*)$ and $\underline{H}^{(\varepsilon)}(\rho) = H(\rho_{*,\varepsilon})$.

A uniform continuity bound for a function H on quantum states bounds $|H(\rho) - H(\sigma)|$ in terms of $\varepsilon \geq \frac{1}{2}\|\rho - \sigma\|_1$ alone. In proofs, this type of bound can be useful to remove a dependence on the particular (possibly unknown) states ρ and σ, and reduce to their trace distance alone. An explicit continuity bound for the von Neumann entropy was obtained by Fannes [Fan73] and later strengthened by Audenaert [Aud07]. It is often referred to as the Audenaert-Fannes (AF) inequality and is an upper bound on the difference of the entropies of two states in terms of their trace distance. Alicki and Fannes found an analogous continuity bound for the conditional entropy for a bipartite state [AF04], which was later strengthened by Winter [Win16]. Continuity bounds for Rényi entropies were obtained in [Che+17; Ras11], although neither bound is known to be sharp, and in the appendix of the published version of [Aud07]. These bounds are "uniform" in the sense that they only depend on the trace distance between the states (and a dimensional parameter) and not on any other property of the states in question. In [Win16], Winter also obtained continuity bounds for the entropy and the conditional entropy for infinite-dimensional systems under an energy constraint. His analysis was extended by Shirokov to the quantum conditional information for finite-dimensional systems in [Shi15], and for infinite-dimensional tripartite systems (under an energy constraint) in [Shi16]. In [Shi15], Shirokov also established continuity bounds for the Holevo quantity, and refined the continuity bounds on the classical and quantum capacities obtained by Leung and Smith [LS09]. Continuity bounds have also been obtained for various other entropic quantities (see e.g. [AE05; AE11; RW15] and references therein).

In comparison, a local continuity bound for a function H at a state ρ is a bound on how far H can vary from $H(\rho)$ when evaluated on states in $B_\varepsilon(\rho)$, in terms of both ε and ρ. A local bound therefore yields sharper results by employing the additional knowledge of ρ. This is the natural setup to control an error $|H(\hat{\rho}) - H(\rho)|$ incurred, when a state $\hat{\rho}$ is produced in an attempt to produce a state ρ, in terms of the error in the state preparation, given by the trace distance $\frac{1}{2}\|\hat{\rho} - \rho\|_1$. For any Schur concave function H, the result (1) immediately yields the bound

$$|H(\omega) - H(\rho)| \leq \max\{H(\rho) - \underline{H}^{(\varepsilon)}(\rho), \bar{H}^{(\varepsilon)}(\rho) - H(\rho)\} \tag{2}$$

for any $\omega \in B_\varepsilon(\rho)$. Since the right-hand side of (2) only depends on ρ and ε, the bound (2) provides a tight local continuity bound for any Schur concave function H (including, e.g. the von Neumann entropy).

We extended these local continuity bounds to uniform continuity bounds for a class of entropic functions, by exploiting the semigroup property of the state ρ_ε^* mentioned above. In fact, our result provides a universal construction of sharp uniform continuity bounds for a class of functions, which include the α-Rényi entropies for $\alpha \in (0,1)$, the Tsallis entropies, and the von Neumann entropy. In the case of the von Neumann entropy, the resulting bound is the Audenaert-Fannes bound, providing a unified proof of this inequality along with the bound on the α-Rényi entropy. This proof also provides necessary and sufficient conditions for two states ρ and ρ to achieve equality in the Audenaert-Fannes bound, which are the same conditions as for any other applicable function H.

The paper is organized as follows. We start with some mathematical preliminaries in Section 2. In Section 3 we state our main results (see Theorem 3.1, Theorem 3.2, and Theorem 3.8 in particular). The proof of Theorem 3.1 entails explicit construction of maximum and minimum entropy states in the ε-ball, which are given in Section 4. In Section 5, we prove further properties of the minimal state in majorization order, and applications to uniform continuity bounds. In Section 6 we investigate the more general framework of maximizing concave functions over the ε-ball, and use the tools of convex optimization to prove Theorem 3.8.

2. Mathematical tools and notation

2.1. Quantum states and majorization

We restrict our attention to finite-dimensional Hilbert spaces. For Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$, we denote the set of linear maps from $\mathcal{H}_1$ to $\mathcal{H}_2$ by $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, and write $\mathcal{B}(\mathcal{H}) \equiv \mathcal{B}(\mathcal{H}, \mathcal{H})$ for the algebra of linear operators on a single Hilbert space $\mathcal{H}$. We denote the set of self-adjoint linear operators on $\mathcal{H}$ by $\mathcal{B}_{\mathrm{sa}}(\mathcal{H}) \subset \mathcal{H}$. Upper case indices label quantum systems: for a Hilbert space $\mathcal{H}_A$ corresponding to a quantum system A, we write $|\psi\rangle_A \in \mathcal{H}_A$ and $\rho_A \in \mathcal{B}(\mathcal{H}_A)$, and we use the notation $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. A *quantum state* (or simply state) is a density matrix, i.e. an operator $\rho_A \in \mathcal{B}(\mathcal{H}_A)$ with $\rho_A \geq 0$ and $\operatorname{Tr} \rho_A = 1$. We denote the set of states on $\mathcal{H}_A$ by $\mathcal{D}(\mathcal{H}_A)$. We say a state ρ is *faithful* if $\rho > 0$, and denote by $\mathcal{D}_+(\mathcal{H}_A)$ the set of faithful states on $\mathcal{H}_A$.

For results which only involve a single Hilbert space, $\mathcal{H}$, we set $\mathcal{B}_{\mathrm{sa}} = \mathcal{B}_{\mathrm{sa}}(\mathcal{H})$, $\mathcal{D} = \mathcal{D}(\mathcal{H})$, $\mathcal{D}_+ = \mathcal{D}_+(\mathcal{H})$ and $d := \dim \mathcal{H}$. A pure state is a rank-1 density matrix; we denote the set of pure states by $\mathcal{D}_{\mathrm{pure}}(\mathcal{H})$. Let $\tau := \frac{1}{d}\mathbb{1}$ denote the completely mixed state.

Note that $\mathcal{B}_{\mathrm{sa}}$ is a real vector space of dimension d^2, which is a (real) Hilbert space when equipped with the Hilbert-Schmidt inner product $\mathcal{B}_{\mathrm{sa}} \times \mathcal{B}_{\mathrm{sa}} \ni (A, B) \mapsto \langle A, B\rangle_{\mathrm{HS}} := \operatorname{Tr}(AB)$, which induces the norm $\|A\|_2 = \sqrt{\operatorname{Tr}(A^2)}$ for $A \in \mathcal{B}_{\mathrm{sa}}$. We also consider the trace norm on $\mathcal{B}_{\mathrm{sa}}$, defined by $\|A\|_1 = \operatorname{Tr}|A|$ for $A \in \mathcal{B}_{\mathrm{sa}}$. The trace norm induces the *trace distance* on $\mathcal{D}(\mathcal{H})$, which is defined as

$$T(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$$

for two quantum states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. We also define the fidelity [Joz94] between two quantum states ρ and σ as

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1.$$

We define the ε-ball around $\sigma \in \mathcal{D}(\mathcal{H})$ as the set

$$B_\varepsilon(\sigma) = \{\omega \in \mathcal{D}(\mathcal{H}) : T(\omega, \sigma) \leq \varepsilon\}. \tag{3}$$

We also define the ε-ball of positive-definite states: $B_\varepsilon^+(\sigma) := B_\varepsilon(\sigma) \cap \mathcal{D}_+$. The following inequality, which follows from [Bha97, eq. (IV.62)], will be useful.

Lemma 2.1. *For any $\rho, \sigma \in \mathcal{D}$,*

$$\|\operatorname{Eig}^\downarrow(\rho) - \operatorname{Eig}^\downarrow(\sigma)\|_1 \leq \|\rho - \sigma\|_1 \tag{4}$$

where for $A \in \mathcal{B}_{sa}$, $\mathrm{Eig}^{\downarrow}(A)$ is the diagonal matrix of eigenvalues of A arranged non-increasing order.

For any $A \in \mathcal{B}_{\mathrm{sa}}(\mathcal{H})$, let $\lambda_+(A)$ and $\lambda_-(A)$ denote the maximum and minimum eigenvalue of A, respectively, and $k_+(A)$ and $k_-(A)$ denote their multiplicities. Let $\lambda_j(A)$ denote the jth largest eigenvalue, counting multiplicity; that is, the jth element of the ordering

$$\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_d(A).$$

We set $\vec{\lambda}(A) := (\lambda_i(A))_{i=1}^d \in \mathbb{R}^d$ and denote the set of eigenvalues of $A \in \mathcal{B}_{\mathrm{sa}}(\mathcal{H})$ by $\operatorname{spec} A \subset \mathbb{R}$, and the kernel of A by $\ker A$.

The notion of *majorization* of vectors will be useful as well. Given $x \in \mathbb{R}^d$, write $x^{\downarrow} = (x_j^{\downarrow})_{j=1}^d$ for the permutation of x such that $x_1^{\downarrow} \geq x_2^{\downarrow} \geq \cdots \geq x_d^{\downarrow}$. Given $x, y \in \mathbb{R}^d$, we say x majorizes y, written $x \succ y$, if

$$\sum_{j=1}^k x_j^{\downarrow} \geq \sum_{j=1}^k y_j^{\downarrow} \quad \forall k = 1, \ldots, d-1, \quad \text{and} \quad \sum_{j=1}^d x_j^{\downarrow} = \sum_{j=1}^d y_j^{\downarrow}. \tag{5}$$

If $x \prec y \prec x$, then x and y are equal up to a permutation (see e.g. [Mar11, p. 18]). A function $f : \mathbb{R}^d \to \mathbb{R}^d$ is *Schur concave* if $f(x) \leq f(y)$ whenever $x \succ y$. The function f is *strictly Schur concave* if $f(x) < f(y)$ whenever $x \succ y$ and $x^{\downarrow} \neq y^{\downarrow}$. If $f : \mathbb{R}^d \to \mathbb{R}^d$ is symmetric and concave, then it is Schur concave; similarly if f is symmetric and strictly concave, then it is strictly Schur concave [Mar11, p. 97]. Moreover, any Schur concave or Schur convex function f is symmetric, as the relation $\Pi x \prec x \prec \Pi x$ which holds for any permutation Π yields $f(x) = f(\Pi x)$.

Given two states $\rho, \sigma \in \mathcal{D}$, if $\vec{\lambda}(\rho) \prec \vec{\lambda}(\sigma)$, we write $\rho \prec \sigma$. Note if $\rho \prec \sigma \prec \rho$, then ρ and σ must have the same eigenvalues with the same multiplicities, and therefore must be unitarily equivalent. We say that $\varphi : \mathcal{D} \to \mathbb{R}$ is *Schur-concave* if $\varphi(\rho) \geq \varphi(\sigma)$ for any $\rho, \sigma \in \mathcal{D}$ with $\rho \prec \sigma$. If $\varphi(\rho) > \varphi(\sigma)$ for any $\rho, \sigma \in \mathcal{D}$ such that $\rho \prec \sigma$, and ρ is not unitarily equivalent to σ, then φ is *strictly Schur-concave*. As in the vector case, φ is *Schur convex* (resp. *strictly Schur convex*) if $-\varphi$ is Schur concave (resp. strictly Schur concave).

It can be shown that $\rho \prec \sigma$ if and only if $\rho = \sum_i p_i U_i \sigma U_i^*$ for some $n \in \mathbb{N}$, $p_i \geq 0$, $\sum_i p_i = 1$, and $U_i \in \mathcal{B}(\mathcal{H})$ unitary for $i = 1, \ldots, n$ [AU82, Theorem 2-2]. Hence, if $\varphi : \mathcal{D} \to \mathbb{R}$ is unitarily invariant and concave, and $\rho \prec \sigma$, then

$$\varphi(\rho) \geq \sum_{i \in I_L} p_i \varphi(U_i \sigma U_i^*) = \varphi(\sigma).$$

Hence any unitarily-invariant concave function φ is Schur concave. Moreover, by the same argument, if φ is strictly concave and $\rho \prec \sigma$ is such that ρ is not unitarily equivalent to σ, we have

$$\varphi(\rho) > \varphi(\sigma) \tag{6}$$

and φ is strictly Schur concave. Moreover, if φ is Schur concave or Schur convex, then it is unitarily invariant: $\varphi(U \rho U^*) = \varphi(\rho)$ for any unitary U. This follows from the relation $\rho \prec U \rho U^* \prec \rho$.

2.2. Entropies

Entropies play a fundamental role in quantum information theory as characterizations of the optimal rates of information theoretic tasks, and as measures of uncertainty. The mathematical properties of entropic functions therefore have important physical implications. The von Neumann entropy S, for instance, as a function of d-dimensional quantum states, is strictly concave, continuous, and is bounded by $\log d$. As the von Neumann entropy characterizes the optimal rate of data compression for a memoryless quantum information source [Sch96], continuity of the von Neumann entropy, for example, implies that the quantum data compression limit is continuous in the source state. The α-Rényi entropies S_α are parametrized by $\alpha \in (0,1) \cup (1,\infty)$, and are a generalization of the von Neumann entropy in the sense that $\lim_{\alpha\to 1} S_\alpha = S$. The α-Rényi entropy has been used to bound the quantum communication complexity of distributed information-theoretic tasks [vH02], can be interpreted in terms of the free energy of a quantum or classical system [Bae11], and is the fundamental quantity defining the entanglement α-Rényi entropy [Wan+16].

In fact, the α-Rényi entropies are members of a large family of entropies called the (h,ϕ)-entropies, which are parametrized by two functions h, ϕ on $\mathbb{R}$ subject to certain constraints (see Section 2). This family includes the Tsallis entropies [Tsa88] and the unified entropies (considered by Rastegin in [Ras11]). Note that the (h,ϕ)-entropy of a quantum state is the classical (h,ϕ)-entropy of its eigenvalues, and therefore the results here apply equally well to probability distributions on finite sets.

Let $h : \mathbb{R} \to \mathbb{R}$ and $\phi : [0,1] \to \mathbb{R}$ with $\phi(0) = 0$ and $h(\phi(1)) = 0$, such that either h is strictly increasing and ϕ strictly concave, or h strictly decreasing and ϕ strictly convex. Then the (h,ϕ)-entropy, $H_{(h,\phi)}$, is defined by

$$H_{(h,\phi)}(\rho) := h(\operatorname{Tr}[\phi(\rho)]) \tag{7}$$

where ϕ is defined on $\mathcal{D}(\mathcal{H})$ by functional calculus, i.e. given the eigen-decomposition $\rho = \sum_i \lambda_i(\rho)\pi_i$, we have $\phi(\rho) = \sum_i \phi(\lambda_i(\rho))\pi_i$. Every (h,ϕ)-entropy is strictly Schur concave and unitarily invariant; moreover, if h is concave, then $H_{(h,\phi)}$ is concave [Bos+16]. Here, we are most interested in the following three examples of (h,ϕ) entropies:

- The von Neumann entropy
$$S(\rho) = -\operatorname{Tr}(\rho \log \rho).$$
S is the (h,ϕ) entropy with $h = \mathrm{id}$, i.e., $h(x) = x$ for $x \in \mathbb{R}$, and with $\phi(x) = -x\log x$ for $x \in [0,1]$.

  An important property of the von Neumann entropy, of particular relevance to us, is its continuity. It is described by the so-called Audenaert-Fannes (AF) bound which is stated in the following lemma [Aud07] (see also Theorem 3.8 of [Pet08]).

  Lemma 2.2 (Audenaert-Fannes bound). *Let $\varepsilon \in [0,1]$, and $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $\frac{1}{2}\|\rho - \sigma\|_1 \le \varepsilon$, and let $d = \dim \mathcal{H}$. Then*
$$|S(\rho) - S(\sigma)| \le \begin{cases} \varepsilon \log(d-1) + h(\varepsilon) & \text{if } \varepsilon < 1 - \frac{1}{d} \\ \log d & \text{if } \varepsilon \ge 1 - \frac{1}{d}, \end{cases} \tag{8}$$
  where $h(\varepsilon) := -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ denotes the binary entropy.

  Without loss of generality, assume $S(\rho) \ge S(\sigma)$. Then equality in (8) occurs if σ is a pure state, and either

  1. *$\varepsilon < 1 - \frac{1}{d}$ and $\vec{\lambda}(\rho) = (1-\varepsilon, \frac{\varepsilon}{d-1}, \ldots \frac{\varepsilon}{d-1})$, or*
  2. *$\varepsilon \ge 1 - \frac{1}{d}$ and $\rho = \tau := \frac{1}{d}\mathbb{1}$.*

  We recall a proof of Lemma 2.2 in Appendix B.

- The q-Tsallis entropy for $q \in (0,1) \cup (1,\infty)$,

$$T_q(\rho) = \frac{1}{1-q}[\mathrm{Tr}(\rho^q) - 1].$$

  T_q can be written as the (h,ϕ)-entropy with $h(x) = x - \frac{1}{1-q}$ for $x \in \mathbb{R}$ and $\phi(x) = \frac{1}{1-q}x^q$. With these choices, h is strictly increasing and affine (and therefore concave) and ϕ is strictly concave, for all $q \in (0,1) \cup (1,\infty)$.

- The α-Rényi entropy for $\alpha \in (0,1) \cup (1,\infty)$,

$$S_\alpha(\rho) = \frac{1}{1-\alpha}\log\left(\mathrm{Tr}\,\rho^\alpha\right).$$

  S_α is the (h,ϕ)-entropy with $h(x) = \frac{1}{1-\alpha}\log x$ for $x \in \mathbb{R}$ and $\phi(x) = x^\alpha$ for $x \in [0,1]$. For $\alpha \in (0,1)$, h is concave and strictly increasing and ϕ is strictly concave. For $\alpha > 1$, h is convex and strictly decreasing, and ϕ is strictly convex. It is known that $\lim_{\alpha\to 1} S_\alpha(\rho) = S(\rho)$.

We will also also consider the min- and max-entropy introduced by Renner [Ren05] which play an important role in one-shot information theory (see e.g. [Tom16] and references therein):

$$H_{\min}(\rho) := -\log\lambda_+(\rho), \qquad H_{\max}(\rho) := S_{1/2}(\rho) = 2\log\mathrm{Tr}[\sqrt{\rho}].$$

Note $\lambda_+(\rho) = \|\rho\|_\infty$ is unitarily invariant and convex, and therefore Schur convex: if $\rho \prec \sigma$, then $\lambda_+(\rho) \leq \lambda_+(\sigma)$. Since $x \mapsto -\log(x)$ is decreasing, $H_{\min}(\rho) \geq H_{\min}(\sigma)$; therefore, $H_{\min}$ is Schur concave. As a Rényi entropy, $H_{\max}$ is Schur concave as well.

We also can define the *conditional entropy* of a bipartite state. Given a bipartite state $\rho_{AB} \in \mathcal{H}_{AB}$, the conditional entropy $S(A|B)_\rho$ is given by

$$S(A|B)_\rho = S(\rho_{AB}) - S(\rho_B), \tag{9}$$

where $\rho_B = \mathrm{Tr}_A\,\rho_{AB}$ denotes the reduced state on the system B. This quantity cannot be written as an (h,ϕ)-entropy.

The Shannon entropy of a classical random variable X (taking values in a discrete alphabet $\mathcal{X}$) with probability mass function (p.m.f.) $(p(x))_{x\in\mathcal{X}}$ is given by $H(X) := -\sum_{x\in\mathcal{X}} p(x)\log p(x)$. The joint- and conditional entropies of two random variables X and Y, with joint probability distribution $(p(x,y))_{x,y}$, are respectively given by

$$H(X,Y) := -\sum_{x,y} p(x,y)\log p(x,y)$$
$$H(X|Y) := H(X,Y) - H(X).$$

Fano's inequality ([FH61, p. 187]) with equality conditions ([Han07, p. 41]), stated as Lemma 2.3 below, provides an upper bound on $H(X|Y)$ and will be employed in our proofs.

Lemma 2.3 (Fano's inequality)**.** *For two random variables X, Y which take values on $\{1,\dots,d\}$, we have, for $\varepsilon := \Pr(X \neq Y)$,*

$$H(X|Y) \leq \varepsilon\log(d-1) + h(\varepsilon),$$

with equality if and only if

$$\Pr(X = i|Y = j) = \begin{cases} 1-\varepsilon & \text{if } i = j \\ \frac{\varepsilon}{d-1} & \text{if } i \neq j \end{cases}$$

for each j such that $\Pr(Y = j) \neq 0$.

In the above, all logarithms are taken to base 2.

3. Main results

Fix $\varepsilon \in (0,1]$ and a state $\rho \in \mathcal{D}(\mathcal{H})$, with $d := \dim \mathcal{H}$. Our first result is that the ε-ball around ρ, the set $B_\varepsilon(\rho)$ defined by (3), admits a minimum and maximum in the majorization order. Note that since majorization is a partial order, a priori one does not know that there are states in $B_\varepsilon(\rho)$ comparable to every other state in $B_\varepsilon(\rho)$.

Theorem 3.1. *Let $\rho \in \mathcal{D}$ and $\varepsilon \in (0,1]$. Then there exists two states ρ_ε^* and $\rho_{*,\varepsilon}$ in $B_\varepsilon(\rho)$ (which are defined by Equation (32) in Section 4.1 and Equation (39) in Section 4.3 respectively) both of which commute with ρ and satisfy*

$$\rho_\varepsilon^* \prec \omega \prec \rho_{*,\varepsilon}, \qquad \forall \omega \in B_\varepsilon(\rho). \tag{10}$$

Moreover, ρ_ε^* is the unique state in $B_\varepsilon(\rho)$ satisfying the left-hand relation of (10), and $\rho_{*,\varepsilon}$ is the unique state in $B_\varepsilon(\rho)$ satisfying the right-hand relation of (10) up to unitary equivalence.

Remark. See Figure 1 for an example of ρ_ε^* and $\rho_{*,\varepsilon}$ for a particular state ρ.

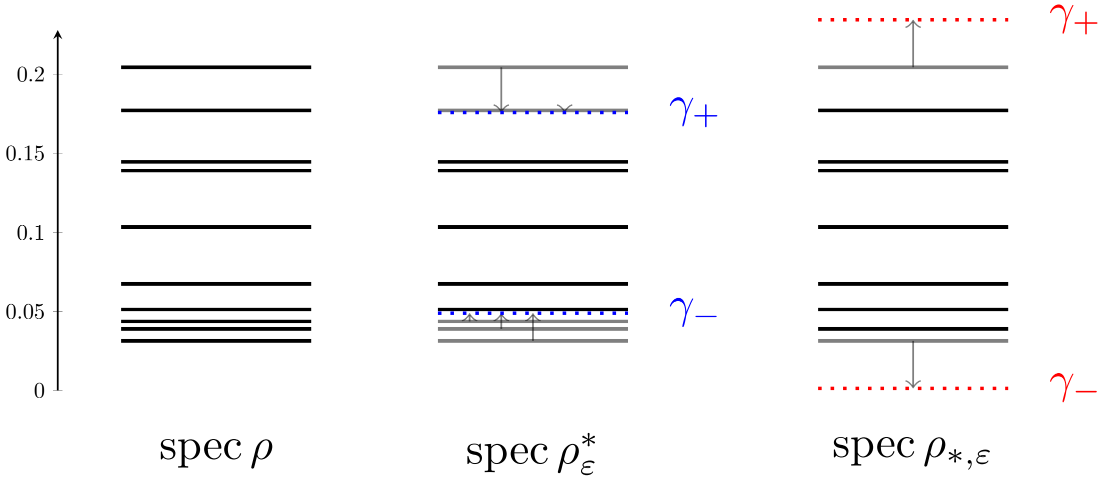


Figure 1: An example. The spectrum of a quantum state ρ is on the left, the spectra of ρ_ε^* in the center and $\rho_{*,\varepsilon}$ on the right, with $\varepsilon = 0.03$. The states are constructed in Section 4.1 and Section 4.3 respectively, and have the structure of raising or lowering some eigenvalues of ρ according to a prescribed formula.

Our characterization of the minimal state ρ_ε^* originates from the following theorem, which is a condensed form of Theorem 6.3 in Section 6. Given a suitable function $\varphi : \mathcal{D}(\mathcal{H}) \to \mathbb{R}$, and any state $\rho \in \mathcal{D}(\mathcal{H})$, the theorem provides a necessary and sufficient condition under which a state maximizes φ in the ε-ball (of positive definite states), $B_\varepsilon^+(\rho)$, of the state ρ.

Theorem 3.2. *Let $\rho \in \mathcal{D}(\mathcal{H})$, $\varepsilon \in (0,1]$, and $\varphi : \mathcal{D}(\mathcal{H}) \to \mathbb{R}$ be a concave, continuous function which is Gâteaux-differentiable*[2] *on $\mathcal{D}_+(\mathcal{H})$. A state $\xi \in B_\varepsilon^+(\rho)$, satisfies $\xi \in \operatorname{argmax}_{B_\varepsilon(\rho)} \varphi$ if and only if* both *of the following conditions are satisfied. Here $L_\xi := \nabla\varphi(\xi)$ denotes the Gâteaux-gradient of φ at ξ.*

1. *Either $\frac{1}{2}\|\xi - \rho\|_1 = \varepsilon$ or $L_\xi = \lambda \mathbb{1}$ for some $\lambda \in \mathbb{R}$, and*

[2]For the definition of Gateaux-differentiability and Gateaux gradient see Section 6.1.

2. *we have*

$$\pi_\pm L_\xi \pi_\pm = \lambda_\pm(L_\xi)\pi_\pm, \tag{11}$$

where $\pi_\pm$ is the projection onto the support of $\Delta_\pm$, and where $\Delta = \Delta_+ - \Delta_-$ is the Jordan decomposition of $\Delta := \xi - \rho$.

A corollary of this theorem concerns the conditional entropy $S(A|B)_\xi$ of a bipartite state ξ_{AB}. It corresponds to the choice $\varphi(\xi_{AB}) = S(A|B)_\xi$ whose Gâteaux gradient $L_\xi := \nabla\varphi(\xi_{AB})$ is given by $L_{\xi_{AB}} = -(\log\xi_{AB} - \mathbb{1}_A \otimes \log\xi_B)$ (see Corollary 6.9).

Corollary 3.3. *Given a state $\rho_{AB} \in \mathcal{D}_+(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\varepsilon \in (0,1]$, a state $\xi_{AB} \in B^+_\varepsilon(\rho_{AB})$ has maximum conditional entropy if and only if*

$$S(A|B)_\xi - S(A|B)_\rho + D(\rho_{AB}\|\xi_{AB}) - D(\rho_B\|\xi_B) = \varepsilon(\lambda_+(L_\xi) - \lambda_-(L_\xi))$$

where $L_\xi = \mathbb{1}_A \otimes \log\xi_B - \log\xi_{AB}$.

Remark. Interestingly, the states ρ^*_ε and $\rho_{*,\varepsilon}$ were derived independently by Horodecki and Oppenheim [OH17], and in the context of thermal majorization by van der Meer and Wehner [Mee16]. They referred to these as the flattest and steepest states. Horodecki and Oppenheim also used these states to obtain expressions for smoothed Schur concave functions.

3.1. Immediate applications of Theorem 3.1

Theorem 3.1 immediately yields maximizers and minimizers over $B_\varepsilon(\rho)$ for any Schur concave function φ. Note that, as stated in the following corollary, the minimizer ρ^*_ε (resp. the maximizer $\rho_{*,\varepsilon}$) in the majorization order (10) is the maximizer (resp. minimizer) of φ over the ε-ball $B_\varepsilon(\rho)$.

Corollary 3.4. *Let $\varphi : \mathcal{D} \to \mathbb{R}$ be Schur concave. Then*

$$\rho^*_\varepsilon \in \operatorname{argmax}_{B_\varepsilon(\rho)} \varphi, \qquad \textit{and} \qquad \rho_{*,\varepsilon} \in \operatorname{argmin}_{B_\varepsilon(\rho)} \varphi.$$

Additionally, if φ is strictly Schur concave, any other state $\rho' \in \operatorname{argmax}_{B_\varepsilon(\rho)} \varphi$ (resp. $\rho' \in \operatorname{argmin}_{B_\varepsilon(\rho)} \varphi$) is unitarily equivalent to ρ^*_ε (resp. $\rho_{*,\varepsilon}$). If φ is strictly concave, then $\operatorname{argmax}_{B_\varepsilon(\rho)} \varphi = \{\rho^*_\varepsilon\}$.

In an experimental setup, one may not know a quantum state exactly, and need to determine an estimate of the state for mathematical analysis. Let us briefly investigate Corollary 3.4's implications for this task. The so-called MaxEnt (or maximum-entropy) principle gives that an appropriate estimate of σ is one which is compatible with the constraints on σ and which has maximum entropy subject to those constraints [Buž+99; Jay57].

Consider an experimental device which attemps to produce a given target pure state σ_{target}, and let σ denote the actual state produced by the device. One can estimate the fidelity between σ and σ_{target} efficiently, using few Pauli measurements of σ [FL11; SLP11]. Using the bound $\frac{1}{2}\|\sigma - \sigma_{\text{target}}\|_1 \leq \varepsilon := \sqrt{1 - F(\sigma, \sigma_{\text{target}})^2}$, one therefore obtains a bound on the trace distance between σ and σ_{target}. Theorem 3.1 gives that the state with maximum entropy in $B_\varepsilon(\sigma_{\text{target}})$ is $\rho^*_\varepsilon(\sigma_{\text{target}})$. Using that σ_{target} is a pure state, one can check using Lemma 4.3 that

$$(\sigma_{\text{target}})^*_\varepsilon = \begin{cases} \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots, \frac{\varepsilon}{d-1}) & \text{if } \varepsilon \leq 1 - \frac{1}{d} \\ \tau := \frac{\mathbb{1}}{d} & \text{else} \end{cases} \tag{12}$$

in the basis in which $\sigma_{\text{target}} = \operatorname{diag}(1, 0, \dots, 0)$, where d is the dimension of the underlying Hilbert space. The maximum-entropy principle therefore implies that $(\sigma_{\text{target}})^*_\varepsilon$ defined by (12) is the appropriate estimate of σ, when given only the condition that $\frac{1}{2}\|\sigma - \sigma_{\text{target}}\|_1 \leq \varepsilon$.

- it may be possible to determine additional constraints on σ by the Pauli measurements performed to estimate $F(\sigma, \sigma_{\text{target}})$. In that case, MaxEnt gives that the appropriate estimate of σ is the state with maximum entropy subject to these constraints as well, and not only the relation $\frac{1}{2}\|\sigma - \sigma_{\text{target}}\|_1 \leq \varepsilon$.

- it may be possible to devise a measurement scheme to estimate $\frac{1}{2}\|\rho - \sigma_{\text{target}}\|_1$ directly, which could be more efficient than first estimating the fidelity and then employing the Fuchs-van de Graaf inequality to bound the trace distance.

Moreover, Corollary 3.4 yields maximizers and minimizers of any (h, ϕ)-entropy, not simply the von Neumann entropy. This allows computation of (trace-ball) "smoothed" Schur-concave functions. Given $\varphi : \mathcal{D} \to \mathbb{R}$, we define

$$\bar{\varphi}^{(\varepsilon)}(\sigma) := \max_{\omega \in B_\varepsilon(\rho)} \varphi(\sigma), \qquad \underline{\varphi}^{(\varepsilon)}(\sigma) := \min_{\omega \in B_\varepsilon(\rho)} \varphi(\sigma). \tag{13}$$

By Corollary 3.4, we therefore obtain explicit formulas: $\bar{\varphi}^{(\varepsilon)}(\sigma) = \varphi(\rho_\varepsilon^*)$, and $\underline{\varphi}^{(\varepsilon)}(\sigma) = \varphi(\rho_{*,\varepsilon})$. In particular, this provides an exact version of Theorem 1 of [Sko16], which formulates approximate maximizers for the smoothed α-Rényi entropy $\bar{S}_\alpha^{(\varepsilon)}$.

Note that by setting $\varphi = H_{\min}$ or $H_{\max}$, the min- and max-entropies, in (13), yields explicit expressions for the min- and max- entropies smoothed over the ε-ball. These choices are of particular interest, due to their relevance in one-shot information theory (see e.g. [Ren05; Tom16] and references therein). In particular, let us briefly consider one-shot classical data compression. Given a source (random variable) X, one wishes to encode output from X using codewords of a fixed length $\log m$, such that the original message may be recovered with probability of error at most ε. It is known that the minimal value of m at fixed ε, denoted $m_*(\varepsilon)$, satisfies

$$\underline{H}_{\max}^{(\varepsilon)}(X) \leq \log m_*(\varepsilon) \leq \inf_{\delta \in (0,\varepsilon)} [\underline{H}_{\max}^{(\varepsilon)}(X) + \log \frac{1}{\delta}] \tag{14}$$

as shown by [Tom16; RR12]. Equation (13) provides the means to explicitly evaluate the quantity $\underline{H}_{\max}^{(\varepsilon)}(X)$ in this bound.

Remark. Although this essay is concerned with the ε-ball defined via the trace distance, these results have some implications for other distance measures. For example, the so-called *purified distance* is often used in defining smoothed entropies [Tom12]. The purified distance between two states ρ and σ, denoted $P(\rho, \sigma)$, satisfies the inequality

$$T(\rho, \sigma) \leq P(\rho, \sigma) \tag{15}$$

by Lemma 3.5 of [Tom16]. Therefore, denoting $\tilde{B}_\varepsilon(\rho)$ by the ε-ball

$$\tilde{B}_\varepsilon(\rho) := \{\omega \in \mathcal{D}(\mathcal{H}) : P(\omega, \sigma) \leq \varepsilon\} \tag{16}$$

one has $\tilde{B}_\varepsilon(\rho) \subset B_\varepsilon(\rho)$. Therefore, for any Schur concave function f,

$$f(\rho_{*,\varepsilon}) \leq \min_{\tilde{B}_\varepsilon(\rho)} f \leq \max_{\tilde{B}_\varepsilon(\rho)} f \leq f(\sigma_\varepsilon^*) \tag{17}$$

by Corollary 3.4. We therefore obtain immediate bounds on the maxima and minima of Schur concave functions over $\tilde{B}_\varepsilon(\rho)$.

Additionally, from Corollary 3.4 we may obtain tight local continuity bounds (with respect to the trace distance) for any Schur concave function.

Proposition 3.5 (Local continuity bounds)**.** *Let f be Schur concave on $\mathcal{D}(\mathcal{H})$. Let $\rho \in \mathcal{D}$ and $\varepsilon \in [0,1]$. Then for any state $\omega \in B_\varepsilon(\rho)$,*

$$|f(\omega) - f(\rho)| \leq \max\{f(\rho_\varepsilon^*) - f(\rho), f(\rho) - f(\rho_{*,\varepsilon})\}. \tag{18}$$

Remark. Note that this inequality indeed gives a local continuity bound: the right-hand side of (18) only depends on ρ and ε. Moreover, the value of the right-hand side may be computed via the definitions of ρ_ε^* and $\rho_{*,\varepsilon}$.

By taking $f = S$, the von Neumann entropy, we obtain a local continuity bound given by inequality (19) in the following corollary. We may compare it to the Audenaert-Fannes bound (stated in Lemma 2.2), which we include as inequality (20) below. By a direct analysis of a proof of the Audenaert-Fannes bound, we may establish necessary and sufficient conditions for the local bound to equal the uniform one[3].

Corollary 3.6 (Local continuity bound for the von Neumann entropy)**.** *Let $\rho \in \mathcal{D}$ and $\varepsilon \in (0,1]$. Then for any state $\omega \in B_\varepsilon(\rho)$,*

$$\begin{aligned} |S(\omega) - S(\rho)| &\leq \max\{S(\rho_\varepsilon^*) - S(\rho), S(\rho) - S(\rho_{*,\varepsilon})\} && (19) \\ &\leq \begin{cases} \varepsilon \log(d-1) + h(\varepsilon) & \text{if } \varepsilon < 1 - \frac{1}{d} \\ \log d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases} && (20) \end{aligned}$$

for $h(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ the binary entropy. Moreover, equality holds in (20) if and only if one of the following holds:

1. *ρ is a pure state; in this case* $\rho_\varepsilon^* = \begin{cases} \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1}) & \text{if } \varepsilon < 1 - \frac{1}{d}, \\ \mathbb{1}/d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases}$
2. *$\varepsilon < 1 - \frac{1}{d}$, and $\rho = \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1})$; in this case $\rho_{\varepsilon,*}(\rho)$ is a pure state,*
3. *$\varepsilon \geq 1 - \frac{1}{d}$ and $\rho = \tau := \mathbb{1}/d$; in this case $\rho_{\varepsilon,*}(\rho)$ is a pure state.*

We defer the proof of this result to Appendix B.

Remark. See Figure 2 for a comparison between the right-hand side of our bound (19) and the right-hand side of the (AF)-bound (20) for 500 random choices of $\rho \in \mathcal{D}$ and $\varepsilon \in (0,1]$. The figure shows the merit of the local continuity bound.

3.2. Further properties of ρ_ε^* and applications

By regarding the transformation from a state $\rho \in \mathcal{D}(\mathcal{H})$ to the maximal state in majorization order in the ε-ball around ρ, namely ρ_ε^*, as a map, we can obtain additional information about the behavior of the majorization order on the set of states. Define the map

$$\begin{aligned} \mathcal{M}_\varepsilon : \qquad \mathcal{D}(\mathcal{H}) &\to \mathcal{D}(\mathcal{H}) \\ \rho &\mapsto \rho_\varepsilon^*. \end{aligned} \tag{21}$$

Note that $\mathcal{M}_\varepsilon$ is a *non-linear* map on the set of states, and therefore, does not represent a physical time evolution. Nevertheless, it has useful mathematical properties.

[3]and in fact, this analysis shows that the known necessary conditions for equality in Lemma 2.2 are, in fact, sufficient.

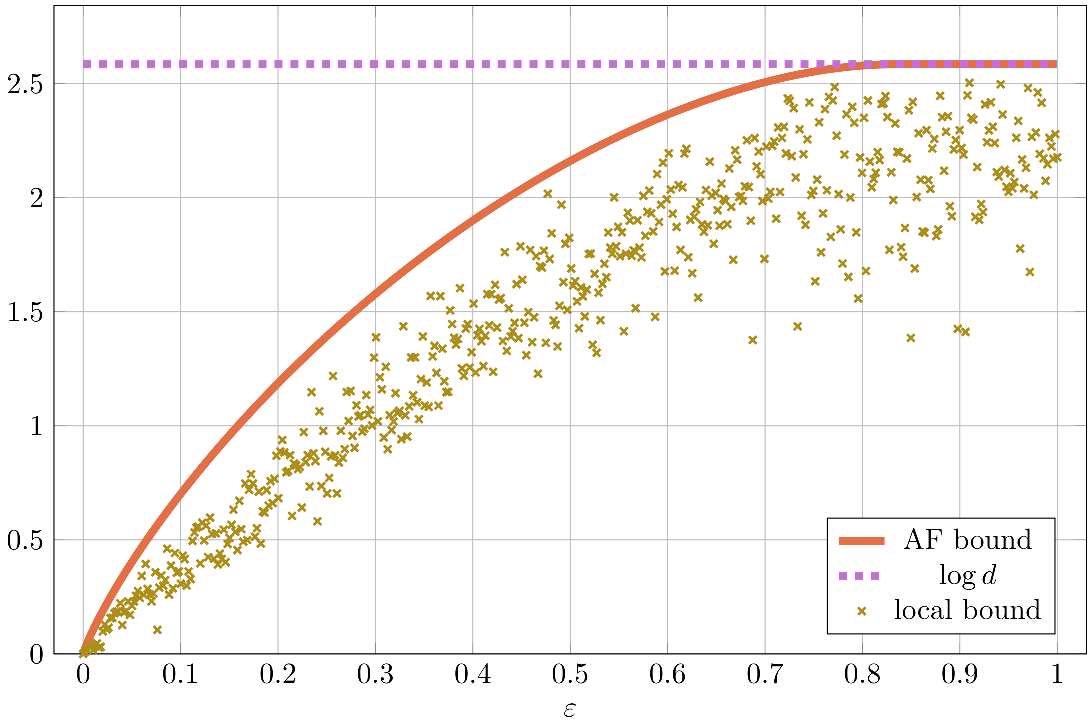


Figure 2: In dimension $d = 6$, the eigenvalues of 500 quantum states ρ and $\varepsilon \in (0, 1]$ were chosen uniformly randomly. For each pair (ρ, ε), the local bound given in the right-hand side of (19) is plotted as a cross. The Audenaert-Fannes bound given by the right-hand side of (20) is plotted as a function of ε. We see for many choices of ρ, the local bound gives a tighter bound than the Audenaert-Fannes bound.

Proposition 3.7 (Properties of $\mathcal{M}_\varepsilon$)**.** *The (non-linear) map $\mathcal{M}_\varepsilon$ has following properties, for any $\varepsilon \in (0, 1]$. Let $\rho \in \mathcal{D}(\mathcal{H})$.*

a. *Maps states to states: $\mathcal{M}_\varepsilon : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H})$.*

b. *Minimal in majorization order: $\mathcal{M}_\varepsilon(\rho) \in B_\varepsilon(\rho)$ and for any $\omega \in B_\varepsilon(\rho)$, we have $\mathcal{M}_\varepsilon(\rho) \prec \omega$.*

c. *Semi-group property: if $\varepsilon_1, \varepsilon_2 \in (0, 1]$ with $\varepsilon_1 + \varepsilon_2 \leq 1$, we have $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \mathcal{M}_{\varepsilon_1} \circ \mathcal{M}_{\varepsilon_2}(\rho)$.*

d. *Majorization-preserving: let $\sigma \in \mathcal{D}(\mathcal{H})$ such that $\rho \prec \sigma$. Then $\mathcal{M}_\varepsilon(\rho) \prec \mathcal{M}_\varepsilon(\sigma)$.*

e. *$\tau = \frac{\mathbb{1}}{d}$ is the unique fixed point of $\mathcal{M}_\varepsilon$, i.e. the unique solution to $\rho = \mathcal{M}_\varepsilon(\rho)$ for $\rho \in \mathcal{D}(\mathcal{H})$. Moreover, for any $\rho \neq \tau$, $\mathcal{M}_\varepsilon(\rho)$ is not unitarily equivalent to ρ.*

f. *For any state $\rho \in B_\varepsilon(\tau)$, we have $\mathcal{M}_\varepsilon(\rho) = \tau$. If $\rho \notin B_\varepsilon(\tau)$, then $\mathcal{M}_\varepsilon(\rho)$ is on the boundary of $B_\varepsilon(\rho)$.*

g. *For any pure state $\psi \in \mathcal{D}_{pure}(\mathcal{H})$, the state $\mathcal{M}_\varepsilon(\psi)$ has the form*

$$\mathcal{M}_\varepsilon(\psi) = \begin{cases} \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1}) & \varepsilon < 1 - \frac{1}{d} \\ \tau := \frac{\mathbb{1}}{d} & \varepsilon \geq 1 - \frac{1}{d}. \end{cases} \tag{22}$$

h. Saturates triangle inequality with $\tau := \frac{\mathbb{1}}{d}$*:*

$$\frac{1}{2}\|\rho - \tau\|_1 = \frac{1}{2}\|\tau - \mathcal{M}_\varepsilon(\rho)\|_1 + \frac{1}{2}\|\mathcal{M}_\varepsilon(\rho) - \rho\|_1. \tag{23}$$

Remark. Item (d) is due to [HOS17].

These properties, and in particular the semi-group property (c) allow the local continuity bounds discussed in the previous section to be upgraded to *uniform* continuity bounds. This generalizes the inequality (20) above, which used that the uniform continuity bound for the von Neumann entropy was already known.

Theorem 3.8 (Uniform continuity bounds)**.** *Let* $H_{(h,\phi)}$ *be an* (h,ϕ)*-entropy, defined through* (7)*) with* h *concave and* ϕ *strictly concave. For* $\varepsilon \in (0,1]$ *and any states* $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ *such that* $\frac{1}{2}\|\rho - \sigma\|_1 \le \varepsilon$*, we have*

$$|H_{(h,\phi)}(\rho) - H_{(h,\phi)}(\sigma)| \le \begin{cases} h(\phi(1-\varepsilon) + (d-1)\phi(\frac{\varepsilon}{d-1})) & \varepsilon < 1 - \frac{1}{d} \\ h(d\phi(\frac{1}{d})) & \varepsilon \ge 1 - \frac{1}{d}, \end{cases} \tag{24}$$

where $d = \dim \mathcal{H}$*. In particular, for* $\alpha \in (0,1)$*,*

$$|S_\alpha(\rho) - S_\alpha(\sigma)| \le \begin{cases} \frac{1}{1-\alpha}\log((1-\varepsilon)^\alpha + (d-1)^{1-\alpha}\varepsilon^\alpha) & \varepsilon < 1 - \frac{1}{d} \\ \log d & \varepsilon \ge 1 - \frac{1}{d} \end{cases} \tag{25}$$

and for $q \in (0,1) \cup (1,\infty)$*,*

$$|T_q(\rho) - T_q(\sigma)| \le \begin{cases} \frac{1}{1-q}((1-\varepsilon)^q + (d-1)^{1-q}\varepsilon^q - 1) & \varepsilon < 1 - \frac{1}{d} \\ \frac{d^{1-q}-1}{1-q} & \varepsilon \ge 1 - \frac{1}{d}. \end{cases} \tag{26}$$

Moreover, equality in (24)*,* (25)*, or* (26) *occurs if and only if one of the two states (say,* σ*) is pure, and either*

1. $\varepsilon < 1 - \frac{1}{d}$ *and* $\vec{\lambda}(\rho) = (1-\varepsilon, \frac{\varepsilon}{d-1}, \dots, \dots, \frac{\varepsilon}{d-1})$*, or*
2. $\varepsilon \ge 1 - \frac{1}{d}$*, and* $\rho = \tau := \frac{\mathbb{1}}{d}$*.*

Remark.

- When (24) is applied to the von Neumann entropy S, one recovers the Audenaert-Fannes bound, (8), with equality conditions. The sufficiency of these equality conditions were shown in [Aud07], and their necessity is derived in Appendix B by a direct analysis of the proof of the bound presented in [Pet08, Thm. 3.8] and [Win16]. The above result establishes that in fact these necessary and sufficient conditions are the same for every (h,ϕ)-entropy with h concave and ϕ strictly concave.
- The inequality (25) reduces to the Audenaert-Fannes bound (8) when the limit $\alpha \to 1$ is taken on both sides of it.
- The inequalities (25) and (26) were in fact obtained in the published version of [Aud07], in an appendix[4]. The necessity of the equality conditions were not established, however. Moreover, the proof of these results was by direct optimization, and would need to be modified for other entropies. In Section 5, we present a unified proof for any (h,ϕ)-entropy (with h concave and ϕ strictly concave) that employs naturally the structure of the majorization order on $\mathcal{D}(\mathcal{H})$.

[4]Additionally, the bound (26) appeared recently in [Che+17], derived along the same lines as the proof in [Aud07].

4. Majorization and the ε-ball

In this section, we consider the ε-ball $B_\varepsilon(\rho)$ around a state $\rho \in \mathcal{D}(\mathcal{H})$, and motivate the construction of maximal and minimal states in the majorization order given in (10) of Section 4.1. We prove Theorem 3.1 by reducing it to the classical case of discrete probability distributions on d symbols, and then constructing explicit states ρ_ε^* (in Section 4.1), and $\rho_{*,\varepsilon}$ (in Section 4.3), whose eigenvalues are respectively given by the probability distributions which are minimal and maximal in majorization order.

The reduction to the classical case is immediate: the state majorization $\rho \prec \sigma$ by definition means that the eigenvalue majorization $\vec{\lambda}(\rho) \prec \vec{\lambda}(\sigma)$ holds. Thus, instead of the set of density matrices $\mathcal{D}$, we consider the simplex of probability vectors

$$\Delta := \{r = (r_i)_{i=1}^d \in \mathbb{R}^d : r_j \geq 0 \text{ for each } j = 1, \dots, d, \text{ and } \sum_{i=1}^d r_i = 1\}.$$

Note that Δ is the polytope (i.e. the convex hull of finitely many points) generated by $(1, 0, \dots, 0)$ and its permutations. Instead of the ball $B_\varepsilon(\rho)$, we consider the 1-norm ball around $p = (p_i)_{i=1}^d := \vec{\lambda}(\rho)$,

$$B_\varepsilon(p) := \{w = (w_i)_{i=1}^d \in \Delta : \frac{1}{2}\|w - p\|_1 := \frac{1}{2}\sum_{i=1}^d |w_i - p_i| \leq \varepsilon\}.$$

The set $\{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$ can be written

$$\{x \in \mathbb{R}^d : \|x\|_1 \leq 1\} = \operatorname{conv}\{e_1, -e_1, \dots, e_d, -e_d\},$$

where $\operatorname{conv}(A)$ denotes the convex hull of a set A, and $e_1, \dots, e_d$ are the vectors of the standard basis (e.g. $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the jth entry), and is therefore a polytope, called the *d-dimensional cross-polytope* (see e.g. [Mat02, p. 82]). As a translation and scaling of the d-dimensional cross-polytope, the set $\{w \in \mathbb{R}^d : \frac{1}{2}\|w - p\|_1 \leq \varepsilon\}$ is a polytope as well. As $B_\varepsilon(p)$ is the intersection of this set and Δ, it too is a polytope (see Figure 3a for an illustration of $B_\varepsilon(p)$ in a particular example).

The existence of ρ_ε^* and $\rho_{*,\varepsilon}$ in $B_\varepsilon(\rho)$ satisfying (10) is equivalent to p_ε^* and $p_{*,\varepsilon}$ in $B_\varepsilon(p)$ satisfying

$$p_\varepsilon^* \prec w \prec p_{*,\varepsilon} \tag{27}$$

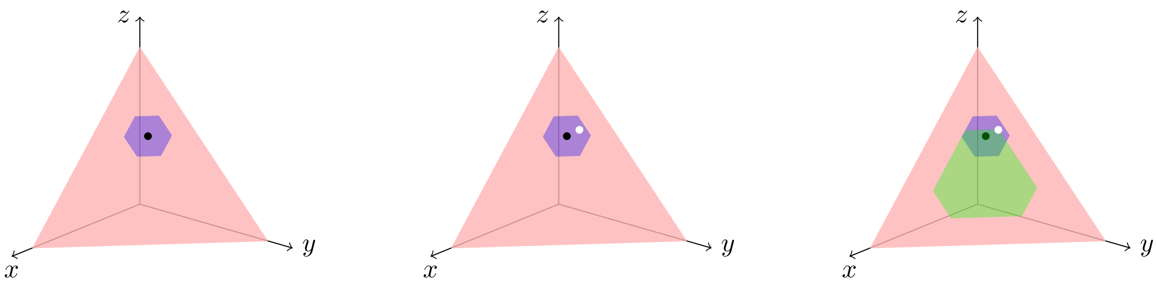
for all $w \in B_\varepsilon(p)$. Using Birkhoff's Theorem (e.g. [NC11, Theorem 12.12]), the set of vectors majorized by a point $w \in \Delta$ can be shown to be given by

$$M_w := \{p \in \Delta : w \succ p\} = \operatorname{conv}\{\pi(w) : \pi \in S_d\}, \tag{28}$$

where S_d is the symmetric group on d letters (see [Mar11, p. 34]). Let us illustrate this with an example in $d = 3$. Let us choose $p = (0.21, 0.24, 0.55)$ and $\varepsilon = 0.1$. The simplex Δ and ball $B_\varepsilon(p)$ are depicted in Figure 3a. A point $w = (0.14, 0.28, 0.58) \in B_\varepsilon(p)$ is shown in Figure 3b, and the set M_w in Figure 3c.

The geometric characterization (28), depicted in Figure 3, requires that $B_\varepsilon(p) \subseteq M_{p_{*,\varepsilon}}$, and conversely, $p_\varepsilon^* \in M_p$ for each $p \in B_\varepsilon(p)$. Figure 3c shows that for the point w, $B_\varepsilon(p) \not\subseteq M_w$, implying that $w \neq p_{*,\varepsilon}$. Moreover, one can check that that e.g. $w \notin M_q$, and hence $w \neq p_\varepsilon^*$.

Next we consider Schur concave functions on Δ, in order to gain insight into the probability distributions $p_{*,\varepsilon}$ and p_ε^* which arise in the majorization order (27). In particular, let us consider



(a) $B_\varepsilon(p)$, with p in black (b) A point $w \in B_\varepsilon(p)$ in white. (c) The set M_w in green.

Figure 3: In dimension $d = 3$, the simplex, Δ, of probability vectors is the shaded triangle shown in (a), along with the ball $B_\varepsilon(p)$ which is the hexagon shown in blue, centered at $p = (0.21, 0.24, 0.55)$ (depicted by a black dot) with $\varepsilon = 0.1$. In (b), a point $w = (0.14, 0.28, 0.58)$ is depicted in white, and in (c), the set M_w is shown in green.

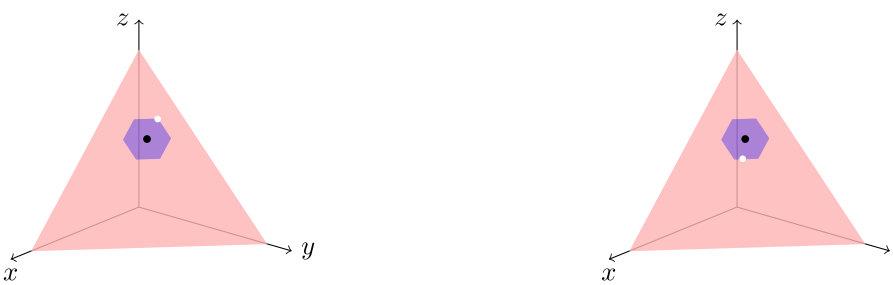
Shannon entropy $H(w) := -\sum_{i=1}^d w_i \log w_i$ of a probability distribution $w = (w_i)_{i=1}^d$. It is known to be strictly Schur concave. Hence, if $p_{*,\varepsilon} \in B_\varepsilon(p)$ satisfying (27) exists, it must satisfy

$$H(p_{*,\varepsilon}) \leq H(w)$$

for any $w \in B_\varepsilon(p)$. Thus, $p_{*,\varepsilon}$ must be a minimizer of H, which is a concave function, over $B_\varepsilon(p)$, a convex set. Similarly, p_ε^* must be a maximizer of H over $B_\varepsilon(p)$. Properties of maximizers of concave functions over a convex sets are well-understood; in particular, any local maximizer is a global maximizer.

The task of minimizing a concave function over a convex set is a priori more difficult; in particular, local minima need not be global minima. There is, however, a *minimum principle* which asserts that the minimum occurs on the boundary of the set; this is formulated more precisely in e.g. [Roc96, Chapter 32]. Since $B_\varepsilon(p)$ is a polytope, H is minimized on one of the finitely many verticies of $B_\varepsilon(p)$. This fact yields a simple solution to the problem of minimizing H over $B_\varepsilon(p)$, as described below by example, and in generality in Section 4.3.

Let us return to the example of Figure 3. We see $B_\varepsilon(p)$ has six vertices; these are $\{p + \pi((\varepsilon, -\varepsilon, 0)) : \pi \in S_d\}$. The vertex which minimizes H is $v := (0.21 - \varepsilon, 0.24, 0.55 + \varepsilon)$, where



(a) A minimum v of H over $B_\varepsilon(p)$, in white. (b) The maximum m of H over $B_\varepsilon(p)$, in white.

Figure 4: For the example of Figure 3, the (unique) maximum and minimum of the Shannon entropy H over $B_\varepsilon(p)$ are shown. Both v and m occur on the boundary of $B_\varepsilon(p)$.

the smallest entry is decreased and the largest entry is increased, as shown in Figure 4b. Moreover, one can check that $w \prec v$ for any $w \in B_\varepsilon(p)$. This leads us to the conjecture that the vertex corresponding to decreasing the smallest entry and increasing the largest entry will yield

$p_{*,\varepsilon}$ satisfying (27) in general. We see in Section 4.3 that this is indeed true, although in some cases more than one entry needs to be decreased.

On the other hand, finding the probability distribution p_ε^* in $B_\varepsilon(p)$ which satisfies (27) is more than a matter of checking the verticies of $B_\varepsilon(p)$, as shown by Figure 4b: in this example, p_ε^* is not a vertex of $B_\varepsilon(p)$. Interestingly, useful insight into this probability distribution can be obtained by using results from convex optimization theory. This is discussed in the following section.

4.1. Constructing the minimal state in the majorization order (10)

Let us assume Theorem 3.2 holds, and that the von Neumann entropy S satisfies the requirements of the function φ of the theorem, with $L_\xi = -\log\xi - \frac{1}{\log_e(2)}\mathbb{1}$. The proof of these facts are given in Section 6. Using Theorem 3.2, we deduce properties of and the form of a maximizer of S in the ε-ball.

Since S is continuous and $B_\varepsilon(\rho)$ is compact, S achieves a maximum over $B_\varepsilon(\rho)$. Moreover, since S is strictly concave, the maximum is unique; otherwise, if $\xi_1, \xi_2 \in B_\varepsilon(\rho)$ were maximizers, $\xi = \frac{1}{2}\xi_1 + \frac{1}{2}\xi_2 \in B_\varepsilon(\rho)$ would have strictly higher entropy.

Now, let ξ be the unique maximizer. Condition 1 of Theorem 3.2 yields that either $\log\xi \propto \mathbb{1}$, so $\xi = \tau = \frac{\mathbb{1}}{d}$, or else $T(\xi,\rho) = \varepsilon$. Since τ is the global maximizer of S over $\mathcal{D}$, we have $\xi = \tau$ whenever $\tau \in B_\varepsilon(\rho)$. If $\tau \notin B_\varepsilon(\rho)$, then this condition yields the first piece of information about the maximizer: it is on the boundary of $B_\varepsilon(\rho)$, in that $T(\xi,\rho) = \varepsilon$, just as shown in Figure 4b in the classical setup.

By Lemma 2.1, working in the basis in which $\rho = \mathrm{Eig}^\downarrow(\rho)$, we have

$$T(\mathrm{Eig}^\downarrow(\xi),\rho) \le T(\xi,\rho) \le \varepsilon$$

and therefore $\mathrm{Eig}^\downarrow(\xi) \in B_\varepsilon(\rho)$. Since S is unitarily invariant, $S(\mathrm{Eig}^\downarrow(\xi)) = S(\xi)$, and by uniqueness of the maximizer, we have $\xi = \mathrm{Eig}^\downarrow(\xi)$. Hence, the maximizer ξ commutes with ρ, and hence with Δ, and the sums of its eigenprojections $\pi_\pm$. Since $[L_\xi, \xi] = 0$ as well, Theorem 3.2 yields

$$L_\xi \pi_\pm = \lambda_\pm(L_\xi)\pi_\pm \tag{29}$$

For any $\psi \in \pi_\pm\mathcal{H}$, we have $L_\xi\psi = \lambda_\pm(L_\xi)\psi$, so (29) is an eigenvalue equation for L_ξ. Since $\xi = \exp(-(L_\xi + \mathbb{1}))$ is a function of L_ξ, it shares the same eigenprojections. In particular,

$$\xi\pi_\pm = \exp\left(-(\lambda_\pm(L_\xi) + 1)\right)\pi_\pm$$

serves as an eigenvalue equation for ξ. Since $[\xi,\rho] = 0$, we can discuss how each acts on each (shared) eigenspace. By definition, ξ and ρ act the same on $\ker\Delta = \ker(\xi - \rho)$. On the other hand, on the subspaces where the eigenvalues of ξ are greater than those of ρ, i.e. on $\pi_+\mathcal{H}$, we see that ξ has the constant eigenvalue $\gamma_- := \exp(-(\lambda_+(L_\xi) + 1))$, and on $\pi_-\mathcal{H}$, $\gamma_+ := \exp(-(\lambda_-(L_\xi) + 1))$. Note that as $x \mapsto -\log x - 1$ is monotone decreasing on $\mathbb{R}$, the subspace $\pi_+\mathcal{H}$ where L_ξ has its largest eigenvalue, ξ has its smallest eigenvalue, and vice-versa. That is, $\lambda_+(\xi) = \gamma_+$ and occurs on the subspace $\pi_-\mathcal{H}$, and $\lambda_-(\xi) = \gamma_-$, and occurs on $\pi_+\mathcal{H}$. Let us briefly remark on the notation: γ_-, for example, has the subscript $-$ as γ_- is the smallest eigenvalue of ξ, whereas π_+ is the projector onto the subspace on which eigenvalues of ρ *increase* to form the eigenvalues of ξ (which are, in fact, γ_-).

Let us summarize the above observations. In $\ker\Delta$, the maximizer ξ has the same eigenvalues as ρ. On the subspace $\pi_-\mathcal{H}$, the state ξ has the constant eigenvalue γ_+, which is the largest eigenvalue of ξ, and $\xi\pi_- = \gamma_+\pi_- \le \rho\pi_-$. In the subspace $\pi_+\mathcal{H}$, ξ has the constant eigenvalue

γ_-, which is its smallest eigenvalue, and $\xi\pi_+ = \gamma_-\pi_+ \geq \rho\pi_+$. It remains to choose subspaces corresponding to $\pi_\pm$, and the associated eigenvalues γ_- and γ_+. Recall that as $T(\xi, \rho) = \varepsilon$, we have $\mathrm{Tr}[(\xi - \rho)_+] = \mathrm{Tr}[(\xi - \rho)_-] = \varepsilon$.

As the entropy is minimized on pure states and maximized on the completely mixed state $\tau := \frac{1}{d}\mathbb{1}$, one can guess that to increase the entropy, one should raise the small eigenvalues of ρ, and lower the large eigenvalues of ρ. That is, π_+ should correspond to the eigenspaces of the m_- smallest eigenvalues of ρ, and π_- should correspond to the eigenspaces of the m_+ largest eigenvalues of ρ, for some $m_-, m_+ \in \{1, \ldots, d-1\}$. Moreover, as γ_+ is the largest eigenvalue of ξ, and γ_- is the smallest one, we must have $\gamma_- \leq \mu \leq \gamma_+$ for any eigenvalue μ of ρ with corresponding eigenspace which is a subspace of $\ker \Delta$. Figure 5 illustrates these ideas in an example at the end of this subsection. In Lemma 4.3, we prove there exists unique γ_-, γ_+, m_- and m_+ which respect these considerations. We will call the resulting state ρ_ε^* (instead of ξ).

Let us recall the task is to construct a state ρ_ε^* satisfying (10). Instead, we have constructed a state in order to maximize the entropy S. However, since any state ρ_ε^* satisfying (10) must maximize S over $B_\varepsilon(\rho)$, and the maximizer is unique, we instead check that the state resulting from this construction satisfies (10) in Section 4.2.

Let $\rho \in \mathcal{D}(\mathcal{H})$ and $\varepsilon \in (0, 1]$. For notational simplicity, we often suppress dependence on ρ and ε in this section, and write $\lambda_j = \lambda_j(\rho)$ so that the eigenvalues of ρ are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. We first define a quantity $\gamma_+^{(m)} \equiv \gamma_+^{(m)}(\rho, \varepsilon)$, for $m \in \{0, 1, \ldots, d-1\}$, as follows

$$\gamma_+^{(m)} := \begin{cases} \frac{1}{m}\left(\sum_{i=1}^m \lambda_i - \varepsilon\right) & \text{if } T(\rho, \tau) > \varepsilon \text{ and } m \neq 0 \\ \frac{1}{d} & \text{else.} \end{cases}$$

Similarly, a quantity $\gamma_-^{(m)} \equiv \gamma_-^{(m)}(\rho, \varepsilon)$ is defined by

$$\gamma_-^{(m)} := \begin{cases} \frac{1}{m}\left(\sum_{i=d-m+1}^d \lambda_i + \varepsilon\right) & \text{if } T(\rho, \tau) > \varepsilon \text{ and } m \neq 0 \\ \frac{1}{d} & \text{else.} \end{cases}$$

Then for $\rho \neq \tau$, we define $m_+ = m_+(\rho, \varepsilon)$ as the unique solution to the following inequalities:

$$\lambda_{m+1} \leq \gamma_+^{(m)} < \lambda_m, \qquad m \in \{1, \ldots, d-1\} \tag{30}$$

and we set $m_+(\tau, \varepsilon) = 0$. Similarly, for $\rho \neq \tau$, we define $m_- = m_-(\rho, \varepsilon)$ as the unique solution to the inequalities:

$$\lambda_{d-m+1} < \gamma_-^{(m)} \leq \lambda_{d-m}, \qquad m \in \{1, \ldots, d-1\} \tag{31}$$

and set $m_-(\tau, \varepsilon) = 0$. Finally, we set $\gamma_+ = \gamma_+(\rho, \varepsilon) := \gamma_+^{(m_+)}$ and $\gamma_- = \gamma_-(\rho, \varepsilon) := \gamma_-^{(m_-)}$. The proof that the above equations have unique solutions follows from Lemma 4.3, which is postponed to the end of this section.

Given the eigen-decomposition $\rho = \sum_{i=1}^d \lambda_i |i\rangle\langle i|$, we define

$$\rho_\varepsilon^* := \sum_{i=1}^{m_+} \gamma_+ |i\rangle\langle i| + \sum_{i=m_++1}^{d-m_-} \lambda_i |i\rangle\langle i| + \sum_{i=d-m_-+1}^{d} \gamma_- |i\rangle\langle i|. \tag{32}$$

We check that $\rho_\varepsilon^* \in B_\varepsilon(\rho)$ as follows. From its definition, its spectrum lies in the interval $(0, 1]$ and ρ_ε^* has trace 1. Additionally, if $T(\rho, \tau) \leq \varepsilon$, then $\rho_\varepsilon^* = \tau \in B_\varepsilon(\rho)$. Otherwise,

$$\|\rho_\varepsilon^* - \rho\|_1 = \sum_{i=1}^{m_+} |\gamma_- - \lambda_i| + \sum_{i=d-m_-+1}^{d} |\gamma_+ - \lambda_i| = \varepsilon + \varepsilon = 2\varepsilon; \tag{33}$$

so for any $\varepsilon \in [0,1]$, we have $\rho_\varepsilon^* \in B_\varepsilon^+(\rho)$.

To summarize, we construct ρ_ε^* as follows: we decrease the m_+ largest eigenvalues of ρ by setting them to γ_+ (where m_+ and γ_+ are related by eq. (30)), increase the m_- smallest eigenvalues of ρ by setting them to γ_- (where m_- and γ_- are related by eq. (31)), and we keep the other eigenvalues of ρ unchanged. This is illustrated in Figure 5, for a state $\rho \in \mathcal{D}(\mathcal{H})$ with $\varepsilon = 0.07$ and $d = 12$.

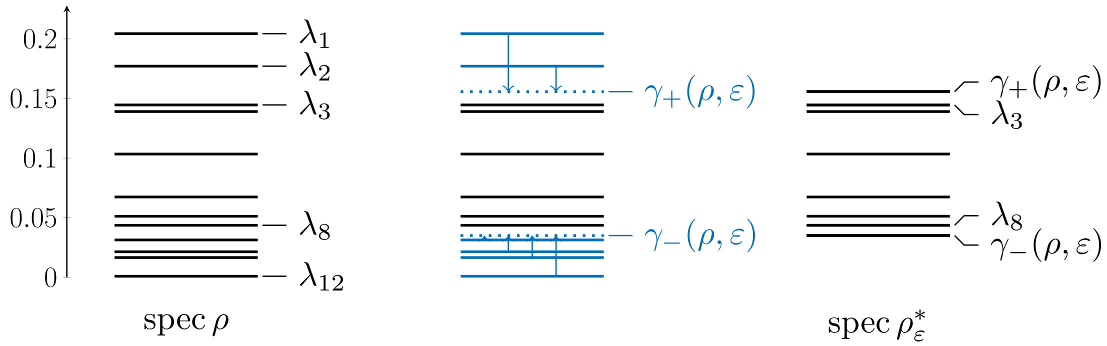


Figure 5: We choose $d = 12$, a state $\rho \in \mathcal{D}(\mathcal{H})$, and $\varepsilon = 0.07$, for which $m_+ = 2$ and $m_- = 4$. Left: the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ of ρ are plotted. Center: the smallest four eigenvalues of ρ are increased to $\gamma_- = \frac{1}{4}[\lambda_d + \cdots + \lambda_{d+4} + \varepsilon]$, and the largest two eigenvalues of ρ decreased to $\gamma_+ = \frac{1}{2}[\lambda_1 + \lambda_2 - \varepsilon]$. Right: the eigenvalues of ρ_ε^* are γ_+ with multiplicity two, $\lambda_3, \lambda_4, \ldots, \lambda_{d-4}$, and γ_- with multiplicity four.

In the next section, we prove that ρ_ε^* satisfies the majorization order (10).

4.2. Minimality of ρ_ε^* in the majorization order (10)

Given $p = \vec{\lambda}(\rho)$, we consider the vector $p_\varepsilon^* := \vec{\lambda}(\rho_\varepsilon^*)$, i.e. the eigenvalues of ρ_ε^*, defined in (32). Let $w \in B_\varepsilon(p)$, and consider its entries arranged in non-increasing order,

$$w_1 \geq \cdots \geq w_d.$$

Let $p_1 \geq \cdots \geq p_d$ be the entries of p in non-increasing order, and $p_1^* \geq \cdots \geq p_d^*$ be the entries of p_ε^* in non-increasing order. In this section, we show $p_\varepsilon^* \prec w$.

1. First, we establish that $p_1^* \leq w_1$.

   To prove this, let us assume the contrary: $p_1^* > w_1$. Then, since $p_1^* = p_2^* = \cdots = p_{m_+}^* = \gamma_+$,

$$m_+\gamma_+ = \sum_{i=1}^{m_+} p_i^* > m_+ w_1 \geq \sum_{i=1}^{m_+} w_i.$$

   We conclude this step with the following lemma.
   Lemma 4.1. *If $m_+\gamma_+ > \sum_{i=1}^{m_+} w_i$, then $w \notin B_\varepsilon(p)$.*

   Proof. Multiplying each side by -1 and adding $\sum_{i=1}^{m_+} q_i$, we have

$$\sum_{i=1}^{m_+}(p_i - p_i^*) < \sum_{i=1}^{m_+}(p_i - w_i) \leq \sum_{i=1}^{m_+}(p_i - w_i)_+ \leq \sum_{i=1}^{d}(p_i - w_i)_+ = \frac{1}{2}\|p - w\|_1. \tag{34}$$

Using $\gamma_+ = \frac{1}{m_+}\left(\sum_{i=1}^{m_+} p_i - \varepsilon\right)$, the far left-hand side is

$$\sum_{i=1}^{m_+} p_i - m_+\gamma_+ = \sum_{i=1}^{m_+} p_i - \left(\sum_{i=1}^{m_+} p_i - \varepsilon\right) = \varepsilon.$$

Then (34) becomes

$$\varepsilon < \frac{1}{2}\|p - w\|_1,$$

contradicting that $w \in B_\varepsilon(p)$. □

2. Next, for $k \in \{1, \dots, m_+\}$, we have $\sum_{i=1}^{k} p_i^* \leq \sum_{i=1}^{k} w_i$.

   We prove this recursively: assume the property holds for some $k \in \{1, \dots, m_+ - 1\}$ but not for $k+1$. Note we have proven the base case of $k = 1$ in the previous step. Then

$$\sum_{i=1}^{k} p_i^* \leq \sum_{i=1}^{k} w_i, \quad \text{and} \quad \sum_{i=1}^{k+1} p_i^* > \sum_{i=1}^{k+1} w_i. \tag{35}$$

   Subtracting the first inequality in (35) from the second, we have

$$p_{k+1}^* > w_{k+1}$$

   yielding $\gamma_+ > w_{k+1} \geq w_k \geq \cdots \geq w_{m_+}$. Summing the inequalities $p_{k+\ell}^* = \gamma_+ > w_{k+\ell}$ for $\ell = 2, 3, \dots, m_+ - k$, we have

$$\sum_{j=k+2}^{m_+} p_j^* > \sum_{j=k+2}^{m_+} w_j.$$

   Adding this to the second inequality of (35), we have

$$\sum_{i=1}^{m_+} p_i^* > \sum_{i=1}^{m_+} w_i.$$

   We thus conclude by Lemma 4.1.

3. Next, let $k \in \{m_+ + 1, \dots, d - m_-\}$. Assume $\sum_{i=1}^{k} p_i^* > \sum_{i=1}^{k} w_i$. Then

$$\sum_{i=1}^{k} p_i - p_i^* < \sum_{i=1}^{k} p_i - w_i$$

   However, the left-hand side is

$$\sum_{i=1}^{k} p_i - p_i^* = \sum_{i=1}^{m_+} p_i - m_+\gamma_+ = \varepsilon.$$

   Hence,

$$\varepsilon < \sum_{i=1}^{k} p_i - w_i \leq \sum_{i=1}^{d} (p_i - w_i)_+ = \frac{1}{2}\|w - p\|_1$$

   which is a contradiction.

4. Finally, we finish with a recursive proof similar to step 2. Assume the property holds for $k \in \{d-m_-, \dots, d-1\}$, but not for $k+1$. For this case too we have proven the base case $k = d - m_-$ in the previous step. We therefore assume

$$\sum_{i=1}^{k} p_i^* \leq \sum_{i=1}^{k} w_i, \qquad \text{and} \qquad \sum_{i=1}^{k+1} p_i^* > \sum_{i=1}^{k+1} w_i. \tag{36}$$

Subtracting the two equations, we have

$$p_{k+1}^* > w_{k+1}.$$

Since $\gamma_- = p_{k+1}^*$ we have $\gamma_- > w_{k+1} \geq w_{k+1} \geq \cdots \geq w_d$. Summing $p_{k+\ell}^* = \gamma_- > w_{k+\ell}$ for $\ell = 2, 3, \cdots, d$, we have

$$\sum_{i=k+2}^{d} p_i^* > \sum_{i=k+2}^{d} w_i.$$

Adding to the second inequality of (36), we find

$$1 = \sum_{i=1}^{d} p_i^* > \sum_{i=1}^{d} w_i$$

which contradicts the assumption that $p \in \Delta$. □

Thus, given $\sigma \in \mathcal{D}$ and $\varepsilon \in (0, 1]$, the state ρ_ε^* defined via (32) has $\rho_\varepsilon^* \prec \omega$ for any $\omega \in B_\varepsilon(\sigma)$, proving the first majorization relation of Theorem 3.1.

4.3. Constructing the maximal state in the majorization order (10)

As mentioned earlier, the minimum principle tells us that the entropy is minimized on a vertex of $B_\varepsilon(p)$. In the example of Figure 3 in $d = 3$, with $p = (0.21, 0.24, 0.55)$ and $\varepsilon = 0.1$, we considered the set M_w of points of Δ majorized by $w \in B_\varepsilon(p)$. In the same example in Figure 4, the minimum of the entropy over $B_\varepsilon(p)$ was found to be $p_* = (0.21 - \varepsilon, 0.24, 0.55 + \varepsilon)$. In Figure 6, we see that in fact $B_\varepsilon(p) \subseteq M_{p_*}$. That is, $w \prec p_*$ for any $w \in B_\varepsilon(p)$. In this section, we provide a general construction of p_*, and show that this property holds. As in the example, the construction will proceed by forming p_* by decreasing the smallest entries of p and increasing the largest entry. Intuitively, one “spreads out” the entries of p to form p_* so that M_{p_*} covers the most area, in order to cover $B_\varepsilon(p)$.

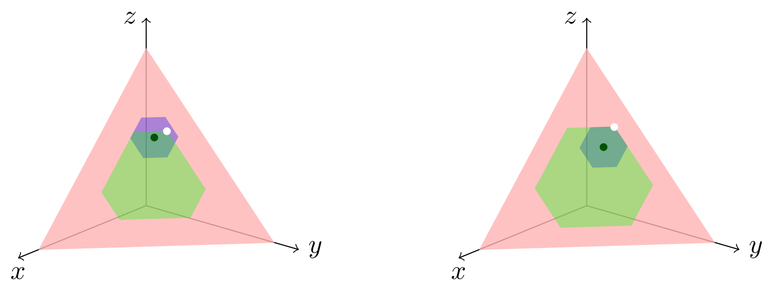


Figure 6: Left: The set M_w for the point $w \in B_\varepsilon(p)$ shown in white. Right: The set M_{p_*} for p_* the minimizer of H over $B_\varepsilon(p)$.

Let $\varepsilon \in (0, 1]$. We construct a probability vector $p_{*,\varepsilon}$ which we show has $M_{p_{*,\varepsilon}} \supseteq B_\varepsilon(p)$ by using the definition of majorization given in (5).

Definition 4.1 ($p_{*,\varepsilon}$)**.** If $p_d > \varepsilon$, let $p_{*,\varepsilon} = (p_1+\varepsilon, p_2, \ldots, p_d-\varepsilon)$. Otherwise, let $\ell \in \{1, \ldots, d-1\}$ be the largest index such that the sum of the ℓ smallest entries $Q_\ell := \sum_{j=d-\ell+1}^d p_j$ has $Q_\ell \leq \varepsilon$. If $\ell = d-1$, set $p_{*,\varepsilon} = (1, 0, \ldots, 0)$. Otherwise, choose $p_{*,\varepsilon} = (p_{j*})_{j=1}^d$ for

$$p_{j*} := \begin{cases} p_1 + \varepsilon & j = 1 \\ p_j & 2 \leq j \leq d-\ell-1 \\ p_{d-\ell+1} - (\varepsilon - Q_\ell) & j = d-\ell \\ 0 & j \geq d-\ell+1. \end{cases} \tag{37}$$

Note that the condition $p_d > \varepsilon$ is equivalent to every vector $w \in B_\varepsilon(p)$ having strictly positive entries. This is the case of Figure 4, and $p_{*,\varepsilon}$ reduces to $(p_1 + \varepsilon, p_2, \ldots, p_d - \varepsilon)$.

Using Definition 4.1, we verify that $p_* \in B_\varepsilon(p)$, as follows. First, if $p_d > \varepsilon$, then

$$\frac{1}{2}\|p_* - p\| = \frac{1}{2}(|p_1 + \varepsilon - p_1| + |p_d - \varepsilon - p_d|) = \varepsilon.$$

If $\ell = d-1$, then $p_* = (1, 0, \ldots, 0)$ and

$$\frac{1}{2}\|p_* - p\| = \sum_{j=1}^d (p_j - p_{j*})_+ = \sum_{j=2}^d p_j = Q_\ell \leq \varepsilon$$

yielding $p_* \in B_\varepsilon(p)$. Otherwise, $(p_{j*})_{j=1}^d$ is defined via (37). For $j \neq d-\ell$, that $p_{j*} \geq 0$ is immediate, and by maximality of ℓ, we have $p_{(d-\ell)*} = Q_{\ell+1} - \varepsilon > 0$. Additionally, $\sum_{j=1}^d p_{j*} = \sum_{j=1}^{d-\ell-1} p_j + Q_\ell - \varepsilon + \varepsilon = \sum_{j=1}^d p_j = 1$. Furthermore,

$$\begin{aligned} \sum_{j=1}^d |p_{j*} - p_j| &= |p_1 + \varepsilon - p_1| + \sum_{j=2}^{d-\ell-1} |p_j - p_j| + |p_{d-\ell} - (\varepsilon - Q_\ell) - p_{d-\ell}| + \sum_{j=d-\ell+1}^d |p_j - 0| \\ &= Q_\ell + (\varepsilon - Q_\ell) + \varepsilon = 2\varepsilon, \end{aligned}$$

so $p_* \in B_\varepsilon(p)$.

The following lemma shows that p_* is indeed the maximal distribution in the majorization order (27).

Lemma 4.2. *We have that $p_* \succ w$ for any $w \in B_\varepsilon(p)$.*

Proof. If $p_* = (1, 0, \ldots, 0)$, then the result is immediate. If $p_* = (p_1 + \varepsilon, \ldots, p_d - \varepsilon)$, then consider $\ell = 0$ and $Q_0 = 0$ in the following. Now, $p_* = (p_{j*})$ for p_{j*} defined via (37). Our task is to show that for any $w = (w_j)_{j=1}^d \in \Delta$,

$$\sum_{j=1}^k w_j \leq \sum_{j=1}^k p_{j*} \tag{38}$$

holds for each $k \in \{1, \ldots, d-1\}$. Equality in (38) obviously holds for $k = d$ since $w, p_* \in \Delta$.

Since $\sum_{j=1}^d (w_j - p_j)_+ = \sum_{j=1}^d (w_j - p_j)_- \leq \varepsilon$, in particular $\sum_{j=1}^k (w_j - p_j) \leq \varepsilon$ and therefore

$$\sum_{j=1}^k w_j \leq \varepsilon + \sum_{j=1}^k p_j.$$

For $k \leq d - \ell - 1$, we have $\sum_{j=1}^{k} p_{j*} = \varepsilon + \sum_{j=1}^{k} p_j$, yielding (38) in this case. On the other hand, for $k \geq d - \ell$ we have $\sum_{j=1}^{k} p_{j*} = \sum_{j=1}^{d} p_{j*} = 1$. Since $\sum_{j=1}^{k} w_j \leq 1$, this completes the proof. $\square$

Given $\varepsilon \in (0, 1]$ and a quantum state $\rho \in \mathcal{D}$, with eigen-decomposition $\rho = \sum_{i=1}^{d} p_i |i\rangle\langle i|$ in the sorted eigenbasis for which $\rho = \mathrm{Eig}^{\downarrow}(\rho)$, we define

$$\rho_{*,\varepsilon} = \sum_{i=1}^{d} p_{i*} |i\rangle\langle i| \tag{39}$$

where $p_{*,\varepsilon}$ is defined via Definition 4.1 where $p = \vec{\lambda}(\rho)$. Lemma 4.2 therefore proves $\rho_{*,\varepsilon} \succ \omega$ for any $\omega \in B_\varepsilon(\rho)$, proving the second majorization relation of Theorem 3.1. The state $\rho_{*,\varepsilon}$ is unique up to unitary equivalence as follows. If another state $\tilde{\rho} \in B_\varepsilon(\rho)$ had $\tilde{\rho} \succ \omega$ for all $\omega \in B_\varepsilon(\rho)$, then in particular, $\tilde{\rho} \succ \rho_{*,\varepsilon} \succ \tilde{\rho}$, which implies that $\tilde{\rho}$ and $\rho_{*,\varepsilon}$ are unitarily equivalent.

4.4. Uniqueness of construction for ρ_ε^*

The uniqueness claims in the construction of ρ_ε^* in Section 4.1 follow from the following lemma. These properties prove useful in Section 5.3.

Lemma 4.3. *Let $\rho \in \mathcal{D}(\mathcal{H})$ have eigenvalues $\vec{\lambda}(\rho) = (\lambda_1, \ldots, \lambda_d)$. Assume $\frac{1}{2}\|\rho - \tau\|_1 > \varepsilon$, where $\tau = \mathbb{1}/d$, i.e. $\tau \notin B_\varepsilon(\rho)$. There is a unique pair $(\gamma_-, m_-) \in [0, 1] \times \{1, \ldots, d-1\}$ such that*

$$\sum_{i=d-m_-+1}^{d} |\gamma_- - \lambda_i| = \varepsilon \quad \text{and} \quad \lambda_{d-m_-+1} < \gamma_- \leq \lambda_{d-m_-}, \tag{40}$$

and similarly, a unique pair $(\gamma_+, m_+) \in [0, 1] \times \{1, \ldots, d-1\}$ with

$$\sum_{j=1}^{m_+} |\gamma_+ - \lambda_i| = \varepsilon \quad \text{and} \quad \lambda_{m_++1} \leq \gamma_+ < \lambda_{m_+}. \tag{41}$$

We define the index sets

$$I_H = \{1, \ldots, m_+\}, \quad I_M = \{m_+ + 1, \ldots, d - m_-\}, \quad I_L = \{d - m_- + 1, \ldots, d\} \tag{42}$$

corresponding to the "highest" m_+ eigenvalues, "middle" $d - m_- - m_+$ eigenvalues, and "lowest" m_- eigenvalues of ρ, respectively.

We have the following properties. The numbers γ_- and γ_+ are such that $\gamma_- < \frac{1}{d} < \gamma_+$, and we have the following characterizations of the pairs (γ_-, m_-) and (γ_+, m_+): For any $m \in \mathbb{N}$, defining the functions

$$\gamma_-^{(m)} := \frac{1}{m}\left(\sum_{j=d-m+1}^{d} \lambda_j + \varepsilon\right), \qquad \gamma_+^{(m)} := \frac{1}{m}\left(\sum_{j=1}^{m} \lambda_j - \varepsilon\right), \tag{43}$$

we have $\gamma_- = \gamma_-^{(m_-)}$ and $\gamma_+ = \gamma_+^{(m_+)}$, where m_- and m_+ are, respectively, the unique solutions of the following:

$$\begin{aligned} \lambda_{d-m'_-+1} < \gamma_- \leq \lambda_{d-m'_-} \quad &: \quad m'_- \in \{1, \ldots, d-1\}, \\ \lambda_{m'_++1} \leq \gamma_+ < \lambda_{m'_+} \quad &: \quad m'_+ \in \{1, \ldots, d-1\}, \end{aligned} \tag{44}$$

and satisfy

$$m_- = \min\{m' \in \{1,\ldots,d-r\} : \gamma_-(m') \leq \lambda_{d-m'_-}\},$$
$$m_+ = \min\{m' \in \{1,\ldots,r\} : \gamma_+^{(m')} \geq \lambda_{m'+1}\}, \tag{45}$$

where r is defined by

$$\lambda_{r+1} < \frac{1}{d} \leq \lambda_r. \tag{46}$$

Proof of Lemma 4.3 Here we prove the results pertaining to the pair (γ_+, m_+). The results for the pair (γ_-, m_-) can be obtained analogously. Note that if any pair $(\gamma_+, m_+) \in [0,1] \times \{1,\ldots,d-1\}$ satisfies (40) then we have

$$\varepsilon = \sum_{i=1}^{m_+} |\gamma_+ - \lambda_i| = \sum_{i=1}^{m_+} (\lambda_i - \gamma_+) = \sum_{i=1}^{m_+} \lambda_i - m_+\gamma_+,$$

implying $\gamma_+ = \frac{1}{m_+}\left(\sum_{i=1}^{m_+} \lambda_i - \varepsilon\right) = \gamma_+^{(m_+)}$. Conversely, if for some $m'_+ \in \{1,\ldots,d-1\}$, the corresponding value $\gamma_+^{(m'_+)}$ satisfies $\lambda_{m'_+ +1} \leq \gamma_+^{(m'_+)} < \lambda_{m'_+}$, then

$$\varepsilon = \sum_{i=1}^{m'_+} \lambda_i - m'_+\gamma_+^{(m'_+)} = \sum_{i=1}^{m'_+} (\lambda_i - \gamma_+^{(m'_+)}) = \sum_{i=1}^{m'_+} |\gamma_+^{(m'_+)} - \lambda_i|$$

and in particular

$$\sum_{i=1}^{m'_+} |\gamma_+^{(m'_+)} - \lambda_i| = \varepsilon.$$

Hence, the existence and uniqueness of (γ_+, m_+) satisfying (40) is equivalent to the existence and uniqueness of $m_+ \in \{1,\ldots,d-1\}$ such that $\gamma_+^{(m_+)}$ satisfies $\lambda_{m_+ +1} \leq \gamma_+^{(m_+)} < \lambda_{m_+}$.

Next, we show that

$$m_+ = \min\{m'_+ \in \{1,\ldots,r\} : \gamma_+^{(m'_+)} \geq \lambda_{m'_+ +1}\}$$

by checking that the minimum exists and uniquely solves $\lambda_{m_+ +1} \leq \gamma_+^{(m_+)} < \lambda_{m_+}$. The proof is then completed by showing that we must have $\gamma_+^{(m_+)} > \frac{1}{d}$. The steps of the construction are elucidated below.

Step 1. $\{m'_+ \in \{1,\ldots,r\} : \gamma_+^{(m'_+)} \geq \lambda_{m'_+ +1}\} \neq \emptyset$.

Let us assume the contrary. Then, in particular, that $\gamma_+^{(r)} < \lambda_{r+1}$. By substituting $\gamma_+^{(r)} = \frac{1}{r}\sum_{i=1}^r \lambda_i - \frac{1}{r}\varepsilon$ in this inequality, we have

$$\sum_{i=1}^{r} \lambda_i < r\lambda_{r+1} + \varepsilon \tag{47}$$

Using $\mathrm{Tr}(\rho - \tau) = 0$, we have $\frac{1}{2}\|\rho - \tau\|_1 = \mathrm{Tr}(\tau - \rho)_-$. Using the definition of r, eq. (46), this can be written as

$$\frac{1}{2}\|\rho - \tau\|_1 = \sum_{i=1}^{r}(\lambda_i - \frac{1}{d}) = \sum_{i=1}^{r}\lambda_i - \frac{r}{d}. \tag{48}$$

Employing (47), and using the definition of r, we find that

$$\frac{1}{2}\|\rho - \tau\|_1 < r\lambda_{r+1} + \varepsilon - \frac{r}{d} = \varepsilon + r(\lambda_{r+1} - \frac{1}{d}) < \varepsilon. \tag{49}$$

That is, $\tau \in B_\varepsilon(\rho)$, which contradicts our assumption.

Step 2. The value $x := \min\{m'_+ \in \{1, \ldots, r\} : \gamma_+^{(m'_+)} \geq \lambda_{m'_+ + 1}\}$ solves $\lambda_{x+1} \leq \gamma_+^{(x)} < \lambda_x$.

If $x = 1$, then using eq. (43) we see that $\lambda_1 > \lambda_1 - \varepsilon = \gamma_+^{(1)}$, and hence $\lambda_{x+1} \leq \gamma_+^{(x)} < \lambda_x$. Otherwise, by minimality of x, we have $\gamma_+^{(x-1)} < \lambda_x$. We first establish that for $m'_+ \in \{2, 3, \ldots, d\}$,

$$\gamma_+^{(m'_+-1)} < \lambda_n \iff \gamma_+^{(m'_+)} < \lambda_n \tag{50}$$

and the result follows by taking $m'_+ = x$. To prove (50), we write

$$\begin{aligned}
\gamma_+^{(m'_+-1)} < \lambda_n \iff & \frac{1}{m'_+ - 1}\Big[\sum_{j=1}^{m'_+-1} \lambda_j - \varepsilon\Big] < \lambda_n \\
\iff & \sum_{j=1}^{m'_+-1} \lambda_j - \varepsilon < (m'_+ - 1)\lambda_n = m'_+\lambda_n - \lambda_n \\
\iff & \sum_{j=1}^{m'_+} \lambda_j - \varepsilon < m'_+\lambda_n \\
\iff & \frac{1}{m'_+}\Big[\sum_{j=1}^{m'_+} \lambda_j - \varepsilon\Big] < \lambda_n \\
\iff & \gamma_+^{(m'_+)} < \lambda_n.
\end{aligned}$$

Step 3. Uniqueness of m_+ satisfying $\lambda_{m_++1} \leq \gamma_+^{(m_+)} < \lambda_{m_+}$.

By substituting the inequality $\lambda_{m'_+-1} \geq \lambda_n$ into (50), we find for $m'_+ \in \{2, 3, \ldots, d\}$,

$$\lambda_n > \gamma_+^{(m'_+)} \implies \lambda_{m'_+-1} > \gamma_+^{(m'_+-1)}. \tag{51}$$

Now, assume that there exists a y satisfying $y > x > 0$ for which $\lambda_y > \gamma_+^{(y)} \geq \lambda_{y+1}$. By applying the implication (51) a total of $(y - x - 1)$ times, we see that

$$\lambda_y > \gamma_+^{(y)} \implies \lambda_{y-1} > \gamma_+^{(y-1)} \implies \cdots \implies \lambda_{x+1} > \gamma_+^{(x+1)}$$

which by (50) is equivalent to $\gamma_+^{(x)} < \lambda_{x+1}$. This contradicts the assumption that $\gamma_+^{(x)} \geq \lambda_{x+1}$. Hence, such a y cannot exist.

Step 4. $\gamma_+^{(m_+)} > \frac{1}{d}$.

We prove this by showing that if $\gamma_+^{(m_+)} \leq \frac{1}{d}$ then we obtain a contradiction to the assumption $\frac{1}{2}\|\rho - \tau\|_1 > \varepsilon$ of Lemma 4.3.

Assume $\gamma_+^{(m_+)} \leq \frac{1}{d}$. Then,

$$\gamma_+^{(m_+)} = \frac{1}{m_+}\sum_{j=1}^{m_+}\lambda_j - \frac{\varepsilon}{m_+} \leq \frac{1}{d} \iff \sum_{j=1}^{m_+}\left(\lambda_j - \frac{1}{d}\right) \leq \varepsilon.$$

Now, since $m_+ \leq r$ by (45), $\lambda_j - \frac{1}{d} \geq 0$ for each $j \in \{1, \dots, m_+\}$.

If $m_+ = r$, then

$$\sum_{j=1}^{r}\left(\lambda_j - \frac{1}{d}\right) \leq \varepsilon. \tag{52}$$

On the other hand, if $m_+ < r$, then $m_+ + 1 \leq r$. Then, using the assumption $\gamma_+^{(m_+)} \leq \frac{1}{d}$,

$$\frac{1}{d} \geq \gamma^{(m_+)} \geq \lambda_{m_++1} \geq \dots \geq \lambda_r \geq \frac{1}{d},$$

so $\lambda_{m_++1} = \lambda_{m_++2} = \dots = \lambda_r = \frac{1}{d}$. Then, (52) holds in this case as well. Then by (48),

$$\frac{1}{2}\|\rho - \tau\|_1 = \sum_{j=1}^{r}\left(\lambda_j - \frac{1}{d}\right) \leq \varepsilon,$$

which contradicts the assumption that $\tau \notin B_\varepsilon(\rho)$. $\square$

5. Further properties of ρ_ε^* and uniform continuity bounds

Proposition 3.7 presents further properties of the state ρ_ε^*. This result is proven in Section 5.3 below. First, let us discuss an application of these properties: uniform continuity bounds for (h, ϕ)-entropies.

5.1. Uniform continuity bounds

Consider an (h, ϕ) entropy $H_{(h,\phi)}$, and let $\varepsilon \in (0, 1]$, and $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with $T(\rho, \sigma) \leq \varepsilon$. If $H_{(h,\phi)}(\rho) \geq H_{(h,\phi)}(\sigma)$, then since $\rho \in B_\varepsilon(\sigma)$,

$$|H(\rho) - H(\sigma)| = H(\rho) - H(\sigma) \leq \max_{\omega \in B_\varepsilon(\sigma)} H(\omega) - H(\sigma) = H(\sigma_\varepsilon^*) - H(\sigma) \tag{53}$$

where the last equality follows from the first majorization relation in eq. (10) and the strict Schur concavity of $H_{(h,\phi)}$. Similarly, if $H_{(h,\phi)}(\sigma) \geq H_{(h,\phi)}(\rho)$, eq. (53) holds with σ (resp. σ_ε^*) replaced by ρ (resp. ρ_ε^*). Hence, in general,

$$|H_{(h,\phi)}(\rho) - H_{(h,\phi)}(\sigma)| \leq \max\{\Delta_\varepsilon(\rho), \Delta_\varepsilon(\sigma)\} \leq \max_{\omega \in \mathcal{D}(\mathcal{H})} \Delta_\varepsilon(\omega), \tag{54}$$

where

$$\begin{aligned} \Delta_\varepsilon : \qquad & \mathcal{D}(\mathcal{H}) \to \mathbb{R}_{\geq 0} \\ & \omega \mapsto H_{(h,\phi)} \circ \mathcal{M}_\varepsilon(\omega) - H_{(h,\phi)}(\omega), \end{aligned} \tag{55}$$

and $\mathcal{M}_\varepsilon$ is the *majorization-minimizer map*,

$$\begin{aligned} \mathcal{M}_\varepsilon : \qquad \mathcal{D}(\mathcal{H}) &\to \mathcal{D}(\mathcal{H}) \\ \omega &\mapsto \omega_\varepsilon^*. \end{aligned} \tag{56}$$

This map is defined explicitly by eq. (32) in Section 4.1. Note that $\Delta_\varepsilon(\omega) \geq 0$ for $\omega \in \mathcal{D}(\mathcal{H})$ follows from the Schur concavity of the (h, ϕ)-entropy. To prove Theorem 3.8, it remains to maximize Δ_ε over $\mathcal{D}(\mathcal{H})$.

We show that for (h, ϕ)-entropies for which h is concave and ϕ (strictly) convex, Δ_ε is a Schur convex function on $\mathcal{D}(\mathcal{H})$. We defer the proof of this to Section 5.2.

Theorem 5.1. *Assume h is concave and ϕ is strictly concave. Let $\varepsilon \in (0, 1]$. Then $\Delta_\varepsilon : \mathcal{D}(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ is Schur convex. That is, if $\rho \prec \sigma$,*

$$\Delta_\varepsilon(\rho) \leq \Delta_\varepsilon(\sigma).$$

Moreover, $\Delta_\varepsilon(\rho) = \Delta_\varepsilon(\sigma)$ implies $\lambda_+(\rho) = \lambda_+(\sigma)$. Lastly, if h is strictly concave, then Δ_ε is strictly Schur convex.

Note that if h is not strictly concave, Δ_ε need not be strictly Schur convex. In fact, for the von Neumann entropy we can find a counterexample to strict Schur convexity of Δ_ε. Setting $\rho = \mathrm{diag}(0.1, 0.2, 0.2, 0.5)$ and $\sigma = \mathrm{diag}(0.1, 0.15, 0.25, 0.5)$ yields $\rho \prec \sigma$ and that ρ and σ are not unitarily equivalent. However, for $\varepsilon \leq 0.05$, we have $\Delta_\varepsilon(\rho) = \Delta_\varepsilon(\sigma)$.

Corollary 5.2. *If h is concave, ϕ strictly concave, and $\varepsilon \in (0, 1]$, then Δ_ε achieves a maximum on $\mathcal{D}(\mathcal{H})$, and moreover* $\operatorname{argmax} \Delta_\varepsilon = \mathcal{D}_{pure}(\mathcal{H})$.

Proof. Since any pure state ψ satisfies $\rho \prec \psi$ for every $\rho \in \mathcal{D}(\mathcal{H})$, we have $\Delta_\varepsilon(\psi) \geq \Delta_\varepsilon(\rho)$ for every $\rho \in \mathcal{D}(\mathcal{H})$. Therefore, $\mathcal{D}_{\text{pure}}(\mathcal{H}) \subset \operatorname{argmax} \Delta_\varepsilon$. On the other hand, if $\omega \in \mathcal{D}(\mathcal{H})$ has $\omega \in \operatorname{argmax} \Delta_\varepsilon$, then

$$\Delta_\varepsilon(\omega) = \Delta_\varepsilon(\psi)$$

for a pure state ψ. Therefore, $\lambda_+(\omega) = \lambda_+(\psi) = 1$, and ω must be a pure state. □

Using these results, the proof of Theorem 3.8 is completed as follows. Let ψ be any pure state, $\psi \in \mathcal{D}_{\text{pure}}(\mathcal{H})$. Then for any $\omega \in \mathcal{D}(\mathcal{H})$, we have $\omega \prec \psi$. Therefore, by Theorem 5.1, we have $\Delta_\varepsilon(\omega) \leq \Delta_\varepsilon(\psi)$, for any $\omega \in \mathcal{D}(\mathcal{H})$, and in particular for $\omega \in \{\rho, \sigma\}$. Therefore, by (54) we have

$$|H_{(h,\phi)}(\rho) - H_{(h,\phi)}(\sigma)| \leq \Delta_\varepsilon(\psi).$$

By computing $\Delta_\varepsilon(\psi)$ using the form given in Proposition 3.7(g), we obtain the right-hand side of eq. (24).

It remains to check under which conditions equality occurs in (24). Assume without loss of generality that $H_{(h,\phi)}(\rho) \geq H_{(h,\phi)}(\sigma)$. Equality in (54) is equivalent to $\sigma \in \mathcal{D}_{\text{pure}}(\mathcal{H})$ by Corollary 5.2. Next, since the (h, ϕ)-entropy is strictly Schur concave and $\sigma_\varepsilon^* \prec \rho$, equality in (53) is equivalent to the fact that ρ is unitarily equivalent to σ_ε^*. The expression for σ_ε^* when $\sigma \in \mathcal{D}_{\text{pure}}(\mathcal{H})$ is given in Proposition 3.7(g). This completes the proof. □

Theorem 5.1 does not extend to the α-Rényi entropy for $\alpha > 1$, in which case h is convex and ϕ strictly convex. This is discussed in the remark following Lemma 5.3, and is illustrated in Figure 7.

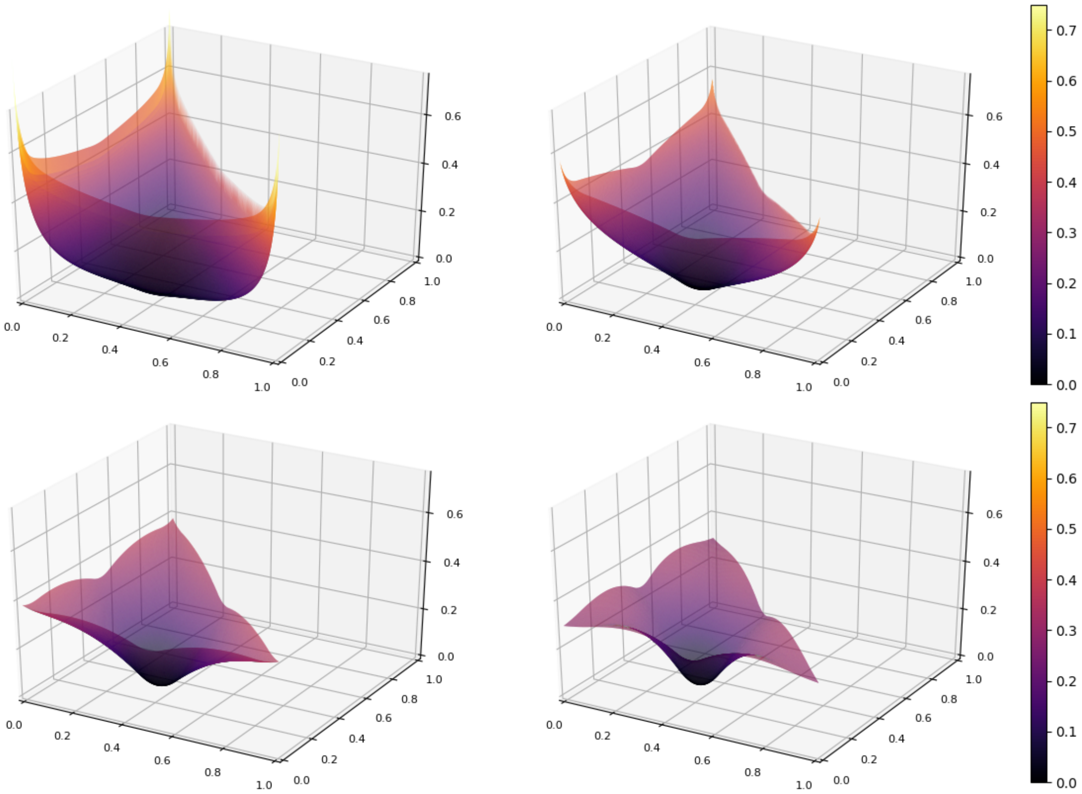


Figure 7: In dimensions $d = 3$, we parametrize $\sigma = \mathrm{diag}(x, y, 1-x-y)$, and plot $(x, y) \mapsto \Delta_\varepsilon(\sigma)$ for $\varepsilon = 0.1$, with $H_{(h,\phi)} = S_\alpha$, the Rényi entropy. That is, above each (x, y) in the xy-plane, the value of $\Delta_\varepsilon(\mathrm{diag}(x, y, 1-x-y))$ is plotted. Top row, left: $\alpha = \frac{1}{2}$, right: $\alpha = 1$. Bottom row, left: $\alpha = 1.5$, right: $\alpha = 2$. The three points $(0, 0)$, $(0, 1)$, $(1, 0)$ in the xy-plane correspond to the pure states $\mathrm{diag}(0, 0, 1)$, $\mathrm{diag}(0, 1, 0)$, and $\mathrm{diag}(1, 0, 0)$, respectively. The central point $(\frac{1}{3}, \frac{1}{3})$ corresponds to the completely mixed state $\tau = \frac{1}{3}\mathbb{1}$. We observe for $\alpha = \frac{1}{2}$ and $\alpha = 1$ the maximum of Δ_ε appears to occur at the pure states. On the other hand, for $\alpha = 1.5$, the maximum is along the boundary (i.e. for a state σ with exactly one zero eigenvalue), and for $\alpha = 2$, the maximum occurs at states without any zero eigenvalues.

5.2. Proof of Theorem 5.1

5.2.1. Reducing to $h = \mathrm{id}$

Our first task is to reduce to the case when $h = \mathrm{id}$, i.e. $h(x) = x$ for all $x \in \mathbb{R}$. Fix $\varepsilon \in (0, 1]$ and $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ such that $\rho \prec \sigma$ and ρ and σ are not unitarily equivalent. Let us define four variables

$$a := H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\rho), \qquad b := H_{(\mathrm{id},\phi)}(\rho), \qquad c := H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\sigma), \qquad d := H_{(\mathrm{id},\phi)}(\sigma)$$

which are non-negative real numbers. Theorem 5.1 is the statement that

$$h(a) - h(b) \leq h(c) - h(d). \tag{57}$$

Lemma 5.3. *Let h be concave, and ϕ strictly concave. If $a - b \leq c - d$, then (57) holds. Moreover, if h is strictly concave, then (57) holds with strict inequality.*

Proof. By the strict Schur concavity of the (id, ϕ)-entropy, we have $b < a$ and $d < c$, and by Proposition 3.7 (b), we have $b > d$ and $a > c$. Therefore, since h is concave, we apply Proposition A.1 to obtain

$$\frac{h(b) - h(d)}{b - d} \geq \frac{h(a) - h(c)}{a - c}.$$

That is,

$$[h(b) - h(d)]\frac{a - c}{b - d} \geq h(a) - h(c)$$

Since we have $a - c \leq b - d$ using the assumption, then $\frac{a-c}{b-d} \leq 1$, and therefore

$$h(b) - h(d) \geq h(a) - h(c)$$

and adding $h(c) - h(b)$ to each side yields (57). □

Therefore, it remains to establish $a - b \leq c - d$, which is Theorem 5.1 when $h = \mathrm{id}$.

Remark. An extension of Theorem 3.8 to treat the α-Rényi entropy for $\alpha > 1$ would need to address the case in which h is convex and strictly decreasing, and ϕ is strictly convex. In this case, $\rho \mapsto \operatorname{Tr}\phi(\rho)$ is Schur convex, and we have $a < b$, $c < d$, $b < d$, and $a < c$. The analog to Lemma 5.3 would be to show that $a - b \geq c - d$ implies (57). However, repeating the proof of Lemma 5.3 in this case yields e.g.

$$[h(b) - h(d)]\frac{c - a}{d - b} \leq h(a) - h(c)$$

which is inconclusive in showing (57) when $a - b \geq c - d$. This is the technical reason this proof does not extend to the α-Rényi entropy for $\alpha > 1$.

In fact, the associated quantity Δ_ε for an α-Rényi entropy with $\alpha > 1$ is not Schur convex. For the example stated after Theorem 5.1, it can be shown that choosing $H_{(h,\phi)} = S_\alpha$ for $\alpha > 1$ yields $\Delta_\varepsilon(\rho) > \Delta_\varepsilon(\sigma)$.

5.2.2. The case $h = \mathrm{id}$

We prove Theorem 5.1 in several steps. First, we use the semigroup property of $\mathcal{M}_\varepsilon$ to decompose Δ_ε for $\varepsilon = \varepsilon_1 + \varepsilon_2$ in terms of ε_1 and ε_2 in Lemma 5.4. Then we define a quantity $\delta(\rho, \sigma)$ in Definition 5.1 such that for $\varepsilon \leq \delta(\rho, \sigma)$, we can show that $\Delta_\varepsilon(\rho) \leq \Delta_\varepsilon(\sigma)$ if $\rho \prec \sigma$ (Lemma 5.6), using properties of $\delta(\rho, \sigma)$ presented in Lemma 5.5. Finally, we show that for arbitrary $\varepsilon \in (0, 1]$, we can use Lemma 5.4 finitely many times to prove Theorem 5.1. We state the lemmas here but defer their proofs to Section 5.2.3.

Lemma 5.4. *Let $\rho \in \mathcal{D}(\mathcal{H})$, and $\varepsilon_1, \varepsilon_2 \in (0, 1]$ with $\varepsilon_1 + \varepsilon_2 \leq 1$. Then*

$$\Delta_{\varepsilon_1+\varepsilon_2}(\rho) = \Delta_{\varepsilon_1} \circ \mathcal{M}_{\varepsilon_2}(\rho) + \Delta_{\varepsilon_2}(\rho).$$

Definition 5.1 ($\delta(\rho, \sigma)$)**.** Let $\rho \in \mathcal{D}(\mathcal{H})$ for $\rho \neq \tau$. Let $\mu_1 > \mu_2 > \cdots > \mu_\ell$ denote the distinct ordered eigenvalues of ρ, and define

$$\delta(\rho) = \min\{k_+(\rho)(\mu_1 - \mu_2), k_-(\rho)(\mu_{\ell-1} - \mu_\ell)\}. \tag{58}$$

For $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with $\rho \neq \tau \neq \sigma$, define

$$\delta(\rho, \sigma) = \min\{\delta(\rho), \delta(\sigma)\}. \tag{59}$$

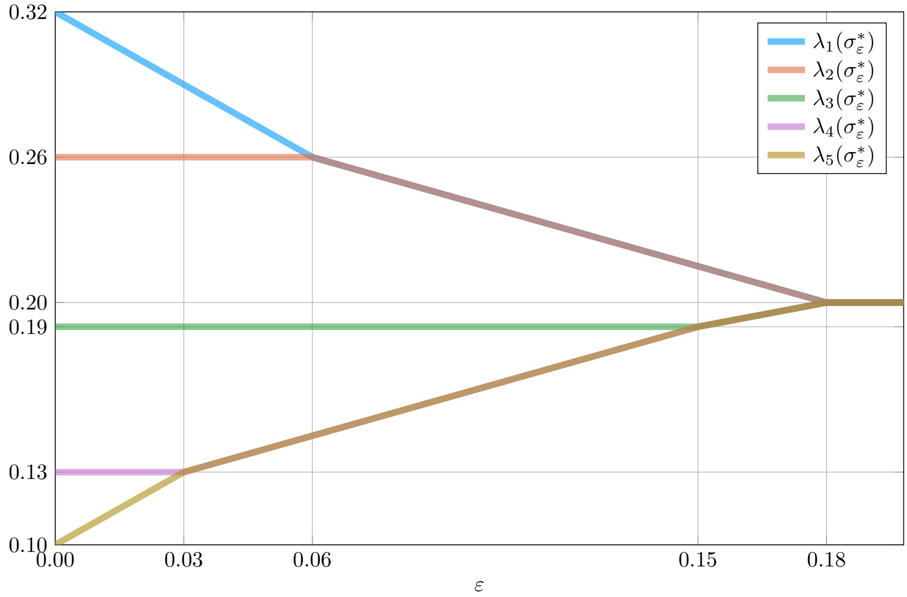


Figure 8: For the 5-dimensional state $\sigma = \operatorname{diag}(0.32, 0.26, 0.19, 0.13, 0.10)$, the spectrum of $\sigma^*_\varepsilon = \mathcal{M}_\varepsilon(\sigma)$ is plotted as a function of ε. This plot is a continuous (in ε) analog to the type of plot shown in Figure 5, which shows the spectrum of σ^*_ε at two discrete points, $\varepsilon = 0$ and $\varepsilon = 0.07$, in a different example. Here, at $\varepsilon = 0$, the five lines correspond to the five eigenvalues of σ, each with multiplicity one. For $\varepsilon \leq 0.03$, $\sigma^*_\varepsilon = \operatorname{diag}(0.32 - \varepsilon, 0.26, 0.19, 0.13, 0.10+\varepsilon)$ and differs from σ only in the smallest and largest eigenvalue. When ε reaches 0.03, the multiplicity of the smallest eigenvalue of σ^*_ε increases to 2. Between $\varepsilon = 0.03$ and $\varepsilon = 0.06$, again only the smallest and largest eigenvalues change, but the smallest eigenvalue has multiplicity 2. This process continues until every eigenvalue reaches $\frac{1}{d} = 0.2$ at $T(\sigma, \tau) = 0.18$.

For any $\varepsilon \leq \delta(\rho, \sigma)$, the map $\mathcal{M}_\varepsilon$ only "moves" the largest and smallest eigenvalue of ρ and of σ, as shown by the following result and illustrated through an example in Figure 8.

Lemma 5.5. *Let $\rho \neq \tau$. For any $\varepsilon \leq \delta(\rho)$, we have*

$$m_+(\rho, \varepsilon) = k_+(\rho), \quad \text{and} \quad m_-(\rho, \varepsilon) = k_-(\rho).$$

Moreover, if $\varepsilon = \delta(\rho)$ then either $k_+(\mathcal{M}_\varepsilon(\rho)) > k_+(\rho)$ or $k_-(\mathcal{M}_\varepsilon(\rho)) > k_-(\rho)$.

Using this result, we can prove the Schur convexity of Δ_ε for ε small enough (depending on ρ and σ).

Lemma 5.6. *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ with $\rho \prec \sigma$. Let $\varepsilon \leq \delta(\rho, \sigma)$, defined by (59). Then*

$$\Delta_\varepsilon(\rho) \leq \Delta_\varepsilon(\sigma). \tag{60}$$

Moreover, equality in (60) implies that $\lambda_\pm(\rho) = \lambda_\pm(\sigma)$.

We can iterate this result using Lemma 5.4 to prove Theorem 5.1 in general.

Proof of Theorem 5.1 Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and $\varepsilon \in (0,1]$. Note that if $\sigma = \tau$, then $\rho \prec \sigma$ implies $\rho = \tau$, and hence $\Delta_\varepsilon(\rho) = 0 = \Delta_\varepsilon(\sigma)$. If $\rho = \tau \neq \sigma$, then $\Delta_\varepsilon(\rho) = 0 < \Delta_\varepsilon(\sigma)$ by the strict Schur concavity of the $H_{(h,\phi)}$ entropy. Therefore, we can assume $\rho \neq \tau \neq \sigma$.

Step 1. Set $\rho_1 = \rho$ and $\sigma_1 = \sigma$. If $\varepsilon \leq \delta_1 := \delta(\rho_1, \sigma_1)$, we conclude via Lemma 5.6. Otherwise, set $\varepsilon_1 = \varepsilon - \delta_1$. Then by Lemma 5.4,

$$\Delta_{\varepsilon_1+\delta_1}(\rho_1) = \Delta_{\varepsilon_1} \circ \mathcal{M}_{\delta_1}(\rho_1) + \Delta_{\delta_1}(\rho_1)$$

and

$$\Delta_{\varepsilon_1+\delta_1}(\sigma_1) = \Delta_{\varepsilon_1} \circ \mathcal{M}_{\delta_1}(\sigma_1) + \Delta_{\delta_1}(\sigma_1).$$

We invoke Lemma 5.6 to find $\Delta_{\delta_1}(\rho_1) \leq \Delta_{\delta_1}(\sigma_1)$; it remains to show

$$\Delta_{\varepsilon_1}(\mathcal{M}_{\delta_1}(\rho_1)) \leq \Delta_{\varepsilon_1}(\mathcal{M}_{\delta_1}(\sigma_1)).$$

Step 2. Set $\rho_2 = \mathcal{M}_{\delta_1}(\rho_1)$ and $\sigma_2 = \mathcal{M}_{\delta_1}(\sigma_1)$. If either $\rho_2 = \tau$ or $\sigma_2 = \tau$ we conclude by the argument presented at the start of the proof. Otherwise, we set $\delta_2 := \delta(\rho_2, \sigma_2)$ and proceed.

If $\varepsilon_1 \leq \delta_2$, we conclude by Lemma 5.6. Otherwise, set $\varepsilon_2 = \varepsilon_1 - \delta_2$, expand e.g.

$$\Delta_{\varepsilon_2+\delta_2}(\rho_2) = \Delta_{\varepsilon_2} \circ \mathcal{M}_{\delta_2}(\rho_2) + \Delta_{\delta_2}(\rho_2), \qquad \Delta_{\varepsilon_2+\delta_2}(\sigma_2) = \Delta_{\varepsilon_2} \circ \mathcal{M}_{\delta_2}(\sigma_2) + \Delta_{\delta_2}(\sigma_2),$$

and conclude $\Delta_{\delta_2}(\rho_2) \leq \Delta_{\delta_2}(\sigma_2)$ by Lemma 5.6. It remains to show

$$\Delta_{\varepsilon_2} \circ \mathcal{M}_{\delta_2}(\rho_2) \leq \Delta_{\varepsilon_2} \circ \mathcal{M}_{\delta_2}(\sigma_2).$$

Step k. We continue recursively: for $k \geq 3$, we define $\rho_k = \mathcal{M}_{\delta_{k-1}}(\rho_{k-1})$, $\sigma_k = \mathcal{M}_{\delta_{k-1}}(\sigma_{k-1})$. If either $\rho_k = \tau$ or $\sigma_k = \tau$, we conclude as before; otherwise, set $\delta_k = \delta(\rho_k, \sigma_k)$. If $\varepsilon_{k-1} \leq \delta_k$, we conclude by Lemma 5.6; otherwise, define $\varepsilon_k = \varepsilon_{k-1} - \delta_k$, expand by Lemma 5.4 to find

$$\Delta_{\varepsilon_k+\delta_k}(\rho_k) = \Delta_{\varepsilon_k} \circ \mathcal{M}_{\delta_k}(\rho_k) + \Delta_{\delta_k}(\rho_k), \qquad \Delta_{\varepsilon_k+\delta_k}(\sigma_k) = \Delta_{\varepsilon_k} \circ \mathcal{M}_{\delta_k}(\sigma_k) + \Delta_{\delta_k}(\sigma_k),$$

and conclude $\Delta_{\delta_k}(\rho_k) \leq \Delta_{\delta_k}(\sigma_k)$ by Lemma 5.6. At the end of step k, it remains to show that

$$\Delta_{\varepsilon_k} \circ \mathcal{M}_{\delta_k}(\rho_k) \leq \Delta_{\varepsilon_k} \circ \mathcal{M}_{\delta_k}(\sigma_k).$$

This process must terminate in less than $4d$ steps, as follows. At each step k for which the process does not conclude, we have either $\delta(\rho_k, \sigma_k) = \delta(\rho_k)$ and therefore $k_+(\rho_k) > k_+(\rho_{k-1})$ or $k_-(\rho_k) > k_-(\rho_{k-1})$ or else $\delta(\rho_k, \sigma_k) = \delta(\sigma_k)$ and therefore $k_+(\sigma_k) > k_+(\sigma_{k-1})$ or $k_-(\sigma_k) > k_-(\sigma_{k-1})$, by Lemma 5.5. Since $k_\pm(\omega) \leq d$ for $\omega \in \mathcal{D}(\mathcal{H})$ and one of each of the four integers $k_\pm(\rho), k_\pm(\sigma)$ increases at each step, there cannot be more than $4d$ steps in total. Note that $\Delta_\varepsilon(\rho) = \Delta_\varepsilon(\sigma)$ implies equality in the use of Lemma 5.6 in Step 1, which requires $\lambda_+(\rho) = \lambda_+(\sigma)$. □

5.2.3. Proof of lemmas

Proof of Lemma 5.4 We expand

$$\Delta_{\varepsilon_1+\varepsilon_2}(\rho) = H_{(h,\phi)} \circ \mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) - H_{(h,\phi)}(\rho).$$

By Proposition 3.7(c), we have $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \mathcal{M}_{\varepsilon_1} \circ \mathcal{M}_{\varepsilon_2}(\rho)$. Thus,

$$\begin{aligned}\Delta_{\varepsilon_1+\varepsilon_2}(\rho) &= H_{(h,\phi)} \circ \mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho)) - H_{(h,\phi)}(\mathcal{M}_{\varepsilon_2}(\rho)) + H_{(h,\phi)}(\mathcal{M}_{\varepsilon_2}(\rho)) - H_{(h,\phi)}(\rho)\\ &= \Delta_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho)) + \Delta_{\varepsilon_2}(\rho). \quad \square\end{aligned}$$

Proof of Lemma 5.5 We use the notation of Definition 5.1. We check that the choice $m = k_+(\rho)$ satisfies the definition of $m_+(\rho,\varepsilon)$, namely that the choice $m = k_+(\rho)$ solves (30).

- If $T(\rho,\tau) > \varepsilon$, then $\gamma_+^{(m)}(\rho,\varepsilon) = \frac{1}{m}\left(\sum_{i=1}^m \lambda_i(\rho) - \varepsilon\right)$. And indeed, taking $m = k_+(\rho)$ we find
$$\lambda_{k_++1}(\rho) = \mu_2 \leq \frac{1}{k_+}\left(\sum_{i=1}^{k_+} \lambda_i(\rho) - \varepsilon\right) = \mu_1 - \frac{\varepsilon}{k_+}$$
since $\frac{\varepsilon}{k_+} \leq \frac{1}{k_+}\delta(\rho,\sigma) \leq \mu_1 - \mu_2$. Additionally, $\mu_1 - \frac{\varepsilon}{k_+} < \mu_1 = \lambda_{k_+}(\rho)$. Therefore, $m = k_+(\rho)$ solves (30), hence $m_+(\rho,\varepsilon) = k_+(\rho)$.

- In the case $0 < T(\rho,\tau) \leq \varepsilon$. Then $\gamma_+(\rho,\varepsilon) = \frac{1}{d}$. Since $\rho \neq \tau$, we have $\lambda_{k_+(\rho)}(\rho) = \mu_1 > \frac{1}{d}$. Moreover,
$$k_+(\rho)\left(\mu_1 - \tfrac{1}{d}\right) \leq \mathrm{Tr}[(\rho-\tau)_+] = T(\rho,\tau) \leq \varepsilon \leq k_+(\rho)(\mu_1 - \mu_2)$$
and therefore $\mu_1 - \frac{1}{d} \leq \mu_1 - \mu_2$, yielding $\mu_2 \leq \frac{1}{d}$. Thus, $m_+(\rho,\varepsilon) = k_+(\rho)$.

Proving that $m_-(\rho,\varepsilon) = k_-(\rho)$ is analogous.

Next, consider $\varepsilon = \delta(\rho)$. If $0 < T(\rho,\tau) \leq \varepsilon$, then $\mathcal{M}_\varepsilon(\rho) = \tau$ (by Proposition 3.7(f)) and $d = k_+(\mathcal{M}_\varepsilon(\rho)) > k_+(\rho)$, by the assumption that $\rho \neq \tau$. Otherwise, without loss of generality, assume $\delta(\rho,\sigma) = k_+(\rho)(\mu_1 - \mu_2)$. We show that $k_+(\mathcal{M}_\varepsilon(\rho)) > k_+(\rho)$. By the above, $m_+(\rho,\varepsilon) = k_+(\rho)$, and therefore
$$\gamma_+(\rho,\varepsilon) = \mu_1 + \frac{\varepsilon}{k_+(\rho)} = \mu_1 + (\mu_1 - \mu_2) = \mu_2.$$
As
$$\mathcal{M}_\varepsilon(\rho) = \sum_{i=1}^{m_+(\rho,\varepsilon)} \gamma_+(\rho,\varepsilon)\,|i\rangle\langle i| + \sum_{i=m+1}^{d-m_-(\rho,\varepsilon)} \lambda_i(\rho)\,|i\rangle\langle i| + \sum_{i=d-m_-(\rho,\varepsilon)+1}^{d} \gamma_-(\rho,\varepsilon)\,|i\rangle\langle i|$$
by eq. (32) and $\lambda_{m_+(\rho,\varepsilon)+1}(\rho) = \mu_2$, we have that $k_+(\mathcal{M}_\varepsilon(\rho))$, the multiplicity of μ_2 for $\mathcal{M}_\varepsilon(\rho)$, is strictly larger than $k_+(\rho) = m_+(\rho,\varepsilon)$. □

Proof of Lemma 5.6 As in the proof of Theorem 5.1, if $\sigma = \tau$, then $\rho \prec \sigma$ implies $\rho = \tau$, and hence $\Delta_\varepsilon(\rho) = 0 = \Delta_\varepsilon(\sigma)$. If $\rho = \tau \neq \sigma$, then $\Delta_\varepsilon(\rho) = 0 < \Delta_\varepsilon(\sigma)$ by the strict Schur concavity of the $H_{\mathrm{id},\phi}$ entropy. Now, assume $\rho \neq \tau \neq \sigma$. We aim to show
$$H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\rho) - H_{(\mathrm{id},\phi)}(\rho) \leq H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\sigma) - H_{(\mathrm{id},\phi)}(\sigma). \tag{61}$$
By two applications of Lemma 5.5, we have $m_+(\rho,\varepsilon) = k_+(\rho)$, $m_-(\rho,\varepsilon) = k_-(\rho)$, $m_+(\sigma,\varepsilon) = k_+(\sigma)$, and $m_-(\sigma,\varepsilon) = k_-(\sigma)$. Therefore, by (7) and (32),
$$H_{(\mathrm{id},\phi)}(\mathcal{M}_\varepsilon(\rho)) = k_+(\rho)\phi(\gamma_+(\rho)) + \sum_{i=k_+(\rho)+1}^{d-k_-(\rho)} \phi(\lambda_i(\rho)) + k_-(\rho)\phi(\gamma_-(\rho))$$
since $h = \mathrm{id}$. The $\phi(\lambda_i(\rho))$ terms for $i = k_+(\rho)+1,\ldots,d-k_-(\rho)$ therefore cancel in $\Delta_\varepsilon(\rho)$ yielding
$$H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\rho) - H_{(\mathrm{id},\phi)}(\rho) = k_+(\rho)[\phi(\gamma_+(\rho,\varepsilon)) - \phi(\lambda_+(\rho))] + k_-(\rho)[\phi(\gamma_-(\rho,\varepsilon)) - \phi(\lambda_-(\rho))] \tag{62}$$

and similarly

$$H_{(\mathrm{id},\phi)} \circ \mathcal{M}_\varepsilon(\sigma) - H_{(\mathrm{id},\phi)}(\sigma) = k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))] + k_-(\sigma)[\phi(\gamma_-(\sigma,\varepsilon)) - \phi(\lambda_-(\sigma))]. \tag{63}$$

We conclude by invoking Lemma 5.7 below, which bounds the first term (resp. second term) of (62) by the first term (resp. second term) of (63). $\square$

Lemma 5.7. *For $\rho \prec \sigma$ with $\rho \neq \tau \neq \sigma$ and $0 < \varepsilon \leq \delta(\rho,\sigma)$, we have*

$$k_\pm(\rho)[\phi(\gamma_\pm(\rho,\varepsilon)) - \phi(\lambda_\pm(\rho))] \leq k_\pm(\sigma)[\phi(\gamma_\pm(\sigma,\varepsilon)) - \phi(\lambda_\pm(\sigma))] \tag{64}$$

and that equality in (64) implies $\lambda_\pm(\rho) = \lambda_\pm(\sigma)$.

To prove this result, we first recall a simple consequence of the majorization order $\rho \prec \sigma$.

Lemma 5.8. *If $\rho \prec \sigma$, then $T(\rho,\tau) \leq T(\sigma,\tau)$.*

Proof. If $\rho \prec \sigma$, then by Theorem 2-2 (b) of [AU82], we have $\rho = \Phi(\sigma)$ for a map $\Phi(\cdot) = \sum_i p_i U_i \cdot U_i^*$ where p_i is a finite probability distribution and each U_i is unitary. Φ is completely positive and trace-preserving (CPTP) as well as unital. Since $\Phi(\tau) = \tau$,

$$T(\rho,\tau) = T(\rho,\Phi(\tau)) = T(\Phi(\sigma),\Phi(\tau)) \leq T(\sigma,\tau)$$

where the inequality follows from the monotonicity of the trace distance under CPTP maps. $\square$

Proof of Lemma 5.7 We prove the case $+$ in eq. (64); the case $-$ is proved analogously. First, we have $\gamma_+(\rho,\varepsilon) \leq \gamma_+(\sigma,\varepsilon)$ and $\lambda_+(\rho) \leq \lambda_+(\sigma)$, using that $\rho \prec \sigma$ and $\mathcal{M}_\varepsilon\rho \prec \mathcal{M}_\varepsilon\sigma$ by Proposition 3.7. Moreover, by definition, $\gamma_+(\rho,\varepsilon) < \lambda_+(\rho)$ and $\gamma_+(\sigma,\varepsilon) < \lambda_+(\sigma)$. Therefore, by applying Proposition A.1 and multiplying by minus one, we have

$$\frac{\phi(\gamma_+(\rho,\varepsilon)) - \phi(\lambda_+(\rho))}{\lambda_+(\rho) - \gamma_+(\rho,\varepsilon)} \leq \frac{\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))}{\lambda_+(\sigma) - \gamma_+(\sigma,\varepsilon)}. \tag{65}$$

and that equality requires $\lambda_+(\rho) = \lambda_+(\sigma)$.

Now, we complete the proof of (64) in three cases.

Case 1: $T(\sigma,\tau) \leq \varepsilon$. Then $T(\rho,\tau) \leq \varepsilon$ as well, and

$$\lambda_+(\rho) - \gamma_+(\rho,\varepsilon) = \lambda_+(\rho) - \frac{1}{d}.$$

As shown in the proof of Lemma 5.5, the second-largest eigenvalue of ρ is less or equal to $\frac{1}{d}$. Therefore,

$$T(\rho,\tau) = \mathrm{Tr}[(\rho-\tau)_+] = k_+(\rho)\big(\lambda_+(\rho) - \tfrac{1}{d}\big),$$

and hence,

$$\lambda_+(\rho) - \gamma_+(\rho,\varepsilon) = \frac{1}{k_+(\rho)} T(\rho,\tau). \tag{66}$$

As $T(\sigma,\tau) \leq \varepsilon$, we likewise have $\lambda_+(\sigma) - \gamma_+(\sigma,\varepsilon) = \frac{1}{k_+(\sigma)} T(\sigma,\tau)$. Then (65) yields

$$\frac{k_+(\rho)[\phi(\gamma_+(\rho,\varepsilon)) - \phi(\lambda_+(\rho))]}{T(\rho,\tau)} \leq \frac{k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))]}{T(\sigma,\tau)}.$$

Since $T(\sigma,\tau) \geq T(\rho,\tau)$, we may bound the right-hand side by $\frac{k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))]}{T(\rho,\tau)}$. Then multiplying by $T(\rho,\tau)$ yields (64).

Case 2: $T(\rho,\tau) \leq \varepsilon < T(\sigma,\tau)$. In this case, (66) holds, and $\gamma_+(\sigma,\varepsilon) = \lambda_+(\sigma) - \frac{\varepsilon}{k_+(\sigma)}$. Therefore, (65) yields

$$\frac{k_+(\rho)[\phi(\gamma_+(\rho,\varepsilon)) - \phi(\lambda_+(\rho))]}{T(\rho,\tau)} \leq \frac{k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))]}{\varepsilon}.$$

Similarly to the previous case, the inequality $\varepsilon \geq T(\rho,\tau)$ bounds the right-hand side by $\frac{k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon))-\phi(\lambda_+(\sigma))]}{T(\rho,\tau)}$, and multiplying by $T(\rho,\tau)$ yields (64).

Case 3: $T(\rho,\tau) > \varepsilon$. Then $\gamma_+(\rho,\varepsilon) = \lambda_+(\rho) - \frac{\varepsilon}{k_+(\rho)}$, and $\gamma_+(\sigma,\varepsilon) = \lambda_+(\sigma) - \frac{\varepsilon}{k_+(\sigma)}$. Therefore, (65) yields

$$\frac{k_+(\rho)[\phi(\gamma_+(\rho,\varepsilon)) - \phi(\lambda_+(\rho))]}{\varepsilon} \leq \frac{k_+(\sigma)[\phi(\gamma_+(\sigma,\varepsilon)) - \phi(\lambda_+(\sigma))]}{\varepsilon}$$

and multiplying by ε yields (64).

Note that in each case, equality in (64) requires equality in (65). $\square$

5.3. Proof of Proposition 3.7

Let us recall Proposition 3.7.

Proposition 3.7 (Properties of $\mathcal{M}_\varepsilon$). *The (non-linear) map $\mathcal{M}_\varepsilon$ has following properties, for any $\varepsilon \in (0,1]$. Let $\rho \in \mathcal{D}(\mathcal{H})$.*

a. Maps states to states: $\mathcal{M}_\varepsilon : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H})$.

b. Minimal in majorization order: $\mathcal{M}_\varepsilon(\rho) \in B_\varepsilon(\rho)$ and for any $\omega \in B_\varepsilon(\rho)$, we have $\mathcal{M}_\varepsilon(\rho) \prec \omega$.

c. Semi-group property: if $\varepsilon_1, \varepsilon_2 \in (0,1]$ with $\varepsilon_1+\varepsilon_2 \leq 1$, we have $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \mathcal{M}_{\varepsilon_1} \circ \mathcal{M}_{\varepsilon_2}(\rho)$.

d. Majorization-preserving: let $\sigma \in \mathcal{D}(\mathcal{H})$ such that $\rho \prec \sigma$. Then $\mathcal{M}_\varepsilon(\rho) \prec \mathcal{M}_\varepsilon(\sigma)$.

e. $\tau = \frac{\mathbb{1}}{d}$ is the unique fixed point of $\mathcal{M}_\varepsilon$, i.e. the unique solution to $\rho = \mathcal{M}_\varepsilon(\rho)$ for $\rho \in \mathcal{D}(\mathcal{H})$. Moreover, for any $\rho \neq \tau$, $\mathcal{M}_\varepsilon(\rho)$ is not unitarily equivalent to ρ.

f. For any state $\rho \in B_\varepsilon(\tau)$, we have $\mathcal{M}_\varepsilon(\rho) = \tau$. If $\rho \notin B_\varepsilon(\tau)$, then $\mathcal{M}_\varepsilon(\rho)$ is on the boundary of $B_\varepsilon(\rho)$.

g. For any pure state $\psi \in \mathcal{D}_{pure}(\mathcal{H})$, the state $\mathcal{M}_\varepsilon(\psi)$ has the form

$$\mathcal{M}_\varepsilon(\psi) = \begin{cases} \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1}) & \varepsilon < 1 - \frac{1}{d} \\ \tau := \frac{\mathbb{1}}{d} & \varepsilon \geq 1 - \frac{1}{d}. \end{cases} \tag{22}$$

h. Saturates triangle inequality with $\tau := \frac{\mathbb{1}}{d}$:

$$\frac{1}{2}\|\rho - \tau\|_1 = \frac{1}{2}\|\tau - \mathcal{M}_\varepsilon(\rho)\|_1 + \frac{1}{2}\|\mathcal{M}_\varepsilon(\rho) - \rho\|_1. \tag{23}$$

The properties (a), (f) and (g) follow from the construction given in Section 4.1. Note in particular, that eq. (33) establishes the statement that $\rho_\varepsilon(\sigma)$ lies on the boundary of $B_\varepsilon(\sigma)$ when $\sigma_\varepsilon^* \neq \tau$. Property (c) is shown by Proposition 5.10 below. Property (b) is due to Theorem 3.1, and is proven in Section 4.1. Property (d) can be found in Lemma 2 of [HOS17]. The property (e) can be shown as follows. $\mathcal{M}_\varepsilon(\rho)$ is not unitarily equivalent to ρ for $\rho \neq \tau$ follows from the construction presented above, in particular, the fact that the eigenvalues of $\mathcal{M}_\varepsilon(\rho)$ differ from ρ. One immediately has that τ is a fixed point of $\mathcal{M}_\varepsilon$, and uniqueness follows from the fact that $\mathcal{M}_\varepsilon(\sigma)$ is not unitarily equivalent to σ for $\sigma \neq \tau$. Lastly, Property (h) is shown by Proposition 5.9 below.

Proposition 5.9. *The state ρ_ε^* saturates the triangle inequality for the completely mixed state $\tau := \frac{1}{d}\mathbb{1}$ and ρ, in that*

$$\frac{1}{2}\|\rho - \tau\|_1 = \frac{1}{2}\|\tau - \rho_\varepsilon^*\|_1 + \frac{1}{2}\|\rho_\varepsilon^* - \rho\|_1.$$

Proposition 5.10. *If $\varepsilon_1, \varepsilon_2 \in (0,1]$ and $\rho \in \mathcal{D}(\mathcal{H})$, we have*

$$\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho)).$$

Proof of Proposition 5.9 Note

$$\frac{1}{2}\|\rho_\varepsilon^* - \rho\|_1 = \min\left(\varepsilon, \frac{1}{2}\|\rho - \tau\|_1\right),$$

from the construction of ρ_ε^*. Now, if $\rho_\varepsilon^* = \tau$, we have the result immediately. Otherwise, for $\vec{\lambda}(\rho) = (\lambda_1, \ldots, \lambda_d)$,

$$\begin{aligned}
\|\tau - \rho_\varepsilon(\rho)\|_1 &= m_-\left(\frac{1}{d} - \gamma_-\right) + m_+\left(\gamma_+ - \frac{1}{d}\right) + \sum_{j=m_++1}^{d-m_-} |\tfrac{1}{d} - \lambda_j| \\
&= \frac{m_-}{d} - \sum_{j=1}^{m_+} \lambda_j - \varepsilon + \sum_{j=d-m_-+1}^{d} \lambda_j - \frac{m_+}{d} - \varepsilon + \sum_{j=m_++1}^{d-m_-} |\tfrac{1}{d} - \lambda_j| \\
&= \sum_{j=1}^{d} |\tfrac{1}{d} - \lambda_j| - 2\varepsilon \\
&= \|\tau - \rho\|_1 - 2\varepsilon = \|\tau - \rho\|_1 - \|\rho_\varepsilon^* - \rho\|_1. \quad \square
\end{aligned}$$

Proof of Proposition 5.10 We first treat the cases in which $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \tau$; that is, when $\frac{1}{2}\|\rho - \tau\|_1 \leq \varepsilon_1 + \varepsilon_2$. If $\frac{1}{2}\|\rho - \tau\|_1 \leq \varepsilon_2$, then $\mathcal{M}_{\varepsilon_2}(\rho) = \tau$ as well; since $\mathcal{M}_{\varepsilon_1}(\tau) = \tau$, we have $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho))$. Next, consider the case $\varepsilon_2 < \frac{1}{2}\|\rho - \tau\|_1 \leq \varepsilon_1 + \varepsilon_2$. To show $\mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho)) = \tau$, we need $\frac{1}{2}\|\mathcal{M}_{\varepsilon_2}(\rho) - \tau\|_1 \leq \varepsilon_1$. By Proposition 5.9,

$$\frac{1}{2}\|\tau - \mathcal{M}_{\varepsilon_2}(\rho)\|_1 = \frac{1}{2}\|\rho - \tau\|_1 - \frac{1}{2}\|\mathcal{M}_{\varepsilon_2}(\rho) - \rho\|_1 \leq \varepsilon_1 + \varepsilon_2 - \frac{1}{2}\|\mathcal{M}_{\varepsilon_2}(\rho) - \rho\|_1$$

but since $\frac{1}{2}\|\mathcal{M}_{\varepsilon_2}(\rho) - \rho\|_1 = \varepsilon_2$, using that $\mathcal{M}_{\varepsilon_2}(\rho) \neq \tau$, we have $\frac{1}{2}\|\tau - \mathcal{M}_{\varepsilon_2}(\rho)\|_1 \leq \varepsilon_1$ as required. This completes the proof of the cases for which $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \tau$.

For the remainder of the proof, we will consider the construction of the states generated by $\mathcal{M}_\varepsilon$ via their characterization in Lemma 4.3. The uniqueness results of that lemma will allow us to show that $\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho)$ indeed is the same state as $\mathcal{M}_{\varepsilon_1} \circ \mathcal{M}_{\varepsilon_2}(\rho)$.

Assume that $\frac{1}{2}\|\rho - \tau\|_1 > \varepsilon_1 + \varepsilon_2$. Let $\rho = \sum_{i=1}^d \lambda_i |i\rangle\langle i|$ and write

$$\mathcal{M}_{\varepsilon_2}(\rho) = \sum_{i \in I_L} \alpha_- |i\rangle\langle i| + \sum_{i \in I_M} \lambda_i |i\rangle\langle i| + \sum_{i \in I_H} \alpha_+ |i\rangle\langle i|$$

for $I_L = \{d-m_-+1, \ldots, d\}$, $I_M = \{m_++1, \ldots, d-m_-\}$, and $I_H = \{1, \ldots, m_+\}$, and (α_-, m_-) and (α_+, m_+) are determined by ρ and ε_2 via Lemma 4.3.

Now, consider

$$\mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho)) = \sum_{i \in I_{L'}} \beta_- |i\rangle\langle i| + \sum_{i \in I_{M'}} \lambda_i |i\rangle\langle i| + \sum_{i \in I_{H'}} \beta_+ |i\rangle\langle i| \tag{67}$$

with $I_{L'} = \{d-m'_-+1, \ldots, d\}$, $I_{M'} = \{m'_++1, \ldots, d-m'_-\}$, and $I_{H'} = \{1, \ldots, m'_+\}$, and where (β_-, m'_-) and (β_+, m_+) are determined by $\mathcal{M}_{\varepsilon_2}(\rho)$ and ε_1 via Lemma 4.3.

We aim to compare the expression (67) of $\mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho))$ to the following:

$$\mathcal{M}_{\varepsilon_1+\varepsilon_2}(\rho) = \sum_{i \in I_{L''}} \gamma_- |i\rangle\langle i| + \sum_{i \in I_{M''}} \lambda_i |i\rangle\langle i| + \sum_{i \in I_{H''}} \gamma_+ |i\rangle\langle i|$$

with $I_{L''} = \{d - m''_- + 1, \ldots, d\}$, $I_{M''} = \{m''_+ + 1, \ldots, d - m''_-\}$, and $I_{H''} = \{1, \ldots, m''_+\}$, and where (γ_-, m''_-) and (γ_+, m''_+) are determined by $(\rho, \varepsilon_1 + \varepsilon_2)$ via Lemma 4.3. That is, we wish to show $(\beta_-, m'_-) = (\gamma_-, m''_-)$, and $(\beta_+, m'_+) = (\gamma_+, m''_+)$. We only consider the first equality here; the second is very similar.

Let $(\nu)_{i=1}^d$ be the eigenvalues of $\mathcal{M}_{\varepsilon_2}(\rho)$. That is, $\nu_i = \alpha_-$ for $i \in I_L$, $\nu_i = \lambda_i$ for $i \in I_M$, and $\nu_i = \alpha_+$ for $i \in I_H$. Equation (44) for $\mathcal{M}_{\varepsilon_1}(\mathcal{M}_{\varepsilon_2}(\rho))$ in Lemma 4.3 yields

$$\nu_{m'_+} > \beta_+ = \frac{1}{m'_+} \left(\sum_{i \in I_{H'}} \nu_i - \varepsilon_1 \right) \geq \nu_{m'_++1}. \tag{68}$$

Thus,

- $m'_+ \geq m_+$. Otherwise, $\beta_+ = \alpha_+ - \frac{\varepsilon_1}{m'_+} < \alpha_+ = \nu_{m'_++1}$.
- $m'_+ \leq d-m_-$. Otherwise, $\beta_+ < \nu_{m'_+} = \alpha_- < \frac{1}{d}$, contradicting that $\beta_+ > \frac{1}{d}$ by Lemma 4.3.

Hence,

$$\begin{aligned}
\beta_+ &= \frac{1}{m'_+} \left(\sum_{i \in I_H} \nu_i + \sum_{m_++1}^{m'_+} \nu_i - \varepsilon_1 \right) = \frac{1}{m'_+} \left(m_+\alpha_+ + \sum_{m_++1}^{m'_+} \lambda_i - \varepsilon_1 \right) \\
&= \frac{1}{m'_+} \left(\sum_{i \in I_H} \lambda_i - \varepsilon_2 + \sum_{m_++1}^{m'_+} \lambda_i - \varepsilon_1 \right) = \frac{1}{m'_+} \left(\sum_{i \in I_{H'}} \lambda_i - \varepsilon_1 - \varepsilon_2 \right).
\end{aligned}$$

That is, $\beta_+ = \gamma_+(m'_+)$. It remains to show $m'_+ = m''_+$.

- If $m'_+ = m_+$, then $\nu_{m'_+} = \alpha_+ < \lambda_{m'_+}$ by equation (41) for $\mathcal{M}_{\varepsilon_2}(\rho)$.
- If $m'_+ > m_+$, then $\nu_{m'_+} = \lambda_{m'_+}$.

In either case, $\nu_{m'_+} \leq \lambda_{m'_+}$. Hence, (68) becomes

$$\lambda_{m_+} \geq \nu_{m'_+} > \beta_+ \geq \nu_{m_++1}.$$

Either $\nu_{m_++1} = \lambda_{m_++1}$, or $\nu_{m_++1} = \alpha_- > \lambda_{m_++1}$; in either case, $\beta_+ \geq \lambda_{m_++1}$. Thus, writing $\beta_+ = \gamma_+(m'_+)$, we have

$$\lambda_{m'_+} > \gamma_+(m'_+) \geq \lambda_{m'_++1}.$$

Equation (44) in Lemma 4.3 defines m''_+ as the unique solution of

$$\lambda_{m''_+} > \gamma_+(m''_+) \geq \lambda_{m''_++1}$$

and therefore $m''_+ = m'_+$. □

6. Beyond the classical case: convex optimization applied to conave functions over the ε-ball

The main results of sections Sections 4 and 5, while formulated in the language of quantum states, are essentially classical in that the first step of the proofs consists of a reduction to the case of probability vectors in $\mathbb{R}^d$. In this section, we motivate the construction of ρ_ε^* by finding necessary and sufficient conditions for maximizing concave functions over the ε-ball. Moreover, we can treat— to some extent— functions beyond the the spectral theorem, which act non-trivially on the eigenprojections as well. In particular, we will find necessary and sufficient conditions for a state ξ to maximize the conditional entropy over the ε-ball (Corollary 3.3). Unfortunately, we will not be able to determine a formula for a specific maximizer, and we don't expect such a maximizer to have the universality of ρ_ε^* (which maximizes any Schur concave function over the ε-ball $B_\varepsilon(\rho)$).

6.1. Tools from convex optimization

Let $\mathcal{X}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be a function. In our applications, we take $\mathcal{X} = \mathcal{B}_{\text{sa}}(\mathcal{H})$ equipped with the Hilbert-Schmidt inner product $\langle \cdot, \cdot \rangle_{\text{HS}}$. Let $\operatorname{dom} f = \{x \in \mathcal{X} : f(x) < \infty\}$ and assume $\operatorname{dom} f \neq \emptyset$. Let $\mathcal{X}^* = \mathcal{B}(\mathcal{X}, \mathbb{R})$ the set of bounded linear maps from $\mathcal{X}$ to $\mathbb{R}$, equipped with the dual norm $\|\ell\|_* = \sup_{\|x\|=1} |\ell(x)|$ for $\ell \in \mathcal{X}^*$. Since $\mathcal{X}$ is a Hilbert space, by the Riesz-Fréchet representation, for each $\ell \in \mathcal{X}^*$ there exists a unique $u_\ell \in \mathcal{X}$ such that $\ell(x) = \langle u_\ell, x \rangle$ for all $x \in \mathcal{X}$. We call u_ℓ the dual vector for ℓ (in particular the Hilbert-Schmidt dual in the case of $\mathcal{X} = (\mathcal{B}_{\text{sa}}(\mathcal{H}), \langle \cdot, \cdot \rangle_{\text{HS}})$). We say f is *lower semicontinuous* if $\liminf_{x \to x_0} f(x) \geq f(x_0)$ for each $x_0 \in \mathcal{X}$. The *directional derivative* of f at $x \in \operatorname{dom} f$ in the direction $h \in \mathcal{X}$ is given by

$$f'(x; h) = \lim_{t \downarrow 0} \frac{1}{t} [f(x + th) - f(x)]. \tag{69}$$

If f is convex, this limit exists in $\mathbb{R} \cup \{\pm\infty\}$. If the map $\phi_x(h) := \mathcal{X} \ni h \mapsto f'(x; h)$ is linear and continuous, then f is called *Gâteaux-differentiable* at $x \in \mathcal{X}$. Moreover, if f is Gâteaux differentiable at every $x \in A \subset \mathcal{X}$, then f is said to be Gâteaux-differentiable on A. If f is convex and continuous at x, it can be shown that the map ϕ_x is finite and continuous. However, a continuous and convex function may not be Gâteaux-differentiable. For example, $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ has $\phi_0(h) = \lim_{t \downarrow 0} \frac{1}{t}[f(x + th) - f(x)] = \lim_{t \downarrow 0} \frac{1}{t}|th| = |h|$ which is nonlinear. If f is Gâteaux-differentiable at x, we call the dual to ϕ_x as the *Gâteaux gradient* of f at x, written $\nabla f(x) \in \mathcal{X}$. That is, $\langle \nabla f(x), h \rangle = \phi_x(h)$ for each $h \in \mathcal{X}$.

The function f is Fréchet-differentiable at x if there is $y \in \mathcal{X}$ such that

$$\lim_{\|r\| \to 0} \frac{1}{\|r\|} |f(x + r) - f(x) - \langle y, r \rangle| = 0 \tag{70}$$

and in this case, one writes $Df(x) \in \mathcal{X}^*$ for the map $\mathcal{X} \ni z \mapsto Df(x)z = \langle y, z \rangle$. If f is Fréchet-differentiable at x, then by taking $r = th$ in (70) we find $\langle y, h \rangle = f'(x, h)$ and therefore f is Gâteaux-differentiable at x with $\langle \nabla f(x), h \rangle = Df(x)h$.

By regarding differentiability as the existence of a linear approximation at a point, we can generalize it by defining a notion of a linear subapproximation at a point.

Definition 6.1. The *subgradient* of a function $f : \mathcal{X} \to \mathbb{R}$ at x is the set

$$\partial f(x) = \{u \in \mathcal{X} : f(y) - f(x) \geq \langle u, y - x \rangle \;\; \forall y \in \mathcal{X}\} \subset \mathcal{X}. \tag{71}$$

For a convex function f, the following properties hold:

- if f is continuous at x, then $\partial f(x)$ is bounded and nonempty.
- if f is Gâteaux-differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

We briefly prove the second point here. Assume f is Gâteax-differentiable at x. Then

$$\langle \nabla f(x), y - x\rangle = f'(x, y-x) = \lim_{t\downarrow 0} \frac{1}{t}[f((1-t)x + ty) - f(x)].$$

By convexity, $f((1-t)x+ty) \leq (1-t)f(x) + tf(y)$. Therefore, $\langle \nabla f(x), y-x\rangle \leq f(y) - f(x)$; hence, $\nabla f(x) \in \partial f(x)$.

On the other hand, given $u \in \partial f$, $h \in \mathcal{X}$, and $t > 0$, we can set $y = th + x$. Then

$$\langle u, th\rangle = \langle u, y-x\rangle \leq f(y) - f(x) = f(x+th) - f(x).$$

Dividing by t and taking the limit $t \downarrow 0$ yields $u = \nabla f(x)$.

Fermat's Rule of convex optimization theory, stated below, provides a simple characterization of the minimum of a function in terms of the zeroes of its subgradient. Moreover, since the subgradient of a Gâteaux differentiable function consists of a single element, namely, its Gâteaux gradient, finding its minimizer amounts to showing that its Gâteaux gradient is equal to zero.

Theorem 6.1 (Fermat's Rule). *Consider a function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ with* dom $f \neq \emptyset$*. Then $\hat{x} \in \mathcal{X}$ is a global minimizer of f if and only if $0_{\mathcal{X}} \in \partial f(\hat{x})$, where $0_{\mathcal{X}}$ is the zero vector of $\mathcal{X}$.*

Proof. $0_{\mathcal{X}} \in \partial f(\hat{x})$ if and only if

$$f(y) - f(\hat{x}) \geq \langle 0, y - \hat{x}\rangle = 0$$

for every $y \in \mathcal{X}$, i.e. if and only if $\hat{x}$ is a global minimizer of f. □

The following result proves useful in computing the subgradient of a sum of convex functions (see e.g. [Pey15, Theorem 3.30] for the proof).

Theorem 6.2 (Moreau-Rockafellar). *Let $f, g : \mathcal{B}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ be convex and lower semi-continuous, with non-empty domains. For each $A \in \mathcal{B}(\mathcal{H})$, we have*

$$\partial f(A) + \partial g(A) := \{a + b : a \in \partial f(A), b \in \partial g(A)\} \subset \partial(f+g)(A). \tag{72}$$

Equality holds for every $A \in \mathcal{B}(\mathcal{H})$ if f is continuous at some $A_0 \in$ dom(g)*.*

6.2. Gâteaux-differentiable functions

Let $\mathcal{F}$ denote the class of functions $\varphi : \mathcal{D}(\mathcal{H}) \to \mathbb{R}$ which are concave and continuous, and Gâteaux-differentiable on $\mathcal{D}_+(\mathcal{H})$. Note that these functions need *not* be Schur concave, and in particular, need not be unitarily invariant. These include the following:

- The von Neumann entropy $\xi \mapsto S(\xi) := -\operatorname{Tr}(\xi \log \xi)$. In Lemma 6.8, we show for $\xi > 0$,

$$\nabla S(\xi) = -\log \xi - \tfrac{1}{\log_e(2)} \mathbb{1}.$$

- The conditional entropy $\xi_{AB} \mapsto S(A|B)_\xi := S(\xi_{AB}) - S(\xi_B)$, for which
$$\nabla S(A|B).(\xi_{AB}) = -(\log \xi_{AB} - \mathbb{1}_A \otimes \log \xi_B)$$
as shown in Corollary 6.9. Note that for $\varphi(\xi) := S(A|B)_\xi$, the conditional entropy satisfies the interesting property that
$$\varphi(\xi) = \langle \nabla\varphi(\xi), \xi\rangle_{\mathrm{HS}}\,. \tag{73}$$
- The α-Rényi entropy for $\alpha \in (0,1)$. The map $\xi \mapsto S_\alpha(\xi)$ is concave for $\alpha \in (0,1)$, continuous, and has Gâteaux gradient
$$\nabla S_\alpha(\xi) = \frac{\alpha}{1-\alpha}\frac{1}{\mathrm{Tr}[\xi^\alpha]}\xi^{\alpha-1}$$
by Lemma 6.10. Note that the α-Rényi entropy for $\alpha > 1$ is not concave.
- The function $-T_\alpha$, for $\xi \mapsto T_\alpha(\xi) := \mathrm{Tr}[\xi^\alpha]$ and $\alpha > 1$. This map is concave for $\alpha > 1$, continuous, and by Lemma 6.10, has Gâteaux gradient
$$\nabla(-T_\alpha)(\xi) = -\alpha\xi^{\alpha-1}.$$
A state $\xi^* \in B_\varepsilon(\rho)$ maximizes $-T_\alpha$ over $B_\varepsilon(\rho)$ if and only if ξ^* minimizes T_α over $B_\varepsilon(\rho)$. For $\alpha > 1$, minimizing T_α is equivalent to maximizing $S_\alpha(\xi)$, using that $x \mapsto \frac{1}{1-\alpha}\log x$ is decreasing. Therefore, ξ^* maximizes $-T_\alpha$ over $B_\varepsilon(\rho)$ if and only if it maximizes S_α over $B_\varepsilon(\rho)$. Thus, by considering the function $f(\xi) = -T_\alpha(\xi)$ for $\alpha > 1$ in Theorem 3.2, one can establish conditions for ξ^* to maximize the α-Rényi entropy for $\alpha > 1$.

6.3. A convex optimization result and its proof

Theorem 3.2 is a condensed form of the following theorem, whose proof we include below.

Theorem 6.3. *Let $f : \mathcal{D}(\mathcal{H}) \to \mathbb{R}$ be a concave, continuous function which is Gâteaux-differentiable on $\mathcal{D}_+(\mathcal{H})$. Moreover, given a state $\rho \in \mathcal{D}(\mathcal{H})$ and $\varepsilon \in (0,1]$, for any $\xi \in B_\varepsilon^+(\rho)$,*
$$\mathrm{Tr}(\nabla f(\xi)(\xi - \rho)) \le \varepsilon(\lambda_+(\nabla f(\xi)) - \lambda_-(\nabla f(\xi))). \tag{74}$$
Furthermore, the following are equivalent:

1. *Equality in (74),*
2. *$\xi \in \mathrm{argmax}_{B_\varepsilon(\rho)} f$,*
3. a) *Either $\frac{1}{2}\|\xi - \rho\|_1 = \varepsilon$ or $\nabla f(\xi) = \lambda\mathbb{1}$ for some $\lambda \in \mathbb{R}$, and*
   b) *we have*
   $$\pi_\pm \nabla f(\xi)\pi_\pm = \lambda_\pm(\nabla f(\xi))\pi_\pm$$
   where $\pi_\pm$ is the projection onto the support of $\Delta_\pm$, and where $\Delta = \Delta_+ - \Delta_-$ is the Jordan decomposition of $\Delta := \xi - \rho$.

We first prove the inequality (74) by considering the Jordan decomposition of $\Delta := \xi - \rho$. Next, we convert the constrained optimization problem $\max_{\xi \in B_\varepsilon(\rho)} f(\xi)$ to an unconstrained optimization problem $\min h$ for a non-Gâteaux differentiable function h defined on all of $\mathcal{B}_{\mathrm{sa}}(\mathcal{H})$ by adding appropriate indicator functions for the sets $\mathcal{D}(\mathcal{H})$ and $\{A \in \mathcal{B}_{\mathrm{sa}}(\mathcal{H}) : \frac{1}{2}\|A - \rho\|_1 \le \varepsilon\}$.

The Moreau-Rockafellar Theorem (Theorem 6.2) allows us to compute $\partial h(\xi)$ in terms of $\partial f(\xi) = \{\nabla f(\xi)\}$ and the subgradients of the indicator functions. We then show $0 \in \partial h(\xi)$ if and only if equality is achieved in (74).

Proof of Theorem 6.3 We start with the following general result, which does not use convex optimization.

Lemma 6.4. *Let $\varepsilon > 0$ and $\Delta \in \mathcal{B}_{sa}$ with $\operatorname{Tr}(\Delta) = 0$, $\frac{1}{2}\|\Delta\|_1 \le \varepsilon$. Let $L \in \mathcal{B}_{sa}$. Then*

$$\operatorname{Tr}(L\Delta) \le \varepsilon(\lambda_+(L) - \lambda_-(L)) \tag{75}$$

with equality if and only if

1. *Either $\frac{1}{2}\|\Delta\|_1 = \varepsilon$ or else $L = \lambda\mathbb{1}$ for $\lambda := \lambda_+(L) = \lambda_-(L)$, and*

2. *we have*
$$\pi_\pm L \pi_\pm = \lambda_\pm(L)\pi_\pm$$
where $\pi_\pm$ is the projection onto the support of $\Delta_\pm$, and where $\Delta = \Delta_+ - \Delta_-$ is the Jordan decomposition of Δ.

Proof. Note $|\Delta| = \Delta_+ + \Delta_-$, and since Δ has trace zero, $\operatorname{Tr}(\Delta_+) = \operatorname{Tr}(\Delta_-)$. We expand

$$\operatorname{Tr}(L\Delta) = \operatorname{Tr}(L\Delta_+) - \operatorname{Tr}(L\Delta_-).$$

Now, we use that L is self-adjoint so that

$$\lambda_-(L)\mathbb{1} \le L \le \lambda_+(L)\mathbb{1}.$$

Since $\Delta_\pm \ge 0$, we have

$$\begin{aligned}
\operatorname{Tr}(L\Delta_+) - \operatorname{Tr}(L\Delta_-) &\le \operatorname{Tr}(L\Delta_+) - \operatorname{Tr}(\lambda_-(L)\Delta_-) \qquad (76)\\
&\le \operatorname{Tr}(\lambda_+(L)\Delta_+) - \operatorname{Tr}(\lambda_-(L)\Delta_-) \qquad (77)\\
&= (\lambda_+(L) - \lambda_-(L))\operatorname{Tr}(\Delta_+),
\end{aligned}$$

where the last line follows from the fact that $\operatorname{Tr}(\Delta_+) = \operatorname{Tr}(\Delta_-)$. However, $\operatorname{Tr}(\Delta_+) \le \varepsilon$, so we have

$$(\lambda_+(L) - \lambda_-(L))\operatorname{Tr}(\Delta_+) \le \varepsilon(\lambda_+(L) - \lambda_-(L)). \tag{78}$$

Thus, (75) follows. To obtain equality in (75), we require equality in (76), (77) and (78). Equality in (78) is equivalent to condition 1 in the statement of the lemma. We now show that equality in (76) and (77) is equivalent to condition 2.

Set $\lambda_+ = \lambda_+(L)$. Then since π_+ is the projection onto the support of Δ_+, we have $\Delta_+ = \pi_+\Delta_+\pi_+$. Using cyclicity of the trace, we obtain

$$\operatorname{Tr}(L\Delta_+) = \operatorname{Tr}(L\pi_+\Delta_+\pi_+) = \operatorname{Tr}(\pi_+L\pi_+\Delta_+).$$

Equality in (76) implies $\operatorname{Tr}(L\Delta_+) = \lambda_+ \operatorname{Tr}(\Delta_+)$, so

$$\operatorname{Tr}((\pi_+L\pi_+)\Delta_+) = \operatorname{Tr}(\lambda_+\Delta_+)$$

and thus

$$\operatorname{Tr}(\Delta_+(\lambda_+\pi_+ - \pi_+L\pi_+)) = 0. \tag{79}$$

Since λ_+ is the largest eigenvalue of L which is self-adjoint, we have $L \le \lambda_+\mathbb{1}$. Since conjugating by any operator preserves the semi-definite order, we have $\lambda_+\pi_+ - \pi_+L\pi_+ \ge 0$. Then since Δ_+ restricted to π_+ is positive definite, (79) implies $\pi_+L\pi_+ = \lambda_+\pi_+$. Similarly, equality in (77) implies $\pi_-L\pi_- = \lambda_-(L)\pi_-$.

Conversely, if $\pi_\pm L\pi_\pm = \lambda_\pm\pi_\pm$, we immediately obtain equality in (76) and (77). □

This lemma with the choices $\Delta = \xi - \rho$ and $L \equiv L_\xi := \nabla f(\xi)$ gives the inequality (74) of Theorem 6.3. It also gives the equivalence between the conditions 1 and 3 of the theorem.

We now turn to the theory of convex optimization to establish the equivalence between conditions 1 and 2. Recall $f : \mathcal{D} \to \mathbb{R}$ is a continuous, concave function which is Gâteaux-differentiable on $\mathcal{D}_+$. Let us write $\tilde{f} = -f$ which is convex, and note $L_\xi = -\nabla \tilde{f}(\xi)$. With this choice, it remains to be shown that $\xi^* \in \operatorname{argmin}_{\xi \in B_\varepsilon(\rho)} \tilde{f}(\xi)$ if and only if

$$\operatorname{Tr}(L_{\xi^*}(\xi^* - \rho)) \geq \varepsilon(\lambda_+(L_{\xi^*}) - \lambda_-(L_{\xi^*})). \tag{80}$$

The Tietze extension theorem (e.g. [Sim15, Theroem 2.2.5]) allows one to extend a bounded continuous function defined on a closed set of a normal topological space (such as a normed vector space) to the entire space, while preserving continuity and boundedness. We use this to extend $\tilde{f}$ (which is bounded as it is a continuous function on the compact set $\mathcal{D}$) to the whole space $\mathcal{B}_{\text{sa}}$, using that $\mathcal{D}$ is a closed set in $\mathcal{B}_{\text{sa}}$. We consider the closed and convex sets $\mathcal{D}$ and

$$T := \{A \in \mathcal{B}_{\text{sa}} : \|A - \rho_{AB}\|_1 \leq 2\varepsilon\}, \tag{81}$$

and note $B_\varepsilon = \mathcal{D} \cap T$. We define $h : \mathcal{B}_{\text{sa}} \to \mathbb{R} \cup \{\infty\}$ as

$$h = \tilde{f} + \delta_{\mathcal{D}} + \delta_T$$

where for $S \subset \mathcal{B}_{\text{sa}}$, the indicator function $\delta_S(A) = \begin{cases} 0 & A \in S \\ +\infty & \text{otherwise.} \end{cases}$

We have the important fact that, by construction,

$$\operatorname{argmin}_{\omega \in B_\varepsilon(\rho)} \tilde{f}(\omega) = \operatorname{argmin}_{A \in \mathcal{B}_{\text{sa}}} h(A). \tag{82}$$

By Theorem 6.1 $\hat{A} \in \mathcal{B}_{\text{sa}}$ is a global minimizer of h if and only if $0 \in \partial h(\hat{A})$. Note each of $\tilde{f}, \delta_{\mathcal{D}}, \delta_T$ is lower semicontinuous, convex, and has non-empty domain. Moreover, both $\tilde{f}$ and δ_T are continuous at any faithful state $\omega \in B_\varepsilon^+(\rho) \subset \operatorname{dom} \delta_{\mathcal{D}}$. Hence, by Theorem 6.2,

$$\partial h = \partial \tilde{f} + \partial \delta_{\mathcal{D}} + \partial \delta_T := \{\ell_f + \ell_{\mathcal{D}} + \ell_T : \ell_f \in \partial \tilde{f},\, \ell_{\mathcal{D}} \in \partial \delta_{\mathcal{D}},\, \ell_T \in \partial \delta_T\}.$$

Hence, to obtain a complete characterization of

$$\operatorname{argmin} h = \{A \in \mathcal{B}_{\text{sa}} : 0 \in \partial h(A)\}$$

we need to evaluate the subgradients of $\tilde{f}$, δ_T, and $\delta_{\mathcal{D}}$.

Since $\tilde{f}$ is Gâteaux-differentiable on $\mathcal{D}_+$, for any $\omega \in \mathcal{D}_+$, we have $\partial \tilde{f}(\omega) = \{-L_\omega\}$ for $L_\omega := -\nabla \tilde{f}(\omega)$. The following two results (proven in Section 6.4) describe the other two relevant subgradients.

Lemma 6.5. *For $A \in \mathcal{B}_{sa}$ with $0 < \frac{1}{2}\|A - \rho_{AB}\|_1 \leq \varepsilon$,*

$$\partial \delta_T(A) = \{u \in \mathcal{B}_{sa} : 2\varepsilon \|u\|_\infty = \langle u, A - \rho_{AB} \rangle\}$$

where

$$\|u\|_\infty := \lambda_+(|u|) = \sup_{A \in \mathcal{B}_{sa},\, \|A\|_1 = 1} |\langle u, A \rangle|.$$

Lemma 6.6. *Let $\omega \in \mathcal{D}_+$. Then $\partial \delta_{\mathcal{D}}(\omega) = \{x\mathbb{1} : x \in \mathbb{R}\}$.*

Putting together these results, we have, for $\xi^* \in B_\varepsilon^+(\rho)$,

$$0 \in \partial h(\xi^*) \iff 0 = -L_{\xi^*} + G + x\mathbb{1}$$

for some $G \in \partial\delta_T(\xi^*)$ and $x \in \mathbb{R}$, where G satisfies

$$2\varepsilon\|G\|_\infty = \operatorname{Tr}(G(\xi^* - \rho)).$$

Then $G = L_{\xi^*} - x\mathbb{1}$, which implies that

$$\operatorname{Tr}(L_{\xi^*}(\xi^* - \rho)) = 2\varepsilon\|L_{\xi^*} - x\mathbb{1}\|_\infty.$$

Set $\alpha = \frac{\operatorname{Tr}(L_{\xi^*}(\xi^*-\rho))}{2\varepsilon}$. Note α and L_{ξ^*} depend on ξ^*. Then we have

$$0 \in \partial h(\xi^*) \iff \exists\, x \in \mathbb{R} : \alpha = \|L_{\xi^*} - x\mathbb{1}\|_\infty.$$

Since $\|L_{\xi^*} - x\mathbb{1}\|_\infty = \max_{\lambda \in \operatorname{spec} L_{\xi^*}} |\lambda - x|$, we have that $0 \in \partial h(\xi^*)$ if and only if

$$\exists\, x \in \mathbb{R} : \alpha = \max_{\lambda \in \operatorname{spec} L_{\xi^*}} |\lambda - x|. \tag{83}$$

Let

$$q(x) = \max_{\lambda \in \operatorname{spec} L_{\xi^*}} |\lambda - x| \tag{84}$$

We immediately see that q is continuous, $q(0) = \lambda_+(|L_{\xi^*}|)$, and $\lim_{z\to\pm\infty} q(z) = +\infty$. In fact, we can write a simple form for $q(x)$ as the following lemma, which is proven in Section 6.4, shows. Set $\lambda_+ \equiv \lambda_+(L_{\xi^*})$ and $\lambda_- \equiv \lambda_-(L_{\xi^*})$ in the following.

Lemma 6.7. *The quantity $q(x)$ defined by (84) can be expressed as follows:*

$$q(x) = \frac{\lambda_+ - \lambda_-}{2} + \left|\frac{\lambda_+ + \lambda_-}{2} - x\right|.$$

Thus, Lemma 6.7 implies that the range of the function q is $[\frac{\lambda_+ - \lambda_-}{2}, \infty)$. Hence, (83) holds if and only if

$$\frac{\lambda_+ - \lambda_-}{2} \leq \alpha.$$

Substituting $\alpha = \frac{\operatorname{Tr}(L_{\xi^*}(\xi^*-\rho))}{2\varepsilon}$ in the above expression yields (80) and therefore concludes the proof of Theorem 6.3. □

Below, we collect some results (proven in Section 6.4) relating to the Gâteaux gradients of relevant entropic functions which are candidates for f in Theorem 6.3.

Lemma 6.8. *The von Neumann entropy $\xi \mapsto S(\xi) := -\operatorname{Tr}(\xi \log \xi)$ is Gâteaux-differentiable at each $\xi > 0$ and $\nabla S(\xi) = -\log\xi - \frac{1}{\log_e(2)}\mathbb{1}$.*

Corollary 6.9. *The conditional entropy $\xi_{AB} \mapsto S(A|B)_\xi := S(\xi_{AB}) - S(\xi_B)$ is Gâteaux-differentiable at each $\xi_{AB} > 0$ and $\nabla S(A|B).(\xi_{AB}) = -(\log\xi_{AB} - \mathbb{1}_A \otimes \log\xi_B)$.*

Lemma 6.10. *For $\alpha \in (0,1) \cup (1,\infty)$, the map T_α and the Rényi entropy S_α are Gâteaux-differentiable at each $\xi \in \mathcal{D}_+$, with Gâteaux gradients*

$$\nabla T_\alpha(\xi) = \alpha\xi^{\alpha-1}, \qquad \nabla S_\alpha(\xi) = \frac{\alpha}{1-\alpha}\frac{1}{\operatorname{Tr}[\xi^\alpha]}\xi^{\alpha-1}.$$

6.4. Proofs of the remaining lemmas and corollaries

Proof of Corollary 3.3 We write the conditional entropy (defined in eq. (9)) as the map

$$
\begin{aligned}
S(A|B). \quad : \quad & \mathcal{D}(\mathcal{H}_{AB}) \to \mathbb{R} \\
& \xi \mapsto S(A|B)_\xi .
\end{aligned} \tag{85}
$$

By Corollary 6.9, for any $\xi_{AB} > 0$ we have

$$\nabla S(A|B).(\xi) = -(\log \xi_{AB} - \mathbb{1}_A \otimes \log \xi_B) = -L_\xi .$$

Substituting this in the left-hand side of (74), we obtain

$$
\begin{aligned}
\operatorname{Tr}(L_\xi(\rho_{AB} - \xi_{AB})) &= \operatorname{Tr}[\rho_{AB} \log \xi_{AB}] - \operatorname{Tr}[\xi_{AB} \log \xi_{AB}] - \operatorname{Tr}[\rho_B \log \xi_B] + \operatorname{Tr}[\xi_B \log \xi_B] \\
&= S(A|B)_\xi + \operatorname{Tr}[\rho_{AB} \log \rho_{AB}] - \operatorname{Tr}[\rho_{AB}(\log \rho_{AB} - \log \xi_{AB})] \\
&\qquad - \operatorname{Tr}[\rho_B \log \rho_B] + \operatorname{Tr}[\rho_B(\log \rho_B - \log \xi_B)] \\
&= S(A|B)_\xi - S(A|B)_\rho - D(\rho_{AB} \| \xi_{AB}) + D(\rho_B \| \xi_B).
\end{aligned}
$$

The statement of Corollary 3.3 then follows from Theorem 6.3. □

Proof of Lemma 6.5 Let $\mathcal{A} := \{u \in \mathcal{B}_{\mathrm{sa}} : 2\varepsilon \|u\|_\infty = \langle u, A - \rho_{AB} \rangle)\}$ and set

$$C := \{B \in \mathcal{B}_{\mathrm{sa}} : \|B - \rho_{AB}\|_1 \leq \varepsilon\} \subset \mathcal{B}_{\mathrm{sa}}.$$

Let $u \in \mathcal{A}$. Then for $y \in C$,

$$
\begin{aligned}
\langle u, y - A \rangle = \langle u, y - \rho_{AB} \rangle + \langle u, \rho_{AB} - A \rangle &= \langle u, y - \rho_{AB} \rangle - 2\varepsilon \|u\|_\infty \\
&\leq \|u\|_\infty \|y - \rho_{AB}\|_1 - 2\varepsilon \|u\|_\infty \\
&= \|u\|_\infty (\|y - \rho_{AB}\|_1 - 2\varepsilon) \leq 0
\end{aligned}
$$

and thus $u \in \partial \delta_T(A)$. On the other hand, take $u \in \partial \delta_T(A)$. Then

$$\|u\|_\infty = \sup_{\|x\|_1 = 1} |\langle u, x \rangle|$$

is achieved at some $x^* \in \mathcal{B}_{\mathrm{sa}}$ since the set $\{x \in \mathcal{B}_{\mathrm{sa}} : \|x\|_1 = 1\}$ is compact, using that $\mathcal{B}_{\mathrm{sa}}$ is finite-dimensional. Then $y = \rho_{AB} + 2\varepsilon x \in C$. Hence,

$$0 \geq \langle u, y - A \rangle = \langle u, \rho_{AB} - A \rangle + 2\varepsilon \langle u, x \rangle = \langle u, \rho_{AB} - A \rangle + 2\varepsilon \|u\|_\infty$$

and thus $\langle u, A - \rho_{AB} \rangle \geq 2\varepsilon \|u\|_\infty$. Then by the bound

$$\langle u, A - \rho_{AB} \rangle \leq \|u\|_\infty \|A - \rho_{AB}\|_1 \leq 2\varepsilon \|u\|_\infty$$

we have $u \in \mathcal{A}$. □

Proof of Lemma 6.6 By definition,

$$\partial\delta_{\mathcal{D}}(\omega) = \{u \in \mathcal{B}_{\mathrm{sa}} : \langle u, y-\omega\rangle \leq 0 \;\text{ for all } y \in \mathcal{D}\}\,.$$

Since u is self-adjoint, we can write its eigen-decomposition as $u = \sum_{k=1}^{d} \alpha_k \,|k\rangle\langle k|$. If $\alpha_k = \alpha_j$ for each j, k, then $u \propto \mathbb{1}$. Conversely, $\alpha_k \mathbb{1} \in \partial\delta_{\mathcal{D}}(\omega)$ since $\mathrm{Tr}[\alpha_k \mathbb{1}(y-\omega)] = \alpha_k(\mathrm{Tr}(y) - \mathrm{Tr}(\omega)) = 0$.

Otherwise, assume $\alpha_k > \alpha_j$ for some $k, j \in \{1, \ldots, d\}$. Let

$$y := \sum_{i,\, i\neq k,\, i\neq j} \langle i|\omega|i\rangle \,|i\rangle\langle i| + (\langle k|\omega|k\rangle + \langle j|\omega|j\rangle)\,|k\rangle\langle k| \in \mathcal{B}_{\mathrm{sa}}.$$

Then $\mathrm{Tr}(y) = \sum_i \langle i|\omega|i\rangle = \mathrm{Tr}(\omega) = 1$, and $y \geq 0$ since all of its eigenvalues are non-negative. Note $|j\rangle$ is an eigenvector of y with eigenvalue zero. Next,

$$\begin{aligned}
\mathrm{Tr}(u(y-\omega)) &= \sum_i \langle i|(y-\omega)|i\rangle \\
&= \sum_{i,\, i\neq k,\, i\neq j} (\langle i|\omega|i\rangle - \langle i|\omega|i\rangle) \\
&\qquad + \alpha_k(\langle k|\omega|k\rangle + \langle j|\omega|j\rangle - \langle k|\omega|k\rangle) + \alpha_j(0 - \langle j|\omega|j\rangle) \\
&= (\alpha_k - \alpha_j)\,\langle j|\omega|j\rangle > 0.
\end{aligned}$$

Thus, $u \notin \partial\delta_{\mathcal{D}}(\omega)$. □

Proof of Lemma 6.7 First, we establish

$$q(x) = \max(|\lambda_+ - x|, |\lambda_- - x|). \tag{86}$$

Given $\lambda \in \operatorname{spec} L$ and $x \in \mathbb{R}$, if $x < \lambda$, then

$$|x-\lambda| = \lambda - x \leq \lambda_+ - x = |\lambda_+ - x|$$

and otherwise

$$|x-\lambda| = x - \lambda \leq x - \lambda_- = |\lambda_- - x|,$$

yielding (86). Next, set $r := \frac{\lambda_+ - \lambda_-}{2}$ and $m := \frac{\lambda_+ + \lambda_-}{2}$. Then $\lambda_\pm = m \pm r$. Therefore, for $x \in \mathbb{R}$,

$$|\lambda_\pm - x| = |m \pm r - x| \leq r + |m - x|,$$

yielding $q(x) = r + |m - x|$ as claimed. □

Proof of Lemma 6.8 Let us introduce the Cauchy integral representation of an analytic function. If g is analytic on a domain $G \subset \mathbb{C}$ and A is a matrix with $\operatorname{spec} A \subset G$, then we can write

$$g(A) = \frac{1}{2\mathrm{i}\pi}\int_\Gamma g(z)(z\mathbb{1} - A)^{-1}\,\mathrm{d}z,$$

and

$$g'(A) = \frac{1}{2\mathrm{i}\pi}\int_\Gamma g(z)(z\mathbb{1} - A)^{-2}\,\mathrm{d}z, \tag{87}$$

where $\Gamma \subset G$ is a simple closed curve with $\operatorname{spec} A \subset \Gamma$, and $g' : G \to \mathbb{C}$ is the derivative of g. [Sti87] shows that with these definitions, the Fréchet derivative of g at A exists, and when applied to a matrix X yields

$$D(g)(A)X = \frac{1}{2\mathrm{i}\pi} \int_\Gamma g(z)(z\mathbb{1} - A)^{-1} X (z\mathbb{1} - A)^{-1} \,\mathrm{d}z.$$

Therefore, by cyclicity of the trace,

$$\operatorname{Tr}(D(g)(A)X) = \operatorname{Tr}\left[\frac{1}{2\mathrm{i}\pi} \int_\Gamma g(z)(z\mathbb{1} - A)^{-2} \,\mathrm{d}z\, X\right] = \operatorname{Tr}(g'(A)X) \tag{88}$$

using (87) for the second equality.

We may write consider von Neumann entropy as the map

$$S : \mathcal{D} \to \mathbb{R}, \qquad \xi \mapsto S(\xi) = -\operatorname{Tr}[\xi \log \xi].$$

Then we may write $S = \operatorname{Tr} \circ \eta$ for $\eta(x) = -x \log x$ which is analytic on $\{\zeta \in \mathbb{C} : \operatorname{Re} \zeta > 0\}$, with derivative $\eta'(x) = -\log x - \frac{1}{\log_e(2)}$. Then for $\xi > 0$,

$$D(S)(\xi) = D(\operatorname{Tr} \circ \eta)(\xi) = D(\operatorname{Tr})(\eta(\xi)) \circ D(\eta)(\xi).$$

by the chain rule for Fréchet derivatives. By the linearity of the trace, $D(\operatorname{Tr})B(X) = \operatorname{Tr}(X)$ for any $X, B \in \mathcal{B}_{\mathrm{sa}}$. So for $X \in \mathcal{B}_{\mathrm{sa}}$,

$$D(S)(\xi)X = \operatorname{Tr}[D(\eta)(\xi)X].$$

By (88) for $g = \eta$, we have

$$D(S)(\xi)X = \operatorname{Tr}[\eta'(\xi)X] = \operatorname{Tr}[(-\log \xi - \tfrac{1}{\log_e(2)}\mathbb{1})X]. \quad \square$$

Proof of Corollary 6.9 We decompose the conditional entropy map (Equation (85)) by writing

$$S(A|B). = S_{AB}(\cdot) - S_B \circ \operatorname{Tr}_A(\cdot) : \mathcal{D}(H_{AB}) \to \mathbb{R}$$

where we have indicated explicitly the domain of each function in the subscript. That is, $S_{AB} : \mathcal{D}(\mathcal{H}_{AB}) \to \mathbb{R}$ is the von Neumann entropy on system AB, and $S_B : \mathcal{D}(\mathcal{H}_B) \to \mathbb{R}$ is the von Neumann entropy on B.

Since $\operatorname{Tr}_A : \mathcal{B}_{\mathrm{sa}}(\mathcal{H}_{AB}) \to \mathcal{B}_{\mathrm{sa}}(\mathcal{H}_B)$ is a bounded linear map and S_B concave and continuous, the chain rule for the composition with a linear map (see Prop. Prop. 3.28 of [Pey15]) gives

$$\nabla(S_B \circ \operatorname{Tr}_A)(\omega_{AB}) = \operatorname{Tr}_A^* \circ \nabla S_B(\omega_B) = \mathbb{1}_A \otimes \nabla S_B(\omega_B)$$

for any $\omega_{AB} \in \mathcal{D}(\mathcal{H}_{AB})$, where Tr_A^* is the dual to the map Tr_A. As $\nabla S_B(\omega_B) = -(\log \omega_B + \frac{1}{\log_e(2)}\mathbb{1}_B)$, we have

$$\begin{aligned}
\nabla S(A|B).(\omega_{AB}) &= \nabla S_{AB}(\omega_{AB}) - \nabla(S_B \circ \operatorname{Tr}_A)(\omega_{AB}) \\
&= -(\log \omega_{AB} + \tfrac{1}{\log_e(2)}\mathbb{1}_{AB}) + \mathbb{1}_A \otimes (\log \omega_B + \tfrac{1}{\log_e(2)}\mathbb{1}_B) \\
&= -(\log \omega_{AB} - \mathbb{1}_A \otimes \log \omega_B). \quad \square
\end{aligned}$$

Proof of Lemma 6.10 The α-Rényi entropy $\alpha \in (0,1) \cup (1,\infty)$ can be described by the map

$$S_\alpha : \mathcal{D} \mapsto \mathbb{R}, \qquad \xi \mapsto S_\alpha(\xi) := \frac{1}{1-\alpha} \log \operatorname{Tr}(\xi^\alpha).$$

Let us use the notation $\mathrm{pow}_\alpha : \mathcal{B}_{\mathrm{sa}} \to \mathcal{B}_{\mathrm{sa}}$, $A \mapsto \mathrm{pow}_\alpha(A) := A^\alpha$. Then

$$T_\alpha = \operatorname{Tr} \circ \mathrm{pow}_\alpha$$

and $(1-\alpha)S_\alpha = \log \circ T_\alpha$. For $\xi \in \mathcal{D}_+$, the function pow_α is analytic on $\{\zeta \in \mathbb{C} : \operatorname{Re}\zeta > 0\}$. As in the proof of Lemma 6.8 then, using the chain rule and linearity of the trace we find

$$D(T_\alpha)(\xi) = \operatorname{Tr}[D(\mathrm{pow}_\alpha)(\xi)X].$$

Invoking (88) for $g(z) = z^\alpha$, with $\Gamma \subset \{\zeta \in \mathbb{C} : \operatorname{Re}\zeta > 0\}$ a simple closed curve enclosing the spectrum of ξ, we have

$$D(T_\alpha)(\xi)X = \alpha \operatorname{Tr}(\xi^{\alpha-1}X). \tag{89}$$

Moreover,

$$\begin{aligned} D(S_\alpha)(\xi)X &= \frac{1}{1-\alpha} D(\log \circ T_\alpha)(\xi) \\ &= \frac{1}{1-\alpha} D(\log)(\operatorname{Tr}(\xi^\alpha)) \circ D(T_\alpha)(\xi) \\ &= \frac{1}{1-\alpha} \frac{\alpha}{\operatorname{Tr}[\xi^\alpha]} \operatorname{Tr}(\xi^{\alpha-1}X). \quad \square \end{aligned}$$

A. An elementary property of concave functions

Given a function $\phi : I \to \mathbb{R}$ defined on an interval $I \subset \mathbb{R}$, we define the “slope function,”

$$s(x_1, x_2) = \frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1}$$

for $x_1, x_2 \in I$ with $x_1 \neq x_2$. Note that s is symmetric in its arguments. It can be shown that ϕ is concave (resp. strictly concave) if and only if s is monotone decreasing (resp. strictly decreasing) in each argument.

Proposition A.1. *Let $I \subset \mathbb{R}$ be an interval and $\phi : I \to \mathbb{R}$ be concave. For any $x_1, x_2, y_1, y_2 \in I$ such that $x_1 \neq x_2$, $y_1 \neq y_2$, $x_1 \leq y_1$ and $x_2 \leq y_2$ we have*

$$s(x_1, x_2) \geq s(y_1, y_2).$$

If ϕ is strictly concave, then equality is achieved if and only if $x_1 = y_1$ and $x_2 = y_2$.

Proof. For ϕ concave, we have $s(x_1, x_2) \geq s(y_1, x_2) \geq s(y_1, y_2)$. Next, assume ϕ is strictly concave. Then equality holds in the first inequality if and only if $x_2 = y_2$, and in the second if and only if $x_1 = y_1$, completing the proof. $\square$

B. Proof of Corollary 3.6

Let us recall Corollary 3.6.

Corollary 3.6 (Local continuity bound for the von Neumann entropy)**.** *Let $\rho \in \mathcal{D}$ and $\varepsilon \in (0,1]$. Then for any state $\omega \in B_\varepsilon(\rho)$,*

$$|S(\omega) - S(\rho)| \leq \max\{S(\rho_\varepsilon^*) - S(\rho), S(\rho) - S(\rho_{*,\varepsilon})\} \tag{19}$$

$$\leq \begin{cases} \varepsilon \log(d-1) + h(\varepsilon) & \text{if } \varepsilon < 1 - \frac{1}{d} \\ \log d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases} \tag{20}$$

for $h(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ the binary entropy. Moreover, equality holds in (20) if and only if one of the following holds:

1. *ρ is a pure state; in this case $\rho_\varepsilon^* = \begin{cases} \mathrm{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1}) & \text{if } \varepsilon < 1 - \frac{1}{d}, \\ \mathbb{1}/d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases}$*
2. *$\varepsilon < 1 - \frac{1}{d}$, and $\rho = \mathrm{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots \frac{\varepsilon}{d-1})$; in this case $\rho_{\varepsilon,*}(\rho)$ is a pure state,*
3. *$\varepsilon \geq 1 - \frac{1}{d}$ and $\rho = \tau := \mathbb{1}/d$; in this case $\rho_{\varepsilon,*}(\rho)$ is a pure state.*

The inequality (19) is an immediate consequence of Corollary 3.3, which in turn follows directly from Theorem 3.2. This can be seen by noting that

$$S(\rho_{*,\varepsilon}) \leq S(\omega) \leq S(\sigma_\varepsilon^*),$$

by Schur concavity of the von Neumann entropy, and the minimality of σ_ε^* and maximality of $\rho_{*,\varepsilon}$ in the majorization order (10).

The inequality (20) follows by applying the (AF)-bound (Lemma 2.2) to the pairs of states $(\sigma_\varepsilon^*, \sigma)$ and $(\sigma, \rho_{*,\varepsilon})$ as follows. The (AF)-bound gives

$$S(\sigma_\varepsilon^*) - S(\sigma) \leq \mathrm{AF}(\varepsilon), \qquad S(\sigma) - S(\sigma_{*,\varepsilon}) \leq \mathrm{AF}(\varepsilon) \tag{90}$$

for

$$\mathrm{AF}(\varepsilon) := \begin{cases} \varepsilon \log(d-1) + h(\varepsilon) & \text{if } \varepsilon < 1 - \frac{1}{d} \\ \log d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases} \tag{91}$$

What remains to be established is the necessary and sufficient condition for equality in (20). A sufficient condition for equality in the (AF)-bound was obtained by Audenaert, and is stated in Lemma 2.2. Here we prove that this condition is also necessary. In order to do this, it is helpful to recall the proof of the (AF)-bound in detail.

Proof of Lemma 2.2 Without loss of generality, assume $S(\rho) \geq S(\sigma)$. Let us first consider $\varepsilon \geq 1 - \frac{1}{d}$. As $0 \leq S(\omega) \leq \log d$ for any state $\omega \in \mathcal{D}$, the left-hand side of (8) is bounded by $S(\rho) - S(\sigma) \leq \log d - 0 = \log d$, which is the right-hand side of the inequality (8) in this range of values of ε. Moreover, equality is achieved if and only if $S(\rho) = \log d$ and $S(\sigma) = 0$. As only the completely mixed state, $\tau = \frac{1}{d}\mathbb{1}$, achieves $S(\tau) = \log d$, and only pure states have zero entropy, we have established the equality conditions for $\varepsilon \geq 1 - \frac{1}{d}$. Note that $\tau \in B_\varepsilon(\sigma)$ for any state $\sigma \in \mathcal{D}$ and $\varepsilon \geq 1 - \frac{1}{d}$.

Next, let $\varepsilon < 1 - \frac{1}{d}$. Note that if $\rho = \mathrm{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \dots, \frac{\varepsilon}{d-1})$, then $S(\rho) = h(\varepsilon) + \varepsilon \log(d-1)$. If σ is a pure state, then $S(\sigma) = 0$. Hence, these choices of ρ and σ are sufficient to attain equality in (8). It remains to show that in this range of values of ε, equality is achieved in the (AF)-bound only for this pair of states. To show this, we first reduce the problem to a classical one. This is because, by comparing the 1-norm on $\mathbb{R}^d$ to the trace norm between mutually diagonal matrices, we have

$$\|\vec{\lambda}(\rho) - \vec{\lambda}(\sigma)\|_1 = \|\operatorname{Eig}^{\downarrow}(\rho) - \operatorname{Eig}^{\downarrow}(\sigma)\|_1 \leq \|\rho - \sigma\|_1 \leq \varepsilon \tag{92}$$

where the first inequality follows from Lemma 2.1. Since $H(\vec{\lambda}(\rho)) = S(\operatorname{Eig}^{\downarrow}(\rho)) = S(\rho)$, and $H(\vec{\lambda}(\sigma)) = S(\operatorname{Eig}^{\downarrow}(\sigma)) = S(\sigma)$, we can reduce to the classical case of probability distributions on $p, q \in \Delta$. This argument is due to [Aud07]. Thus the (AF)-bound is equivalently stated in terms of probability distributions as in the following lemma.

Lemma B.1. *Let $\varepsilon \in (0,1]$ and $p, q \in \Delta$ such that $\frac{1}{2}\|p-q\|_1 \leq \varepsilon$, then*

$$|H(p) - H(q)| \leq \begin{cases} \varepsilon \log(d-1) + h(\varepsilon) & \text{if } \varepsilon < 1 - \frac{1}{d} \\ \log d & \text{if } \varepsilon \geq 1 - \frac{1}{d}. \end{cases} \tag{93}$$

Without loss of generality, assume $H(p) \geq H(q)$. Then equality in (93) occurs if and only if for some $\pi \in S_d$, we have $\pi(q) = (1, 0, \ldots, 0)$, and either

1. *$\varepsilon < 1 - \frac{1}{d}$, and $\pi(p) = (1-\varepsilon, \frac{\varepsilon}{d-1}, \ldots, \frac{\varepsilon}{d-1})$, or*
2. *$\varepsilon \geq 1 - \frac{1}{d}$, and $p = (\frac{1}{d}, \ldots, \frac{1}{d})$.*

Using this result, by setting $p := \vec{\lambda}(\rho)$, and $q := \vec{\lambda}(\sigma)$, by (92) and Lemma B.1, we obtain inequality (8). Moreover, to attain equality in (8), we require equality in (93). This fixes $\rho = \operatorname{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \ldots, \frac{\varepsilon}{d-1})$ in some basis, and σ a pure state. $\square$

It remains to prove the equality condition in Lemma B.1 for the range $\varepsilon < 1 - \frac{1}{d}$.

Proof of Lemma B.1 To establish the inequality (93), we recall the proof presented in [Win16] in detail[5]. This also helps us to investigate when equality occurs, and deduce the form of p and q.

Without loss of generality, assume $H(p) \geq H(q)$. A coupling of (p, q) is a probability measure ω on $[d] \times [d]$, for $[d] = \{1, \ldots, d\}$, such that $\sum_{i \in [d]} \omega(i,j) = q(j)$ and $\sum_{j \in [d]} \omega(i,j) = p(i)$. For any coupling ω, it is known that, if (X, Y) are a pair of random variables with joint measure ω, i.e. $(X, Y) \sim \omega$, then $\Pr[X \neq Y] \geq \frac{1}{2}\|p - q\|_1$. Moreover, there exist optimal couplings ω^* which achieve equality: if $(X, Y) \sim \omega^*$, then $\Pr[X \neq Y] = \frac{1}{2}\|p-q\|_1$. In fact, we can choose

$$\omega^*(i,j) := \begin{cases} \min(p(i), q(i)) & \text{if } i = j \\ \frac{(p(i)-q(i))_+ \cdot (q(j)-p(j))_+}{\frac{1}{2}\|p-q\|_1} & \text{otherwise,} \end{cases} \tag{94}$$

using the notation $(z)_+ = \max(z, 0)$ for $z \in \mathbb{R}$.

Lemma B.2. *ω^* is a maximal coupling of (p, q).*

Proof. First, $\sum_i \omega^*(i,i) = \sum_i \min(p(i), q(i)) = \frac{1}{2}\sum_i [p(i) + q(i) - |p(i) - q(i)|] = 1 - \frac{1}{2}\|p-q\|_1$, and therefore $\Pr[X \neq Y] = 1 - \Pr[X = Y] = \frac{1}{2}\|p - q\|_1$. Moreover, ω^* is a coupling of p and q. We have

$$\sum_j w^*(i,j) = \min(p(i), q(i)) + \frac{(p(i)-q(i))_+}{\frac{1}{2}\|p-q\|_1} \sum_{j \neq i} (q(j) - p(j))_+.$$

[5] This proof is originally due to Csiszár (see [Pet08, Theorem 3.8]).

If $p(i) \leq q(i)$, then $(p(i) - q(i))_+ = 0$, and hence $\sum_j w^*(i,j) = p(i)$. Otherwise,

$$\sum_j w^*(i,j) = q(i) + \frac{p(i) - q(i)}{\frac{1}{2}\|p - q\|_1} \sum_j (q(j) - p(j))_+.$$

However, $\sum_j (q(j) - p(j))_+ = \frac{1}{2}\|p - q\|_1$, yielding $\sum_j w^*(i,j) = p(i)$. Checking $\sum_i w^*(i,j) = q(j)$ is similar. □

Let $(X,Y) \sim \omega^*$. As $H(X,Y) \geq H(X)$, we have

$$H(X) - H(Y) \leq H(X,Y) - H(Y) = H(X|Y) \leq h(\varepsilon) + \varepsilon \log(d-1) \tag{95}$$

where the second inequality is Fano's inequality, Lemma 2.3. Since $H(X) = H(p)$ and $H(Y) = H(q)$, this concludes the proof of the inequality (93).

Now, assume we attain equality in (93). Note first that we must have $\frac{1}{2}\|p - q\|_1 = \varepsilon$. Otherwise, if $\frac{1}{2}\|p - q\|_1 = \varepsilon' < \varepsilon < 1 - \frac{1}{d}$, then by (93),

$$H(p) - H(q) \leq \varepsilon' \log(d-1) + h(\varepsilon') < \varepsilon \log(d-1) + h(\varepsilon)$$

by the strict monotonicity of $\varepsilon \mapsto \varepsilon \log(d-1) + h(\varepsilon)$ for $\varepsilon \in [0, 1 - \frac{1}{d})$, which can be confirmed by differentiation. As $H(X,Y) = H(X) + H(Y|X)$, to have equality in $H(X,Y) \geq H(X)$, we require $H(Y|X) = 0$. The second inequality in (95) is Fano's inequality, and to obtain equality, we require

$$\frac{\omega^*(i,j)}{q(j)} = \begin{cases} 1 - \varepsilon & i = j \\ \frac{\varepsilon}{d-1} & i \neq j \end{cases} \tag{96}$$

whenever $q(j) \neq 0$, by Lemma 2.3. Let k be such that $q(k) > 0$. Then by (94) and (96),

$$\omega^*(k,k) = \min(p(k), q(k)) = (1 - \varepsilon)q(k).$$

Since $q(k) > 0$, we cannot have $q(k) = (1 - \varepsilon)q(k)$, and therefore $p(k) = (1 - \varepsilon)q(k)$. Next, by (96), for $i \neq k$,

$$q(k)\frac{\varepsilon}{d-1} = \omega^*(i,k) = \frac{((p(i) - q(i))_+ \cdot (q(k) - p(k))_+}{\varepsilon}$$

using (94) for the second equality. Since $(q(k) - p(k))_+ = \varepsilon q(k)$, we have $(p(i) - q(i))_+ = \frac{\varepsilon}{d-1}$, and thus

$$p(i) = q(i) + \frac{\varepsilon}{d-1}. \tag{97}$$

Now, if we sum (97) over all i such that $i \neq k$, we obtain

$$1 - p(k) = \sum_{i \neq k} p(i) = \sum_{i \neq k} q(i) + \varepsilon = 1 - q(k) + \varepsilon.$$

Substituting $p(k) = (1 - \varepsilon)q(k)$, we have

$$\varepsilon q(k) = 1 - q(k) + \varepsilon$$

That is, $q(k) = 1$, and therefore $p(k) = 1 - \varepsilon$. This fixes $q(i) = 0$ for $i \neq k$, so (97) yields $p(i) = \frac{\varepsilon}{d-1}$ for $i \neq k$. This completes the proof.

For completeness, one can note that although we did not use the assumption $H(Y|X) = 0$ directly, the deductions above from equality in Fano's inequality yield

$$\omega^*(i,j) = \begin{cases} 0 & \text{if } i = j \neq k \text{ or } j \neq i = k, \\ \frac{\varepsilon}{d-1} & \text{if } i \neq j = k, \\ 1 - \varepsilon & \text{if } i \neq j = k \end{cases} \implies \frac{\omega^*(i,j)}{p(i)} = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k, \end{cases}$$

and therefore $H(Y|X) = 0$, as required. □

Proof of Proposition 3.6 From the discussion at the beginning of the Section, it only remains to establish the equality conditions for Equation (20). By (90), if either $S(\rho_\varepsilon^*(\sigma)) - S(\sigma) = \mathrm{AF}(\varepsilon)$ or $S(\sigma) - S(\sigma_{*,\varepsilon}) = \mathrm{AF}(\varepsilon)$, then we achieve equality in (20). Here, $\mathrm{AF}(\varepsilon)$ is as given by (91).

If σ is a pure state, in some basis we can write $\sigma = |0\rangle\langle 0|$. Then

$$\sigma_\varepsilon^* = (1-\varepsilon)\,|0\rangle\langle 0| + \sum_{i=1}^{d-1} \frac{\varepsilon}{d-1}\,|i\rangle\langle i|$$

yielding $S(\sigma_\varepsilon^*) = -(d-1)\frac{\varepsilon}{d-1}\log(\frac{\varepsilon}{d-1}) - (1-\varepsilon)\log(1-\varepsilon) = \mathrm{AF}(\varepsilon)$. Likewise, if $\sigma = (1-\varepsilon)\,|0\rangle\langle 0| + \sum_{i=1}^{d-1}\frac{\varepsilon}{d-1}\,|i\rangle\langle i|$, then $|0\rangle\langle 0| \in B_\varepsilon(\sigma)$. Hence, $\rho_{*,\varepsilon}(\sigma) = |0\rangle\langle 0|$, and by the same computation, as for the pure state case, we recover $S(\sigma) - S(\sigma_{*,\varepsilon}) = \mathrm{AF}(\varepsilon)$.

On the other hand, let us assume equality in (20). Then in particular, either $S(\rho_\varepsilon^*(\sigma)) - S(\sigma) = \mathrm{AF}(\varepsilon)$ or $S(\sigma) - S(\sigma_{*,\varepsilon}) = \mathrm{AF}(\varepsilon)$.

1. Case 1: $S(\rho_\varepsilon^*(\sigma)) - S(\sigma) = \mathrm{AF}(\varepsilon)$. Since $S(\rho_\varepsilon^*(\sigma)) \geq S(\sigma)$, we have by Lemma 2.2 that σ must be a pure state.
2. Case 2: $S(\sigma) - S(\sigma_{*,\varepsilon}) = \mathrm{AF}(\varepsilon)$. Then by Lemma 2.2, as $S(\sigma) \geq S(\rho_{*,\varepsilon})$, if $\varepsilon < 1 - \frac{1}{d}$, then $\sigma = \mathrm{diag}(1-\varepsilon, \frac{\varepsilon}{d-1}, \ldots, \frac{\varepsilon}{d-1})$, and if $\varepsilon \geq 1 - \frac{1}{d}$, then $\sigma = \tau$.

This completes the proof. □

References

[AE05] K. M. R. Audenaert and J. Eisert. "Continuity bounds on the quantum relative entropy". *Journal of Mathematical Physics* 46.10 (Oct. 2005), p. 102104.

[AE11] K. M. R. Audenaert and J. Eisert. "Continuity bounds on the quantum relative entropy — II". *Journal of Mathematical Physics* 52.11 (Nov. 2011), p. 112201. arXiv: 1105.2656 [math-ph].

[AF04] R. Alicki and M. Fannes. "Continuity of quantum conditional information". *Journal of Physics A: Mathematical and General* 37.5 (2004), p. L55.

[AU82] P. Alberti and A. Uhlmann. *Stochasticity and partial order : doubly stochastic maps and unitary mixing*. Berlin: Dt. Verl. d. Wiss. u.a, 1982. ISBN: 978-90-277-1350-6.

[Aud07] K. M. R. Audenaert. "A sharp continuity estimate for the von Neumann entropy". *Journal of Physics A: Mathematical and Theoretical* 40.28 (2007), p. 8127.

[Bae11] J. C. Baez. "Renyi Entropy and Free Energy". *ArXiv e-prints* (Feb. 2011). arXiv: 1102.2098 [quant-ph].

[Bha97] R. Bhatia. *Matrix Analysis*. Vol. 169. Springer, 1997. ISBN: 978-0-387-94846-1.

[Bos+16] G. M. Bosyk, S. Zozor, F. Holik, M. Portesi, and P. W. Lamberti. "A family of generalized quantum entropies: definition and properties". *Quantum Information Processing* 15.8 (2016), pp. 3393–3420. ISSN: 1573-1332.

[Buž+99] V. Bužek, G. Drobný, R. Derka, G. Adam, and H. Wiedemann. "Quantum State Reconstruction From Incomplete Data". *Chaos, Solitons & Fractals* 10.6 (1999), pp. 981–1074. ISSN: 0960-0779.

[Che+17] Z. Chen, Z. Ma, I. Nikoufar, and S.-M. Fei. "Sharp continuity bounds for entropy and conditional entropy". *Science China Physics, Mechanics, and Astronomy* 60, 020321 (Feb. 2017), p. 020321. arXiv: 1701.02398 [quant-ph].

[Fan73] M. Fannes. "A continuity property of the entropy density for spin lattice systems". *Communications in Mathematical Physics* 31.4 (1973), pp. 291–294. ISSN: 1432-0916.

[FH61] R. M. Fano and D. Hawkins. "Transmission of Information: A Statistical Theory of Communications". *American Journal of Physics* 29 (Nov. 1961), pp. 793–794.

[FL11] S. T. Flammia and Y.-K. Liu. "Direct Fidelity Estimation from Few Pauli Measurements". *Phys. Rev. Lett.* 106 (23 June 2011), p. 230501.

[Han07] T. Han. *Mathematics of Information and Coding.* City: Amer. Mathematical Society, 2007. ISBN: 978-0-8218-4256-0.

[HD17a] E. P. Hanson and N. Datta. "Maximum and minimum entropy states yielding local continuity bounds". *ArXiv e-prints* (June 2017). arXiv: `1706.02212 [quant-ph]`.

[HD17b] E. P. Hanson and N. Datta. "Tight uniform continuity bound for a family of entropies". *ArXiv e-prints* (July 2017). arXiv: `1707.04249 [quant-ph]`.

[HLP29] G. H. Hardy, J. E. Littlewood, and G. Pólya. "Some simple inequalities satisfied by convex functions". *Messenger Mathematics* 58 (1929), pp. 145–152.

[Hor54] A. Horn. "Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix". *American Journal of Mathematics* 76.3 (1954), pp. 620–630. ISSN: 00029327, 10806377.

[HOS17] M. Horodecki, J. Oppenheim, and C. Sparaciari. "Approximate Majorization". *ArXiv e-prints* (June 2017). arXiv: `1706.05264 [quant-ph]`.

[Jay57] E. T. Jaynes. "Information Theory and Statistical Mechanics. II". *Phys. Rev.* 108 (2 Oct. 1957), pp. 171–190.

[Joz94] R. Jozsa. "Fidelity for Mixed Quantum States". *Journal of Modern Optics* 41.12 (1994), pp. 2315–2323.

[LS09] D. Leung and G. Smith. "Continuity of Quantum Channel Capacities". *Communications in Mathematical Physics* 292.1 (2009), pp. 201–215. ISSN: 1432-0916.

[Mar11] A. Marshall. *Inequalities: Theory of majorization and its applications.* New York: Springer Science+Business Media, LLC, 2011. ISBN: 978-0-387-40087-7.

[Mat02] J. Matoušek. *Lectures on discrete geometry.* New York: Springer, 2002. ISBN: 978-0-387-95374-8.

[Mee16] R. van der Meer. "The Properties of Thermodynamical Operations". Bachelor's thesis. Delft University of Technology, 2016.

[MNW17] R. van der Meer, N. H. Y. Ng, and S. Wehner. "Smoothed generalized free energies for thermodynamics". *arXiv preprint arXiv:1706.03193* (June 2017). arXiv: `1706.03193 [quant-ph]`.

[NC11] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* 10th. New York, NY, USA: Cambridge University Press, 2011. ISBN: 978-1-107-00217-3.

[OH17] J. Oppenheim and M. Horodecki. *Smoothing majorization.* Private communication. 2017.

[Pet08] D. Petz. *Quantum Information Theory and Quantum Statistics.* 1st ed. Theoretical and mathematical physics. Springer-Verlag Berlin Heidelberg, 2008.

[Pey15] J. Peypouquet. *Convex Optimization in Normed Spaces.* Springer International Publishing, 2015.

[Ras11] A. E. Rastegin. "Some General Properties of Unified Entropies". *Journal of Statistical Physics* 143 (June 2011), pp. 1120–1135. arXiv: 1012.5356 [quant-ph].

[Ren05] R. Renner. "Security of Quantum Key Distribution". PhD thesis. ETH Zürich, 2005. arXiv: quant-ph/0512258.

[Roc96] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1996. ISBN: 9780691015866,0691015864.

[RR12] J. M. Renes and R. Renner. "One-Shot Classical Data Compression With Quantum Side Information and the Distillation of Common Randomness or Secret Keys". *IEEE Transactions on Information Theory* 58.3 (Mar. 2012), pp. 1985–1991. ISSN: 0018-9448.

[RW15] D. Reeb and M. M. Wolf. "Tight Bound on Relative Entropy by Entropy Difference". *IEEE Transactions on Information Theory* 61.3 (Mar. 2015), pp. 1458–1473. ISSN: 0018-9448.

[Sch23] I. Schur. "Uber eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie". *Sitzungsberichte der Berliner Mathematischen Gesellschaft* 22 (1923), pp. 9–20.

[Sch96] B. Schumacher. "Sending entanglement through noisy quantum channels". *Phys. Rev. A* 54 (4 Oct. 1996), pp. 2614–2628.

[Shi15] M. E. Shirokov. "Tight continuity bounds for the quantum conditional mutual information, for the Holevo quantity and for capacities of quantum channels". *ArXiv e-prints* (Dec. 2015). arXiv: 1512.09047 [quant-ph].

[Shi16] M. E. Shirokov. "Squashed entanglement in infinite dimensions". *Journal of Mathematical Physics* 57.3 (2016), p. 032203.

[Sim15] B. Simon. *A comprehensive course in analysis*. Providence, Rhode Island: American Mathematical Society, 2015. ISBN: 978-1-4704-1099-5.

[Sko16] M. Skorski. "How to Smooth Entropy?" *Proceedings of the 42Nd International Conference on SOFSEM 2016: Theory and Practice of Computer Science - Volume 9587*. New York, NY, USA: Springer-Verlag New York, Inc., 2016, pp. 392–403. ISBN: 978-3-662-49191-1.

[SLP11] M. P. da Silva, O. Landon-Cardinal, and D. Poulin. "Practical Characterization of Quantum Devices without Tomography". *Phys. Rev. Lett.* 107 (21 Nov. 2011), p. 210404.

[Sti87] E. Stickel. "On the Fréchet derivative of matrix functions". *Linear Algebra and its Applications* 91 (1987), pp. 83–88. ISSN: 0024-3795.

[Tom12] M. Tomamichel. "A framework for non-asymptotic quantum information theory". PhD thesis. ETH Zürich, 2012. arXiv: 1203.2142.

[Tom16] M. Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. arXiv: 1504.00233.

[Tsa88] C. Tsallis. "Possible generalization of Boltzmann-Gibbs statistics". *Journal of Statistical Physics* 52.1 (July 1988), pp. 479–487. ISSN: 1572-9613.

[vH02] W. van Dam and P. Hayden. "Renyi-entropic bounds on quantum communication". *eprint arXiv:quant-ph/0204093* (Apr. 2002). eprint: quant-ph/0204093.

[Wan+16] Y.-X. Wang, L.-Z. Mu, V. Vedral, and H. Fan. “Entanglement Rényi α entropy”. *Phys. Rev. A* 93 (2 Feb. 2016), p. 022324.

[Win16] A. Winter. “Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints”. *Communications in Mathematical Physics* 347 (Oct. 2016), pp. 291–313. arXiv: 1507.07775 [quant-ph].