DTU

DTU Library

From Classical to Quantum
Uniform Continuity Bounds on Entropies in Infinite Dimensions

Becker, Simon; Datta, Nilanjana; Jabbour, Michael G.

Link back to DTU Orbit

From Classical to Quantum: Uniform Continuity bounds on entropies in Infinite Dimensions

Simon Becker, Nilanjana Datta and Michael G Jabbour

Abstract—We prove a variety of improved uniform continuity bounds for entropies of both classical random variables on an infinite state space and of quantum states of infinite-dimensional systems. We obtain the first tight continuity estimate on the Shannon entropy of random variables with a countably infinite alphabet. The proof relies on a new mean-constrained Fano-type inequality. We then employ this classical result to derive a tight energy-constrained continuity bound for the von Neumann entropy. To deal with more general entropies in infinite dimensions, *e.g.* α-Rényi and α-Tsallis entropies, we develop a novel approximation scheme based on operator Hölder continuity estimates. Finally, we settle an open problem raised by Shirokov [1], [2] regarding the characterisation of states with finite entropy.

I. INTRODUCTION

IT has been known, at least since the thorough study of entropies in infinite dimensions in [3], that the von Neumann entropy is a discontinuous function of density operators with respect to trace distance in any infinite-dimensional Hilbert space. In the same article, it has been observed that continuity can be restored by imposing an additional energy constraint on the density operators. Since then continuity bounds under energy constraints for entropies in infinite dimensions have been widely studied and a particularly comprehensive and practicable list of continuity bounds has been obtained by Winter [4]. Most noteworthy for our purposes is his Lemma 15 in which he shows that for density operators satisfying a uniform energy constraint with respect to a Hamiltonian satisfying the so-called *Gibbs hypothesis* (cf. Def. III.2), the von Neumann entropy becomes a continuous function of the density operator. Similar bounds for general α-Rényi entropies or α-Tsallis entropies, in the regime $\alpha \in (0,1)$, are missing and existing continuity bounds are limited to the finite-dimensional case [5], [6]. We also provide estimates in the technically simpler regime for $\alpha > 1$.

While the derivation of continuity bounds for classical entropies of random variables with infinite state space is interesting in itself, our ultimate goal is to provide new continuity bounds for the entropies of states of infinite-dimensional quantum systems. This would prove particularly useful in the context of continuous-variable (CV) quantum systems *e.g.* collections of electromagnetic modes travelling along an optical fibre [7], [8]. The natural Hamiltonian of these systems is the so-called boson *number operator*. Such systems are of immense technological and experimental relevance since promising proposed protocols for quantum communication and computation rely on them. Consequently they have been the subject of extensive research in recent years. Continuity bounds for quantum entropies of states of CV systems which satisfy an energy constraint with respect to the number operator are of particular importance since they would lead to bounds on optimal communication rates in protocols which employ them.

Shannon and von Neumann entropy. Our first result is concerned with providing a tight version of Winter's bound in [4] on the difference of von Neumann entropies for two states, when the Hamiltonian imposing an energy constraint on the input states is the number operator. The bound obtained by Winter is asymptotically tight, see also [9], [10], [11].

In contrast, for any given energy threshold E, our bound (cf. Theorem 5) is tight for all values of the trace distances (ε) such that $\varepsilon \in [0, E/(E+1)]$. The proof of this new estimate builds upon a new Fano inequality for classical random variables on the natural numbers. Let X, Y be random variables with finite state space. Fano's inequality relates the conditional Shannon entropy $H(X|Y)$ and the error probability $\mathbb{P}(X \neq Y)$, and is one of the most elementary and yet important examples of entropic inequalities which are of fundamental importance in information theory. If X and Y take values in a finite alphabet $\mathcal{A}$, and $\mathbb{P}(X \neq Y) = \varepsilon$, then Fano's inequality states that

$$H(X|Y) \leq \varepsilon \log(|\mathcal{A}| - 1) + h(\varepsilon), \tag{I.1}$$

where $h(\varepsilon) := -\varepsilon \log \varepsilon - (1 - \varepsilon) \log(1 - \varepsilon)$ is the binary entropy. Throughout the manuscript $\log$ denotes the natural logarithm.

However, if the alphabet $\mathcal{A}$ is countably infinite, then the above inequality does not hold. In fact, it is even possible for $H(X|Y)$ to remain non-zero in the limit $\varepsilon \to 0$. This is a consequence of the discontinuity of the Shannon entropy for such alphabets. Hence, it is interesting to derive generalized forms of Fano's inequality in this case, under suitable constraints, which ensure that $H(X|Y)$ tends to zero as $\varepsilon \to 0$. One such generalization was obtained by Ho and Verdu [12] in which the marginal p_X of the joint distribution p_{XY} was fixed. See also [13] for further generalizations. In this paper we consider a different generalization of Fano's inequality for a countably infinite alphabet, namely, one in which the means of X and Y are constrained to be below a prescribed threshold

S. Becker is in the Mathematics Department at ETH Zurich (e-mail: simon.becker@math.ethz.ch).

N. Datta is in the Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom (e-mail: n.datta@statslab.cam.ac.uk).

M. G. Jabbour is in the Department of Physics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark (e-mail: mgija@dtu.dk).

value. Hence, we refer to this inequality as a *mean-constrained Fano's inequality*.

We then employ this inequality to derive a tight uniform continuity bound for the Shannon entropy of random variables with a countably infinite alphabet by using the notion of maximal coupling, thus extending the work of Sason in [14]. Next we use this classical continuity bound to obtain a tight uniform continuity bound on the von Neumann entropy of states of an infinite-dimensional quantum system satisfying an energy constraint, in the case in which the Hamiltonian is the number operator.

Rényi and Tsallis entropies. We then turn to α-Rényi- and α-Tsallis entropies, for which in infinite dimensional quantum systems[1] no general continuity bounds have been established so far when $\alpha \in (0, 1)$. We also discuss the case that $\alpha > 1$ where dimension-independent estimates have only been established for the Tsallis entropy [16], [15], [6]. The corresponding bounds for the α-Rényi entropy [15], [17], [6] are dimension-dependent and therefore do not apply to infinite dimensional quantum systems. We would like to emphasize that the study of continuity bounds for α-Rényi- and α-Tsallis entropies is rather different from the von Neumann setting. For example, while the von Neumann entropy is a continuous map on the set of states with uniformly bounded energy with respect to the number operator, this fails to be true for general α-Rényi and α-Tsallis entropies. To illustrate this, consider a state ρ with eigenvalues

$$\mathrm{Spec}(\rho) = \left\{ \frac{1}{(i+1)^{1/\alpha}} \frac{1}{\zeta(1/\alpha)}; i \in \mathbb{N}_0 \right\},$$

where ζ is the Riemann zeta function. For $\alpha < 1/2$, this state has bounded energy with respect to the number operator $\hat{N}$, as $\mathrm{Tr}(\rho \hat{N}) < \infty$ for $\alpha < 1/2$. However, for this range of α, $\mathrm{Tr}(\rho^{\alpha}) = \infty$ which implies that neither the α-Rényi entropy nor the α-Tsallis entropy exists. Therefore, simple energy constraints by the number operator are in general insufficient to obtain continuity bounds for such entropies and we must use a different approach, as the one we pursue for the von Neumann entropy. Using recent results on operator Hölder continuous functions, we are able to obtain continuity estimates for α-Rényi and Tsallis entropies. In fact, we identify technical spectral conditions on the Hamiltonian under which an energy constraint by the Hamiltonian gives rise to continuity bounds for any $\alpha \in (0, 1)$ in Lemma A.1 in the appendix.

In this article, we obtain such bounds, after first deriving them for discrete and continuous random variables in Section III-C. We do this by outlining a simple procedure in Theorem 7, based on the proof of the gentle measurement lemma [18, Lemma 9], to obtain continuity bounds for Hölder continuous functions, but also for functions f with different regularity, of density operators in infinite-dimensional Hilbert spaces under more refined energy constraints on the state. For instance, α-Rényi and α-Tsallis entropies, for $\alpha \in (0, 1)$, depend on functions $f_{\alpha}(x) = x^{\alpha}$ of the density operator.

---
[1] For finite-dimensional systems, a tight uniform continuity bound for α-Rényi entropies for $\alpha \in (0, 1)$ was proved by Audenaert in [15].

For states ρ, σ, our procedure allows us to obtain bounds on the trace distance $\|f_{\alpha}(\rho) - f_{\alpha}(\sigma)\|_1$ which easily leads to continuity bounds for entropies. In a nutshell, we identify constraints, cf. Theorem 7, under which we can approximate the operator $f(\rho)$ by a finite rank operator. We then utilize a continuity result on the level of the finite-rank operator to derive a continuity estimate for the function of the density operator.

FA-property. In a recent series of papers [1], [2], Shirokov identified a property that he coined the *Finite-dimensional Approximation (FA) property* on density operators. He showed that the set of density operators satisfying FA contains almost all states of finite entropy but left open the following question: *Does any state with finite entropy necessarily satisfy the FA-property?* We give a negative answer to this question. The strength of the newly introduced FA-property is due to various approximation, continuity, and stability estimates obtained for states satisfying that property in the papers [1], [2]. The FA property for a state is equivalent, see [1, Theo.1] to the existence of a positive semi-definite Hamiltonian $\hat{H}$ with discrete spectrum such that the state has finite energy $\mathrm{Tr}(\rho \hat{H}) < \infty$. In addition, one requires $e^{-\beta \hat{H}}$ to be a trace-class operator with controlled limiting behaviour $\lim_{\beta \downarrow 0} \left(\mathrm{Tr}(e^{-\beta \hat{H}}) \right)^{\beta} = 1$. Using the work by Wehrl [3], one could already conclude from this that any such state has finite entropy, but the converse implication was left as an open problem. We address this is open problem in Theorem 8.

A summary of our main results and layout of the paper. Note that through all this paper we consider random variables with infinite state spaces and quantum states of infinite-dimensional systems.

- *Shannon entropies*: In Section III-A we consider classical random variables with state space $\mathbb{N}_0$ and state a tight mean-constrained Fano's inequality in Theorem 1. Employing this, we obtain a tight mean-constrained continuity bound for the Shannon entropy of such random variables. This is stated in Theorem 3.
- *von Neumann entropies*: In Section III-B, we obtain a tight energy-constrained continuity bound for von Neumann entropies of states of infinite-dimensional quantum systems, when the Hamiltonian is the number operator. This is stated in Theorem 5.
- *Classical α-Rényi and α-Tsallis entropies*: In Section III-C we derive continuity estimates for the classical α-Rényi and α-Tsallis entropies for random variables with discrete and continuous state spaces in Corollaries III.4 and III.6, respectively, without any restrictions on the state space.
- *Quantum α-Rényi and α-Tsallis entropies:* In Section III-D we then introduce a general approximation scheme for functions of quantum states in Theorem 7 that allows us to obtain continuity estimates for the α-Rényi and α-Tsallis entropies with $\alpha \in (0, 1)$ in Corollary III.7. Analogous continuity bounds for the range $\alpha > 1$ are given in Proposition III.8.

- *FA-property*: The final section is on the FA property. In Theorem 8 we answer a question raised by Shirokov in [1], [2] by showing that there exist states of finite entropy that do not satisfy the FA-property.

II. MATHEMATICAL PRELIMINARIES

Notation The countably infinite set of non-negative integers is denoted by $\mathbb{N}_0$ and the set of strictly positive ones by $\mathbb{N}$. We consider random variables X, Y on some probability space $(\Omega, \Sigma, \mathbb{P})$ and, if they are integrable, denote their expectation by $\mathbb{E}(X)$. For X, Y taking values in a countably infinite state space (also referred to as alphabet) $Z := \bigcup_{i \in \mathbb{N}}\{z_i\}$ and positive weights $w = (w_i)$, the total variation distance is given by by

$$\|X - Y\|_{\mathrm{TV}(w)} := \mathrm{TV}_{(w)}(X, Y)$$
$$:= \frac{1}{2}\sum_{i \in \mathbb{N}} w_i |\mathbb{P}(X = z_i) - \mathbb{P}(Y = z_i)|.$$
$$(\mathrm{II.1})$$

The spaces of p-summable and p-integrable functions with weights $w = (w_i)$ or $w(x)$ are denoted by $\ell^p(w)$ and $L^p(w)$ as usual. In the unweighted case, we just omit the argument (w). We denote by $\mathbb{1}_A$ the indicator function of a set A.

We denote by $\mathcal{H}$ a separable infinite-dimensional Hilbert space. The pth Schatten class on $\mathcal{H}$ is denoted by $\mathcal{S}_p$ with norm $\|\cdot\|_p$, and the operator norm is denoted by $\|\cdot\|$. In particular, $\mathcal{T}(\mathcal{H}) := \mathcal{S}_1$ is the Banach space of trace class operators. A state (or density operator) ρ is a positive trace-class operator of unit trace on $\mathcal{H}$. The trace is denoted by Tr. The spectrum of a linear operator T is denoted by $\mathrm{Spec}(T)$. Let ρ, σ be states, the fidelity between them is defined as $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$. The Fuchs-van de Graaf inequality then states that

$$1 - F(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F^2(\rho, \sigma)}. \qquad (\mathrm{II.2})$$

Let $\hat{H}$ be an unbounded positive semi-definite operator and ρ a state. Let $\mathbb{1}_{[0,n]}(\hat{H})$ be the spectral projection of $\hat{H}$ onto energies of at most n, we then define

$$\mathrm{Tr}(\hat{H}\rho) := \sup_{n \in \mathbb{N}} \mathrm{Tr}\left(\hat{H}\,\mathbb{1}_{[0,n]}(\hat{H})\rho\right) \in [0, \infty].$$

A single mode of a continuous variable quantum system can be described by bosonic annihilation and creation operators, $\hat{a}$ and $\hat{a}^\dagger$, respectively. They satisfy the canonical commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$. The so-called bosonic number operator is then defined as $\hat{N} = \hat{a}^\dagger\hat{a}$. In this paper, we focus on infinite-dimensional quantum systems governed by the Hamiltonian $\hat{H} = \hat{N}$.

The entropies considered in this paper are defined in Section II-A below. In addition to them, the binary entropy is defined as $h(\varepsilon) := -\varepsilon \log \varepsilon - (1 - \varepsilon)\log(1 - \varepsilon)$. We also use the functions $f_1(x) = -x\log(x)$ and $f_\alpha(x) = x^\alpha$ for $\alpha \in (0, 1)$.

We denote by $C^\alpha(I)$ the space of α-Hölder continuous functions on I. We denote by Λ_ω the space of functions continuous with respect to the modulus of continuity ω. Definitions can be found in Section II-B. The integrated modulus of continuity ω^* is defined in Section II-C.

We write $f = \mathcal{O}(g)$ to indicate that there is $C > 0$ such that $|f(x)| \leq C|g(x)|$ for all x in the common domain of functions f and g. We write $f = o(g)$ as x to a if $\lim_{x \to a}\left|\frac{f(x)}{g(x)}\right| = 0$. The error function is denoted by $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$. The cardinality of a set A is denoted by $|A|$.

Finally, we recall the concept of *asymptotic tightness:* As explained, for example in [2], a continuity bound $\sup_{x,y \in S_a} |f(x) - f(y)| \leq C_a(x, y)$ depending on a parameter a (with S_a being a set), is called asymptotically tight for large a if

$$\limsup_{a \to \infty} \sup_{x,y \in S_a} \frac{|f(x) - f(y)|}{C_a(x, y)} = 1.$$

A. Entropies

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Let $X, Y : \Omega \to Z$ be discrete random variables, with $Z = \bigcup_{i \in \mathbb{N}_0}\{z_i\}$ and probabilities $p_X(i) := \mathbb{P}(X = z_i)$ for $i \in \mathbb{N}_0$.

Definition II.1 (Entropies: Discrete random variables). *The Shannon entropy of X is then defined as*

$$H(X) = -\sum_{i=0}^\infty p_X(i)\log(p_X(i)),$$

and for $\alpha \in (0, 1) \cup (1, \infty)$, we introduce the α-Tsallis entropy

$$T_\alpha(X) = \frac{\sum_{i=0}^\infty p_X(i)^\alpha - 1}{1 - \alpha},$$

and α-Rényi entropy

$$R_\alpha(X) = \frac{\log\left(\sum_{i=0}^\infty p_X(i)^\alpha\right)}{1 - \alpha}.$$

Let $p_{XY}(i, j) := \mathbb{P}(X = z_i, Y = z_j)$ be the joint distribution, then the conditional entropy of X given Y is defined as

$$H(X|Y)_p = -\sum_{(i,j) \in \mathbb{N}_0^2} p_{XY}(i, j)\log(p_{XY}(i, j)/p_X(i)).$$

Definition II.2 (Quantum entropies). *The von Neumann entropy of a state ρ of an infinite-dimensional quantum system with Hilbert space $\mathcal{H}$ is defined as*

$$S(\rho) = -\mathrm{Tr}(\rho\log(\rho)),$$

and for $\alpha \in (0, 1)$, we introduce the quantum α-Tsallis entropy

$$T_\alpha(\rho) = \frac{\mathrm{Tr}(\rho^\alpha) - 1}{1 - \alpha},$$

and the quantum α-Rényi entropy

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

B. Quantitative continuity measures for functions

A general quantitative measure of continuity for functions is the so-called *modulus of continuity* $\omega : [0, \infty] \rightarrow [0, \infty]$. We say that a function $f : (X, |\cdot|_X) \rightarrow (Z, |\cdot|_Z)$, where X, Z are subsets of a normed space, admits ω as a modulus of continuity if $|f|_{\Lambda_\omega} := \sup_{x \neq y} \frac{|f(x) - f(y)|_Z}{\omega(|x - y|_X)}$ is finite, in which case

$$|f(x) - f(y)|_Z \leq |f|_{\Lambda_\omega} \omega(|x - y|_X) \text{ for all } x, y \in X.$$

The function ω is assumed to be monotonically-increasing, positive-definite, and subadditive. The vector space of such functions f for which $|f|_{\Lambda_\omega}$ is finite is denoted by Λ_ω. The space of functions $C^\alpha := \Lambda_{\omega_\alpha}$, with $\alpha \in (0, 1)$ and modulus of continuity $\omega_\alpha(t) = t^\alpha$ is called the space of α-*Hölder continuous functions*. We also introduce the space of functions Λ_{AL} that are characterized by a modulus of continuity

$$\omega_{\mathrm{AL}}(t) = \begin{cases} -t \log(t) & \text{for } t \in [0, 1/e] \text{ and} \\ 1/e & \text{otherwise.} \end{cases} .$$

This is the class of local *almost Lipschitz functions*. Here, we employ a cut-off at $t = 1/e$, which is the t at which $-t \log(t)$ attains its maximum, as the maximum distance between discrete probability distributions in total variation distance and states in trace norm is always bounded by two. Many functions related to entropies fall in some of these two spaces C^α or Λ_{AL}.

The function $f_1(x) = -x \log(x)$, related to the Shannon entropy is almost Lipschitz and the functions associated with α-Rényi entropies $f_\alpha(x) = x^\alpha \in C^\gamma([0, \tau])$ for $\gamma \leq \alpha < 1, \tau > 0$ are Hölder continuous. Indeed, that ω_{AL} is a modulus of continuity for f_1 follows from [19, Theo. 17.3.3]

$$|f_1(x) - f_1(y)| \leq f_1(|x - y|) \text{ for all } x, y \in [0, 1/e]. \quad \text{(II.3)}$$

For functions f_α we find that since $|x^\alpha - y^\alpha| \leq |x - y|^\alpha$ then $|f_\alpha|_{\Lambda_{\omega_\alpha}} = 1$. Since f_1 is smooth on $[1/e, \infty)$ this clearly implies that f_1 is almost Lipschitz and therefore in particular Hölder continuous. That almost Lipschitz functions are always Hölder continuous follows from the limit $\lim_{t \to 0} \omega_{\mathrm{AL}}(t)/t^\alpha = 0$.

C. Quantitative continuity measures for functions of operators

The analysis of continuity estimates for functions of self-adjoint operators is more subtle as the regularity of functions is usually not preserved at the operator level. By this we mean that for instance Lipschitz functions f are in general not operator Lipschitz, *i.e.* functions $f : \mathbb{R} \rightarrow \mathbb{R}$ for which there is $C > 0$ such that

$$|f(x) - f(y)| \leq C|x - y| \text{ for all } x, y \in \mathbb{R}$$

do not satisfy $\|f(A) - f(B)\| \leq C'\|A - B\|$ for bounded self-adjoint A, B for any $C' > 0$. An example of a function that is Lipschitz but not operator Lipschitz is the function $f(x) = |x|$.

However, the study of continuity estimates, for functions in Λ_ω, can be transformed to estimate functions of bounded self-adjoint operators. This has been established in a series of papers starting from [20], [21]. The figure of merit for

estimates on the operator level is then the *integrated modulus of continuity* $\omega^*(t) = t \int_t^\infty \frac{\omega(x)}{x^2} \, dx$, with $t \geq 0$ which for the cases we considered before reads

$$\omega_\alpha^*(t) = \frac{t^\alpha}{1 - \alpha} \quad \text{(II.4)}$$

and

$$\omega_{\mathrm{AL}}^*(t) = \begin{cases} \left(e^2 - \frac{1}{2}\right)t + \frac{\log(t)^2 t}{2}, \text{ for } t \in [0, 1/e] \text{ and} \\ e, \text{ otherwise.} \end{cases} \quad \text{(II.5)}$$

Details on the specifics of the continuity bounds obtained in [20], [21] are provided in the beginning of Section III-D.

D. Analytical background

Proposition II.3 (Courant-Fisher)**.** *Let $\hat{H}$ be a self-adjoint operator that is bounded from below with purely discrete spectrum. Let $E_1 \leq E_2 \leq E_3 \leq \cdots$ be the eigenvalues of $\hat{H}$; then*

$$E_n = \min_{\psi_1, \ldots, \psi_n} \max\{\langle \psi, \hat{H}\psi \rangle : \psi \in \mathrm{lin}\{\psi_1, \ldots, \psi_n\}, \|\psi\| = 1\},$$

where lin *denotes the linear hull. In particular, this implies that for $\pi_1, .., \pi_n$ being the first n eigen-projections of $\hat{H}$, counting multiplicity, and $\tilde{\pi}_1, .., \tilde{\pi}_n$ any other mutually orthogonal one-dimensional projections, then we have for any $\lambda_1 \geq ... \geq \lambda_n \geq 0$ that*

$$\mathrm{Tr}\left(\hat{H} \sum_{i=1}^n \lambda_i \pi_i\right) \leq \mathrm{Tr}\left(\hat{H} \sum_{i=1}^n \lambda_i \tilde{\pi}_i\right).$$

We also recall the following simple fact:

Lemma II.4. *Let $f : (0, 1) \rightarrow [0, \infty)$ be a measurable function such that $\int_0^1 \frac{f(t)}{t} dt < \infty$. Then there is a sequence of $t_n \downarrow 0$ such that $f(t_n) \downarrow 0$.*

Proof. If such a sequence does not exist, then $f(t) \geq \epsilon > 0$ for $t \in (0, \delta)$ and $\delta > 0$. Thus, $\int_0^1 \frac{f(t)}{t} \, dt \geq \int_0^\delta \frac{\epsilon}{t} \, dt = \infty$ which contradicts our assumption. $\square$

III. Main Results

A. Shannon entropies

1) A mean-constrained Fano's inequality: Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{N}_0$ be a pair of random variables. Let $\mathcal{P}$ denote the set of joint probability distributions p_{XY} on $\mathbb{N}_0^2$. Since the alphabet is infinite, we need to impose a constraint on the random variables in order for the entropies of the distributions in $\mathcal{P}$ to be finite. We choose a constraint on the means of the marginals, *i.e*, for $p_{XY} \in \mathcal{P}$,

$$\mathbb{E}(X) = \sum_{n=0}^\infty n p_X(n) \leq E, \quad \mathbb{E}(Y) = \sum_{m=0}^\infty m p_Y(m) \leq E, \quad \text{(III.1)}$$

for some finite $E > 0$.

Theorem 1 (Mean-constrained Fano's inequality on $\mathbb{N}_0$)**.** *Let X and Y be a pair of random variables taking values in $\mathbb{N}_0$, with joint probability distribution p_{XY}, satisfying the*

constraint $\mathbb{E}(X) \leq E$ *for some* $0 < E < \infty$. *Then for all* $\varepsilon \in [0, E/(E+1)]$ *the following inequality holds:*

$$H(X|Y)_p \leq h(\varepsilon) + Eh(\varepsilon/E), \tag{III.2}$$

where $\varepsilon := \mathbb{P}(X \neq Y)$. *Furthermore, for any given* E, *the inequality is tight for all* $\varepsilon \in [0, E/(E+1)]$.

Proof. Let us define the set $Z := \{m \in \mathbb{N}_0 : p_{XY}(m,m) \geq p_{XY}(0,m)\}$ and the new random variable X' such that the joint probability distribution $p_{X'Y}$ is defined as follows: for all $m \in \mathbb{N}_0$,

$$p_{X'Y}(0,m) := \begin{cases} p_{XY}(m,m), & \forall m \in Z, \\ p_{XY}(0,m), & \text{else,} \end{cases}$$

for all $m \in \mathbb{N}$,

$$p_{X'Y}(m,m) := \begin{cases} p_{XY}(0,m), & \forall m \in Z, \\ p_{XY}(m,m), & \text{else,} \end{cases}$$

and $p_{X'Y}(n,m) := p_{XY}(n,m)$, otherwise. First note that $\mathbb{E}(X') \leq \mathbb{E}(X) \leq E$ since

$$\begin{aligned}
\mathbb{E}(X') &= \sum_{n \in \mathbb{N}_0} n p_{X'}(n) = \sum_{n \in \mathbb{N}_0} n \sum_{m \in \mathbb{N}_0} p_{X'Y}(n,m) \\
&= \sum_{n \in Z} n\, p_{X'Y}(n,n) + \sum_{n \in \mathbb{N}_0 \setminus Z} n\, p_{X'Y}(n,n) \\
&\quad + \sum_{n \in \mathbb{N}_0} n \sum_{\substack{m \in \mathbb{N}_0 \\ m \neq n}} p_{X'Y}(n,m) \\
&= \sum_{n \in Z} n\, p_{XY}(0,n) + \sum_{n \in \mathbb{N}_0 \setminus Z} n\, p_{XY}(n,n) \\
&\quad + \sum_{n \in \mathbb{N}_0} n \sum_{\substack{m \in \mathbb{N}_0 \\ m \neq n}} p_{XY}(n,m) \\
&\leq \sum_{n \in Z} n\, p_{XY}(n,n) + \sum_{n \in \mathbb{N}_0 \setminus Z} n\, p_{XY}(n,n) \\
&\quad + \sum_{n \in \mathbb{N}_0} n \sum_{\substack{m \in \mathbb{N}_0 \\ m \neq n}} p_{XY}(n,m) \\
&= \mathbb{E}(X) \leq E,
\end{aligned} \tag{III.3}$$

where the first inequality follows from the definition of the set Z, while the second follows from the constraint on the mean: $\mathbb{E}(X) \leq E$. Secondly, note that $H(X'|Y) = H(X|Y)$, since $H(X'Y) = H(XY)$, which can be easily checked.

Using the fact that conditioning decreases entropy, we end up with

$$H(X|Y) = H(X'|Y) \leq H(X'). \tag{III.4}$$

Hence, to complete the proof, it suffices to find an upper bound on $H(X')$. Define $\varepsilon' := 1 - p_{X'}(0)$ and note that

$$\begin{aligned}
\varepsilon' &= 1 - \sum_{m \in \mathbb{N}_0} p_{X'Y}(0,m) \\
&= 1 - \sum_{m \in Z} p_{X'Y}(0,m) - \sum_{m \in \mathbb{N}_0 \setminus Z} p_{X'Y}(0,m) \\
&= 1 - \sum_{m \in Z} p_{XY}(m,m) - \sum_{m \in \mathbb{N}_0 \setminus Z} p_{XY}(0,m).
\end{aligned} \tag{III.5}$$

Since $p_{XY}(m,m) \geq p_{XY}(0,m)$ if and only if $m \in Z$, we have that $p_{XY}(m,m) < p_{XY}(0,m)$ for all $m \in \mathbb{N}_0 \setminus Z$, so that

$$\begin{aligned}
\varepsilon' &\leq 1 - \sum_{m \in Z} p_{XY}(m,m) - \sum_{m \in \mathbb{N}_0 \setminus Z} p_{XY}(m,m) \\
&= 1 - \sum_{m \in \mathbb{N}_0} p_{XY}(m,m) = \mathbb{P}(X \neq Y) = \varepsilon,
\end{aligned} \tag{III.6}$$

which in turn implies that $\sum_{n=1}^{\infty} p_{X'}(n) = \varepsilon' \leq \varepsilon$. Hence,

$$\begin{aligned}
H(X') &= -(1-\varepsilon')\log(1-\varepsilon') - \sum_{n=1}^{\infty} p_{X'}(n)\log p_{X'}(n) \\
&= h(\varepsilon') + \varepsilon'\log\varepsilon' - \sum_{n=1}^{\infty} p_{X'}(n)\log p_{X'}(n) \\
&= h(\varepsilon') - \sum_{n=1}^{\infty} p_{X'}(n)\log\frac{p_{X'}(n)}{\varepsilon'} \\
&= h(\varepsilon') - \varepsilon'\sum_{n=0}^{\infty} r(n)\log r(n) = h(\varepsilon') + \varepsilon' H(R)
\end{aligned} \tag{III.7}$$

where R is a random variable taking values in $\mathbb{N}_0$ with distribution

$$\mathbb{P}(R = n) := \frac{p_{X'}(n+1)}{\varepsilon'}, \quad \forall n \in \mathbb{N}_0.$$

To find an upper bound on $H(R)$, we estimate $\mathbb{E}(R)$ using (III.3)

$$\mathbb{E}(R) = \frac{1}{\varepsilon'}\sum_{n=1}^{\infty}(n-1)p_{X'}(n) = \frac{1}{\varepsilon'}\mathbb{E}(X') - 1 \leq \frac{E}{\varepsilon'} - 1. \tag{III.8}$$

It is known that the geometric distribution achieves maximum entropy among all distributions of a given mean on $\mathbb{N}_0$. This allows us to upper bound $H(R)$ with the entropy of an appropriate geometric random variable. Let Z denote a geometric random variable with parameter $p \in (0,1)$, that is $\mathbb{P}(Z = k) = (1-p)^k p$ for $k \in \mathbb{N}_0$. Its mean and its Shannon entropy are respectively given by

$$\mathbb{E}(Z) = \frac{1-p}{p}, \quad \text{and} \quad H(Z) = \frac{h(p)}{p}, \tag{III.9}$$

and the entropy is a decreasing function of the parameter p. By setting $\mathbb{E}(Z) = E/\varepsilon' - 1$ (which is the upper bound on $\mathbb{E}(R)$) we obtain

$$p = \frac{1}{\mathbb{E}(Z)+1} = \frac{\varepsilon'}{E} \text{ and hence } H(R) \leq H(Z) = \frac{E}{\varepsilon'}h(\varepsilon'/E). \tag{III.10}$$

From (III.4), (III.7) and (III.10), it follows that

$$H(X|Y)_p \leq h(\varepsilon') + Eh(\varepsilon'/E), \tag{III.11}$$

with $\varepsilon' \leq \varepsilon$. By studying its derivative, it is easy to see that the right hand side of (III.11) is an increasing function of ε' for all $\varepsilon' \in [0, E/(E+1)]$. As a result, we end up with

$$H(X|Y)_p \leq h(\varepsilon) + Eh(\varepsilon/E), \tag{III.12}$$

for all $\varepsilon \in [0, E/(E+1)]$, which proves (III.2).

In order to see that the above inequality is tight for $\varepsilon \in [0, E/(E+1)]$, consider the random variables (X^*, Y^*)

characterized by the joint probability distribution $p_{X^*Y^*}$ which is defined as follows:

$$p_{X^*Y^*}(n,m) := \begin{cases} 1-\varepsilon & \text{if } n=0, m=0, \\ \varepsilon w(n-1) & \forall n \in \mathbb{N}, \text{ for } m=0, \\ 0, & \text{else}, \end{cases}$$
(III.13)

where $\{w(n)\}_{n \in \mathbb{N}_0}$ is the probability distribution of the geometric random variable W with mean $\mathbb{E}(W) = E/\varepsilon - 1$. First note that

$$\mathbb{E}(X^*) = \sum_{n \in \mathbb{N}_0} n \sum_{m \in \mathbb{N}_0} p_{X^*Y^*}(n,m) = \sum_{n \in \mathbb{N}_0} n p_{X^*Y^*}(n,0)$$
$$= \varepsilon \sum_{n \in \mathbb{N}_0} (n+1)w(n) = \varepsilon \left(\mathbb{E}(W) + 1 \right) = E.$$
(III.14)

Secondly, note that

$$\mathbb{P}(X^* \neq Y^*) = 1 - \mathbb{P}(X^* = Y^*) = 1 - \sum_{n \in \mathbb{N}_0} p_{X^*Y^*}(n,n) = \varepsilon.$$

Finally, since $H(Y^*) = 0$,

$$H(X^*|Y^*)_p = H(X^*Y^*)$$
$$= -(1-\varepsilon)\log(1-\varepsilon) - \sum_{n \in \mathbb{N}_0} \varepsilon w(n) \log(\varepsilon w(n))$$
$$= h(\varepsilon) - \varepsilon \sum_{n \in \mathbb{N}_0} w(n) \log w(n)$$
$$= h(\varepsilon) + \varepsilon H(W) = h(\varepsilon) + Eh(\varepsilon/E).$$
(III.15)

This proves the theorem. □

Consider a function $f : \mathbb{N}_0 \to \mathbb{R}_+$ that satisfies the following property:

$$\text{for any } \lambda < 0, \quad e^{-\lambda_0} := \sum_{x \in \mathbb{N}_0} e^{\lambda f(x)} < \infty, \quad \text{(III.16)}$$

with $\lambda_0 \in \mathbb{R}$, so that $\{e^{\lambda_0 + \lambda f(x)}\}_{x \in \mathbb{N}_0}$ is a valid probability distribution with finite Shannon entropy. Then for every $E > \inf_{x \in \mathbb{N}_0} f(x)$ we consider the set of all random variables with probability distributions $\{p(x)\}_{x \in \mathbb{N}_0}$ satisfying the constraint $\sum_{x \in \mathbb{N}_0} f(x)p(x) \leq E$. It is known that the maximal Shannon entropy is achieved (uniquely) by the random variable W_E taking values in $\mathbb{N}_0$ with probability distribution $\{w_E(x)\}_{x \in \mathbb{N}_0}$ given by

$$w_E(x) := e^{\lambda_0(E) + \lambda(E)f(x)}, \quad \forall x \in \mathbb{N}_0. \quad \text{(III.17)}$$

Here, $\lambda_0(E)$ and $\lambda(E)$ are chosen such that

$$\sum_{x \in \mathbb{N}_0} w_E(x) = 1, \qquad \mathbb{E}[f(W_E)] = E. \quad \text{(III.18)}$$

We then refer to W_E as the random variable that achieves the maximum Shannon entropy for $\mathbb{E}[f(W_E)] = E$. Similarly, define the function $\tilde{f} : \mathbb{N}_0 \to \mathbb{R}_+$ as

$$\tilde{f}(x) = f(x+1). \quad \text{(III.19)}$$

In that case, $\tilde{f}$ also satisfies (III.16) with f being replaced by $\tilde{f}$. This then guarantees the existence of the random variable $\tilde{W}_E$ that achieves the maximum Shannon entropy for $\mathbb{E}\left[\tilde{f}(\tilde{W}_E)\right] = E$. With these considerations, Theorem 1

can be generalized by considering a constraint of the form $\mathbb{E}[f(X)] \leq E$ instead of a mean constraint.

Theorem 2 (Fano's inequality for countably infinite alphabet with a general constraint). *Let X and Y be a pair of random variables taking values in $\mathbb{N}_0$, having a joint distribution p_{XY}, and satisfying the constraint $\mathbb{E}[f(X)] \leq E$ for some $E \in (0, \infty)$. Then the following inequality holds:*

$$H(X|Y)_p \leq h(\varepsilon) + \varepsilon H(\tilde{W}_{E/\varepsilon}), \quad \text{(III.20)}$$

for $\varepsilon \in \mathcal{E}(E)$, where $\varepsilon := \mathbb{P}(X \neq Y)$, $\mathcal{E}(E) \subseteq [0,1]$ contains the values of ε for which the right-hand side of (III.20) is a non-decreasing function of ε, and $\tilde{W}_{E/\varepsilon}$ is the random variable that achieves the maximum Shannon entropy for $\mathbb{E}\left[\tilde{f}(\tilde{W}_{E/\varepsilon})\right] = E/\varepsilon$, with $\tilde{f}$ being defined in (III.19). Furthermore, the inequality is tight for $\varepsilon \in \mathcal{E}(E)$.

Since the proof of Theorem 2 closely follows the lines of that of Theorem 1, we omit its proof. Note that the right-hand side of (III.20) can be shown to be a concave function of ε that reaches 0 when $\varepsilon = 0$ and is non-negative for all $\varepsilon \in [0,1]$. Consequently, there always exists some $\varepsilon^* \in (0,1]$ such that $\mathcal{E}(E) = [0, \varepsilon^*]$.

2) A mean-constrained continuity bound for the entropy of random variables with a countably infinite alphabet: In [14], Sason exploited the concept of maximal coupling of random variables in order to rederive the standard continuity bound for the Shannon entropy (*i.e.*, for random variables with a finite alphabet) from Fano's inequality. In this section, we extend his proof to obtain a continuity bound for the Shannon entropy of random variables with a countably infinite alphabet, under mean constraints. In order to do so, we make use of the mean-constrained Fano inequality given in Theorem 1, as well as the concept of maximum coupling.

Recall that a *coupling* of a pair of random variables (X,Y) is a pair of random variables $(\hat{X}, \hat{Y})$ with the same marginal probability distributions as of (X,Y).

Definition III.1 (Maximal coupling). *For a pair of random variables (X,Y), a coupling $(\hat{X}, \hat{Y})$ is called a* maximal coupling *if $\mathbb{P}(\hat{X} = \hat{Y})$ attains its maximal value among all the couplings of (X,Y).*

Theorem 3 (Mean-constrained Shannon entropy continuity bound). *Let X and Y be a pair of random variables taking values in $\mathbb{N}_0$, with respective probability distributions p_X and p_Y, and satisfying the constraints $\mathbb{E}(X) \leq E$ and $\mathbb{E}(Y) \leq E$ for some $0 < E < \infty$. Then for all $\varepsilon \in [0, E/(E+1)]$ the following inequality holds:*

$$|H(X)_p - H(Y)_p| \leq h(\varepsilon) + Eh(\varepsilon/E), \quad \text{(III.21)}$$

where $\varepsilon := \|X - Y\|_{\text{TV}}$. Furthermore, for any given $0 < E < \infty$, the inequality is tight for all $\varepsilon \in [0, E/(E+1)]$.

Proof. Again, we take inspiration from the proof of Theorem 3 in [14]. Let $(\hat{X}, \hat{Y})$ be a maximal coupling of (X,Y), and p_{XY} be the corresponding joint probability distribution. Since

$H(X)_p = H(\hat{X})_p$ and $H(Y)_p = H(\hat{Y})_p$, we have, for all $\varepsilon \in [0, E/(E+1)]$,

$$
\begin{aligned}
|H(X)_p - H(Y)_p| &= \left| H(\hat{X})_p - H(\hat{Y})_p \right| \\
&= \left| H(\hat{X}|\hat{Y})_p - H(\hat{Y}|\hat{X})_p \right| \\
&\leq \max \left\{ H(\hat{X}|\hat{Y})_p, H(\hat{Y}|\hat{X})_p \right\} \\
&\overset{(1)}{\leq} h(\varepsilon') + E h\left(\varepsilon'/E\right) \\
&\overset{(2)}{=} h(\varepsilon) + E h\left(\varepsilon/E\right)
\end{aligned}
$$

where $\varepsilon' := \mathbb{P}(\hat{X} \neq \hat{Y})$, and (1) follows from Theorem 1, while (2) is a consequence of the fact that if $(\hat{X}, \hat{Y})$ is a maximal coupling of (X, Y) then

$$
\mathbb{P}(\hat{X} \neq \hat{Y}) = \|X - Y\|_{\mathrm{TV}}.
$$

A proof of this for random variables with finite state space has been given in [14] and the proof for infinite state alphabets is a straightforward adaptation of the argument.

In order to see that the above inequality is tight for $\varepsilon \in [0, E/(E+1)]$, consider the random variables X^* and Y^* characterized by the probability distributions p_{X^*} which is defined as follows:

$$
p_{X^*}(n) := \begin{cases} 1 - \varepsilon & \text{if } n = 0, \\ \varepsilon\, w(n-1) & \text{else,} \end{cases} \tag{III.22}
$$

where $\{w(n)\}_{n \in \mathbb{N}_0}$ is the probability distribution of the geometric random variable W characterized by a mean-value $\mathbb{E}(W) = E/\varepsilon - 1$, and p_{Y^*} which is defined as follows:

$$
p_{Y^*}(m) := \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{else.} \end{cases}
$$

Note that p_{X^*} and p_{Y^*} correspond to the two marginals of $p_{X^*Y^*}$ defined in (III.13). From this and (III.14), we know that $\mathbb{E}(X^*) = E$. From (III.15), we have that $H(X^*) = h(\varepsilon) + E h(\varepsilon/E)$. Obviously, we also have that $\mathbb{E}(Y^*) = 0 < E$ and $H(Y^*) = 0$. Finally, it is easy to see that $\|X^* - Y^*\|_{\mathrm{TV}} = \varepsilon$. This proves the theorem. $\qquad\square$

If a function $f : \mathbb{N}_0 \to \mathbb{R}_+$ satisfies (III.16) and $f(0) = 0$, one has the following generalization of Theorem 3 for a constraint of the form $\mathbb{E}[f(X)] \leq E$. Since its proof is analogous to that of Theorem 3, we omit the proof.

Theorem 4 (Constrained continuity bound for the Shannon entropy of random variables with countably infinite alphabets). *Let X and Y be a pair of random variables taking values in $\mathbb{N}_0$, with respective probability distributions p_X and p_Y, and satisfying the constraints $\mathbb{E}[f(X)] \leq E$ and $\mathbb{E}[f(Y)] \leq E$ for some $E \in (0, \infty)$. Then the following inequality holds:*

$$
|H(X)_p - H(Y)_p| \leq h(\varepsilon) + \varepsilon H(\tilde{W}_{E/\varepsilon}), \tag{III.23}
$$

for $\varepsilon \in \mathcal{E}(E)$, where $\varepsilon := \|X - Y\|_{\mathrm{TV}}$, $\mathcal{E}(E) \subseteq [0, 1]$ contains the values of ε for which the right-hand side of (III.23) is a non-decreasing function of ε, and $\tilde{W}_{E/\varepsilon}$ is the random variable that achieves the maximum Shannon entropy for $\mathbb{E}\left[\tilde{f}(\tilde{W}_{E/\varepsilon}) \right] = E/\varepsilon$, with $\tilde{f}$ being defined in (III.19).

B. von Neumann entropies

We now lift the classical entropy continuity estimate in Theorem 3 to general density operators, i.e. positive trace-class operators on a separable, infinite-dimensional Hilbert space with unit trace. The assumption that the state space of the classical random variables is $\mathbb{N}_0$ enforces us now to take as the Hamiltonian the single-mode number operator, $\hat{N} = a^\dagger a$.

Theorem 5 (von Neumann entropy continuity bound). *Let $\hat{H}$ denote the number operator, and let ρ and σ be two quantum states on a separable, infinite-dimensional Hilbert space $\mathcal{H}$, satisfying the energy constraints $\mathrm{Tr}(\hat{H}\rho), \mathrm{Tr}(\hat{H}\sigma) \leq E$, for some $0 < E < \infty$, such that*

$$
\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon. \tag{III.24}
$$

Then for all $\varepsilon \in [0, E/(1+E)]$ the following inequality holds:

$$
|S(\rho) - S(\sigma)| \leq h(\varepsilon) + E h\left(\varepsilon/E\right). \tag{III.25}
$$

Furthermore, for any given $0 < E < \infty$, the inequality is tight for all $\varepsilon \in [0, E/(E+1)]$.

Before providing the proof of the above theorem, let us recall the known uniform energy-constrained continuity bound for the von Neumann entropy of infinite-dimensional quantum states obtained in [22] by Winter. In his paper, Winter considers a Hamiltonian $\hat{H}$ which has a discrete spectrum, is bounded from below[2] and satisfies the following so-called Gibbs Hypothesis.

Definition III.2. *[22]. A Hamiltonian $\hat{H}$ is said to satisfy the Gibbs Hypothesis if for any $\beta > 0$, $Z(\beta) := \mathrm{Tr}(e^{-\beta\hat{H}}) < \infty$, so that $e^{-\beta\hat{H}}/Z(\beta)$ is a valid quantum state with finite entropy.*

If $\hat{H}$ satisfies the Gibbs hypothesis, then for every $E > \inf \mathrm{Spec}(\hat{H})$, among all states satisfying the energy constraint $\mathrm{Tr}(\hat{H}\rho) \leq E$, the maximal entropy is achieved (uniquely) by the Gibbs state:

$$
\gamma(E) = \frac{e^{-\beta(E)\hat{H}}}{Z(\beta(E))}, \tag{III.26}
$$

where the parameter $\beta(E)$ is decreasing with E and is determined by the equality

$$
\mathrm{Tr}(e^{-\beta\hat{H}}(\hat{H} - E)) = 0. \tag{III.27}
$$

For such a Hamiltonian, Winter proved the following energy-constrained continuity bound [22]: for any two states ρ and σ on a separable, infinite-dimensional Hilbert space with $\mathrm{Tr}(\rho\hat{H}), \mathrm{Tr}(\sigma\hat{H}) \leq E$ and $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$,

$$
|S(\rho) - S(\sigma)| \leq h(\varepsilon) + 2\varepsilon S\left(\gamma(E/\varepsilon)\right). \tag{III.28}
$$

In the case in which the Hamiltonian is the number operator $\hat{N}$ corresponding to a single mode, the bound (III.28) reduces to

$$
|S(\rho) - S(\sigma)| \leq h(\varepsilon) + 2(E + \varepsilon)\, h\left(\frac{\varepsilon}{E + \varepsilon}\right). \tag{III.29}
$$

---

[2]In fact, he fixes the ground state energy to be zero.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3248228
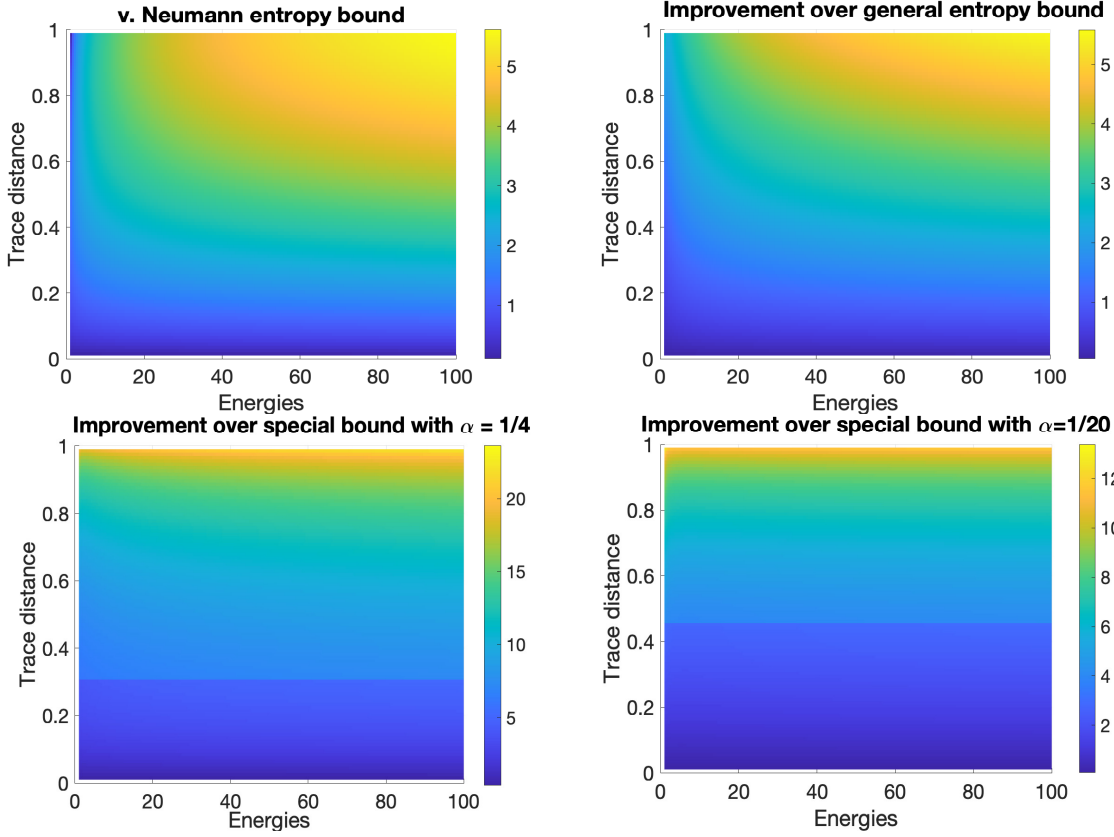
8

Fig. 1. The upper left figure illustrates the right-hand side of our tight bound (III.25). The upper right figure illustrates the difference of the bound (III.28) obtained by Winter [4] to our bound on the von Neumann entropy. The lower two figures compare the bound (III.30) found by Winter, for fixed α, to our bound. The improvement is in all cases particularly significant for high energies and large trace distances.
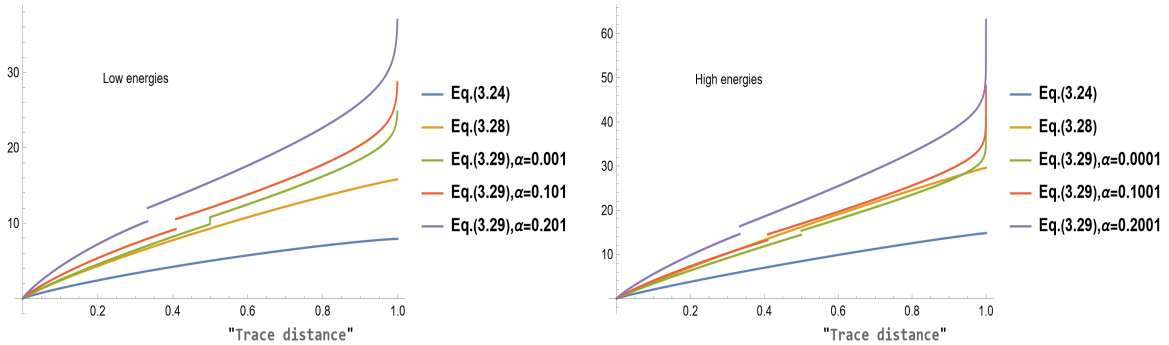

Fig. 2. We compare our tight bound (III.25) to the bound (III.29) by Winter for general Hamitonians specialized to the single-mode number operator and also to (III.30) for different values of α. Due to the piecewise definition of (III.30) the latter curves also show a jump discontinuity. We see that for a wide range of α and low energies, (III.29) outperforms (III.30), but for high energies, there exist values of α for which (III.30) outperforms (III.29).

Winter also gets an additional bound for the single mode number operator $\hat{N}$ (see Lemma 18 of [22]):

$$|S(\rho) - S(\sigma)|$$
$$\leq \varepsilon \left(\frac{1+\alpha}{1-\alpha} + 2\alpha \right) \left[\log(E+1) + \log \frac{\varepsilon}{\alpha(1-\varepsilon)} \right]$$
$$+ 3 \left(\frac{1+\alpha}{1-\alpha} + 2\alpha \right) \tilde{h} \left(\frac{1+\alpha}{1-\alpha} \varepsilon \right) =: K(\varepsilon, \alpha, E),$$
(III.30)

where $\alpha \in (0, 1/2)$ and $\tilde{h}(x) = h(x)$ for $x \leq 1/2$ and $\tilde{h}(x) = 1$ for $x \geq 1/2$. As we explain below and as has been already

observed in [4], the bound (III.30) is asymptotically tight for a suitable joint limit of $\alpha \to 0$ and $E \to \infty$ for every $\varepsilon > 0$ fixed. In addition, it is unknown so far how to optimize the choice of α in (III.30) and for many choices of α, the bound (III.28) seems to be better than the bound (III.30); see Fig. III-B.

In contrast, our bound given in Theorem 5 is tight for all values of the trace distance $\varepsilon \in [0, E/(E+1)])$ for any given $0 < E < \infty$.

Remark 1 (Asymptotic tightness). *We now address the issue*

of asymptotic tightness: the difference of entropies for states ρ, σ with eigenvalues according to the probability distributions optimizing the estimate in the proof of Theorem 3 is given by (III.15)

$$|S(\rho) - S(\sigma)| = h(\varepsilon) + Eh(\varepsilon/E).$$

We then study the behaviour of (III.30) *for a one-parameter family* $E(n) = e^n, \alpha(n) = n^{-1}$ *for any fixed* $\varepsilon > 0$*. Thus, using Taylor expansion*

$$
\begin{aligned}
&E(n)h(\varepsilon/E(n)) \\
&= -\varepsilon \log(\varepsilon e^{-n}) - e^n(1 - \varepsilon e^{-n}) \log(1 - \varepsilon e^{-n}) \\
&= -\varepsilon \log(\varepsilon) + n\varepsilon - (e^n - \varepsilon) \log(1 - \varepsilon e^{-n}) \\
&= n\varepsilon + o(n) \text{ as } n \to \infty.
\end{aligned}
\tag{III.31}
$$

In particular, since $h(\varepsilon)$ is just a constant, we have

$$|S(\rho) - S(\sigma)| = n\varepsilon + o(n) \text{ as } n \to \infty. \tag{III.32}$$

Now observe that

$$\varepsilon \left(\frac{1 + \frac{1}{n}}{1 - \frac{1}{n}} + \frac{2}{n} \right) \log(e^n + 1) = \varepsilon n + o(n) \text{ as } n \to \infty.$$

Since this is the leading order term in the right-hand side of (III.30)*, we thus also have that*

$$K(\varepsilon, \alpha(n), E(n)) = \varepsilon n + o(n) \text{ as } n \to \infty. \tag{III.33}$$

Hence, by combining (III.32) *with* (III.33)

$$\lim_{n \to \infty} \frac{|S(\rho) - S(\sigma)|}{K(\varepsilon, \alpha(n), E(n))} = 1$$

which shows asymptotic tightness.

We now provide a proof of Theorem 5.

Proof. Let the spectral decompositions of ρ and σ be given by

$$\rho = \sum_{n=0}^{\infty} r(n) |\phi_n\rangle \langle \phi_n|, \qquad \sigma = \sum_{n=0}^{\infty} s(n) |\psi_n\rangle \langle \psi_n|. \tag{III.34}$$

Consider the passive states

$$\rho^{\downarrow} = \sum_{n=0}^{\infty} r^{\downarrow}(n) |n\rangle \langle n|, \qquad \sigma^{\downarrow} = \sum_{n=0}^{\infty} s^{\downarrow}(n) |n\rangle \langle n|, \tag{III.35}$$

where the states $|n\rangle$ for $n = 0, 1, \cdots$ are eigenstates of the Hamiltonian $\hat{H}$, $\{r^{\downarrow}(n)\}_{n \in \mathbb{N}_0}$ represents the distribution containing the non-zero elements of $\{r(n)\}_{n \in \mathbb{N}_0}$ arranged in non-increasing order, *i.e.*, $r^{\downarrow}(n) \geq r^{\downarrow}(n+1)$ for all $n \in \mathbb{N}_0$, and similarly for $\{s^{\downarrow}(n)\}_{n \in \mathbb{N}_0}$. We obviously have that $S(\rho^{\downarrow}) = S(\rho)$ and $S(\sigma^{\downarrow}) = S(\sigma)$, so that

$$|S(\rho) - S(\sigma)| = |S(\rho^{\downarrow}) - S(\sigma^{\downarrow})|. \tag{III.36}$$

Furthermore, from the Courant-Fischer theorem in Proposition II.3, we have $\mathrm{Tr}(\hat{H}\rho^{\downarrow}) \leq \mathrm{Tr}(\hat{H}\rho) \leq E$, $\mathrm{Tr}(\hat{H}\sigma^{\downarrow}) \leq \mathrm{Tr}(\hat{H}\sigma) \leq E$ [3] and

$$\varepsilon' := \frac{1}{2} ||\rho^{\downarrow} - \sigma^{\downarrow}||_1 \leq \frac{1}{2} ||\rho - \sigma||_1 \leq \varepsilon. \tag{III.37}$$

[3] This can also be proved using Ky Fan's Maximum Principle [23, Lemma IV.9]

The above inequality also follows from Mirsky's inequality [24], see also [25, (1.22)] for a version in infinite dimensions. Let X and Y denote random variables on $\mathbb{N}_0$, with probability distributions $r \equiv \{r^{\downarrow}(n)\}_{n \in \mathbb{N}_0}$ and $s \equiv \{s^{\downarrow}(n)\}_{n \in \mathbb{N}_0}$, respectively. Then $\mathbb{E}(X) = \mathrm{Tr}(\hat{H}\rho^{\downarrow}) \leq E$, $\mathbb{E}(Y) = \mathrm{Tr}(\hat{H}\sigma^{\downarrow}) \leq E$, $H(X) = S(\rho^{\downarrow})$, $H(Y) = S(\sigma^{\downarrow})$ and $\|X - Y\|_{\mathrm{TV}} = \varepsilon' \leq \varepsilon$. Using Theorem 3, we have

$$|S(\rho) - S(\sigma)| = |H(X) - H(Y)| \leq h(\varepsilon') + Eh(\varepsilon'/E). \tag{III.38}$$

As mentioned before, it is easy to see by analyzing its derivative that the right-hand side of the last inequality in the above equation is an increasing function of ε' for all $\varepsilon' \in [0, E/(E+1)]$. As a result, we end up with

$$|S(\sigma) - S(\rho)| \leq h(\varepsilon) + Eh(\varepsilon/E), \tag{III.39}$$

for all $\varepsilon \in [0, E/(E+1)]$.

In order to see that the above inequality is tight for $\varepsilon \in [0, E/(E+1)]$, consider the quantum states $\rho^* := \sum_{n=0}^{\infty} p_{X^*}(n) |n\rangle \langle n|$ and $\sigma^* = |0\rangle \langle 0|$ where p_{X^*} is the probability distribution defined in (III.22). From this, we have that $S(\rho) = h(\varepsilon) + Eh(\varepsilon/E)$ and $\mathrm{Tr}(\hat{H}\rho) \leq E$. Obviously, we also have that $S(\sigma) = 0$ and $\mathrm{Tr}(\hat{H}\sigma) = 0 < E$. Finally, it is trivial to see that $\frac{1}{2}||\rho^* - \sigma^*||_1 = \varepsilon$. This proves the theorem. $\square$

Consider a Hamiltonian $\hat{H}$ on a separable infinite-dimensional Hilbert space $\mathcal{H}$, with ground state energy 0, that satisfies the Gibbs Hypothesis. As a consequence, it has a discrete spectrum of finite multiplicity and can be represented as [26]

$$\hat{H} = \sum_{n=0}^{\infty} f(n) |e_n\rangle \langle e_n|, \tag{III.40}$$

where $\{|e_n\rangle\}_{n \in \mathbb{N}_0}$ is an orthonormal basis and the eigenvalues $f(n)$ for $n \in \mathbb{N}_0$ are such that $f(n) \leq f(n+1)$. Define the function $\tilde{f} : \mathbb{N}_0 \to \mathbb{R}_+$ as $\tilde{f}(x) = f(x+1)$ and the Hamiltonian $\hat{H}_+$ on $\mathcal{H}$ as

$$\hat{H}_+ = \sum_{n=0}^{\infty} \tilde{f}(n) |e_n\rangle \langle e_n|. \tag{III.41}$$

The following generalization of Theorem 5 can be shown using a similar proof.

Theorem 6 (von Neumann entropy continuity bound for general Hamiltonian)**.** *Let the Hamiltonian $\hat{H}$ on $\mathcal{H}$, with ground state energy 0, satisfy the Gibbs Hypothesis. Let ρ and σ be two quantum states on $\mathcal{H}$, satisfying the energy constraints $\mathrm{Tr}(\hat{H}\rho), \mathrm{Tr}(\hat{H}\sigma) \leq E$, for some $E \in (0, \infty)$, such that*

$$\frac{1}{2} ||\rho - \sigma||_1 \leq \varepsilon, \tag{III.42}$$

with $\varepsilon \in [0, 1]$. Then the following inequality holds:

$$|S(\rho) - S(\sigma)| \leq h(\varepsilon) + \varepsilon S(\tilde{\gamma}(E/\varepsilon)), \tag{III.43}$$

for $\varepsilon \in \mathcal{E}(E)$, where $\mathcal{E}(E) \subseteq [0, 1]$ contains the values of ε for which the right-hand side of (III.43) *is a non-decreasing function of ε, and $\tilde{\gamma}(E/\varepsilon)$ denotes the Gibbs state of energy*

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3248228

10

E/ε corresponding to the Hamiltonian $\hat{H}_+$. Furthermore, the inequality is tight for $\varepsilon \in \mathcal{E}(E)$.

Note that the right-hand side of (III.43) is the same function as the right-hand side of (III.20). Consequently, there always exists some $\varepsilon^* \in (0, 1]$ such that $\mathcal{E}(E) = [0, \varepsilon^*]$ in the above theorem. In order to better understand the size and scaling of the interval $\mathcal{E}(E)$, it is at least necessary to understand the behaviour of the entropy of the Gibbs state $\tilde{\gamma}(E/\varepsilon)$ for small values of ε. This question has been addressed in [11, Theorem 3]. This result states that for any Hamiltonian satisfying a certain spectral condition that is virtually met for all common Hamiltonians satisfying the Gibbs hypothesis, there exists a parameter η such that

$$S(\tilde{\gamma}(E)) = \eta \log(E)(1 + o(1)) \text{ as } E \to \infty. \quad \text{(III.44)}$$

Since for $\varepsilon \in [0, 1/2]$, $\varepsilon \mapsto h(\varepsilon)$ is monotone, the right hand side of (III.43) is increasing, at least for $\varepsilon \in [0, 1/2]$ if $[0, 1/2] \ni \varepsilon \mapsto \varepsilon S(\tilde{\gamma}(E/\varepsilon))$ increases.

Using the asymptotic result (III.44), we see that for either large energies or η sufficiently small, the monotonicity of the right-hand side of (III.43) is expected to hold on a large interval $\mathcal{E}(E)$ of admissible ε. For Schrödinger operators with bounded potential on bounded domains $\Omega \subset \mathbb{R}^n$, it was shown in [11] that $\eta = n/2$ for example. Therefore, one might expect. in addition, that the interval does not increase for Hamiltonians on multi-dimensional domains.

In principle, one could hope to turn (III.44) into a quantitative estimate that provides quantitative estimates on the size of $\mathcal{E}(E)$. In practice, it seems more effective to directly analyze concrete Hamiltonians and exploit specific features about their spectra, such as their high-energy limits, to understand the size of $\mathcal{E}(E)$.

C. Classical α-Rényi and α-Tsallis entropies

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. We start by considering two random variables $X, Y : \Omega \to Z$ where we shall assume that Z is either a discrete countably infinite set or a measurable subset $D \subset \mathbb{R}^d$. In the latter case, we assume that X and Y possess probability densities μ_X, μ_Y, respectively.

1) Discrete random variables: Let $Z := \bigcup_{i \in \mathbb{N}} \{z_i\}$ and $p_X(z_i) := \mathbb{P}(X = z_i)$. We are interested in studying for α-Hölder continuous functions $f : [0, 1] \to \mathbb{R}$ the quantities with $f(p_X) := (f(p_X(z_1)), f(p_X(z_2)), f(p_X(z_3)), ...) \in \ell^1(\mathbb{N})$ with

$$(Tf)(X) := \sum_{i \in \mathbb{N}} f(p_X(z_i)), \quad \text{(III.45)}$$

where $H(X) = Tf_1(X)$ is the *Shannon entropy* of the random variable X, $R_\alpha(X) = \frac{\log(Tf_\alpha(X))}{1-\alpha}$ its *α-Rényi entropy*, and $T_\alpha(X) = \frac{Tf_\alpha(X)-1}{1-\alpha}$ its *α-Tsallis entropy*.

Proposition III.3. *Let $f \in C^\alpha([0, 1])$, $\alpha \in (0, 1)$, and X, Y random variables with discrete countably infinite state space Z. Let $w = (w_i)$ be a sequence of positive weights such that $\left(w_i^{-\frac{\beta}{1-\alpha}} \right) \in \ell^1(\mathbb{N})$ for some β such that $0 < \beta < \alpha$ and*

$(w_i p_X(z_i)), (w_i p_Y(z_i)) \in \ell^1(\mathbb{N})$. *Then we have the following continuity bound:*

$$\|f(p_X) - f(p_Y)\|_{\ell^1}$$
$$\leq 2^\alpha |f|_{\Lambda_{\omega_\alpha}} \|X - Y\|_{\text{TV}(w)}^\beta \|X - Y\|_{\text{TV}}^{\alpha-\beta} \|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha}, \quad \text{(III.46)}$$

where by the triangle inequality $|Tf(X) - Tf(Y)| \leq \|f(p_X) - f(p_Y)\|_{\ell^1}$.

Note that the conditions $\left(w_i^{-\frac{\beta}{1-\alpha}} \right) \in \ell^1(\mathbb{N})$ for some $\beta < \alpha$ and $(w_i p_X(z_i)), (w_i p_Y(z_i)) \in \ell^1(\mathbb{N})$ replace the moment constraint in Eq. III.21.

Proof. Using Hölder continuity of f, we find

$$\|f(p_X) - f(p_Y)\|_{\ell^1} \leq |f|_{\Lambda_{\omega_\alpha}} \sum_{i \in \mathbb{N}} w_i^{-\beta} w_i^\beta |p_X(z_i) - p_Y(z_i)|^\alpha.$$

Choosing $p = 1/\alpha$ and its Hölder conjugate $q = 1/(1-\alpha)$, we have by Hölder's inequality

$$\|f(p_X) - f(p_Y)\|_{\ell^1}$$
$$\leq |f|_{\Lambda_{\omega_\alpha}} \left(\sum_{i \in \mathbb{N}} w_i^{\frac{\beta}{\alpha}} |p_X(z_i) - p_Y(z_i)| \right)^\alpha \left(\sum_{i \in \mathbb{N}} w_i^{-\frac{\beta}{1-\alpha}} \right)^{1-\alpha}. \quad \text{(III.47)}$$

Applying Hölder's inequality to the second summand on the right hand side of the above line with the choice $p = \alpha/\beta$ and $q = \alpha/(\alpha - \beta)$, we obtain

$$\sum_{i \in \mathbb{N}} w_i^{\beta/\alpha} |p_X(z_i) - p_Y(z_i)| \leq \left(\sum_{i \in \mathbb{N}} w_i |p_X(z_i) - p_Y(z_i)| \right)^{\frac{\beta}{\alpha}}$$
$$\times \left(\sum_{i \in \mathbb{N}} |p_X(z_i) - p_Y(z_i)| \right)^{\frac{\alpha-\beta}{\alpha}}, \quad \text{(III.48)}$$

which together with (III.47) yields

$$\|f(p_X) - f(p_Y)\|_{\ell^1}$$
$$\leq |f|_{\Lambda_{\omega_\alpha}} \|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha} \times \|(p_X(z_i) - p_Y(z_i))\|_{\ell^1(w_i)}^\beta$$
$$\times \|(p_X(z_i) - p_Y(z_i))\|_{\ell^1}^{\alpha-\beta}. \quad \text{(III.49)}$$

Expressing the ℓ^1 distances in terms of the total variation distances using (II.1) then yields the claim. $\qquad \square$

Remark 2. *Let $Z \subset \mathbb{Z}$ and $w_i := \max\{|z_i|, 1\}$ then the condition $(w_i p_X(z_i)), (w_i p_Y(z_i)) \in \ell^1(\mathbb{N})$, is equivalent to the existence of a first moment. If $Z = \mathbb{N}$, then $\|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha} = \zeta(\frac{\beta}{1-\alpha})^{1-\alpha}$ where ζ is the Riemann zeta function and we may choose $\beta \in (1-\alpha, \alpha)$. In particular, the constraints restrict us to choosing $\alpha \in (1/2, 1)$.*

As an immediate corollary from Proposition III.3, we then find:

Corollary III.4. *Let $\alpha \in (0, 1)$ and X, Y be random variables with discrete countably infinite state space Z. Let (w_i) be a sequence of positive weights such that $\left(w_i^{-\frac{\beta}{1-\alpha}} \right) \in \ell^1(\mathbb{N})$ for*

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3248228

11

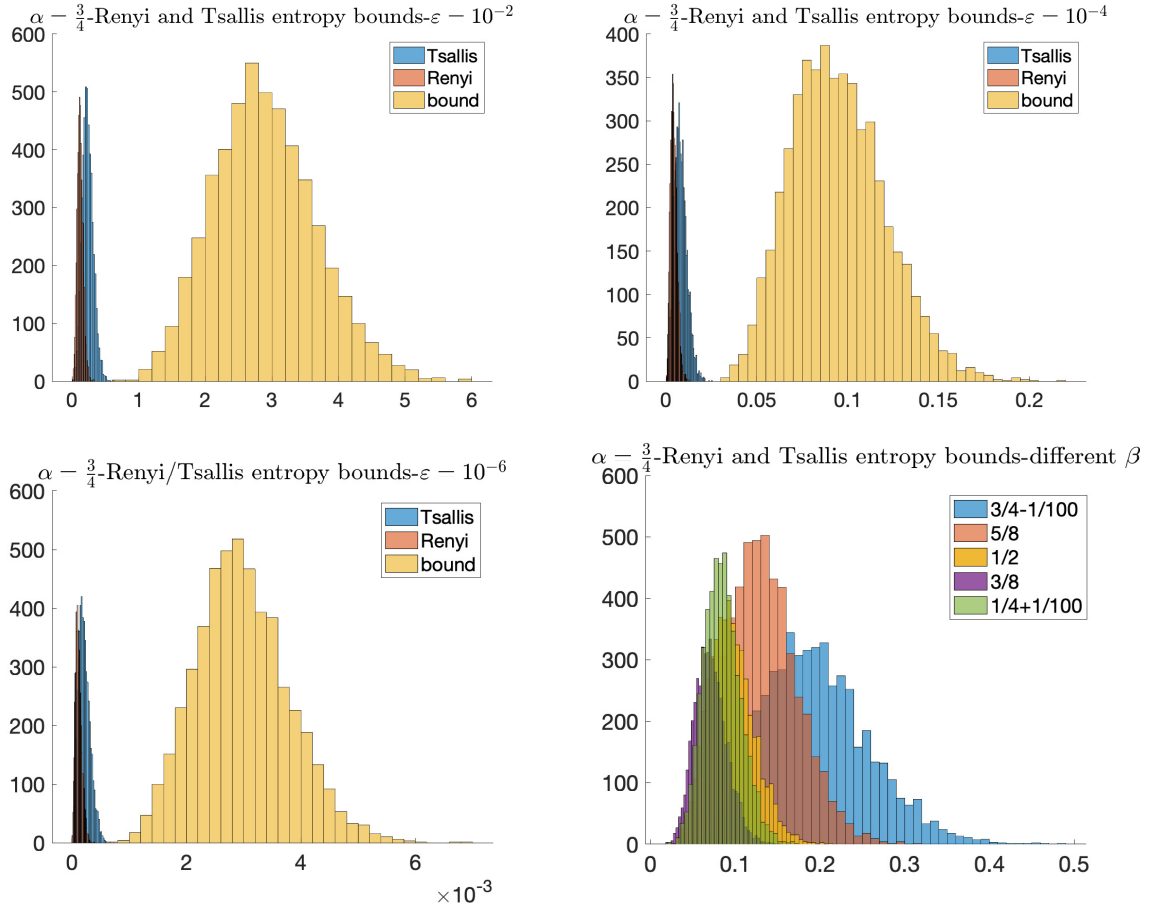Fig. 3. These figures illustrate our findings of Corollary III.4. We compute for 5000 random distributions p on $\{1, 2, ..., 1000\}$ the α-Tsallis and α-Rényi entropy by perturbing the distribution by εq where q is another random vector. In the fourth plot we compute the bounds for different values of β and observe that low values of β yield better bounds. We choose the weights $w_i = i$. In this histogram, the x-axis depicts the absolute value of the difference of the α-Tsallis and α-Rényi entropy for the realizations of our 5000 sample distributions and also the value of our bound for these realizations. The y-axis shows the number of times this value on the x-axis was achieved among the 5000 realizations.

some $\beta < \alpha$ and $(w_i p_X(z_i)), (w_i p_Y(z_i)) \in \ell^1(\mathbb{N})$. Then we have the following continuity bounds: The α-Tsallis entropy satisfies

$$|T_\alpha(X) - T_\alpha(Y)|$$
$$\leq \frac{2^\alpha}{1-\alpha}\|X - Y\|_{\mathrm{TV}((w_i))}^\beta \|X - Y\|_{\mathrm{TV}}^{\alpha-\beta}\|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha}.$$
(III.50)

The α-Rényi entropy satisfies

$$|R_\alpha(X) - R_\alpha(Y)|$$
$$\leq \frac{2^\alpha}{1-\alpha}\|X - Y\|_{\mathrm{TV}((w_i))}^\beta \|X - Y\|_{\mathrm{TV}}^{\alpha-\beta}\|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha}.$$
(III.51)

Proof. Recall that $(Tf_\alpha)(X) \geq 1$, which allows us to use that $|\log(x) - \log(y)| \leq |x - y|$ for $x, y \geq 1$, for the Rényi entropy. The result then immediately follows from the triangle inequality and Proposition III.3. □

2) Continuous random variables: In a recent paper [27], a continuity bound for the differential entropy has been obtained, but analogous bounds for the α-Rényi and α-Tsallis entropies are missing. Therefore, we now turn to the study of continuous random variables X, Y with densities μ_X, μ_Y on some domain

D and derive continuity bounds for the α-Rényi and Tsallis entropies. This leads us then to the study of α-Hölder continuous functions $f : [0, \max\{\|\mu_X\|_\infty, \|\mu_Y\|_\infty\}] \to \mathbb{R}$ defining the quantities

$$(Tf)(X) \coloneqq \int_D f(\mu_X(x)) \; dx.$$

Analogously, to the discrete case, $H(X) = Tf_1(X)$ is the differential *Shannon entropy* of the random variable X, $R_\alpha(X) = \frac{\log(Tf_\alpha(X))}{1-\alpha}$ its differential *α-Rényi entropy*, and $T_\alpha(X) = \frac{Tf_\alpha(X)-1}{1-\alpha}$ its differential *α-Tsallis entropy*. A straightforward adaptation of Proposition III.3 yields

Proposition III.5. *Let D be some domain and $f \in C^\alpha([0,\tau])$, $\alpha \in (0, 1)$, with $\tau = \max\{\|\mu_X\|_{L^\infty}, \|\mu_Y\|_{L^\infty}\} \in (0, \infty]$ where μ_X, μ_Y are probability densities on D associated with random variables X, Y respectively. Let w be a positive weight function such that $w^{-\frac{\beta}{1-\alpha}} \in L^1(D)$ for some $\beta < \alpha$ and $w\mu_X, w\mu_Y \in L^1(D)$. Then we have the following continuity*

bound:

$$\|f(\mu_X) - f(\mu_Y)\|_{L^1(D)}$$
$$\leq |f|_{\Lambda_{\omega_\alpha}} \|\mu_X - \mu_Y\|_{L^1(w)}^\beta \|\mu_X - \mu_Y\|_{L^1}^{\alpha-\beta} \|w^{-\frac{\beta}{1-\alpha}}\|_{L^1}^{1-\alpha}.$$
$$(\text{III.52})$$

As in the discrete case, we can therefore conclude the following:

Corollary III.6. *Let D be a domain, $\alpha \in (0,1)$, and μ_X, μ_Y be probability densities on D associated with random variables X, Y respectively. Let w be a positive weight function such that $w^{-\frac{\beta}{1-\alpha}} \in L^1(D)$ for some $\beta < \alpha$ and $w\mu_X, w\mu_Y \in L^1(D)$. In addition, let $(Tf_\alpha)(X), (Tf_\alpha)(Y) \geq \delta^{-1}$, for some $\delta > 0$. Then we have the following continuity bounds: The α-Tsallis entropy satisfies*

$$|T_\alpha(X) - T_\alpha(Y)|$$
$$\leq \frac{1}{1-\alpha} \|X-Y\|_{\text{TV}((w_i))}^\beta \|X-Y\|_{\text{TV}}^{\alpha-\beta} \|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha}.$$
$$(\text{III.53})$$

The α-Rényi entropy satisfies

$$|R_\alpha(X) - R_\alpha(Y)|$$
$$\leq \frac{\delta}{1-\alpha} \|X-Y\|_{\text{TV}((w_i))}^\beta \|X-Y\|_{\text{TV}}^{\alpha-\beta} \|(w_i^{-\frac{\beta}{1-\alpha}})\|_{\ell^1}^{1-\alpha}.$$
$$(\text{III.54})$$

Proof. As in the discrete case, we can now use that $|\log(x) - \log(y)| \leq \delta|x-y|$ for all $x, y \geq \delta^{-1} > 0$, for the α-Rényi entropy, which we apply to the integrals $Tf_\alpha(X), Tf_\alpha(Y)$ appearing in the definition of R_{alpha}. The result then immediately follows from the triangle inequality and Proposition III.5 □

Remark 3. *Continuity estimates for $\alpha > 1$ for classical α-Rényi and Tsallis entropies can be obtained along the lines of the corresponding quantum mechanical result that we state as Proposition III.8.*

D. Quantum Rényi and Tsallis entropies

Let ρ be a quantum state, *i.e.* a positive trace-class operator on a separable Hilbert space with unit trace. The spectral theorem implies that for any Borel function $f : \mathbb{R} \to \mathbb{C}$ we can write, in terms of rank 1-projections π_k, just $f(\rho) = \sum_k f(\lambda_k)\pi_k$. Our first theorem of this section, Theorem 7, is a general continuity result for functions of density operators under moment constraints. The moment constraints stated in the theorem follow for the entropies already from energy constraints on the states themselves. We elaborate on this in Lemma A.1 in the appendix.

The theorem crucially relies on two results obtained by Aleksandrov and Peller, [20, Theo. 5.8] and [21, Theo. 7.1]. The first important result that is crucial for our purposes is a continuity bound for Schatten norms and Hölder continuous functions: Fix $\alpha \in (0,1)$ and $p \in (1,\infty)$. There exists a universal constant $c > 0$ such that for any function $f \in C^\alpha(\mathbb{R})$ and A, B self-adjoint operators with $A - B \in \mathcal{S}_p$, the operator $f(A) - f(B) \in \mathcal{S}_{p/\alpha}$ and

$$\|f(A) - f(B)\|_{p/\alpha} \leq c|f|_{\Lambda_{\omega_\alpha}} \|A - B\|_p^\alpha. \quad (\text{III.55})$$

It is important to observe here that the result does not allow us to take $p/\alpha = 1$ immediately. In fact, a bound for $p/\alpha = 1$ has been found, but requires a higher level of regularity [20, Theo. 6.2]. Thus, we cannot directly apply the above estimates in trace distance but need to work in weaker Schatten norms, first.

The second result is a continuity bound that is merely in operator norm but for arbitrary moduli of continuity: Similarly, for every function ω^* associated to a modulus of continuity ω as defined in Section II-B, there exists some $c > 0$ such that for self-adjoint A and B and any $f \in \Lambda_\omega$ we have

$$\|f(A) - f(B)\| \leq c|f|_{\Lambda_\omega} \omega^*(\|A - B\|). \quad (\text{III.56})$$

Now, we have all the prerequisites to state the approximation theorem:

Theorem 7 (Approximation theorem). *Let ρ, σ be states, f a measurable function and $\hat{H}$ a positive Hamiltonian with compact resolvent such that for some $\beta > 0$ we have $\text{Tr}(\hat{H}^\beta |f|(\rho)), \text{Tr}(\hat{H}^\beta |f|(\sigma)) \leq \mu < \infty$. Let $\varepsilon > 0$, $f \in C^\alpha$, and take the spectral projection $P := \mathbb{1}_{[0,\mu/\varepsilon]}(\hat{H}^\beta)$. Then, there is $c > 0$ such that for all $w/q < 1$ and p conjugate to q*

$$\|f(\rho) - f(\sigma)\|_1 \leq \sqrt{8\varepsilon}\left(\sqrt{\|f(\rho)\|_1} + \sqrt{\|f(\sigma)\|_1}\right) + c|f|_{C^{w/q}} \|P\|_p \|\rho - \sigma\|_w,$$
$$(\text{III.57})$$

and for general $f \in \Lambda_\omega$, modulus of continuity ω, and integrated modulus of continuity ω^*

$$\|f(\rho) - f(\sigma)\|_1 \leq \sqrt{8\varepsilon}\left(\sqrt{\|f(\rho)\|_1} + \sqrt{\|f(\sigma)\|_1}\right) + c|f|_{\Lambda_\omega} \|P\|_1 \omega^*(\|\rho - \sigma\|).$$
$$(\text{III.58})$$

Proof. We then find as in the proof of the Gentle Measurement Lemma [18, Lemma 9] that for any projection P

$$\|f(\rho) - Pf(\rho)P\|_1^2$$
$$\overset{(i)}{\leq} \left(\sum_k |f(\lambda_k)| \|\pi_k - P\pi_k P\|_1\right)^2$$
$$\overset{(ii)}{\leq} \sum_k |f(\lambda_k)| \|\pi_k - P\pi_k P\|_1^2 \|f(\rho)\|_1$$
$$\overset{(iii)}{\leq} 4\sum_k |f(\lambda_k)| (1 - \text{Tr}(\pi_k P\pi_k P)) \|f(\rho)\|_1$$
$$\overset{(iv)}{\leq} 8\sum_k |f(\lambda_k)| (1 - \text{Tr}(\pi_k P)) \|f(\rho)\|_1$$
$$\leq 8(\|f(\rho)\|_1 - \text{Tr}(|f|(\rho)P)) \|f(\rho)\|_1,$$
$$(\text{III.59})$$

using (i) the spectral decomposition, (ii) the Cauchy-Schwarz inequality, (iii) the Fuchs-van de Graaf inequality, cf. (II.2), and (iv) $(1-x^2) \leq 2(1-x)$. For a general α-Hölder continuous function f, assuming that $\text{Tr}(\hat{H}^\beta |f|(\rho)) \leq \mu < \infty$ and P as in the statement, we obtain

$$\mu \geq \text{Tr}(\hat{H}^\beta |f|(\rho))$$
$$= \text{Tr}(\hat{H}^\beta P|f|(\rho)) + \text{Tr}(\hat{H}^\beta (1-P)|f|(\rho)) \quad (\text{III.60})$$
$$\geq \mu/\varepsilon \, \text{Tr}((1-P)|f|(\rho)).$$

This implies that $\mathrm{Tr}((1 - P)|f|(\rho)) \leq \varepsilon \Rightarrow \|f(\rho)\|_1 - \mathrm{Tr}(P|f|(\rho)) \leq \varepsilon$. From this, we get in (III.59) that $\|f(\rho) - Pf(\rho)P\|_1 \leq \sqrt{8\|f(\rho)\|_1\varepsilon}$. Putting it all together, we find for any admissible $w/q < 1$, using [20, Theo. 5.8], and p being the Hölder conjugate exponent to q

$$
\begin{aligned}
&\|f(\rho) - f(\sigma)\|_1 \\
&\leq \|f(\rho) - Pf(\rho)P\|_1 + \|P(f(\rho) - f(\sigma))P\|_1 \\
&\quad + \|Pf(\sigma)P - f(\sigma))\|_1 \\
&\leq \sqrt{8\varepsilon}\Big(\sqrt{\|f(\rho)\|_1} + \sqrt{\|f(\sigma)\|_1}\Big) \\
&\quad + \|P\|_p\|(f(\rho) - f(\sigma))\|_q \\
&\leq \sqrt{8\varepsilon}\Big(\sqrt{\|f(\rho)\|_1} + \sqrt{\|f(\sigma)\|_1}\Big) \\
&\quad + c|f|_{C^{w/q}}\|P\|_p\|\rho - \sigma\|_w^{w/q},
\end{aligned}
\tag{III.61}
$$

where we used (III.55) in the last step. For the general case, we thus find, using (III.56),

$$
\begin{aligned}
\|f(\rho) - f(\sigma)\| &\leq c\sqrt{8\varepsilon}\Big(\sqrt{\|f(\rho)\|_1} + \sqrt{\|f(\sigma)\|_1}\Big) \\
&\quad + c|f|_{\Lambda_\omega}\|P\|_1\omega^*(\|\rho - \sigma\|).
\end{aligned}
\tag{III.62}
$$

$\square$

Let $(Tf)(\rho) \coloneqq \mathrm{Tr}(f(\rho))$, then $S(\rho) = Tf_1(\rho)$ is the *von Neumann entropy* of the density operator ρ, $R_\alpha(\rho) = \frac{\log(Tf_\alpha(\rho))}{1-\alpha}$ its α-*Rényi entropy*, and $T_\alpha(\rho) = \frac{Tf_\alpha(\rho)-1}{1-\alpha}$ its α-*Tsallis entropy*. Using that $Tf_\alpha(\rho) \geq 1$, we thus find the following immediate corollary, since $|\log(x) - \log(y)| \leq |x - y|$ for $x, y \geq 1$.

Corollary III.7. *Let* $\alpha \in (0,1)$ *and* ρ, σ *be states and* $\hat{H}$ *a positive Hamiltonian with compact resolvent such that for some* $\beta > 0$ *we have* $\mathrm{Tr}(\hat{H}^\beta f_\alpha(\rho)), \mathrm{Tr}(\hat{H}^\beta f_\alpha(\sigma)) \leq \mu < \infty$. *Let* $\varepsilon > 0$ *and take the spectral projection* $P \coloneqq \mathbb{1}_{[0,\mu/\varepsilon]}(\hat{H}^\beta)$. *Then for all* $\alpha = w/q < 1$ *and* p *conjugate to* q,

$$
\begin{aligned}
&|T_\alpha(\rho) - T_\alpha(\sigma)| \\
&\leq \frac{\sqrt{8\varepsilon}\Big(\sqrt{\|f_\alpha(\rho)\|_1} + \sqrt{\|f_\alpha(\sigma)\|_1}\Big) + c\|P\|_p\|\rho - \sigma\|_{\alpha q}^\alpha}{1-\alpha}
\end{aligned}
\tag{III.63}
$$

and

$$
\begin{aligned}
&|R_\alpha(\rho) - R_\alpha(\sigma)| \\
&\leq \frac{\sqrt{8\varepsilon}\Big(\sqrt{\|f_\alpha(\rho)\|_1} + \sqrt{\|f_\alpha(\sigma)\|_1}\Big) + c\|P\|_p\|\rho - \sigma\|_{\alpha q}^\alpha}{1-\alpha}.
\end{aligned}
\tag{III.64}
$$

Estimates on $\|f_\alpha(\rho)\|_1, \|f_\alpha(\sigma)\|_1$ *can be found in Lemma A.1.*

E. The case $\alpha > 1$

The case $\alpha > 1$ is fundamentally different from the case $\alpha \in (0,1)$ studied before. For completeness, we state the following Proposition on α-Rényi and α-Tsallis entropies for $\alpha > 1$. For $\alpha > 1$, the α-Tsallis entropy in fact becomes Lipschitz continuous, as has already been observed in [16]. This is different from the α-Rényi entropy which is not uniformly continuous for any $\alpha > 1$.

Proposition III.8. *Let* $\alpha > 1$. *The* α-*Tsallis entropy is always Lipschitz continuous with respect to the* α-*Schatten distance:*

$$
|T_\alpha(\rho) - T_\alpha(\sigma)| = \frac{|\|\rho\|_\alpha^\alpha - \|\sigma\|_\alpha^\alpha|}{\alpha - 1} \leq \frac{\alpha}{\alpha - 1}\|\rho - \sigma\|_\alpha.
$$

The α-*Rényi entropy is not uniformly continuous on the set of states unless an energy constraint is imposed. Thus, let* ρ, σ *be states such that at least one of the following two conditions holds,*

1) $\mathrm{Tr}(\rho^\alpha), \mathrm{Tr}(\sigma^\alpha) \geq \delta^{-1} > 0$ *for some* $\delta > 0$ *or*
2) *There exists a positive definite Hamiltonian such that* $\mathrm{Tr}(\hat{H}\rho), \mathrm{Tr}(\hat{H}\sigma) \leq E < \infty$ *and for some* $\beta \in (0,1)$, *we have* $\tau \coloneqq \mathrm{Tr}\left(\hat{H}^{-\frac{\alpha\beta}{(1-\beta)(\alpha-1)}}\right) < \infty$ *such that we can define* $\delta \coloneqq E^{\frac{\alpha\beta}{1-\alpha}}\tau^{\alpha-1} > 0$,

then

$$
|R_\alpha(\rho) - R_\alpha(\sigma)| \leq \frac{\alpha\delta}{1-\alpha}\|\rho - \sigma\|_\alpha.
$$

Proof. For the α-Rényi entropy, we proceed as follows. Let $\mathrm{Spec}(\rho) = \{\lambda_n; n \in \mathbb{N}\}$ and $\mathrm{Spec}(\hat{H}) = \{\mu_n; n \in \mathbb{N}\}$, then we find

$$
\begin{aligned}
1 &= \sum_{i=1}^\infty \lambda_i = \sum_{i=1}^\infty \mu_i^\beta \lambda_i^\beta \lambda_i^{1-\beta} \mu_i^{-\beta} \\
&\overset{(1)}{\leq} \left(\sum_{i=1}^\infty \mu_i\lambda_i\right)^\beta \left(\sum_{i=1}^\infty \frac{1}{\mu_i^{\frac{\beta}{1-\beta}}}\lambda_i\right)^{1-\beta} \\
&\overset{(2)}{\leq} (\mathrm{Tr}(\hat{H}\rho))^\beta (\mathrm{Tr}(\rho^\alpha))^{\frac{1-\beta}{\alpha}} \left(\sum_{i=1}^\infty \mu_i^{-\frac{\alpha\beta}{(1-\beta)(\alpha-1)}}\right)^{\frac{(1-\beta)(\alpha-1)}{\alpha}},
\end{aligned}
\tag{III.65}
$$

where in (1) and (2) we use Hölder's inequality, and for the first term in (2) we use the Courant-Fischer theorem, Proposition II.3. This implies by rearranging (III.65) that

$$
\begin{aligned}
\|\rho\|_\alpha^\alpha &\geq \left((\mathrm{Tr}(\hat{H}\rho))^{\frac{\alpha\beta}{1-\beta}}(\mathrm{Tr}(\hat{H}^{-\frac{\alpha\beta}{(1-\beta)(\alpha-1)}}))^{\alpha-1}\right)^{-1} \\
&\geq \left(E^{\frac{\alpha\beta}{1-\beta}}\left(\mathrm{Tr}\left(\hat{H}^{-\frac{\alpha\beta}{(1-\beta)(\alpha-1)}}\right)\right)^{\alpha-1}\right)^{-1} =: \frac{1}{\delta}.
\end{aligned}
\tag{III.66}
$$

Thus, we have proven that $\mathrm{Tr}(\rho^\alpha) \geq \delta^{-1} > 0$ for some $\delta > 0$. Hence, for the α-Rényi entropy we get,

$$
\begin{aligned}
|R_\alpha(\rho) - R_\alpha(\sigma)| &\overset{(1)}{\leq} \frac{\delta}{\alpha - 1}|\|\rho\|_\alpha^\alpha - \|\sigma\|_\alpha^\alpha| \\
&\overset{(2)}{\leq} \alpha\frac{\delta}{\alpha - 1}|\|\rho\|_\alpha - \|\sigma\|_\alpha| \\
&\leq \alpha\frac{\delta}{1-\alpha}\|\rho - \sigma\|_\alpha,
\end{aligned}
\tag{III.67}
$$

where in (1) we use that $|\log(x) - \log(y)| \leq \delta|x - y|$ for all $x, y \geq \delta^{-1}$ and in (2) we used that $|x^\alpha - y^\alpha| \leq \alpha|x - y|$ for $x, y \in [0,1]$. Both estimates are readily verified using the mean value theorem.

For the α-Tsallis entropy we end up directly estimating as in the last part of (III.67)

$$
|T_\alpha(\rho) - T_\alpha(\sigma)| = \frac{|\|\rho\|_\alpha^\alpha - \|\sigma\|_\alpha^\alpha|}{\alpha - 1} \leq \frac{\alpha}{\alpha - 1}\|\rho - \sigma\|_\alpha.
$$

$\square$

F. The Finite-dimensional Approximation (FA) property

As has been discussed in the papers [3], [4] and has also in this article, continuity bounds for states on infinite-dimensional Hilbert spaces often rely on constraints on the energy of states by some Hamiltonian to make the entropy functional continuous. It is now tempting to turn this question around and ask if there always exists a natural Hamiltonian, defining a Gibbs state, for any state of finite entropy.

There are various functionals that are not continuous with respect to trace distance. For instance, in any arbitrarily small neighbourhood of a state in an infinite-dimensional Hilbert space, there exists a state of infinite entropy. The following condition that excludes many such pathological states, for various quantities of information theoretic interest, was recently proposed by Shirokov [1], [2]:

Definition III.9 (FA-property). *A state $\rho = \sum_{k \in \mathbb{N}} \lambda_k \pi_k$ satisfies the* FA*-property if there is a sequence (g_k) with $g_k \geq 0$ such that*

$$\sum_{k \in \mathbb{N}} \lambda_k g_k < \infty \qquad \text{(III.68)}$$

$$\text{and} \quad \lim_{\beta \downarrow 0} \left(\sum_{k \in \mathbb{N}} e^{-\beta g_k} \right)^{\beta} = 1. \qquad \text{(III.69)}$$

It was observed in [1, Theo. 1] that if a state ρ satisfies the FA property, then ρ has finite von Neumann entropy. In that same article, the question was raised whether any state ρ that has finite von Neumann entropy, necessarily satisfies the FA property.

In particular, for $\lambda_k \in \text{Spec}(\rho)$ it has been shown that once $\sum_{k \geq 1} \lambda_k \log(k)^q < \infty$ for some $q > 2$, then ρ satisfies the FA-property. Yet, there clearly exist states that do not fall in this category. Let ρ_α be a state whose eigenvalues $\lambda_{\alpha,k}$ are, up to a normalizing constant

$$\nu_\alpha := \sum_{k \geq 2} \frac{1}{k \log(k)^\alpha}, \qquad \text{(III.70)}$$

given as $\lambda_{\alpha,k} = \frac{1}{\nu_\alpha (k \log(k)^\alpha)}$ for $\alpha \in (2,3)$. Any such ρ has finite entropy. This can be seen as follows:

$$S(\rho_\alpha) = H(\{\lambda_{\alpha,k}\}) = -\sum_{k \geq 2} \lambda_{\alpha,k} \log \lambda_{\alpha,k} = \sum_{k \geq 2} a_k, \qquad \text{(III.71)}$$

where

$$a_k = \frac{\log(\nu_\alpha k \log(k)^\alpha)}{\nu_\alpha k \log(k)^\alpha}.$$

The terms in the sequence (a_k) are all positive and decrease for k large enough. Note that

$$\sum_{k \geq 2} 2^k a_{2^k} = a \sum_{k \geq 2} \frac{1}{k^\alpha} + b \sum_{k \geq 2} \frac{1}{k^{\alpha-1}} + c \sum_{k \geq 2} \frac{\log k}{k^\alpha} < \infty,$$

for some constants a, b, c. By Cauchy's condensation test this implies that $\sum_k a_k$ converges, and hence $S(\rho) < \infty$. We now want to argue that any such ρ cannot satisfy the FA-property answering the question raised in [1] in a negative way:

Theorem 8. *Any state ρ with spectrum $\text{Spec}(\rho) = \{\lambda_n; n \in \mathbb{N}\}$ such that $\lambda_n \leq \nu/(n \log(n)^3)$ for almost all $n \geq 2$, with normalizing constant $\nu > 0$, does not satisfy the FA-property. In particular, the set of states satisfying the FA-property is strictly smaller than the set of finite entropy states.*

Proof. The comparison test for the sequence implies that for a positive sequence λ_n satisfying the condition of Theorem 8, $\sum_{k \geq 2} \lambda_k g_k < \infty$ implies

$$\nu \sum_{k \geq 4} \frac{g_k}{k \log(k)^2 (\log(k) - 1)}$$
$$= \nu \sum_{k \geq 4} \frac{g_k}{k \log(k)^3} + \nu \sum_{k \geq 4} \frac{g_k}{k \log(k)^3 (\log(k) - 1)} < \infty, \qquad \text{(III.72)}$$

where the last sum converges by the majorant criterion. Continuity of the logarithm, implies that the Gibbs part of the condition of the FA-property can be rewritten as

$$\lim_{\beta \downarrow 0} \beta \log \left(\sum_{k \in \mathbb{N}} e^{-\beta g_k} \right) = 0. \qquad \text{(III.73)}$$

Our aim is now to show that for any sequence (g_k) such that the first part, (III.68), of the FA-property holds, *i.e.* $\sum_k \lambda_k g_k < \infty$, the second one, (III.69) in the form of (III.73), is necessarily violated. Writing $A_+ := \{k; g_k \leq \log(k)^2\}$ and A_- for its complement, we find by monotonicity and $\mathbb{N} = A_+ \dot{\cup} A_-$ in (1) and again monotonicity in (2) to estimate the series by an integral

$$\sum_{k \in A_+} e^{-\beta g_k} \overset{(1)}{\geq} \sum_{k \in \mathbb{N}} e^{-\beta \log(k)^2} - \sum_{k \in A_-} e^{-\beta \log(k)^2}$$
$$\overset{(2)}{\geq} \int_1^\infty e^{-\beta \log(x)^2} \, dx - \sum_{k \in A_-} e^{-\beta \log(k)^2}. \qquad \text{(III.74)}$$

Quite explicitly, we observe that by substituting $x = e^u$, we have

$$\int_1^\infty e^{-\beta \log(x)^2} \, dx = \int_0^\infty e^{u(1-\beta u)} \, du$$
$$= e^{\frac{1}{4\beta}} \frac{\sqrt{\pi} \left(1 + \text{erf} \left(\frac{1}{2\sqrt{\beta}} \right) \right)}{2\sqrt{\beta}}. \qquad \text{(III.75)}$$

Monotonicity of the logarithm yields that for all $k \in A_-$ such that $\log(k) \in [n, n+1)$ we have $e^{-\beta \log(k)^2} \leq e^{-\beta n^2}$

$$\sum_{k \in A_-} e^{-\beta \log(k)^2} \leq \sum_{n \in \mathbb{N}} |\log(A_-) \cap [n, n+1)| e^{-\beta n^2}, \quad \text{(III.76)}$$

where $\log(S) := \{\log(s); s \in S\}$ for a suitable set S. We then introduce renormalized coefficients $\alpha_k := K(k)/e^k \leq (e-1)$, with $K(k) := |\log(A_-) \cap [k, k+1)|$, where we used for the last inequality the simple worst case estimate by replacing A_- in the definition of $K(k)$ by $\mathbb{N}$, in which case $K(k)$ may

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3248228

15

be replaced by $e^{k+1} - e^k$, as they obey the same asymptotic scaling. This way, we find that for any $\delta > 0$

$$\sum_{k \in A_-} e^{-\beta \log(k)^2} \overset{\text{(III.76)}}{\leq} \sum_{k \in \mathbb{N}} \alpha_k e^{k(1-\beta k)}$$

$$\overset{(1)}{\leq} \delta \int_0^\infty e^{u(1-\beta u)} \, du \qquad \text{(III.77)}$$
$$+ \delta e^{\frac{1}{4\beta}} + \sum_{k \in \mathbb{N}; \alpha_k \geq \delta} \alpha_k e^{k(1-\beta k)},$$

where in (1) we split the sum into parts $\{\alpha_k \geq \delta\}$ and its complement and estimated the latter series by an integral and its maximum value. We then estimate in terms of the cardinality $\gamma_\beta := \left| k \in [\frac{1-\sqrt{\beta}}{2\beta}, \frac{1+\sqrt{\beta}}{2\beta}] \cap \mathbb{N}; \alpha_k \geq \delta \right|$, by just using monotonicity and $|\alpha_k| \leq e - 1$

$$\sum_{k \in \mathbb{N}; \alpha_k \geq \delta} \alpha_k e^{k(1-\beta k)}$$

$$\leq (e-1) \left(\left(\int_0^{\frac{1-\sqrt{\beta}}{2\beta}} + \int_{\frac{1+\sqrt{\beta}}{2\beta}}^\infty \right) e^{x(1-\beta x)} \, dx + \gamma_\beta e^{\frac{1}{4\beta}} \right)$$

$$= e^{\frac{1}{4\beta}} \frac{\sqrt{\pi}}{2\sqrt{\beta}} \Big((e-1)(1 - 2\operatorname{erf}(\frac{1}{2}) + \operatorname{erf}(\frac{1}{2\sqrt{\beta}})) $$
$$+ \frac{2(e-1)}{\sqrt{\pi}} \sqrt{\beta} \gamma_\beta \Big), \qquad \text{(III.78)}$$

where we used in addition that by differentiation we see that $e^{k(1-\beta k)} \leq e^{1/(4\beta)}$ for $k \in [\frac{1-\sqrt{\beta}}{2\beta}, \frac{1+\sqrt{\beta}}{2\beta}]$. Applying all the above inequalities to (III.74), we find for $\delta > 0$ sufficiently small, which we shall assume from now on,

$$\sum_{k \in \mathbb{N}} e^{-\beta g_k} \geq \sum_{k \in A_+} e^{-\beta g_k} \geq e^{\frac{1}{4\beta}} \left[A(\beta)(1-\delta) - \delta - B(\beta) \right],$$
$$\text{(III.79)}$$

where

$$A(\beta) := \frac{\sqrt{\pi}(1 + \operatorname{erf}((2\sqrt{\beta})^{-1}))}{2\sqrt{\beta}} \qquad \text{(III.80)}$$

and

$$B(\beta) := \frac{\sqrt{\pi}}{2\sqrt{\beta}} \Big((e-1)(1 - 2\operatorname{erf}(\frac{1}{2}) + \operatorname{erf}(\frac{1}{2\sqrt{\beta}})) $$
$$+ \frac{2(e-1)}{\sqrt{\pi}} \beta^{1/2} \gamma_\beta \Big). \qquad \text{(III.81)}$$

Recall that we want to show $\lim_{\beta \downarrow 0} \beta \log \left(\sum_{k \in \mathbb{N}} e^{-\beta g_k} \right) \neq 0$. So upon taking the logarithm and multiplying by β the exponential function $e^{\frac{1}{4\beta}}$ contributes a term $\beta \log(e^{\frac{1}{4\beta}}) = \frac{1}{4}$ which is non-zero. To show that (III.73) does not hold, it suffices therefore to show that, for fixed $\delta > 0$ small, $A(\beta)(1-\delta) - \delta - B(\beta)$ is bounded uniformly away from zero for choices of $\beta = \beta_n$ with a sequence (β_n) that tends to zero. In fact, we have that

$$A(\beta)(1-\delta) - \delta - B(\beta) = \frac{1}{\sqrt{\beta}} (C(\beta) - (e-1)\beta^{1/2} \gamma_\beta),$$
$$\text{(III.82)}$$

with

$$C(\beta) = \frac{\sqrt{\pi}}{2} \Bigg[(1 + \operatorname{erf}(\frac{1}{2\sqrt{\beta}}))(1-\delta) - \frac{2\sqrt{\beta}}{\sqrt{\pi}} \delta $$
$$- \Big((e-1)(1 - 2\operatorname{erf}(\frac{1}{2}) + \operatorname{erf}(\frac{1}{2\sqrt{\beta}})) \Big) \Bigg]. \qquad \text{(III.83)}$$

Since with our standing assumption that $\delta > 0$ is sufficiently small, one verifies that $C(\beta) \geq \varepsilon > 0$ for some sufficiently small $\varepsilon > 0$ and all β small enough, it suffices therefore, in order to show that (III.82) is uniformly positive, that $\beta^{1/2} \gamma_\beta$ is strictly smaller than $\frac{\varepsilon}{2(e-1)}$ for a sequence of β_n tending to zero. In fact, this implies that $C(\beta_n) - (e-1)\beta_n^{1/2} \gamma_{\beta_n} \geq c > 0$ for some $c > 0$ and all $n \in \mathbb{N}$ which appears in the right hand side of (III.82). Thus by applying Lemma II.4 to $f(t) = t^{1/2} \gamma_t$, it suffices to show the finiteness of the following integral, where we without loss of generality restrict ourselves to $k \geq 4$,

$$\int_0^1 \gamma_\beta \frac{d\beta}{\beta^{1/2}} \overset{(1)}{=} 2 \int_1^\infty \gamma_{\frac{1}{\alpha^2}} \frac{d\alpha}{\alpha^2}$$

$$\overset{(2)}{=} 2 \sum_{k; \alpha_k \geq \delta} \int_1^\infty \mathbb{1}(k) \Big[\frac{\alpha(\alpha-1)}{2}, \frac{\alpha(\alpha+1)}{2} \Big] \frac{d\alpha}{\alpha^2}$$

$$\overset{(3)}{=} 2 \sum_{k; \alpha_k \geq \delta} \int_1^\infty \mathbb{1}(\alpha) \Big[\frac{\sqrt{8k+1}-1}{2}, \frac{\sqrt{8k+1}+1}{2} \Big] \frac{d\alpha}{\alpha^2}$$

$$\overset{(4)}{=} \sum_{k; \alpha_k \geq \delta} \frac{1}{k}$$

$$\overset{(5)}{\leq} \sum_k \frac{\alpha_k}{\delta k}$$

$$\overset{(6)}{=} \sum_k \frac{K(k)}{\delta e^k k}$$

$$\overset{(7)}{=} \sum_k \sum_{\substack{r \in A_-: \\ \log(r) \in \log(A_-) \cap [k, k+1]}} \frac{e}{\delta e^{k+1} k}$$

$$\overset{(8)}{\leq} \sum_k \sum_{\substack{r \in A_-: \\ \log(r) \in \log(A_-) \cap [k, k+1]}} \frac{e}{\delta r (\log(r) - 1)}$$

$$\overset{(9)}{=} \sum_{r \in A_-} \frac{e}{\delta r (\log(r) - 1)}$$

$$\overset{(10)}{\leq} \sum_{r \in A_-} \frac{e g_r}{\delta r \log(r)^2 (\log(r) - 1)} < \infty,$$

where we
1) substituted $\beta = \alpha^{-2}$,
2) used the definition of γ_β,
3) used an equivalent representation of the indicator function,
4) evaluated the integral,
5) used that for the k we are summing over $\alpha_k / \delta > 1$,
6) used the definition of α_k,
7) partitioned the summation according to the definition of $K(k)$,

8) used that $r(\log(r) - 1) \leq e^{k+1}k$ on the respective partitions $\{\log(r) \in \log(A_-) \cap [k, k+1)\}$,
9) dropped the partitioning over k,
10) used the definition of A_- and that the final sum has to be finite by the first condition of the FA property, (III.68) which in our case is (III.72). □

IV. Conclusion and Open questions

Conclusion

In this article, we provide for the first time a tight continuity estimate for the classical Shannon entropy for random variables on $\mathbb{N}_0$ and von Neumann entropy of quantum states on a separable, infinite-dimensional Hilbert space whose energy is constrained by the number operator.

We also provide for the first time continuity estimates for α-Rényi and α-Tsallis entropies for $\alpha \in (0, 1)$ both in classical probability theory and in quantum mechanics for infinite-dimensional Hilbert spaces. By doing so, we provide a tool to derive general continuity bounds for Hölder continuous functions of density operators.

Finally, we show that the finiteness of the von Neumann entropy of a state does not imply the FA-property.

Future directions

As possible future directions, we would like to propose the following list of open problems:

1) Provide a tight continuity estimate for the Shannon entropy of random variables on general countable alphabets.
2) Generalize the tight continuity bounds for the von Neumann entropy to general Hamiltonians satisfying the Gibbs hypothesis, see Def. III.2. Our method of proof seems to apply to this more general framework as well, but requires the solution of an optimization problem.
3) Derive continuity estimates for different moments. Due to its fundamental relevance in the uncertainty principle, it seems also reasonable to request a bound on the variance of the energy or other appropriate quantum observables instead of the energy.
4) Provide a tight continuity estimate for the differential entropy of random variables with densities.
5) We give sufficient criteria for the finiteness of α-Rényi and α-Tsallis entropies for states satisfying certain energy constraints. Do there exist also necessary criteria?
6) Investigate the tightness of the continuity bounds for α-Rényi and α-Tsallis entropies.
7) The set of states of finite von Neumann entropy is strictly larger than the set of states satisfying the FA property. Can one obtain analogous results to the ones obtained in [1], [2] only assuming the finiteness of the entropy?
8) In [4] similar continuity estimates as for the von Neumann entropy have also been obtained for the conditional von Neumann entropy, cf. Lemma 17. Can one provide a tight version of that Lemma too?
9) As in the preceding open problem, similar questions about tightness can be asked for estimates on quantum

conditional mutual information, the Holevo quantity and for capacities of quantum channels, as investigated in [28].

Acknowledgements

Appendix

We now state some moment bounds on f_α and f_1 using energy constraints on the state. The proofs give rise to slightly sharper bounds, that are less concise to state, than the ones we outline in the statement of the Lemma. Therefore, the reader might want to consult the proof of the following Lemma for slightly improved estimates. For notational simplicity, we use the notation $(\mathrm{Tr}(A))^k$ to denote $\mathrm{Tr}(A)^k$ below.

Lemma A.1. *Let $\hat{H}$ be a positive Hamiltonian with compact resolvent and ρ a state. Then, as soon as the right-hand side is finite, we have the following moment bounds for the function $f_1(x) = -x \log(x)$*

$$\mathrm{Tr}(\hat{H}^{1/2} f_1(\rho)) \leq \mathrm{Tr}(\hat{H}\rho) + \mathrm{Tr}\left(e^{-\sqrt{\hat{H}}}(1 + \hat{H})\right).$$

Similarly, for $f_\alpha(x) = x^\alpha$, we have

$$\mathrm{Tr}\left(\hat{H}^{1/2} f_\alpha(\rho)\right) \leq \mathrm{Tr}(\hat{H}\rho) + \mathrm{Tr}\left(\hat{H}^{\frac{2\alpha-1}{2(\alpha-1)}}\right) \quad \text{(A.1)}$$

for $\alpha \in (1/2, 1)$, and

$$\mathrm{Tr}(\hat{H}^\beta f_\alpha(\rho)) \leq \mathrm{Tr}(\hat{H}\rho)^\alpha \, \mathrm{Tr}(\hat{H}^{-\frac{r}{1-\alpha}}) \quad \text{(A.2)}$$

for $\beta = \alpha - r$ and $r \in (0, \alpha]$. Furthermore, in terms of the number operator, we find $\mathrm{Tr}(\hat{H}^{-\frac{r}{1-\alpha}}) < \infty$ for $\hat{H} = \hat{N} + 1$ and $\alpha > 1/2$ as well as $\hat{H} = (\hat{N} + 1)^\kappa$ for $\kappa < \frac{1-\alpha}{r}$, with $r = \alpha - \varepsilon$ for $1/2 \geq \alpha > \varepsilon > 0$.

Proof. For the following computations, we recall the Hölder inequality $\mathrm{Tr}(\hat{H}^\beta \pi_k) \leq \mathrm{Tr}(\hat{H}\pi_k)^\beta$ for $\beta \in (0, 1)$. Then writing $I := \{k; 1 \leq \mathrm{Tr}(\hat{H}^{1/2}\pi_k)\}$, $I^c := \mathbb{N} \setminus I$, $J := \{k; \log(\lambda_k) > -\mathrm{Tr}(\hat{H}^{1/2}\pi_k)\}$ and $J^c := \mathbb{N} \setminus J$, we have

$$\mathrm{Tr}(\hat{H}^{1/2} f_1(\rho)) = -\sum_{k\in\mathbb{N}} \lambda_k \log(\lambda_k)\, \mathrm{Tr}(\pi_k \hat{H}^{1/2})$$

$$\overset{(1)}{\leq} \sum_{k\in J} \lambda_k \, \mathrm{Tr}(\pi_k \hat{H}^{1/2})^2$$

$$+ \sum_{k\in J^c} e^{-\mathrm{Tr}(\hat{H}^{1/2}\pi_k)} \Big(\mathrm{Tr}(\pi_k \hat{H}^{1/2})^2 \, \mathbb{1}_I(k)$$

$$+ \mathrm{Tr}(\pi_k \hat{H}^{1/2}) \, \mathbb{1}_{I^c} \Big)$$

$$\overset{(2)}{\leq} \mathrm{Tr}(\hat{H}\rho) + \sum_{k\in J^c} e^{-\mathrm{Tr}(\hat{H}^{1/2}\pi_k)} \Big(\mathrm{Tr}(\pi_k \hat{H}^{1/2})^2 \, \mathbb{1}_I(k)$$

$$+ \mathrm{Tr}(\pi_k \hat{H}^{1/2}) \, \mathbb{1}_{I^c}(k) \Big),$$

$$\text{(A.3)}$$

where (1) follows by separately estimating different eigenvalues and using that $x^2 \geq x$ for $x \geq 1$ and (2) follows from Hölder's inequality. We find using a similar splitting for $f_\alpha(x) = x^\alpha$,

$$\mathrm{Tr}(\hat{H}^{1/2} f_\alpha(\rho))$$

$$\leq \sum_{k\in\mathbb{N}} \lambda_k \lambda_k^{\alpha-1} \, \mathrm{Tr}(\hat{H}\pi_k)^{1/2}$$

$$\overset{(1)}{\leq} \sum_{\substack{k \\ \lambda_k^{\alpha-1} < \mathrm{Tr}(\hat{H}\pi_k)^{1/2}}} \lambda_k \, \mathrm{Tr}(\hat{H}\pi_k)$$

$$+ \sum_{\substack{k \\ \lambda_k^{\alpha-1} > \mathrm{Tr}(\hat{H}\pi_k)^{1/2}}} \mathrm{Tr}(\hat{H}\pi_k)^{\frac{1}{2}\left(1-\frac{\alpha}{1-\alpha}\right)} \qquad \text{(A.4)}$$

$$\leq \mathrm{Tr}(\hat{H}\rho) + \sum_{\substack{k \\ \lambda_k^{\alpha-1} > \mathrm{Tr}(\hat{H}\pi_k)^{1/2}}} \frac{1}{\mathrm{Tr}(\hat{H}\pi_k)^{\frac{2\alpha-1}{2(1-\alpha)}}},$$

where in (1) we use that $\lambda^{\alpha-1} \leq \mathrm{Tr}(\hat{H}\pi_k)$ in the first term and $\lambda^\alpha \leq \mathrm{Tr}(\hat{\pi}_k)^{-\frac{\alpha}{2(1-\alpha)}}$ in the second term. Aside from special cases, this bound is only saleable for $\alpha > 1/2$. To satisfactorily treat also the cases $\alpha \in (0,1)$, by choosing $p = 1/\alpha$ and its Hölder conjugate $q = 1/(1-\alpha)$, we have by Hölder's inequality that for $\beta = \alpha - r$ and any $r \in (0,\alpha)$

$$\mathrm{Tr}(\hat{H}^\beta f_\alpha(\rho))$$

$$= \sum_{k\in\mathbb{N}} \lambda_k^\alpha \, \mathrm{Tr}(\hat{H}^\beta \pi_k) \, \mathrm{Tr}(\hat{H}\pi_k)^r \, \mathrm{Tr}(\hat{H}\pi_k)^{-r}$$

$$\leq \left(\sum_{k\in\mathbb{N}} \lambda_k \, \mathrm{Tr}(\hat{H}\pi_k)^{\frac{\beta+r}{\alpha}} \right)^\alpha \left(\sum_{k\in\mathbb{N}} \mathrm{Tr}(\hat{H}\pi_k)^{-\frac{r}{1-\alpha}} \right)^{1-\alpha}$$

$$= \mathrm{Tr}(\hat{H}\rho)^\alpha \left(\sum_{k\in\mathbb{N}} \mathrm{Tr}(\hat{H}\pi_k)^{-\frac{r}{1-\alpha}} \right)^{1-\alpha}.$$

$$\text{(A.5)}$$

Finally, we have using in (1) that $e^{-x}\max\{x^2,x\} \leq e^{-x}(x^2+1)$ and in (2) that $e^{-x}(x^2+1)$ is monotonically decreasing together with the Courant-Fischer theorem stated in Proposition II.3, that for the eigenbasis of the Hamiltonian with

rank-1 projections p_k corresponding to the ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq ...$ of the Hamiltonian

$$\sum_{k\in J^c} e^{-\mathrm{Tr}(\hat{H}^{1/2}\pi_k)} \Big(\mathrm{Tr}(\pi_k \hat{H}^{1/2})^2 \, \mathbb{1}_I(k)$$

$$+ \mathrm{Tr}(\pi_k \hat{H}^{1/2}) \, \mathbb{1}_{I^c}(k) \Big)$$

$$\overset{(1)}{\leq} \sum_{k\in J^c} e^{-\mathrm{Tr}(\hat{H}^{1/2}\pi_k)} \left(1 + \mathrm{Tr}(\pi_k \hat{H}^{1/2})^2 \right) \qquad \text{(A.6)}$$

$$\overset{(2)}{\leq} \sum_{k\in J^c} e^{-\mathrm{Tr}(\hat{H}^{1/2} p_k)} \left(1 + \mathrm{Tr}(\hat{H}^{1/2} p_k)^2 \right)$$

$$\overset{(3)}{\leq} \mathrm{Tr}(e^{-\sqrt{\hat{H}}}(\hat{H}+1)),$$

where we dropped the constraint on k in (3).

Analogously, we have again used monotonicity and the Courant Fischer theorem

$$\sum_k \frac{1}{\mathrm{Tr}(\hat{H}\pi_k)^{\frac{2\alpha-1}{2(1-\alpha)}}} \leq \mathrm{Tr}\left(\hat{H}^{-\frac{2\alpha-1}{2(1-\alpha)}} \right),$$

$$\sum_{k\in\mathbb{N}} \mathrm{Tr}(\hat{H}\pi_k)^{-\frac{r}{1-\alpha}} \leq \mathrm{Tr}\left(\hat{H}^{-\frac{r}{1-\alpha}} \right).$$

$$\text{(A.7)}$$

$$\square$$

REFERENCES

[1] M. E. Shirokov, "On quantum states with a finite-dimensional approximation property," *Lobachevskii Journal of Mathematics*, vol. 42, no. 10, pp. 2437–2454, Oct 2021. [Online]. Available: https://doi.org/10.1134/S1995080221100206

[2] ——, "Approximation of multipartite quantum states and the relative entropy of entanglement," *arXiv*, vol. 2103.12111, 2021.

[3] A. Wehrl, "General properties of entropy," *Rev. Mod. Phys.*, vol. 50, 1978.

[4] A. Winter, "Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints," *Commun. Math. Phys.*, vol. 347(1), pp. 291–313, 2016.

[5] K. Audenaert, "A sharp continuity estimate for the von neumann entropy," *Journal of Physics A: Mathematical and Theoretical*, vol. 40(28), pp. 1827–1836, 2007.

[6] E. Hanson and N. Datta, "Tight uniform continuity bound for a family of entropies." *arXiv*, vol. 1707.04249, 2021.

[7] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, 2017.

[8] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, "Gaussian quantum information," *Rev. Mod. Phys.*, vol. 84, pp. 621–669, May 2012. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.84.621

[9] M. E. Shirokov, "Advanced Alicki-Fannes-Winter method for energy-constrained quantum systems and its use," *Quantum Information Processing*, vol. 19, no. 5, p. 164, Apr. 2020.

[10] ——, "Uniform continuity bounds for characteristics of multipartite quantum systems," *arXiv e-prints*, p. arXiv:2007.00417, Jul. 2020.

[11] S. Becker and N. Datta, "Convergence Rates for Quantum Evolution and Entropic Continuity Bounds in Infinite Dimensions," *Communications in Mathematical Physics*, vol. 374, no. 2, pp. 823–871, Nov. 2019.

[12] S. Ho and S. Verdú, "On the interplay between conditional entropy and error probability," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 5930–5942, 2010.

[13] Y. Sakai, "Generalized Fano-type inequality for countably infinite systems with list-decoding," arXiv:1801.02876, 2018.

[14] I. Sason, "Entropy bounds for discrete random variables via maximal coupling," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7118–7131, 2013.

[15] K. M. R. Audenaert, "A sharp continuity estimate for the von Neumann entropy," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 28, pp. 8127–8136, Jun 2007. [Online]. Available: https://doi.org/10.1088%2F1751-8113%2F40%2F28%2Fs18

[16] G. A. Raggio, "Properties of q-entropies," *Journal of Mathematical Physics*, vol. 36, no. 9, pp. 4785–4791, Sep. 1995.

[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*, vol. 60, no. 2, Nov 2016. [Online]. Available: http://dx.doi.org/10.1007/s11433-016-0367-x

[18] A. Winter, "Coding theorem and strong converse for quantum channels," *Journal of Physics A: Mathematical and Theoretical*, vol. 40(28), pp. 1827–1836, 1999.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory, 2nd ed.* New York: John Wiley & Sons, 2006.

[20] A.B.Aleksandrov and V.V.Peller, "Functions of operators under perturbations of class s_P," *Journal of Functional Analysis*, no. 258, pp. 3675–3724, 2010.

[21] ——, "Operator Hölder–Zygmund functions," *Advances in Mathematics*, vol. 224, pp. 910–966, 2010.

[22] A. Winter, "Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints," *Communications in Mathematical Physics*, vol. 347, no. 1, pp. 291–313, Oct 2016. [Online]. Available: https://doi.org/10.1007/s00220-016-2609-8

[23] G. D. Palma, D. Trevisan, and V. Giovannetti, "Passive states optimize the output of bosonic gaussian quantum channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2895–2906, 2016. [Online]. Available: https://doi.org/10.1109/TIT.2016.2547426

[24] L. Mirsky, "Symmetric Gauge Functions and Unitarily Invariant Norms," *The Quarterly Journal of Mathematics*, vol. 11, no. 1, pp. 50–59, 01 1960.

[25] B. Simon, *Trace Ideals and Their Applications*, ser. Mathematical surveys and monographs. American Mathematical Society, 2005. [Online]. Available: https://books.google.ch/books?id=9UJJxsobYR8C

[26] M. E. Shirokov, "Estimates for discontinuity jumps of information characteristics of quantum systems and channels," *Problems of Information Transmission*, vol. 52, no. 3, pp. 239–264, Jul 2016. [Online]. Available: https://doi.org/10.1134/S0032946016030030

[27] H. Ghourchian, A. Gohari, and A. Amini, "Existence and continuity of differential entropy for a class of distributions," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1469–1472, 2017. [Online]. Available: https://doi.org/10.1109/LCOMM.2017.2689770

[28] M. E. Shirokov, "Tight continuity bounds for the quantum conditional mutual information, for the Holevo quantity and for capacities of quantum channels," *J. Math. Phys.*, p. 58, Dec. 2015.

Michael G. Jabbour received the B.S. and M.S. degrees in physics engineering from École polytechnique de Bruxelles, Université libre de Bruxelles, Brussels, in 2013 and the Ph.D. degree in engineering sciences from École polytechnique de Bruxelles, Université libre de Bruxelles, Brussels, in 2018.

From 2019 to 2020, he was a Postdoctoral Researcher in the Department of Applied Mathematics and Theoretical Physics, in the Centre for Mathematical Sciences of the University of Cambridge. Since 2021, he has been a Postdoctoral Researcher in the Department of Physics of the Technical University of Denmark. His research interests include quantum optics, quantum information theory and mathematical physics.

Simon Becker is a mathematical physicist who has completed his undergraduate studies in mathematics and physics at the Free University of Berlin and Ludwig Maximilian University in Munich. After obtaining his master's degree, he obtained a PhD in applied mathematics from the University of Cambridge in 2021.

Simon spent a year at the Courant Institute at New York University, following the completion of his PhD, where he continued his research in mathematical aspects of condensed matter physics. Currently, he is a postdoctoral researcher in the mathematics department at ETH Zurich.

Nilanjana Datta obtained a PhD in mathematical physics from ETH Zurich, Switzerland, in 1996. From 1997 to 2000, she was a Postdoctoral Researcher at the Dublin Institute of Advanced Studies, C.N.R.S. Marseille, and EPFL in Lausanne. In 2001 she joined the University of Cambridge, U.K., as a Lecturer in Mathematics of Pembroke College, and a member of the Statistical Laboratory in the Centre for Mathematical Sciences. She is currently a Professor of Quantum Information Theory in the Department of Applied Mathematics and Theoretical Physics of the University of Cambridge, and a Fellow of Pembroke College. Her scientific interests include quantum information theory and mathematical physics.