



**HAL**  
open science

# Continuity of Quantum Entropic Quantities via Almost Convexity

Andreas Bluhm, Ángela Capel, Paul Gondolf, Antonio Pérez-Hernández

► **To cite this version:**

Andreas Bluhm, Ángela Capel, Paul Gondolf, Antonio Pérez-Hernández. Continuity of Quantum Entropic Quantities via Almost Convexity. *IEEE Transactions on Information Theory*, 2023, 69 (9), pp.5869-5901. 10.1109/TIT.2023.3277892 . hal-04438926

**HAL Id: hal-04438926**

**<https://hal.science/hal-04438926>**

Submitted on 5 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Continuity of quantum entropic quantities via almost convexity\*

Andreas Bluhm<sup>†1,2</sup>, Ángela Capel<sup>‡3</sup>, Paul Gondolf<sup>§3</sup>, and Antonio Pérez-Hernández<sup>¶4</sup>

<sup>1</sup>QMATH, Department of Mathematical Sciences, University of Copenhagen,  
Universitetsparken 5, 2100 Copenhagen, Denmark

<sup>2</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

<sup>3</sup>Fachbereich Mathematik, Universität Tübingen, 72076 Tübingen, Germany

<sup>4</sup>Departamento de Matemática Aplicada I, Escuela Técnica Superior de Ingenieros Industriales, Universidad Nacional de Educación a Distancia, calle Juan del Rosal 12, 28040 Madrid (Ciudad Universitaria), Spain

February 5, 2024

## Abstract

Based on the proofs of the continuity of the conditional entropy by Alicki, Fannes, and Winter, we introduce in this work the almost locally affine (ALAFF) method. This method allows us to prove a great variety of continuity bounds for the derived entropic quantities. First, we apply the ALAFF method to the Umegaki relative entropy. This way, we recover known almost tight bounds, but also some new continuity bounds for the relative entropy. Subsequently, we apply our method to the Belavkin-Staszewski relative entropy (BS-entropy). This yields novel explicit bounds in particular for the BS-conditional entropy, the BS-mutual and BS-conditional mutual information. On the way, we prove almost concavity for the Umegaki relative entropy and the BS-entropy, which might be of independent interest. We conclude by showing some applications of these continuity bounds in various contexts within quantum information theory.

---

\*This paper was presented at Beyondiid10 2022 and is a long version of [19].

<sup>†</sup>andreas.bluhm@univ-grenoble-alpes.fr

<sup>‡</sup>angela.capel@uni-tuebingen.de

<sup>§</sup>paul.gondolf@student.uni-tuebingen.de

<sup>¶</sup>antperez@ind.uned.es

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Main results</b>	<b>4</b>
<b>3</b>	<b>Preliminaries</b>	<b>5</b>
3.1	Notation and basic concepts . . . . .	5
3.2	Entropies and derived quantities . . . . .	6
<b>4</b>	<b>From almost convexity to continuity bounds</b>	<b>8</b>
<b>5</b>	<b>Almost concavity and continuity bounds for the Umegaki relative entropy</b>	<b>11</b>
5.1	Almost concavity for the relative entropy . . . . .	12
5.2	Reduction to almost tight previously-known continuity bounds for the relative entropy	15
5.2.1	Uniform continuity for the conditional entropy . . . . .	15
5.2.2	Uniform continuity for the mutual information . . . . .	15
5.2.3	Uniform continuity for the conditional mutual information . . . . .	16
5.3	New continuity bounds for the relative entropy . . . . .	16
5.3.1	Divergence bounds for the relative entropy . . . . .	17
5.3.2	Continuity bounds for the relative entropy . . . . .	19
<b>6</b>	<b>Almost concavity and continuity bounds for the Belavkin-Staszewski entropy</b>	<b>22</b>
6.1	Almost concavity for the BS-entropy . . . . .	22
6.2	Continuity bounds for the BS-entropy . . . . .	26
6.2.1	Uniform continuity for the BS-conditional entropy . . . . .	27
6.2.2	Uniform continuity for the BS-mutual information . . . . .	29
6.2.3	Uniform continuity for the BS-conditional mutual information . . . . .	31
6.2.4	Divergence bound for the BS-entropy . . . . .	31
<b>7</b>	<b>Applications</b>	<b>32</b>
7.1	Quantum hypothesis testing . . . . .	32
7.2	Free energy . . . . .	33
7.3	Approximate Quantum Markov Chains . . . . .	34
7.4	Difference between relative entropy and BS-entropy . . . . .	36
7.5	Weak quasi-factorization of the relative entropy . . . . .	38
7.6	Minimal distance to separable states . . . . .	40
7.7	Rains information . . . . .	44
<b>8</b>	<b>Outlook</b>	<b>46</b>
<b>A</b>	<b>Numerical investigation of the variational definition of the BS-conditional entropy</b>	<b>54</b>
<b>B</b>	<b>Supplements to the proof of theorem 5.1</b>	<b>54</b>
<b>C</b>	<b>Proof of proposition 5.2</b>	<b>56</b>
<b>D</b>	<b>Proof of lemma 5.12</b>	<b>57</b>

<b>E</b>	<b>Proof of proposition 6.5</b>	<b>58</b>
<b>F</b>	<b>Proof of proposition 6.6</b>	<b>59</b>
<b>G</b>	<b>Behavior of <math>g_d</math></b>	<b>61</b>

## 1 Introduction

Entropic quantities have proven essential in characterizing the information-processing capabilities both of classical and quantum systems. As the real world cannot be measured to infinite precision, such quantities need to be continuous to contain meaningful information about physical systems. Often, however, we do not only want to know whether an entropic quantity is continuous but also to quantify this continuity. That means we are interested in estimating for an entropic quantity  $f$

$$\sup\{|f(\rho) - f(\sigma)| : \rho, \sigma \in \mathcal{S}_0, d(\rho, \sigma) \leq \varepsilon\}.$$

for some subset  $\mathcal{S}_0$  of the quantum states and some appropriate distance measure  $d$  such as the trace distance, for example.

Already in 1973, Fannes [30] proved that the von Neumann entropy is uniformly continuous and gave a concrete dimension-dependent bound, which was later improved to a sharp version in [4, 64]. Similar results in the line of almost concavity for the von Neumann entropy were provided in [50], [26], [49] or [5], among others. Another example of a concrete continuity estimate is the Alicki-Fannes inequality for the conditional entropy [1], which was subsequently improved to an almost tight version by Winter [88]. Applications of this kind of continuity bounds include, but are not limited to, entanglement measures [60] and the capacities of quantum channels [52, 69]. We refer the reader to textbooks such as [87] for a thorough discussion of continuity bounds and their applications.

The importance of the Alicki-Fannes result in [1] goes beyond its quantification of the continuity of the conditional entropy, but their method and its improved versions [59, 79, 88] work quite generally for entropic quantities. Most clearly, this has been realized by Shirokov, who has named this approach the *Alicki-Fannes-Winter method* [72, 74]. We continue this line of work by generalising the Shirokov approach further to what we call the *almost locally affine (ALAFF) method*. The aim of this generalization is to apply it to entropic quantities beyond those derived from the Umegaki relative entropy [83], such as the conditional entropy. In particular, we are interested in the Belavkin-Staszewski relative entropy (BS-entropy) [13] and its derived entropic quantities. As the Umegaki relative entropy, it generalizes the Kullback-Leibler relative entropy of classical systems [51], but it is less well studied (see [16, 18, 43, 56, 57] for some recent results). The BS-entropy and the related family of geometric Rényi divergences have recently found an application for estimating channel capacities [29]. Moreover, generalizations of the mutual information and other entropic quantities based on the BS-entropy have been defined [68, 89]. The BS-mutual information has been instrumental in proving that the mutual information in one-dimensional quantum Gibbs states of finite-range, translation-invariant Hamiltonians decays exponentially fast [17] and that Davies generators in one dimension which converge to those Gibbs states, in the commuting case, satisfy a positive modified logarithmic Sobolev inequality at every temperature, and thus rapid mixing [9, 10].

A short version of the current manuscript, with new applications in the context of quantum entropic uncertainty relations, has been published in [19].

## 2 Main results

This section summarizes the main results of this article. The focus of this work is not so much on the continuity bounds themselves, but more on the introduction of the method which allows deriving all of them in a systematic way (section 4). Our approach is summarized in fig. 1. For a given divergence, in this paper either the Umegaki relative entropy [83] or the BS-entropy [13], we need to prove two properties: its (joint) convexity and its almost (joint) concavity. Both of these properties, under certain conditions on the remainder function, then directly translate into almost local affinity (definition 4.5) of the entropic quantities derived from the divergence at hand on a suitably defined subset  $\mathcal{S}_0$  of  $\mathcal{S}(\mathcal{H})$ . Serving as input to the ALAFF method, the remainder estimates get translated into continuity bounds for said quantities. The entropic quantities include, for example, versions of the conditional entropy and the (conditional) mutual information, as defined in fig. 1. The necessity of  $\mathcal{S}_0$  as a restriction of  $\mathcal{S}(\mathcal{H})$  becomes obvious when trying to prove continuity bounds for the Umegaki relative entropy, for example. It is known not to be continuous on the set of all pairs of states  $(\rho, \sigma)$ , which makes a careful choice of  $\mathcal{S}_0$  inevitable. To this end, we define  $s$ -perturbed  $\Delta$ -invariant convex subsets of  $\mathcal{S}(\mathcal{H})$  (definition 4.3) for which we can show that the ALAFF method works and which are general enough to capture all situations of interest. For the formal statement of the ALAFF method, we refer the reader to theorem 4.6.

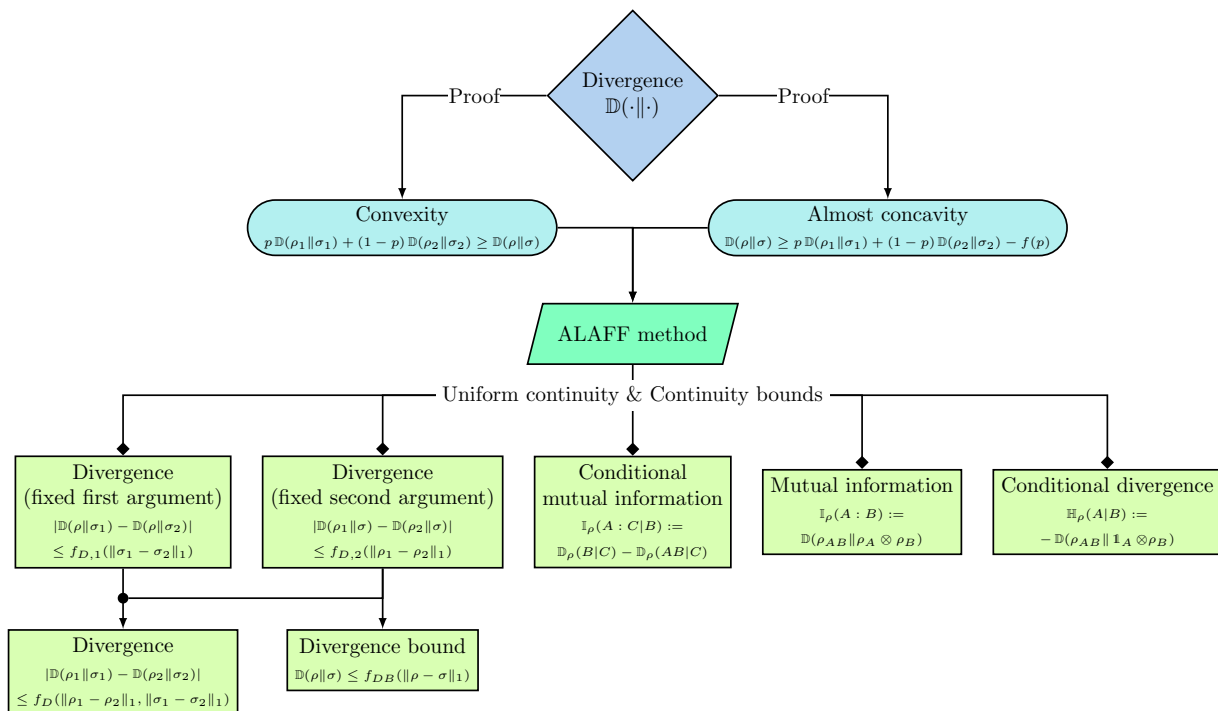


Figure 1: A flow chart demonstrating how convexity and almost concavity of a divergence can be used to obtain uniform continuity and explicit continuity bounds on entropic quantities derived from that divergence. The subscripts of the functions  $f_{D,1/2}$  and  $f_{DB}$  stand for divergence first, second argument and divergence bound respectively.

Thus, we are left with proving convexity and almost concavity for the divergences we are interested in, namely the Umegaki relative entropy (section 5) and the Belavkin-Staszewski entropy (section 6), and deriving the precise continuity estimates. For the convexity, we can rely on well-known results from the literature both for the Umegaki relative entropy [55] and the BS-entropy

[43, 56]. For the Umegaki relative entropy, given by

$$D(\rho\|\sigma) := \text{Tr} [\rho(\log \rho - \log \sigma)] \quad \text{if } \ker \sigma \subseteq \ker \rho ,$$

or  $+\infty$  otherwise, we prove almost concavity in theorem 5.1 and find that it is tight. The application of the ALAFF method then allows us to recover in section 5.2 the almost tight results for the conditional entropy [88] and the mutual and conditional mutual information (which can be derived from the conditional entropy [87]), but also to derive in section 5.3 new versions of what we call divergence bounds [7, 8, 21, 86], i.e. bounds on  $D(\rho\|\sigma)$  in terms of  $\frac{1}{2}\|\rho - \sigma\|_1$ . Furthermore, our technique produces a new result, which is the uniform continuity of the relative entropy itself (in both arguments, on a suitable set  $\mathcal{S}_0$ ), as well as an explicit continuity bound.

For the BS-entropy, given by

$$\widehat{D}(\rho\|\sigma) := \text{Tr} \left[ \rho \log(\rho^{1/2} \sigma^{-1} \rho^{1/2}) \right] \quad \text{if } \ker \sigma \subseteq \ker \rho ,$$

or  $+\infty$  otherwise, we prove the almost concavity in theorem 6.3.

The ALAFF method yields novel explicit bounds in particular for the BS-conditional entropy, the BS-mutual and BS-conditional mutual information that we gather in section 6.2. We expect these new continuity bounds and those provided for quantities derived from the relative entropy to find applications in proving the continuity of various quantities in diverse fields related to quantum information theory. In particular, we provide here a number of applications of our results in the context of quantum hypothesis testing (section 7.1), to show that states that are hard to discriminate have almost the same performance in terms of hypothesis testing, as well as in quantum thermodynamics (section 7.2), to show continuity of the distillable athermality. We also reprove that a state is an approximate quantum Markov chain if and only if it is close to being recovered by the Petz recovery map (section 7.3), and use our most general continuity bounds for the relative entropy to obtain bounds for the difference between the relative entropy and the BS-entropy of two quantum states (section 7.4). Additionally, we show a new result of weak quasi-factorization for the relative entropy, i.e. with an additive error term and no multiplicative error term (section 7.5). Finally, we include continuity bounds for the relative entropy of entanglement as well as the analogously defined BS-entropy of entanglement (section 7.6), and subsequently lift these results to show continuity of the Rains information induced by the relative and the BS-entropy respectively (section 7.7).

## 3 Preliminaries

### 3.1 Notation and basic concepts

We denote a Hilbert space by  $\mathcal{H}$  which, throughout this paper is assumed to be finite. The dimension of such a Hilbert space will be called  $d$  and for its elements, we use  $|\varphi\rangle$ ,  $|\psi\rangle$  and  $|i\rangle$  for  $i \in \mathbb{N}$ , possibly with additional indices. If we are concerned with a bipartite or tripartite system, we will always use capital letters in the index to identify objects associated with the respective subsystems. If we have, for example, the bipartite space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  and consider the dimension of  $\mathcal{H}_A$ , we write  $d_A$ .

The set of (bounded) linear operators on a Hilbert space  $\mathcal{H}$  is  $\mathcal{B}(\mathcal{H})$  and the subspace of positive semi-definite operators with trace one, i.e., the quantum states or density matrices, is denoted by  $\mathcal{S}(\mathcal{H})$ . If we want to restrict this set even further, we indicate this with a subindex. Thus, the set of positive definite quantum states becomes  $\mathcal{S}_+(\mathcal{H})$ , or if we want to restrict moreover to the set of quantum states that have minimal eigenvalue greater than  $m$ , we write  $\mathcal{S}_{\geq m}(\mathcal{H})$ . On the set of

quantum states as well as on the set of self-adjoint operators, the relation  $\leq$  is meant to be the partial order in the Löwner sense. That is,  $\rho \geq \sigma$  if and only if  $\rho - \sigma$  is positive semidefinite.

We use  $\text{Tr}[\cdot]$  for the usual matrix trace and  $\|\cdot\|_1 = \text{Tr}[\|\cdot\|]$  and  $\|\cdot\|_\infty$  to denote the trace norm and the spectral norm on  $\mathcal{B}(\mathcal{H})$ , respectively. Quantum states in general are denoted by lower Greek letters such as  $\rho, \sigma$  and  $\tau$ , for example. For Hermitian operators in  $\mathcal{B}(\mathcal{H})$  we usually use upper Latin letters such as  $X, Y$ . For any such  $X$ , we denote by  $[X]_+$  and  $[X]_-$  its positive and negative parts, respectively.

As we later want to formally control the dependence on the states  $\rho$  and  $\sigma$  that are given as arguments to the divergences, we further introduce  $\mathcal{H} \times \mathcal{H}$  the cartesian product of the Hilbert space  $\mathcal{H}$  with itself. Moreover, on a bipartite system  $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ , we set  $\rho_A$  to be the state on  $\mathcal{H}_A$  that  $\rho \in \mathcal{S}(\mathcal{H}_{AB})$  is mapped to under the partial trace with respect to the subsystem  $B$  which is a completely positive trace-preserving (CPTP) map. Furthermore, we denote by  $\mathbb{1}_A$  the identity matrix on  $A$  and, in a slight abuse of notation, we denote by  $\text{Tr}_A[\cdot]$  both the partial trace with respect to  $A$  as well as the complemented map on  $\mathcal{H}_{AB}$  by tensorizing with  $\mathbb{1}_A$ .

### 3.2 Entropies and derived quantities

The *von Neumann entropy* of  $\rho \in \mathcal{S}(\mathcal{H})$  is given by

$$S(\rho) := -\text{Tr}[\rho \log(\rho)] .$$

For two quantum states  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , their (*Umegaki*) *relative entropy* [83] is defined as

$$D(\rho\|\sigma) := \begin{cases} \text{Tr}[\rho \log \rho - \rho \log \sigma] & \text{if } \ker \sigma \subseteq \ker \rho, \\ +\infty & \text{otherwise,} \end{cases}$$

and their *Belavkin-Staszewski (BS) entropy* [13] by

$$\widehat{D}(\rho\|\sigma) := \begin{cases} \text{Tr}[\rho \log \rho^{1/2} \sigma^{-1} \rho^{1/2}] & \text{if } \ker \sigma \subseteq \ker \rho, \\ +\infty & \text{otherwise.} \end{cases}$$

In the event of  $\rho$  and  $\sigma$  commuting, the two entropies coincide. Otherwise, the BS-entropy is strictly larger than the relative entropy [43]. We further note that both can also be defined in terms of positive semi-definite operators  $A, B$  (without normalisation), by just replacing  $\rho$  with  $A$  and  $\sigma$  with  $B$ . We make use of this alternative definition when we define the conditional entropy and the BS-conditional entropy, for example. Using this notation we can write the *conditional entropy* of  $\rho$  as

$$H_\rho(A|B) := S(\rho_{AB}) - S(\rho_B) = -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B),$$

with the last equality being a straightforward calculation. The subscript  $AB$  in  $\rho_{AB} = \rho$  just emphasises the fact that  $\rho$  stems from  $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  and to distinguish it from its partial trace  $\rho_B$ , for example. It is noteworthy that the conditional entropy admits the following variational expression

$$H_\rho(A|B) = \max_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} -D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B). \quad (1)$$

Furthermore, in a similar manner as for the conditional entropy, one obtains the representation of the *mutual information* in terms of the von Neumann entropy and the conditional entropy

$$I_\rho(A : B) := S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = S(\rho_A) - H_\rho(A|B) = D(\rho_{AB}\|\rho_A \otimes \rho_B).$$

Finally, on a tripartite system  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  the *conditional mutual information* of a state  $\rho \in \mathcal{S}(\mathcal{H})$  is given by

$$\begin{aligned} I_\rho(A : B|C) &:= S(\rho_{AC}) + S(\rho_{BC}) - S(\rho_C) - S(\rho_{ABC}) \\ &= H_\rho(A|C) - H_\rho(A|BC) \\ &= I_\rho(A : BC) - I_\rho(A : C). \end{aligned} \quad (2)$$

The last equalities are again straightforward. One easily checks that both the mutual information and the conditional mutual information are symmetric under the exchange of the  $A$  and  $B$  system.

Let us proceed now to introduce the analogous quantities from the BS instead of the relative entropy. In this framework, we cannot construct them as sums and differences of von Neumann entropies, which, for every BS-entropic quantity, gives rise to a zoo of different possible definitions. Some of them have already appeared before in [18, 68, 89]. For a bipartite state  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , inspired by the notion of conditional entropy, we define the *BS-conditional entropy* as

$$\widehat{H}_\rho(A|B) := -\widehat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B), \quad (3)$$

and building on the mutual information, we define the *BS-mutual information* as

$$\widehat{I}_\rho(A : B) := \widehat{D}(\rho_{AB} \| \rho_A \otimes \rho_B).$$

Finally, the analogue of the conditional mutual information in this setting is a more subtle matter. Indeed, two natural ways to construct such a quantity would be either as a difference of BS-conditional entropies or of BS-mutual information, as shown in eq. (2), which in general do not coincide. Given  $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ :

- We define the (*one-sided*) *BS-conditional mutual information* (os BS-CMI in short) by

$$\widehat{I}_\rho^{\text{os}}(A : B|C) := \widehat{H}_\rho(A|C) - \widehat{H}_\rho(A|BC) = \widehat{D}(\rho_{ABC} \| \mathbb{1}_A \otimes \rho_{BC}) - \widehat{D}(\rho_{AC} \| \mathbb{1}_A \otimes \rho_C). \quad (4)$$

- We define the (*two-sided*) *BS-conditional mutual information* (ts BS-CMI in short) by

$$\widehat{I}_\rho^{\text{ts}}(A : B|C) := \widehat{I}_\rho(A : BC) - \widehat{I}_\rho(A : C) = \widehat{D}(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) - \widehat{D}(\rho_{AC} \| \rho_A \otimes \rho_C).$$

Note that both notions are clearly non-negative, as a consequence of the data processing inequality for the BS-entropy. In this project, we focus for simplicity on the first definition, i.e. the one-sided one. We will therefore drop the “os” notation, as there is no possible confusion.

Let us emphasize at this stage that the difference between the aforementioned two definitions of BS-conditional mutual information is partly related to the pathological behaviour of the BS-entropy with respect to continuity in general, and more specifically to the fact that the BS-conditional entropy is discontinuous on the set of positive semi-definite quantum states (cf. proposition 6.7). We suspect that as a consequence thereof, the variational definition of the BS-conditional entropy (generalizing eq. (1)) does not agree with the one we have given in eq. (3), namely

$$\widehat{H}_\rho(A|B) \leq \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} -\widehat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B). \quad (5)$$

We have numerical results that suggest that the inequality in the eq. (5) is strict, at least in some cases. A plot of those numerics can be found in appendix A. Moreover, we will indeed formally show that both quantities are different in general in remark 7.17.



## 4 From almost convexity to continuity bounds

As depicted in fig. 1, our approach is based on the convexity and almost concavity of a divergence. More precisely, it is based on its joint convexity and almost joint concavity, but for the sake of better readability, we will just speak of convexity and almost concavity.

It is immediately clear what is meant by convexity and this property is often even a defining property of a divergence [44] or a direct consequence thereof<sup>1</sup> [81, Proposition 4.2]. The almost (joint) concavity, however, needs yet to be defined.

### Definition 4.1 (Almost (joint) concavity of a divergence)

A divergence  $\mathbb{D}(\cdot\|\cdot)$  is called *almost (jointly) concave* on a convex set  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H})$  if, for  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_0$ , there exists a continuous function  $f : [0, 1] \rightarrow \mathbb{R}$  with  $f(0) = f(1) = 0$  such that, for all  $p \in [0, 1]$ ,

$$\mathbb{D}(\rho\|\sigma) \geq p \mathbb{D}(\rho_1\|\sigma_1) + (1 - p) \mathbb{D}(\rho_2\|\sigma_2) - f(p) \quad (6)$$

holds. Here,  $\rho = p\rho_1 + (1 - p)\rho_2$  and  $\sigma = p\sigma_1 + (1 - p)\sigma_2$ . It is important to emphasise that  $f$  in general depends on the states involved.

**Remark 4.2** *We note that the definition of almost concavity presented above is not itself a very strong property. For example, one could just choose  $f$  to be the remainders that give equality in eq. (6). It is the behaviour of the remainder functions that is pivotal, i.e., it becomes independent of  $\rho_i, \sigma_i, i = 1, 2$  under certain restrictions on the states, e.g. requiring that  $\sigma_i$  is a marginal of  $\rho_i$ .*

Our approach, therefore, does not only need joint convexity but a well-behaved remainder function. If we find such a function and combine it with the boundedness of the divergence (or underlying entropic quantity), ALAFF directly gives uniform continuity through explicit continuity bounds.

As we already discussed in the introduction, the predecessor of ALAFF was developed and used by Alicki and Fannes [1], as well as Winter [88], to prove uniform continuity and give an explicit continuity bound for the conditional entropy. Shirokov then noticed the potential beyond this specific application and moulded a method that can be applied to functions defined on convex and  $\Delta$ -invariant subsets of  $\mathcal{S}(\mathcal{H})$  [72, 74]. Independently, similar techniques were already used in [59]. In short,  $\Delta$ -invariance means that for two elements their normalised positive and negative part again lies in the set (see also Definition 4.3). This definition of  $\Delta$ -invariance will, however, turn out to be a limitation when trying to prove the uniform continuity of the relative entropy, while in the case of the BS-entropy, it is unfitting even from the beginning, i.e., even for the BS-conditional entropy. The problem is due to  $\Delta$ -invariance being a rather strong property that sets like  $\mathcal{S}_{\geq m}(\mathcal{H})$  or  $\{(\rho, \sigma) : \ker \sigma \subseteq \ker \rho\}$  do not have. Yet, those sets, or modified versions thereof, are the relevant sets for the relative and, in particular, the BS-entropy.

In light of those problems and in an effort to make our approach as general as possible, we propose the almost locally affine (ALAFF) method, a generalisation of the Alicki-Fannes-Winter-Shirokov method that reduces to one implication of the former in a special case. First of all, we define a perturbed version of the  $\Delta$ -invariant subset, with the perturbation controlled by a parameter  $s$ .

---

<sup>1</sup>Some authors define divergences as functions on two density operators fulfilling a data processing inequality; however, note that convexity for a divergence implies a data processing inequality and follows from it together with additional properties, as shown in [44, Corollary 4.7].

**Definition 4.3 (Perturbed  $\Delta$ -invariant subset)**

Let  $s \in [0, 1)$ . A subset  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$  is called  $s$ -perturbed  $\Delta$ -invariant, if for  $\rho, \sigma \in \mathcal{S}_0$  with  $\rho \neq \sigma$  there exists  $\tau \in \mathcal{S}(\mathcal{H})$  such that the two states

$$\Delta^\pm(\rho, \sigma, \tau) = s\tau + (1-s)\varepsilon^{-1}[\rho - \sigma]_\pm \quad (7)$$

lie again in  $\mathcal{S}_0$ . Here  $\varepsilon := \frac{1}{2} \|\rho - \sigma\|_1$  and  $[A]_\pm$  denotes the negative and positive part of a self-adjoint operator, respectively. For  $s = 0$ , we recover the definition of  $\Delta$ -invariant subset used in [74].

We want to give the reader some intuition about those  $s$ -perturbed  $\Delta$ -invariant sets.

**Remark 4.4** 1. Let  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$  be  $s$ -perturbed  $\Delta$ -invariant convex set. Then for  $t \in [s, 1)$  it is  $t$ -perturbed  $\Delta$ -invariant as well. In particular, being 0-perturbed is the strongest condition.

2. If  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$  has non-empty interior with respect to the 1-norm, then it is  $s$ -perturbed for some  $s \in [0, 1)$ .

3. If  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$  is  $s$ -perturbed  $\Delta$ -invariant containing more than one state, then there exist  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 = 1 - s$ . This follows directly from the definition.

In order to get well-behaved remainder functions, we define a stronger property that we call “almost local affinity”.

**Definition 4.5 (Almost locally affine (ALAFF) function)**

Let  $f$  be a real-valued function on the convex set  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$ , fulfilling

$$-a_f(p) \leq f(p\rho + (1-p)\sigma) - pf(\rho) - (1-p)f(\sigma) \leq b_f(p) \quad (8)$$

for all  $p \in [0, 1]$  and  $\rho, \sigma \in \mathcal{S}_0$ . The functions  $a_f : [0, 1] \rightarrow \mathbb{R}$  and  $b_f : [0, 1] \rightarrow \mathbb{R}$  are required to vanish as  $p \rightarrow 0^+$ , to be non-decreasing on  $[0, \frac{1}{2}]$ , continuous in  $p$  and independent of  $\rho, \sigma \in \mathcal{S}_0$ . We then call  $f$  an *almost locally affine (ALAFF) function*.

The notion of almost locally affine functions has appeared previously in the literature, also under the name “approximate affinity” (see e.g. [20]). We can now formulate the following theorem, whose proof is inspired by Shirokov [74].

**Theorem 4.6 (Almost locally affine (ALAFF) method)**

Let  $s \in [0, 1)$  and  $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$  be an  $s$ -perturbed  $\Delta$ -invariant convex subset of  $\mathcal{S}(\mathcal{H})$  containing more than one element. Let further  $f$  be an ALAFF function. We then find that  $f$  is uniformly continuous if

$$C_f^s := \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1-s}} |f(\rho) - f(\sigma)| < +\infty.$$

In this case, we have for  $\varepsilon \in (0, 1]$

$$\sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon}} |f(\rho) - f(\sigma)| \leq C_f^s \frac{\varepsilon}{1-s} + \frac{1-s+\varepsilon}{1-s} E_f^{\max} \left( \frac{\varepsilon}{1-s+\varepsilon} \right), \quad (9)$$

with

$$E_f^{\max} : [0, 1) \rightarrow \mathbb{R}, \quad p \mapsto E_f^{\max}(p) = (1-p) \max \left\{ \frac{E_f(t)}{1-t} : 0 \leq t \leq p \right\},$$

where  $E_f = a_f + b_f$ . Note that on  $\varepsilon \in (0, 1-s]$   $E_f$  and  $E_f^{\max}$  coincide.

*Proof.* Let  $s \in [0, 1)$  and  $\varepsilon \in (0, 1]$ . Let further  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 = \varepsilon$ . Then by the property of  $s$ -perturbed  $\Delta$ -invariance there exists  $\tau \in \mathcal{S}(\mathcal{H})$  such that  $\gamma_{\pm} := \Delta^{\pm}(\rho, \sigma, \tau) \in \mathcal{S}_0$  defined as in eq. (7). For every such  $\gamma_{\pm}$  with a representation in terms of  $\rho, \sigma \in \mathcal{S}_0$  and a  $\tau \in \mathcal{S}(\mathcal{H})$  we have that

$$\frac{1-s}{1-s+\varepsilon}\rho + \frac{\varepsilon}{1-s+\varepsilon}\gamma_- = \omega^* = \frac{1-s}{1-s+\varepsilon}\sigma + \frac{\varepsilon}{1-s+\varepsilon}\gamma_+,$$

which can be easily checked by inserting the explicit form of  $\gamma_{\pm}$  and using that  $[\rho - \sigma]_+ - [\rho - \sigma]_- = \rho - \sigma$ . Now  $\omega^* \in \mathcal{S}_0$  as  $\mathcal{S}_0$  is convex, which allows us to evaluate  $f$  at  $\omega^*$  and use eq. (8) for both of the representations we have for the state in question. This gives us

$$\begin{aligned} -a_f(p) &\leq f(\omega^*) - (1-p)f(\rho) - pf(\gamma_-) \leq b_f(p), \\ -a_f(p) &\leq f(\omega^*) - (1-p)f(\sigma) - pf(\gamma_+) \leq b_f(p), \end{aligned}$$

where we set  $p = p(\varepsilon) = \frac{\varepsilon}{1-s+\varepsilon}$  for better readability. Note that  $p \in (0, \frac{1}{2-s}] \subseteq [0, 1)$  as  $\varepsilon \in (0, 1]$  and  $s \in [0, 1)$  and further that  $p(\varepsilon)$  is monotone with respect to  $\varepsilon$ . We recombine the above to get

$$\begin{aligned} (1-p)(f(\rho) - f(\sigma)) &\leq p(f(\gamma_+) - f(\gamma_-)) + a_f(p) + b_f(p), \\ (1-p)(f(\sigma) - f(\rho)) &\leq p(f(\gamma_-) - f(\gamma_+)) + a_f(p) + b_f(p). \end{aligned}$$

Those two inequalities immediately give us

$$(1-p)|f(\rho) - f(\sigma)| \leq p|f(\gamma_+) - f(\gamma_-)| + (a_f + b_f)(p).$$

If we now insert  $E_f = a_f + b_f$ , we obtain

$$|f(\rho) - f(\sigma)| \leq \frac{p}{1-p}|f(\gamma_+) - f(\gamma_-)| + \frac{1}{1-p}E_f(p).$$

In the case that  $C_f^s$  is finite, we can take the supremum over all  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 = \varepsilon$  of the last equation and even extend to  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$  in two steps. The first step is upper bounding  $\frac{1}{1-p}E_f(p)$  with  $\frac{1}{1-p}E_f^{\max}(p)$  and the second one using that  $\frac{1}{1-p}E_f^{\max}(p)$  is engineered to be non-decreasing on  $[0, 1)$  and thereby for the specific  $p = \frac{\varepsilon}{1-s+\varepsilon} \in [0, \frac{1}{2-s}] \subset [0, 1)$ , is non-decreasing in  $\varepsilon$  as well. Since the  $\gamma_+$  and  $\gamma_-$  created from  $\rho$  and  $\sigma$  obviously fulfill  $\gamma_{\pm} \in \mathcal{S}_0$  and  $\frac{1}{2} \|\gamma_+ - \gamma_-\|_1 = 1-s$ , we immediately get the upper bound in eq. (9). The reduction of  $E_f^{\max}$  to  $E_f$  on  $\varepsilon \in (0, 1-s]$  follows immediately from  $E_f$  being non-decreasing on  $[0, \frac{1}{2}]$  meaning further that  $E_f^{\max}$  inherits the vanishing property of  $E_f$  as  $p \rightarrow 0^+$ . This directly translates into  $E_f^{\max}(p(\varepsilon)) \rightarrow 0$  if  $\varepsilon \rightarrow 0^+$ , hence concluding the proof of uniform continuity.  $\square$

The method presented in theorem 4.6 is named the ‘‘ALAFF method’’ to highlight the required ALAFF property necessary, for this technique to be applicable. We will refer to this theorem by that name in subsequent sections.

**Remark 4.7** For  $s = 0$ , one recovers one implication of the method by Shirokov, i.e., the definitions for perturbed  $\Delta$ -invariance and  $\Delta$ -invariance coincide,  $E_f^{\max}$  reduces to  $E_f$  on the relevant domain  $\varepsilon \in [0, 1]$ , and eq. (9) becomes

$$\sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon}} |f(\rho) - f(\sigma)| \leq C_f^{\perp} \varepsilon + (1 + \varepsilon)E_f\left(\frac{\varepsilon}{1 + \varepsilon}\right)$$

with

$$C_f^0 = \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1}} |f(\rho) - f(\sigma)| = \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \text{Tr}[\rho\sigma] = 0}} |f(\rho) - f(\sigma)| =: C_f^{\perp},$$

as states with maximal trace distance have orthogonal support.

In the next sections, we will use theorem 4.6 together with the almost concavity of the relative entropy and the BS-entropy, respectively, to derive a plethora of results of uniform continuity and continuity bounds for entropic quantities defined through them. Depending on the case, we will sometimes have to employ the whole machinery devised in theorem 4.6, whereas at other times the simplification provided in remark 4.7 will be enough.

## 5 Almost concavity and continuity bounds for the Umegaki relative entropy

In this section, we apply the ALAFF method introduced in section 4 for the particular case of the relative entropy, as well as some other entropic quantities derived from it.

All the results provided in this section are summarized in fig. 2.

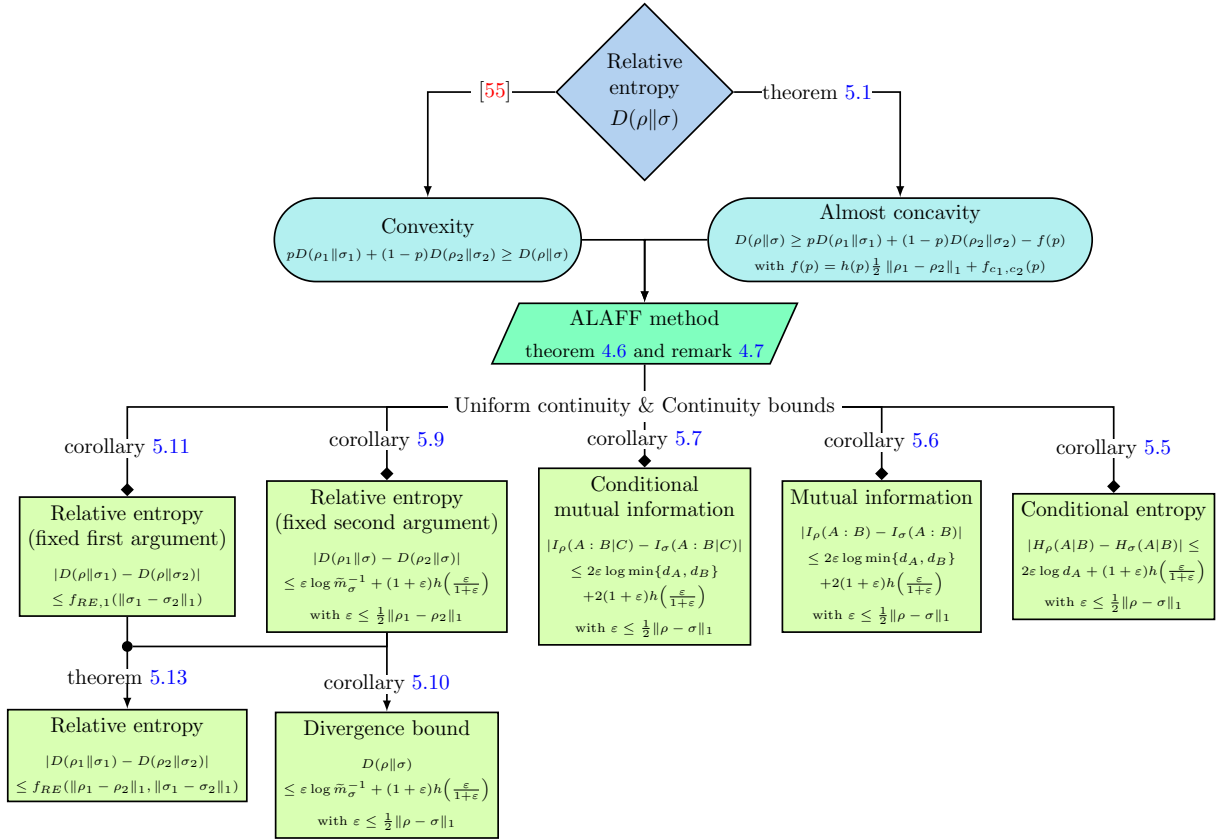


Figure 2: In this flow chart we collect the main results from this chapter, starting with the almost concavity of the relative entropy, which together with the ALAFF method outputs a collection of continuity bounds for related entropic quantities. For the convexity and almost concavity, we are setting  $\rho = p\rho_1 + (1-p)\rho_2$  and  $\sigma = p\sigma_1 + (1-p)\sigma_2$ , with  $p \in [0, 1]$ . We denote by  $\tilde{m}_\sigma$  the minimal non-zero eigenvalue of  $\sigma$ . The specific bounds obtained for the relative entropy fixing the first argument and in the general case (modifying both arguments) are omitted due to their technicality.

## 5.1 Almost concavity for the relative entropy

The (joint) convexity of the relative entropy is a well-established result with proofs found for example in [87]. In this section, we complement this result with almost concavity and further prove that the bound we obtain is tight.

### Theorem 5.1 (Almost concavity of the relative entropy)

Let  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_{\ker}$  with

$$\mathcal{S}_{\ker} := \{(\rho, \sigma) \in \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H}) : \ker \sigma \subseteq \ker \rho\}$$

and  $p \in [0, 1]$ . Then, for  $\rho = p\rho_1 + (1-p)\rho_2$  and  $\sigma = p\sigma_1 + (1-p)\sigma_2$ ,

$$D(\rho\|\sigma) \geq pD(\rho_1\|\sigma_1) + (1-p)D(\rho_2\|\sigma_2) - h(p)\frac{1}{2}\|\rho_1 - \rho_2\|_1 - f_{c_1, c_2}(p). \quad (10)$$

Here,

$$\begin{aligned} h(p) &= -p \log(p) - (1-p) \log(1-p), \\ f_{c_1, c_2}(p) &= p \log(p + (1-p)c_1) + (1-p) \log((1-p) + pc_2), \end{aligned}$$

with the first one being the binary entropy. The constants in  $f_{c_1, c_2}$  are non-negative real numbers and are given by

$$\begin{aligned} c_1 &:= \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma_2 \sigma_1^{\frac{-it-1}{2}} \right] < \infty, \\ c_2 &:= \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{Tr} \left[ \rho_2 \sigma_2^{\frac{it-1}{2}} \sigma_1 \sigma_2^{\frac{-it-1}{2}} \right] < \infty. \end{aligned}$$

Here,  $\beta_0$  is a probability density on  $\mathbb{R}$  (see eq. (13) for a concrete expression). It is noteworthy that  $f_{1,1}(\cdot) = 0$  and  $f_{c_1, c_2}(0) = f_{c_1, c_2}(1) = 0$ .

*Proof.* It is clear that  $\mathcal{S}_{\ker}$  is a convex set and that the bound holds trivially for  $p = 0$  and  $p = 1$ . Hence let  $p \in (0, 1)$  and  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_{\ker}$  in the following. We find that

$$\begin{aligned} pD(\rho_1\|\sigma_1) + (1-p)D(\rho_2\|\sigma_2) - D(\rho\|\sigma) &= -pS(\rho_1) - (1-p)S(\rho_2) + S(\rho) \\ &\quad + (1-p)\operatorname{Tr}[\rho_2(\log \sigma - \log \sigma_2)] \\ &\quad + p\operatorname{Tr}[\rho_1(\log \sigma - \log \sigma_1)] \\ &\leq h(p)\frac{1}{2}\|\rho_1 - \rho_2\|_1 + f_{c_1, c_2}(p), \end{aligned}$$

where we split the relative entropies and used that the von Neumann entropy fulfils [5, Theorem 14]

$$S(\rho) \leq \frac{1}{2}\|\rho_1 - \rho_2\|_1 h(p) + pS(\rho_1) + (1-p)S(\rho_2). \quad (11)$$

Furthermore, we upper bound the remaining terms by  $f_{c_1, c_2}(p)$ , estimating the two separately. We will only demonstrate the derivation for the second term, as it is completely analogous to the first one. We have

$$\begin{aligned} p\operatorname{Tr}[\rho_1(\log(\sigma) - \log(\sigma_1))] &= p\operatorname{Tr}[\exp(\log(\rho_1))(\log(\sigma) - \log(\sigma_1))] \\ &\leq p\log \operatorname{Tr}[\exp(\log(\rho_1) + \log(\sigma) - \log(\sigma_1))] \\ &\leq p\log \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right]. \end{aligned} \quad (12)$$

The first estimate follows immediately using the well-known Peierls-Bogoliubov inequality [62]. The second one involves a generalisation of the Araki-Lieb-Thirring inequality [3, 54] by Sutter et al. [76, Corollary 3.3] with

$$\beta_0(t) = \frac{\pi}{2} \frac{1}{\cosh(\pi t) + 1} \quad (13)$$

a probability density on  $\mathbb{R}$ . In the above steps, i.e. eq. (12), we relied on  $\rho_1, \sigma_1$  and  $\sigma$  to be full rank. If this is not the case one obtains the same result, however, the procedure is more involved. A thorough discussion can be found in appendix B. Note here that in the most general case,  $\cdot^{-1}$  in the RHS of eq. (12) is the Moore-Penrose pseudoinverse. The trace in the integral can now be estimated for each  $t$  by

$$\mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right] = p + (1-p) \mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma_2 \sigma_1^{\frac{-it-1}{2}} \right]. \quad (14)$$

Here, we just split  $\sigma$  and used that  $\mathrm{Tr}[\rho_1] = 1$ . To see that  $c_1 < \infty$ , we upper bound  $\sigma_2$  by  $\mathbb{1}$  and  $\sigma_1^{-1}$  by  $\tilde{m}_{\sigma_1}^{-1} \mathbb{1}$  where  $\tilde{m}_{\sigma_1}$  is the smallest non-zero eigenvalue of  $\sigma_1$ . This can be done, since  $\ker \sigma_1 \subseteq \ker \rho_1$ . We end up with  $c_1 \leq \tilde{m}_{\sigma_1}^{-1} < \infty$ . Inserting eq. (14) into eq. (12), we obtain the first part of  $f_{c_1, c_2}(p)$  and repeating the steps for  $(1-p) \mathrm{Tr}[\rho_2(\log(\sigma) - \log(\sigma_2))]$  the second one as well. This concludes the proof.  $\square$

We remark that eq. (10) provides a result of almost concavity for the relative entropy in the sense of definition 4.1. Indeed, the additive ‘‘correction’’ term obtained for such an inequality to hold behaves well enough, in the sense that it reduces to the previously known bounds for quantities derived from the relative entropy, e.g. the von Neumann entropy or the conditional entropy, and it is almost tight in general. To illustrate that, we provide now two propositions that put the almost concavity of the relative entropy into perspective.

**Proposition 5.2 (Almost concavity estimate of the relative entropy is well behaved)**

*The function  $f_{c_1, c_2} + h \frac{1}{2} \|\rho_1 - \rho_2\|_1$  obtained in theorem 5.1 is well behaved in the following sense: For  $j = 1, 2$  and  $(\rho_j, \sigma_j) \in \mathcal{S}_{\ker}$ , we have the following:*

1. *If  $\sigma_1 = \sigma_2$ , then  $c_1 = c_2 = 1$ , resulting in  $f_{c_1, c_2} + \frac{1}{2} \|\rho_1 - \rho_2\|_1 h \leq h$ .*
2. *If for  $\tilde{m} > 0$  we have  $\tilde{m} \rho_j \leq \sigma_j$ , then  $f_{c_1, c_2} + h \frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq f_{\tilde{m}^{-1}, \tilde{m}^{-1}} + h$ .*
3. *If  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is a bipartite space and furthermore  $\sigma_j = d_A^{-1} \mathbb{1}_A \otimes \rho_{j, B}$ , then  $f_{c_1, c_2} + h \frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq h$ .*
4. *For  $m_1, m_2 \geq 1$  we find that both  $p \mapsto \frac{1}{1-p} f_{m_1, m_2}(p)$  and  $p \mapsto \frac{1}{1-p} h(p)$  are non-decreasing on  $[0, 1)$ .*

We hence find that in the cases listed above the bound becomes independent of the states and further that the remainder functions have a desirable non-decreasing property. The proof is straightforward and can be found in appendix C.

**Remark 5.3** *The different cases discussed in proposition 5.2 are used in the following to find almost concavity results with a function that does not depend on the specifics of the states involved, as necessary for applying the ALAFF method.*

- *If  $\sigma_1 = \sigma_2$ , we are reducing eq. (10) to a result of almost concavity only in the first input. This case was addressed in [20], where they obtained  $h(p)$  as a correction for almost concavity, a bound we are tightening here. Moreover, this case will yield a continuity bound for the relative entropy with fixed second input as shown in corollary 5.9.*

- Point 3 of proposition 5.2 can be interpreted as a result of almost convexity for the conditional entropy. Moreover, it will yield a continuity bound for the conditional entropy in corollary 5.5. Since the latter result is almost tight, this shows the good behaviour of the bound obtained in theorem 5.1.
- Point 2 of proposition 5.2 is the most general setting for full-rank states  $\sigma_j$ , with  $j = 1, 2$ , and will be essential for deriving the most general continuity bounds for the relative entropy in theorem 5.13.

Finally, we conclude this subsection with some discussion of our almost concave bound.

**Proposition 5.4 (Almost concavity estimate of the relative entropy is tight)**

The bound presented in theorem 5.1 is tight. More specifically, there are some density operators  $\rho_1, \rho_2, \sigma_1, \sigma_2$  on  $\mathcal{S}(\mathcal{H})$  which saturate the inequality in eq. (10).

*Proof.* We can assume that the dimension of the underlying Hilbert space is  $d_{\mathcal{H}} \geq 2$ . We then find two orthonormal states  $|0\rangle, |1\rangle \in \mathcal{H}$  that we use to create

$$\begin{aligned}\rho_1 &:= |0\rangle\langle 0|, \\ \rho_2 &:= |1\rangle\langle 1|, \\ \sigma_1 &:= t|0\rangle\langle 0| + (1-t)|1\rangle\langle 1|, \\ \sigma_2 &:= (1-t)|0\rangle\langle 0| + t|1\rangle\langle 1|,\end{aligned}$$

for  $t \in (0, 1)$ . We find, as of the orthonormality, that for  $p \in [0, 1]$  and

$$\begin{aligned}\rho &:= p\rho_1 + (1-p)\rho_2, \\ \sigma &:= p\sigma_1 + (1-p)\sigma_2,\end{aligned}$$

the relative entropy between the states given by the convex combinations takes the value

$$\begin{aligned}D(\rho\|\sigma) &= \text{Tr}[\rho \log(\rho) - \rho \log(\sigma)] \\ &= -h(p) - p \log(pt + (1-p)(1-t)) - (1-p) \log((1-p)t + p(1-t)),\end{aligned}$$

and

$$\begin{aligned}D(\rho_1\|\sigma_1) &= -\log(t), \\ D(\rho_2\|\sigma_2) &= -\log(t).\end{aligned}$$

This gives us

$$\begin{aligned}pD(\rho_1\|\sigma_1) + (1-p)D(\rho_2\|\sigma_2) - D(\rho\|\sigma) \\ = h(p) + p \log\left(p + (1-p)\frac{1-t}{t}\right) + (1-p) \log\left((1-p) + p\frac{1-t}{t}\right).\end{aligned}\tag{15}$$

As  $[\rho_i, \sigma_j] = 0$  for  $i, j = 1, 2$  and further  $[\rho_i \sigma_j, \sigma_i] = 0$  we find that the constants in theorem 5.1 are given by

$$c_i = \text{Tr}[\rho_i \sigma_i^{-1} \sigma_j] = \frac{1-t}{t},$$

for  $i, j = 1, 2, i \neq j$ . Since  $\rho_1$  and  $\rho_2$  orthogonal we get  $\frac{1}{2}\|\rho_1 - \rho_2\|_1 = 1$ . We hence obtain the RHS of eq. (15) from the almost concavity estimate in eq. (10). This concludes the claim.  $\square$

## 5.2 Reduction to almost tight previously-known continuity bounds for the relative entropy

In this section, we will show that a number of almost tight previously-known continuity bounds for quantities derived from the relative entropy can be obtained as corollaries of the results of almost concavity in theorem 5.1 and proposition 5.2 in combination with the results concerning the ALAFF method, i.e. theorem 4.6 and remark 4.7.

### 5.2.1 Uniform continuity for the conditional entropy

Let us first consider a bipartite space and the conditional entropy of a density matrix with respect to one of the subsystems. Note that, in this case, we are able to prove a result of uniform continuity for any positive semidefinite matrix (with trace one), but we do not require positive definiteness. The following coincides with the result of Winter [88], which he proved to be almost tight.

#### Corollary 5.5 (Uniform continuity of the conditional entropy)

The conditional entropy over the bipartite Hilbert space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}(\mathcal{H})$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ , it holds that

$$|H_\rho(A|B) - H_\sigma(A|B)| \leq 2\varepsilon \log d_A + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

*Proof.* First of all,  $\mathcal{S}_0$  is clearly 0-perturbed  $\Delta$ -invariant. Setting  $f(\cdot) = H(\cdot|B)$ , we find that it is ALAFF with  $a_{H(\cdot|B)} = 0$  as  $H(\cdot|B)$  is concave, and  $b_{H(\cdot|B)} = h$  since the result in theorem 5.1 becomes independent of the states as we go to  $H(\cdot|B)$  using point 3 of proposition 5.2. Finally, we find that

$$C_f^\perp = \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \text{Tr}[\rho\sigma] = 0}} |H_\rho(A|B) - H_\sigma(A|B)| \leq 2 \log d_A,$$

where we used  $-\log d_X \leq H(X|Y) \leq \log d_X$  shown, for example, in [87]. Using theorem 4.6 in the form of remark 4.7, we can infer the claimed continuity bound.  $\square$

### 5.2.2 Uniform continuity for the mutual information

For the mutual information, it is straightforward to derive a continuity bound for such a quantity just by combining the bounds of [4] and [88] for the von Neumann and conditional entropy, respectively:

$$|I_\rho(A : B) - I_\sigma(A : B)| \leq 3\varepsilon \log \min\{d_A, d_B\} + 2(1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right),$$

where  $\varepsilon := \frac{1}{2} \|\rho - \sigma\|_1$ . For an early version, see [41, Exercise 5.40]. The multiplicative factor in the first term of the right-hand side was subsequently improved to  $2\sqrt{2}$  in [70] and to 2 in [72]. Moreover, we can adapt corollary 5.5 to obtain the following bound on the mutual information, which coincides with the tightest previously-known continuity bound for the mutual information (see e.g. [72]).

#### Corollary 5.6 (Continuity bound for the mutual information)

The mutual information on a bipartite Hilbert space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}(\mathcal{H})$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ , we find that

$$|I_\rho(A : B) - I_\sigma(A : B)| \leq 2\varepsilon \log \min\{d_A, d_B\} + 2(1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$



*Proof.* First of all,  $\mathcal{S}_0$  is clearly 0-perturbed  $\Delta$ -invariant. With  $f(\cdot) = I.(A : B) = S(\cdot_A) - H.(A|B)$  one can immediately conclude almost local affinity of  $I.(A : B)$  as  $S(\cdot_A)$  is concave and fulfills eq. (11) and  $-H.(A|B)$  is almost locally affine with  $a_{-H.(A|B)} = 0$  and  $b_{-H.(A|B)} = h$ . Combined we get  $a_f = h$  and  $b_f = h$ . We further have that

$$C_f^\perp = \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \text{Tr}[\rho\sigma]=0}} |I_\rho(A : B) - I_\sigma(A : B)| \leq \sup_{\rho \in \mathcal{S}_0} I_\rho(A : B) \leq 2 \log \min\{d_A, d_B\},$$

where we used that  $0 \leq I.(A : B)$  and  $I.(A : B) \leq 2 \log \min\{d_A, d_B\}$  [87]. Applying theorem 4.6 in the form of remark 4.7, we can conclude the claim and obtain the given continuity bound.  $\square$

### 5.2.3 Uniform continuity for the conditional mutual information

We can also provide a continuity bound for the conditional mutual information of two tripartite states  $\rho_{ABC}, \sigma_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$  from corollary 5.5, by viewing it as the difference between two conditional entropies. The following result coincides with the best previously-known bound for the named quantity and appeared explicitly in [71, Lemma 4], and with a worsening of a factor 2 previously in [77] and [41, Exercise 5.41].

#### Corollary 5.7 (Uniform continuity of the conditional mutual information)

*The conditional mutual information with respect to  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  is uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}(\mathcal{H})$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ , we find that*

$$|I_\rho(A : B|C) - I_\sigma(A : B|C)| \leq 2\varepsilon \log \min\{d_A, d_B\} + 2(1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

*Proof.* The procedure is now familiar. We first note that  $\mathcal{S}_0$  is 0-perturbed  $\Delta$ -invariant. Without loss of generality, we can assume that  $d_A \leq d_B$  and rewrite  $f(\cdot) = I.(A : B|C) = H.(A|C) - H.(A|BC)$ . With this representation, we can immediately conclude that  $I.(A : B|C)$  is ALAFF with  $a_f = h$  and  $b_f = h$ . Finally, we have that

$$\begin{aligned} C_f^\perp &= \sup_{\substack{\rho, \sigma \in \mathcal{S}_0 \\ \text{Tr}[\rho\sigma]=0}} |I_\rho(A : B|C) - I_\sigma(A : B|C)| \\ &\leq \sup_{\rho \in \mathcal{S}_0} I_\rho(A : B|C) \\ &= \sup_{\rho \in \mathcal{S}_0} H_\rho(A|BC) - H_\rho(A|C) \\ &\leq 2 \log d_A = 2 \log \min\{d_A, d_B\}, \end{aligned}$$

as the conditional mutual information is non-negative and again  $-\log d_X \leq H.(X|Y) \leq \log d_X$ . Using theorem 4.6 in the form of remark 4.7, we can conclude the claim and obtain the given continuity bound.  $\square$

### 5.3 New continuity bounds for the relative entropy

Now, we prove some new continuity bounds for further quantities derived from the relative entropy as a consequence of the results of almost concavity in theorem 5.1 and proposition 5.2 in combination with the results concerning the ALAFF method, i.e. theorem 4.6 and remark 4.7. All bounds in this section can be simplified using the following lemma:

**Lemma 5.8** *Using the notations introduced in theorem 5.1 and remark 5.3, we have the following estimates for the error bounds obtained in all results of this section:*

$$(1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right) \leq \sqrt{2\varepsilon}, \quad (16)$$

$$\frac{l_{\tilde{m}} + \varepsilon}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{\varepsilon}{l_{\tilde{m}} + \varepsilon}\right) \leq \frac{\varepsilon}{l_{\tilde{m}}} \log \tilde{m}^{-1} + \log\left(1 + \frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} \frac{1}{\tilde{m}}\right). \quad (17)$$

*Proof.* The first inequality appeared in [75] before and its proof follows from some elementary calculus. For the second inequality note that  $\varepsilon \in [0, 1]$  and  $\tilde{m} \in (0, 1)$ , it holds that  $l_{\tilde{m}} = 1 - \tilde{m} \in (0, 1)$ , allowing us to estimate  $1 \leq \frac{1}{\tilde{m}}$  and  $\frac{l_{\tilde{m}}}{l_{\tilde{m}} + \varepsilon} \leq 1$ . This results in:

$$\begin{aligned} \frac{l_{\tilde{m}} + \varepsilon}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{\varepsilon}{l_{\tilde{m}} + \varepsilon}\right) &= \frac{\varepsilon}{l_{\tilde{m}}} \log\left(\frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} + \frac{l_{\tilde{m}}}{l_{\tilde{m}} + \varepsilon} \frac{1}{\tilde{m}}\right) + \log\left(\frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} \frac{1}{\tilde{m}} + \frac{l_{\tilde{m}}}{l_{\tilde{m}} + \varepsilon}\right) \\ &\leq \frac{\varepsilon}{l_{\tilde{m}}} \log \tilde{m}^{-1} + \log\left(1 + \frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} \frac{1}{\tilde{m}}\right). \end{aligned}$$

□

### 5.3.1 Divergence bounds for the relative entropy

In this section, we prove an upper bound on the relative entropy  $D(\rho\|\sigma)$  which involves the trace norm distance of  $\rho$  and  $\sigma$ . The literature calls these bounds upper continuity bounds [6, 67, 86], for which we would expect an upper bound of  $|D(\rho_1\|\sigma_1) - D(\rho_2\|\sigma_2)|$  in terms of the norm distance of  $\rho_1$  and  $\rho_2$ , and  $\sigma_1$  and  $\sigma_2$ , respectively. We hence propose the name ‘‘divergence bound’’ for this kind of bound, a fitting name, since we are relating the strength of divergence (between  $\rho$  and  $\sigma$ ) to a fixed distance measure (the trace norm).

We now give the divergence bound we obtain when using the convexity and almost concavity of  $D(\rho\|\sigma)$  together with theorem 4.6 by going through uniform continuity of the relative entropy in its first argument.

#### Corollary 5.9 (Uniform continuity of the relative entropy in the first argument)

*Let  $\sigma \in \mathcal{S}(\mathcal{H})$  be fixed. Then  $D(\cdot\|\sigma)$  is uniformly continuous on  $\mathcal{S}_0 = \{\rho \in \mathcal{S}(\mathcal{H}) : \ker \sigma \subseteq \ker \rho\}$  and, for  $\rho_1, \rho_2 \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq \varepsilon \leq 1$ , it holds that*

$$|D(\rho_1\|\sigma) - D(\rho_2\|\sigma)| \leq \varepsilon \log \tilde{m}_\sigma^{-1} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right), \quad (18)$$

*with  $\tilde{m}_\sigma$  the minimal non-zero eigenvalue of  $\sigma$ .*

*Proof.*  $\mathcal{S}_0$  is clearly convex and 0-perturbed  $\Delta$ -invariant as for two operators  $A, B$ ,  $\ker A \cap \ker B \subseteq \ker(A - B)$  and  $[A - B]_\pm$  are orthogonal. We set  $f(\cdot) = D(\cdot\|\sigma)$ . Using theorem 5.1 and point 1 of proposition 5.2, we find that  $D(\cdot\|\sigma)$  is ALAFF with  $a_f = h$  and  $b_f = 0$ . At last, we have that

$$C_f^\perp = \sup_{\substack{\rho_1, \rho_2 \in \mathcal{S}_0 \\ \frac{1}{2} \|\rho_1 - \rho_2\|_1 = 1}} |D(\rho_1\|\sigma) - D(\rho_2\|\sigma)| \leq \sup_{\rho \in \mathcal{S}(\mathcal{H})} D(\rho\|\sigma) \leq \log \tilde{m}_\sigma^{-1}.$$

In the first inequality, we used that  $D(\rho\|\sigma) \geq 0$ , and in the second one that  $\tilde{m}_\sigma \rho \leq \sigma$  hence  $D(\rho\|\sigma) \leq \log \tilde{m}_\sigma^{-1}$ . Using theorem 4.6 in the form of remark 4.7 concludes the claim. □

We can compare eq. (18) with the findings of [35, Eq. (43) and (44)], based on the previous [36], where it was shown that

$$|D(\rho_1\|\sigma) - D(\rho_2\|\sigma)| \leq \max_{i=1,2} \log \left( 1 + \frac{\|\rho_1 - \rho_2\|_\infty}{m_{\rho_i} m_\sigma} \right), \quad (19)$$

whenever  $\rho_i > 0$  and  $\min_{i=1,2} m_{\rho_i} > \|\rho_1 - \rho_2\|_\infty$ . Here  $m_{\rho_i}$  is the minimal eigenvalue of  $\rho_i$  for  $i = 1, 2$  and correspondingly  $m_\sigma$  the one of  $\sigma$ . This expression presents the advantage with respect to ours of depending on the operator norm of the difference of  $\rho_1$  and  $\rho_2$ , instead of the trace norm. However, when  $\rho_1 \approx \rho_2$ , the upper bound in eq. (19) can be approximated by  $\frac{\|\rho_1 - \rho_2\|_\infty}{m_{\rho_i} m_\sigma}$ , and thus the dependence with  $m_\sigma^{-1}$  is linear, instead of logarithmic as in eq. (18). Further in eq. (19) one needs  $\rho_1$  and  $\rho_2$  to be full rank and has a condition on their minimal eigenvalues.

We can subsequently use the corollary 5.9 to prove a divergence bound for the relative entropy.

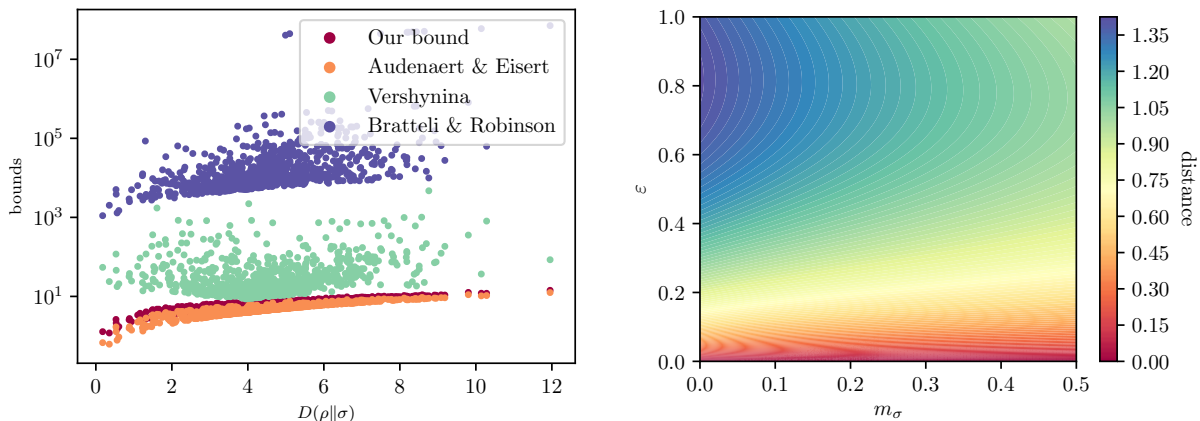
**Corollary 5.10 (Divergence bound for the relative entropy)**

Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with  $\ker \sigma \subseteq \ker \rho$  and  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ , we have

$$D(\rho\|\sigma) \leq \varepsilon \log \tilde{m}_\sigma^{-1} + (1 + \varepsilon) h \left( \frac{\varepsilon}{1 + \varepsilon} \right) \leq \left( 1 + \frac{\log \tilde{m}_\sigma^{-1}}{\sqrt{2}} \right) \varepsilon^{1/2}.$$

with  $\tilde{m}_\sigma$  the minimal non-zero eigenvalue of  $\sigma$ . The second inequality follows from (16) and the fact that  $\varepsilon \leq \sqrt{\varepsilon}$  for any  $\varepsilon \in [0, 1]$ .

*Proof.* In the context of corollary 5.9, we just set  $\rho_1 = \rho$  and  $\rho_2 = \sigma$ , giving us that  $\frac{1}{2} \|\rho_1 - \rho_2\|_1 = \frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ . Furthermore,  $D(\rho_2\|\sigma) = D(\sigma\|\sigma) = 0$  and  $|D(\rho_1\|\sigma)|$  loses the absolute value, as  $D(\cdot\|\cdot) \geq 0$ . The bound follows immediately.  $\square$



(a) The magnitude of the different bounds plotted over the relative entropy. We sampled thousand different pairs of qubits and controlled the minimal eigenvalue of  $\sigma$  in a range from  $10^{-4}$  to  $10^{-8}$ . The explicit bounds can be found in table 1. (b) The difference between the bound from corollary 5.10 and the one of Audenaert & Eisert [8, Theorem 1]. On the x-axis we plot the minimal eigenvalue of  $\sigma$  and on the y-axis  $\varepsilon = \frac{1}{2} \|\rho - \sigma\|_1$ . The minimal eigenvalue of  $\rho$  is set to the minimal eigenvalue of  $\sigma$ , thereby strengthening the bound of Audenaert & Eisert. Both were varied between  $10^{-20}$  and  $\frac{1}{2}$ .

Figure 3: Two plots comparing the divergence bounds from table 1.

Bound by	not full rank $\rho$	not full rank $\sigma$	Bound on $D(\rho\ \sigma)$
corollary 5.10	✓	✓	$\varepsilon \log \tilde{m}_\sigma^{-1} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1+\varepsilon}\right)$
Audenaert & Eisert [8, Theorem 1]	✓	x	$(m_\sigma + \varepsilon) \log\left(\frac{m_\sigma + \varepsilon}{m_\sigma}\right) - m_\rho \log\left(\frac{m_\rho + \varepsilon}{m_\rho}\right)$
Vershynina [86]	x	x	$2\varepsilon \lambda_\rho \frac{\log m_\rho - \log m_\sigma}{m_\rho - m_\sigma}$
Bratteli & Robinson [21]	x	x	$m_\sigma^{-1} \ \rho - \sigma\ _\infty$

Table 1: A comparison of different divergence bounds. Here  $\varepsilon = \frac{1}{2} \|\rho - \sigma\|_1$  and  $m_\cdot$  and  $\tilde{m}_\cdot$  are the minimal and the minimal non-zero eigenvalue of the quantum state in the index, respectively. Further  $\lambda_\rho$  is the maximal eigenvalue of  $\rho$ . The bound of Audenaert & Eisert in the case  $m_\rho = 0$  has to be understood as the limit  $m_\rho \rightarrow +0$ .

There exist results on divergence bounds in the literature which predate ours. In [7, 86], the authors present some linear bounds for the relative entropy in terms of the trace norm difference between those states, with some multiplicative factors depending on the eigenvalues of the states involved, whereas in [21] a similar bound is provided in terms of the operator norm of the difference between the states. One of the bounds in [7] is further generalised in [8] and is closely related to our bound as both of them are non-linear in the trace norm (resp. operator norm) difference between the involved states, and show a dependence on the inverse of the minimal eigenvalue of  $\sigma$  only logarithmically. This is partly an advantage over the bounds in [21, 86]. There further exists a bound in [37, Proposition 5.81] which has an explicit dependence on the dimension in addition to the dependence on the minimal eigenvalue of  $\sigma$  and therefore was not investigated. The bound might have an advantage in low-dimensional settings. In table 1 and fig. 3 we compare the aforementioned bounds from [8, 21, 86]. From fig. 3a it is clear that our bound, in the majority of the cases, outperforms the bound by Vershynina and the one by Bratteli & Robinson. This is because of the logarithmic scaling with the inverse minimal eigenvalue of  $\sigma$  of our bound versus the linear scaling with the inverse minimal eigenvalue of  $\sigma$  of theirs. We hence reduce the discussion to a comparison between Audenaert & Eisert's and our bound. From the first fig. 3a and second plot fig. 3b we conclude a slight advantage of theirs. The numerical experiments suggest, however, that the difference between both bounds is bounded by two, hence as the minimal eigenvalue decreases both bounds should converge asymptotically. Furthermore, our bound has the advantage that it does not need  $\sigma$  nor  $\rho$  to be full rank. This fact and its simple representation might give some advantages in applications.

### 5.3.2 Continuity bounds for the relative entropy

We conclude our section on continuity bounds with the most involved continuity bound until now. It concerns the relative entropy and regards it in all its power as a function of two variables, i.e., it constitutes a continuity bound both for the first and the second input simultaneously. This presents some challenges that need to be dealt with, as the relative entropy exhibits discontinuity whenever the kernel of the second input is not contained in that of the first one. To overcome these issues, we need to employ the ALAFF method in its full generality.

In the first step, we fix the first input of the relative entropy and provide a continuity bound for the relative entropy in the second argument.

**Corollary 5.11 (Uniform continuity of the relative entropy in the second argument)**

Let  $\rho \in \mathcal{S}(\mathcal{H})$  be fixed and  $1 > \tilde{m} > 0$ . Then,  $D(\rho\|\cdot)$  is uniformly continuous on

$$\mathcal{S}_0 := \{\sigma \in \mathcal{S}(\mathcal{H}) : \tilde{m}\rho \leq \sigma\}.$$

We further get that, for  $\sigma_1, \sigma_2 \in \mathcal{S}_0$  with  $\frac{1}{2}\|\sigma_1 - \sigma_2\|_1 \leq \varepsilon$ ,

$$\begin{aligned} |D(\rho\|\sigma_1) - D(\rho\|\sigma_2)| &\leq \frac{\varepsilon}{l_{\tilde{m}}} \log(\tilde{m}^{-1}) + \frac{l_{\tilde{m}} + \varepsilon}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}} \left( \frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} \right) \\ &\leq 2 \frac{\varepsilon}{l_{\tilde{m}}} \log \tilde{m}^{-1} + \log \left( 1 + \frac{\varepsilon}{l_{\tilde{m}} + \varepsilon} \frac{1}{\tilde{m}} \right), \end{aligned} \quad (20)$$

where  $l_{\tilde{m}} = 1 - \tilde{m}$ . The second inequality follows from (17).

*Proof.* We have that  $\mathcal{S}_0$  is clearly convex as, for  $\sigma_1, \sigma_2 \in \mathcal{S}_0$  and  $\lambda \in [0, 1]$ ,

$$\lambda\sigma_1 + (1 - \lambda)\sigma_2 \geq \lambda\tilde{m}\rho + (1 - \lambda)\tilde{m}\rho = \tilde{m}\rho,$$

giving the kernel inclusion as well as the condition for the smallest eigenvalue on the support of  $\rho$ . Furthermore,  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant with  $s = \tilde{m}$ . This is because one can perturb with  $\tau = \rho$  and get subminorization by  $\tilde{m}\rho$ . Employing point 2 of proposition 5.2 we further find that  $f(\cdot) = D(\rho\|\cdot)$  satisfies eq. (8) with  $b_f = 0$  and  $a_f = f_{\tilde{m}^{-1}, \tilde{m}^{-1}}$ , hence  $E_f = f_{\tilde{m}^{-1}, \tilde{m}^{-1}}$ . Using again proposition 5.2 (point 4, since  $\tilde{m} \leq 1$ ) we conclude  $E_f^{\max} = f_{\tilde{m}^{-1}, \tilde{m}^{-1}}$ . At last, we have that

$$\begin{aligned} C_f^{\tilde{m}} &= \sup_{\substack{\sigma_1, \sigma_2 \in \mathcal{S}_0 \\ \frac{1}{2}\|\sigma_1 - \sigma_2\|_1 = 1 - \tilde{m}}} |D(\rho\|\sigma_1) - D(\rho\|\sigma_2)| \\ &\leq \sup_{\sigma \in \mathcal{S}_0} D(\rho\|\sigma) \\ &\leq \log(\tilde{m}^{-1}), \end{aligned}$$

where we used that  $D(\rho\|\cdot) \geq 0$  and for the last inequality that  $\tilde{m}\rho \leq \sigma$  for all  $\sigma \in \mathcal{S}_0$ . Employing now theorem 4.6 we obtain uniform continuity and the claimed continuity bound.  $\square$

As in the case of the continuity bound for the relative entropy in the first input, we can compare eq. (20) with [35, Eq. (39) and (40)], as an extension of the previous [36], in which it was shown that

$$|D(\rho\|\sigma_1) - D(\rho\|\sigma_2)| \leq \max_{i=1,2} -\log \left( 1 - \frac{\|\sigma_1 - \sigma_2\|_\infty}{m_{\sigma_i}} \right), \quad (21)$$

whenever  $\min_{i=1,2} m_{\sigma_i} > \|\sigma_1 - \sigma_2\|_\infty$ , where  $m_{\sigma_i}$  is the minimal eigenvalue of  $\sigma_i$  for  $i = 1, 2$ . In the low  $\varepsilon$  regime the bound in eq. (21) as well as the bound in eq. (20) scale linearly in  $m_\sigma^{-1}$ .

In the above corollary, two choices need some more justification. The first choice is  $1 > \tilde{m}$  and the second one is  $s = \tilde{m}$ . We want to put them into context by the following lemma, demonstrating that these assumptions are necessary to obtain a non-trivial  $\mathcal{S}_0$ .

**Lemma 5.12**

Let  $\rho \in \mathcal{S}(\mathcal{H})$  and  $s \in [0, 1)$  with  $\text{rank } \rho \geq 2$ , further  $\tilde{m} \in (0, \infty)$  and

$$\mathcal{S}_0 := \{\sigma \in \mathcal{S}(\mathcal{H}) : \ker \sigma \subseteq \ker \rho, \tilde{m}\rho \leq \sigma\}.$$

Then, the following is true:

1. If  $1 > \tilde{m}$ , then  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant if and only if  $s \geq \tilde{m}$ .
2. If  $1 = \tilde{m}$ , then  $\mathcal{S}_0 = \{\rho\}$ .
3. If  $1 < \tilde{m}$ ,  $\mathcal{S}_0 = \emptyset$ .

We will only give proof for the first one in appendix D and leave the last two for the reader. Next, we proceed to state and prove the main result of this subsection on continuity bounds, namely the uniform continuity bound for the relative entropy in both arguments on a suitable subspace. Since we have already explored the cases in which we either fix the second (corollary 5.9) or first (corollary 5.11) density operator, we now combine both results in the proof of the next theorem.

**Theorem 5.13 (Uniform continuity of the relative entropy)**

Let  $1 > 2\tilde{m} > 0$  and

$$\mathcal{S}_0 = \{(\rho, \sigma) : \rho, \sigma \in \mathcal{S}(\mathcal{H}), \ker \sigma \subseteq \ker \rho, 2\tilde{m} \leq \tilde{m}_\sigma\},$$

with  $\tilde{m}_\sigma$  the minimal non-zero eigenvalue of  $\sigma$ . Then,  $D(\cdot\|\cdot)$  is uniformly continuous on  $\mathcal{S}_0$  and we find that for  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_0$  with  $\frac{1}{2}\|\rho_1 - \rho_2\| \leq \varepsilon \leq 1$  and  $\frac{1}{2}\|\sigma_1 - \sigma_2\|_1 \leq \delta \leq 1$

$$\begin{aligned} |D(\rho_1\|\sigma_1) - D(\rho_2\|\sigma_2)| &\leq \left(\varepsilon + \frac{\delta}{l_{\tilde{m}}}\right) \log(\tilde{m}^{-1}) + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right) + 2\frac{l_{\tilde{m}} + \delta}{l_{\tilde{m}}}f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{\delta}{l_{\tilde{m}} + \delta}\right) \\ &\leq \left(\sqrt{2} + \log \tilde{m}^{-1}\right)\varepsilon^{1/2} + 3\frac{\delta}{l_{\tilde{m}}}\log \tilde{m}^{-1} + 2\log\left(1 + \frac{\delta}{l_{\tilde{m}} + \delta}\frac{1}{\tilde{m}}\right), \end{aligned} \quad (22)$$

with  $l_{\tilde{m}} = 1 - \tilde{m}$ . The second inequality follows from (17) and  $\varepsilon \leq \sqrt{\varepsilon}$  for  $\varepsilon \in [0, 1]$ .

*Proof.* We will prove the uniform continuity by proving that the bound eq. (22) holds. Therefore, let  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_0$  with  $\frac{1}{2}\|\rho_1 - \rho_2\| \leq \varepsilon \leq 1$  and  $\frac{1}{2}\|\sigma_1 - \sigma_2\| \leq \delta \leq 1$ . We define

$$\bar{\sigma} = \frac{1}{2}\sigma_1 + \frac{1}{2}\sigma_2, \quad (23)$$

and obtain

$$\begin{aligned} \frac{1}{2}\|\bar{\sigma} - \sigma_1\|_1 &= \frac{1}{4}\|\sigma_1 - \sigma_2\|_1 \leq \frac{\delta}{2} \leq 1, \\ \frac{1}{2}\|\bar{\sigma} - \sigma_2\|_1 &= \frac{1}{4}\|\sigma_1 - \sigma_2\|_1 \leq \frac{\delta}{2} \leq 1. \end{aligned}$$

Using this, we get

$$|D(\rho_1\|\sigma_1) - D(\rho_2\|\sigma_2)| \leq |D(\rho_1\|\sigma_1) - D(\rho_1\|\bar{\sigma})| + |D(\rho_1\|\bar{\sigma}) - D(\rho_2\|\bar{\sigma})| + |D(\rho_2\|\bar{\sigma}) - D(\rho_2\|\sigma_2)|.$$

The middle term can be bounded using corollary 5.9 and the fact that

$$\log \tilde{m}_{\bar{\sigma}}^{-1} \leq \log(2 \max\{\tilde{m}_{\sigma_1}^{-1}, \tilde{m}_{\sigma_2}^{-1}\}) \leq \log \tilde{m}^{-1}.$$

One obtains

$$|D(\rho_1\|\bar{\sigma}) - D(\rho_2\|\bar{\sigma})| \leq \varepsilon \log \tilde{m}^{-1} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

The other two terms are bound using corollary 5.11 and the fact that  $\tilde{m}\rho_1 \leq \frac{1}{2}\sigma_1 \leq \bar{\sigma}$  and  $\tilde{m}\rho_2 \leq \frac{1}{2}\sigma_2 \leq \bar{\sigma}$  by construction of  $\mathcal{S}_0$  and the definition of  $\bar{\sigma}$ , respectively. We therefore obtain

$$\begin{aligned} |D(\rho_1\|\sigma_1) - D(\rho_1\|\bar{\sigma})| &\leq \frac{\delta}{2l_{\tilde{m}}}\log(\tilde{m}^{-1}) + \frac{l_{\tilde{m}} + 2^{-1}\delta}{l_{\tilde{m}}}f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{2^{-1}\delta}{l_{\tilde{m}} + 2^{-1}\delta}\right), \\ |D(\rho_2\|\bar{\sigma}) - D(\rho_2\|\sigma_2)| &\leq \frac{\delta}{2l_{\tilde{m}}}\log(\tilde{m}^{-1}) + \frac{l_{\tilde{m}} + 2^{-1}\delta}{l_{\tilde{m}}}f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{2^{-1}\delta}{l_{\tilde{m}} + 2^{-1}\delta}\right). \end{aligned}$$

Combining the bounds and further using that

$$\frac{l_{\tilde{m}} + 2^{-1}\delta}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{2^{-1}\delta}{l_{\tilde{m}} + 2^{-1}\delta}\right) \leq \frac{l_{\tilde{m}} + \delta}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{\delta}{l_{\tilde{m}} + \delta}\right),$$

we obtain the claimed bound, and thereby also uniform continuity.  $\square$

Let us conclude this section by emphasizing that there might be some room for improvement in the previous result. For instance, it should be possible to improve the interpolation between  $\sigma_1$  and  $\sigma_2$  considered in eq. (23) by optimizing over the interpolation parameter instead of setting it to 1/2. However, we believe this would not change the appearance of the bound drastically and thus the reason for not performing this optimization.

## 6 Almost concavity and continuity bounds for the Belavkin-Staszewski entropy

Following the same lines as in the previous section, now we apply the ALAFF method introduced in section 4 for the particular case of the BS-entropy. For that, we need to prove a result of almost concavity for the BS-entropy, which we do in section 6.1. However, in contrast to the case of the relative entropy, our result for the BS-entropy is not tight. We leave the discussion on the almost concavity bound and the difficulties that appear in the BS-entropy case to the next subsection.

Subsequently, we combine our result of almost concavity for the BS-entropy with the ALAFF method to provide certain results of uniform continuity and explicit continuity bounds for entropic quantities derived from the BS-entropy in section 6.2. All the results provided in this section are summarized in fig. 4.

### 6.1 Almost concavity for the BS-entropy

In this section we prove the almost concavity of the BS-entropy and thereby complement the established result of convexity [56, Theorem 4.4], [44, Corollary 4.7]. We first want to give some auxiliary results that will be needed later. The first of those concerns an operator inequality for the term inside the trace in the definition of the BS-entropy.

#### Lemma 6.1

Let  $A_1, A_2 \in \mathcal{B}(\mathcal{H})$  positive semi-definite,  $p \in [0, 1]$  and

$$A = pA_1 + (1 - p)A_2.$$

Then

$$-A \log(A) \leq -pA_1 \log(A_1) - (1 - p)A_2 \log(A_2) + h_{A_1, A_2}(p) \mathbb{1}$$

with  $h_{A_1, A_2}(p) = -p \log(p) \text{Tr}[A_1] - (1 - p) \log(1 - p) \text{Tr}[A_2]$  a distorted binary entropy.

*Proof.* It holds that

$$\begin{aligned} -A \log(A) + pA_1 \log(A_1) + (1 - p)A_2 \log(A_2) \\ \leq \|-A \log(A) + pA_1 \log(A_1) + (1 - p)A_2 \log(A_2)\|_1 \mathbb{1} . \end{aligned} \tag{24}$$

Now, since  $x \mapsto -x \log(x)$  is operator concave [25, Theorem 2.6], we have

$$-A \log(A) \geq -pA_1 \log(A_1) - (1 - p)A_2 \log(A_2),$$

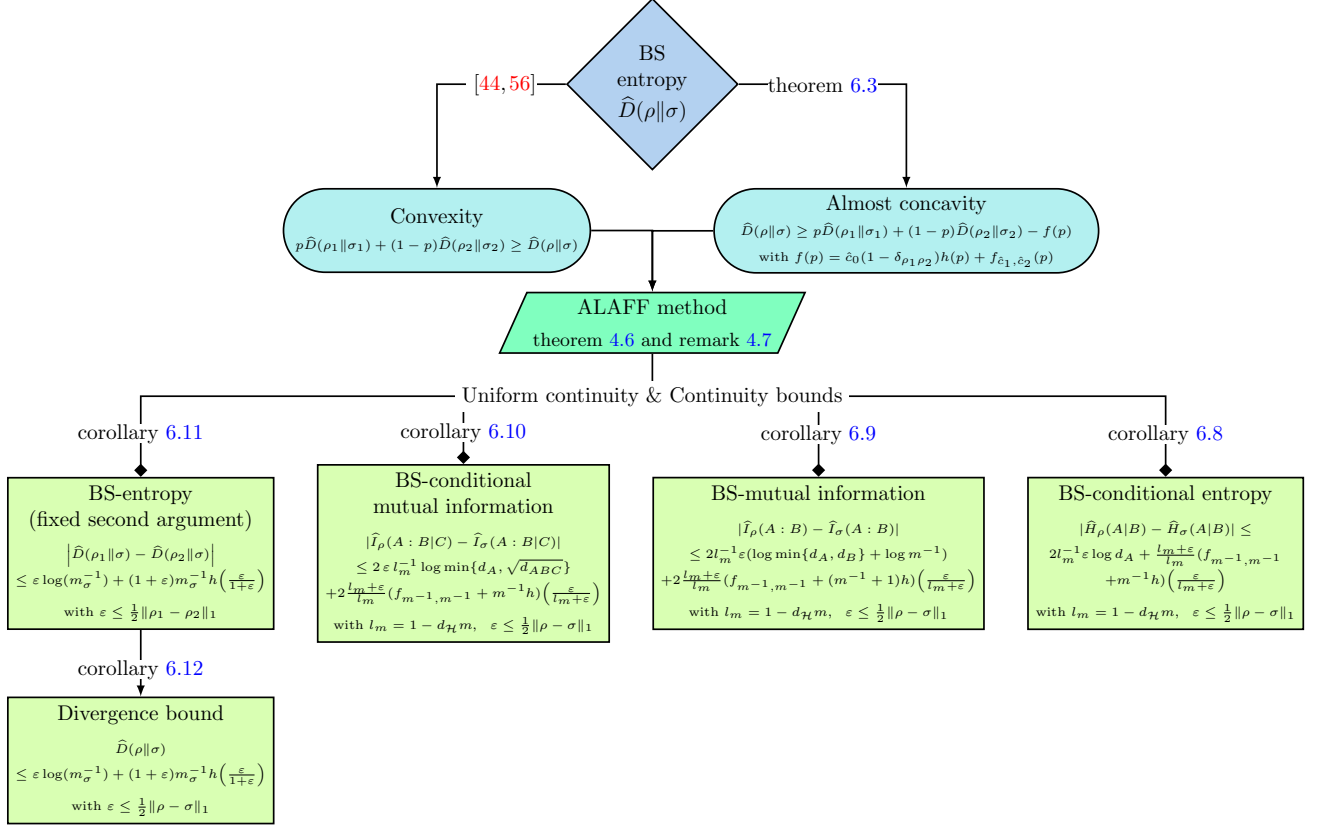


Figure 4: In this flow chart we collect the main results from this section, starting with the almost concavity for the BS-entropy, which together with the ALAFF method outputs a plethora of continuity bounds for related entropic quantities. For the convexity and almost concavity of the BS-entropy we are setting  $\rho = p\rho_1 + (1-p)\rho_2$  and  $\sigma = p\sigma_1 + (1-p)\sigma_2$ , with  $p \in [0, 1]$ . We denote by  $m_\sigma$  the minimal eigenvalue of  $\sigma$ . In the almost concavity bound,  $\hat{c}_0$  is the maximum of  $\|\sigma_1^{-1}\|_\infty$  and  $\|\sigma_2^{-1}\|_\infty$ . Additionally, we assume in all the continuity bounds that  $m \leq \|\eta^{-1}\|_\infty$ , for  $\eta = \sigma, \rho$ .

giving us that

$$-A \log(A) + pA_1 \log(A_1) + (1-p)A_2 \log(A_2) \geq 0,$$

and hence

$$\begin{aligned} & \| -A \log(A) + pA_1 \log(A_1) + (1-p)A_2 \log(A_2) \|_1 \\ &= \text{Tr} [ -A \log(A) + pA_1 \log(A_1) + (1-p)A_2 \log(A_2) ]. \end{aligned} \quad (25)$$

We now use the operator monotonicity of the logarithm to find

$$\begin{aligned} -\text{Tr} [A \log(A)] &= -p\text{Tr} [A_1 \log(A)] - (1-p)\text{Tr} [A_2 \log(A)] \\ &\leq -p\text{Tr} [A_1 \log(pA_1)] - (1-p)\text{Tr} [A_2 \log((1-p)A_2)] \\ &= -p\text{Tr} [A_1 \log(A_1)] - (1-p)\text{Tr} [A_2 \log(A_2)] + h_{A_1, A_2}(p) \end{aligned}$$

Inserting this into eq. (25) and then into eq. (24) yields the claimed result.  $\square$

The next auxiliary result concerns an equivalent formulation for the BS-entropy constructed from the function  $x \mapsto x \log x$  and has already appeared in the literature (see e.g. [62, Eq. (7.35)]). We include here a short proof of this result for completeness.



**Lemma 6.2**

Let  $\rho \in \mathcal{S}(\mathcal{H})$  and  $\sigma \in \mathcal{S}_+(\mathcal{H})$ , then

$$\widehat{D}(\rho\|\sigma) = \text{Tr} \left[ \sigma(\sigma^{-1/2}\rho\sigma^{-1/2}) \log(\sigma^{-1/2}\rho\sigma^{-1/2}) \right].$$

*Proof.* Slightly misusing notation, we can replace the regular log with a log that evaluates to 0 at 0 and thereby artificially add 0 to the domain. This changes neither the RHS nor the LHS but allows us to derive

$$\begin{aligned} \widehat{D}(\rho\|\sigma) &= \text{Tr} \left[ \rho \log(\rho^{1/2}\sigma^{-1}\rho^{1/2}) \right] = \text{Tr} \left[ \log(\rho^{1/2}\sigma^{-1/2}\sigma^{-1/2}\rho^{1/2})\rho^{1/2}\sigma^{-1/2}\sigma^{1/2}\rho^{1/2} \right] \\ &= \text{Tr} \left[ \rho^{1/2}\sigma^{-1/2} \log(\sigma^{-1/2}\rho\sigma^{-1/2})\sigma^{1/2}\rho^{1/2} \right] \\ &= \text{Tr} \left[ \sigma(\sigma^{-1/2}\rho\sigma^{-1/2}) \log(\sigma^{-1/2}\rho\sigma^{-1/2}) \right]. \end{aligned}$$

We used the cyclicity of the trace several times, and the well-known fact that we have  $f(L^*L)L^* = L^*f(LL^*)$  in case the spectrum of  $L^*L$  and  $LL^*$  lie in the domain of  $f$  [47, Lemma 61].  $\square$

Building on the previous results from this section, we proceed to prove now the main result, namely the almost concavity for the BS-entropy. This falls in the line of results of almost concavity discussed in definition 4.1.

**Theorem 6.3 (Almost concavity of the BS-entropy)**

Let  $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_{\text{ker},+}$  with

$$\mathcal{S}_{\text{ker},+} := \{(\rho, \sigma) \in \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H}) : \sigma \in \mathcal{S}_+(\mathcal{H})\}$$

and  $p \in [0, 1]$ . Then, for  $\rho = p\rho_1 + (1-p)\rho_2$ ,  $\sigma = p\sigma_1 + (1-p)\sigma_2$ , we have

$$\widehat{D}(\rho\|\sigma) \geq p\widehat{D}(\rho_1\|\sigma_1) + (1-p)\widehat{D}(\rho_2\|\sigma_2) - \hat{c}_0(1 - \delta_{\rho_1\rho_2})h(p) - f_{\hat{c}_1, \hat{c}_2}(p),$$

with

$$\begin{aligned} h(p) &= -p \log(p) - (1-p) \log(1-p), \\ f_{\hat{c}_1, \hat{c}_2}(p) &= p \log(p + \hat{c}_1(1-p)) + (1-p) \log((1-p) + \hat{c}_2p), \\ \delta_{\rho_1\rho_2} &= \begin{cases} 1 & \text{if } \rho_1 = \rho_2 \\ 0 & \text{otherwise} \end{cases}, \end{aligned}$$

and the constants

$$\begin{aligned} \hat{c}_0 &:= \max\{\|\sigma_1^{-1}\|_\infty, \|\sigma_2^{-1}\|_\infty\}, \\ \hat{c}_1 &:= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_1(\rho_1^{1/2}\sigma_1^{-1}\rho_1^{1/2})^{\frac{it+1}{2}} \rho_1^{-1/2}\sigma_2\rho_1^{-1/2}(\rho_1^{1/2}\sigma_1^{-1}\rho_1^{1/2})^{\frac{-it+1}{2}} \right], \\ \hat{c}_2 &:= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_2(\rho_2^{1/2}\sigma_2^{-1}\rho_2^{1/2})^{\frac{it+1}{2}} \rho_2^{-1/2}\sigma_1\rho_2^{-1/2}(\rho_2^{1/2}\sigma_2^{-1}\rho_2^{1/2})^{\frac{-it+1}{2}} \right], \end{aligned} \tag{26}$$

with the probability density  $\beta_0$  defined as in eq. (13).

*Proof.* The formula for  $p = 0, 1$  is trivial, hence let  $p \in (0, 1)$ . We find that

$$\begin{aligned} p\widehat{D}(\rho_1\|\sigma_1) + (1-p)\widehat{D}(\rho_2\|\sigma_2) - \widehat{D}(\rho\|\sigma) \\ \leq p(\widehat{D}(\rho_1\|\sigma_1) - \widehat{D}(\rho_1\|\sigma)) + (1-p)(\widehat{D}(\rho_2\|\sigma_2) - \widehat{D}(\rho_2\|\sigma)) + \hat{c}_0 h(p). \end{aligned}$$

Indeed, as of lemma 6.2 and then lemma 6.1 with  $A_1 = \sigma^{-1/2}\rho_1\sigma^{-1/2}$ ,  $A_2 = \sigma^{-1/2}\rho_2\sigma^{-1/2}$  respectively, we can prove

$$\begin{aligned} -\widehat{D}(\rho\|\sigma) &= \text{Tr} \left[ \sigma \left( -\sigma^{-1/2}\rho\sigma^{-1/2} \log(\sigma^{-1/2}\rho\sigma^{-1/2}) \right) \right] \\ &\leq p \text{Tr} \left[ \sigma \left( -\sigma^{-1/2}\rho_1\sigma^{-1/2} \log(\sigma^{-1/2}\rho_1\sigma^{-1/2}) \right) \right] \\ &\quad + (1-p) \text{Tr} \left[ \sigma \left( -\sigma^{-1/2}\rho_2\sigma^{-1/2} \log(\sigma^{-1/2}\rho_2\sigma^{-1/2}) \right) \right] + h_{A_1, A_2}(p) \\ &= -p\widehat{D}(\rho_1\|\sigma) - (1-p)\widehat{D}(\rho_2\|\sigma) + h_{A_1, A_2}(p). \end{aligned}$$

At last we can estimate  $\text{Tr}[A_j] = \text{Tr}[\sigma^{-1}\rho_j] \leq \|\sigma^{-1}\|_\infty \leq \hat{c}_0$  for  $j = 1, 2$  using Hölder's inequality, giving us  $h_{A_1, A_2}(p) \leq \hat{c}_0 h(p)$ .

We now have to estimate terms of the form  $\widehat{D}(\rho_j\|\sigma_j) - \widehat{D}(\rho_j\|\sigma)$  for  $j = 1, 2$ . This is done using the Peierls-Bogoliubov inequality [62] and the multivariate trace inequalities of Sutter et al. [76]:

$$\begin{aligned} \widehat{D}(\rho_j\|\sigma_j) - \widehat{D}(\rho_j\|\sigma) &= \text{Tr} \left[ \rho_j \left( \log(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}) - \log(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}) \right) \right] \\ &\leq \text{Tr} \left[ \exp \left( \log(\rho_j) + \log(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}) - \log(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}) \right) \right] \\ &\leq \text{Tr} \left[ \exp \left( \log(\rho_j) + \log(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}) + \log(\rho_j^{-1/2}\sigma\rho_j^{-1/2}) \right) \right] \\ &\leq \log \left( \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_j (\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2})^{\frac{it+1}{2}} \rho_j^{-1/2}\sigma\rho_j^{-1/2} (\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2})^{-\frac{it+1}{2}} \right] \right) \\ &= \begin{cases} \log(p + (1-p)\hat{c}_1) & j = 1 \\ \log((1-p) + p\hat{c}_2) & j = 2 \end{cases}. \end{aligned} \tag{27}$$

In the third line, we use that

$$-\log(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}) \leq \log(\rho_j^{-1/2}\sigma\rho_j^{-1/2})$$

which is true since for  $P_{\rho_j}$  the projection on the support of  $\rho_j$ , we have

$$P_{\rho_j}(P_{\rho_j}\sigma P_{\rho_j})^{-1}P_{\rho_j} \leq P_{\rho_j}\sigma^{-1}P_{\rho_j},$$

as  $x \rightarrow x^{-1}$  is operator convex and hence fulfills the Sherman-Davis inequality [25, Theorem 4.19]. Note that  $\sigma$  is invertible and that by  $(P_{\rho_j}\sigma P_{\rho_j})^{-1}$  we mean the Moore-Penrose pseudoinverse. We find

$$\begin{aligned} -\log(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}) &= -\log(\rho_j^{1/2}P_{\rho_j}\sigma^{-1}P_{\rho_j}\rho_j^{1/2}) \\ &\leq -\log(\rho_j^{1/2}P_{\rho_j}(P_{\rho_j}\sigma P_{\rho_j})^{-1}P_{\rho_j}\rho_j^{1/2}) \\ &= \log(\rho_j^{-1/2}P_{\rho_j}\sigma P_{\rho_j}\rho_j^{-1/2}) \\ &= \log(\rho_j^{-1/2}\sigma\rho_j^{-1/2}). \end{aligned}$$

The argument why the inequalities in eq. (27) hold in the case of  $\rho_j$  not being full rank is simpler than in the case of the corresponding inequality for the relative entropy (cf. theorem 5.1 and appendix B). For the BS-entropy, we can already restrict eq. (27) to the support of  $\rho_j$  as all operators involved,  $\rho_j$ ,  $\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2}$  and  $\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2}$ , commute with the projection onto this support. In the last step we split  $\sigma$  and evaluated the first term to  $p$  in case  $j = 1$  or the second term in case  $j = 2$  to  $(1 - p)$  and left the other one untouched, respectively. This concludes the proof.  $\square$

**Remark 6.4** *We strongly suspect that theorem 6.3 can be improved because of two reasons. The first one is that we would expect the results of almost concavity of the relative and the BS-entropy to coincide in the case that the involved states commute. The reason is that in this case, both quantities reduce to the classical relative entropy. A straightforward calculation shows that then  $\hat{c}_1 = c_1$  and  $\hat{c}_2 = c_2$ , hence  $f_{c_1, c_2} = f_{\hat{c}_1, \hat{c}_2}$ , but  $h \leq \hat{c}_0 h$  with equality if, and only if,  $\sigma_1$  and  $\sigma_2$  are pure, which in addition to  $\sigma_1$  and  $\sigma_2$  being full rank means  $\mathcal{H} = \mathbb{C}$ .*

*The other reason is given by the continuity bound we obtain for the BS-conditional entropy in corollary 6.8. Numerics suggest an almost convex bound that is independent of the minimal eigenvalue (cf. fig. 5) if the inputs are full rank<sup>2</sup>. Hence we would also suspect that an optimal almost concave remainder of the BS-entropy reduces to an eigenvalue independent bound in the case of the BS-conditional entropy.*

Analogous to the case of the relative entropy we provide an additional proposition to give context to the above result, i.e. to provide simpler expressions if the involved states satisfy specific conditions.

**Proposition 6.5 (Almost concavity estimate of the BS-entropy is well behaved)**

*The function  $\hat{c}_0 h + f_{\hat{c}_1, \hat{c}_2}$  obtained in theorem 6.3 is well behaved in the following sense: Let  $j = 1, 2$  and  $(\rho_j, \sigma_j) \in \mathcal{S}_{\ker, +}$ . We have the following:*

1. *If  $\sigma_1 = \sigma_2$ , then  $\hat{c}_j = 1$ , resulting in  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h = \hat{c}_0 h$ .*
2. *If the  $\sigma_j$  have a minimal eigenvalue that is bounded from below by  $m > 0$  respectively, then  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h \leq f_{m^{-1}, m^{-1}} + m^{-1} h$ .*
3. *If  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is a bipartite space,  $\rho_j$  has a minimal eigenvalue bounded from below by  $m > 0$ , and further  $\sigma_j = d_A^{-1} \mathbb{1}_A \otimes \rho_{j, B}$ , then  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h \leq f_{m^{-1}, m^{-1}} + m^{-1} h$ .*
4. *We find for  $m_1, m_2 \geq 1$ ,  $p \mapsto \frac{1}{1-p} f_{m_1, m_2}(p)$  and  $p \mapsto \frac{1}{1-p} \hat{c}_0 h(p)$  are non-decreasing on  $[0, 1)$ .*

This result should be compared to proposition 5.2, its analogue for the relative entropy. The proof can be found in appendix E. We will use the reductions from proposition 6.5 to simplify the terms in theorem 6.3 for the various applications presented in the subsequent section.

## 6.2 Continuity bounds for the BS-entropy

In this section, we will use the almost concavity for the BS-entropy from theorem 6.3 together with the ALAFF method in its full generality.

Before we dive into the continuity and divergence bounds, we want to collect some lower and upper estimates of entropic quantities derived from the BS-entropy (see section 3.1 for the specific definitions).

---

<sup>2</sup>The full rank requirement is necessary, as we will show in proposition 6.7 that the BS-conditional entropy is discontinuous in the presence of vanishing eigenvalues

**Proposition 6.6 (Bounds on BS-entropic quantities)**

For  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , we find:

1. For the BS-conditional entropy:

$$-\log \min\{d_A, d_B\} \leq \widehat{H}_\rho(A|B) \leq \log d_A. \quad (28)$$

2. For the BS-mutual information:

$$0 \leq \widehat{I}_\rho(A : B) \leq \log \min\{d_A, d_B\} + \log \min\{\|\rho_A^{-1}\|_\infty, \|\rho_B^{-1}\|_\infty\}, \quad (29)$$

with  $\cdot^{-1}$  the Moore-Penrose pseudoinverse.

3. For  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ , we find that the BS-conditional mutual information satisfies:

$$0 \leq \widehat{I}_\rho(A : B|C) \leq \log \min\{d_A^2, d_{ABC}\}.$$

The first two bounds are shown to be tight. For the third one, we expect that similar reasoning should also show its tightness.

The proof can be found in appendix F. We further want to remark that the scaling of the bound with respect to  $\|\rho_A^{-1}\|_\infty$  or  $\|\rho_B^{-1}\|_\infty$  is justified. The reasoning can be found in appendix F as well.

**6.2.1 Uniform continuity for the BS-conditional entropy**

We encounter a slight complication when it comes to the uniform continuity of the BS-conditional entropy compared to the uniform continuity of the conditional entropy that we have covered in corollary 5.5. This is because the almost concave bound of the BS-entropy depends on the minimal eigenvalue of the second argument (see eq. (26)), i.e. it has to be full rank. This means the input to the BS-conditional entropy has to be full rank as well. Although we think that the result of almost concavity for the BS-entropy can be improved, we know that there is no extension of uniform continuity nor continuity for the BS-conditional entropy to positive semi-definite states, as this quantity is not continuous on those. This is the content of the next proposition. We also refer the reader to [31, Remark 3.3] for a similar behaviour of the sharp quantum Rényi divergences.

**Proposition 6.7 (Discontinuity of the BS-conditional entropy)**

The BS-conditional entropy is discontinuous on the set of positive semi-definite operators over  $\mathcal{H}_A \otimes \mathcal{H}_B$  if  $d_A, d_B \geq 2$ .

*Proof.* Since  $d_A \geq 2$  as well as  $d_B \geq 2$ , we find orthogonal  $|i_A\rangle \in \mathcal{H}_A$ ,  $|i_B\rangle \in \mathcal{H}_B$ ,  $i = 0, 1$ . For  $\varepsilon \in (0, 1)$  we then define

$$|\varepsilon_B\rangle = \sqrt{1-\varepsilon}|0_B\rangle + \sqrt{\varepsilon}|1_B\rangle,$$

which is clearly normalised. Furthermore,

$$\begin{aligned} \rho_0 &:= \frac{1}{2}(|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|) \otimes |0_B\rangle\langle 0_B|, \\ \rho_\varepsilon &:= \frac{1}{2}|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + \frac{1}{2}|1_A\rangle\langle 1_A| \otimes |\varepsilon_B\rangle\langle \varepsilon_B|, \end{aligned}$$

The above are states and fulfil

$$\begin{aligned} \|\rho_0 - \rho_\varepsilon\|_1 &= \frac{1}{2} \left\| |1_A\rangle\langle 1_A| \otimes (|0_B\rangle\langle 0_B| - |\varepsilon_B\rangle\langle \varepsilon_B|) \right\|_1 \\ &= \frac{1}{2} \left\| |0_B\rangle\langle 0_B| - |\varepsilon_B\rangle\langle \varepsilon_B| \right\|_1 = \sqrt{\varepsilon}. \end{aligned} \quad (30)$$

To see the last equality, we can identify the subspace spanned by  $|0_B\rangle$  and  $|1_B\rangle$  with  $\mathbb{C}^2$  and then get that

$$|0_B\rangle\langle 0_B| \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad |\varepsilon_B\rangle\langle \varepsilon_B| \rightarrow \begin{pmatrix} 1-\varepsilon & \sqrt{\varepsilon}\sqrt{1-\varepsilon} \\ \sqrt{\varepsilon}\sqrt{1-\varepsilon} & \varepsilon \end{pmatrix}. \quad (31)$$

Calculating the eigenvalues of the difference and taking the sum of their absolute value gives  $2\sqrt{\varepsilon}$  and thereby eq. (30). Since clearly  $[\rho_0, \mathbb{1} \otimes \text{Tr}_A[\rho_0]] = 0$ , the BS and conditional entropy coincide and we find

$$\begin{aligned} \hat{H}_{\rho_0}(A|B) &= \text{Tr} [|0_B\rangle\langle 0_B| \log |0_B\rangle\langle 0_B|] \\ &\quad - \text{Tr} \left[ \frac{1}{2}(|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|) \otimes |0_B\rangle\langle 0_B| \log \frac{1}{2}(|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|) \otimes |0_B\rangle\langle 0_B| \right] \\ &= 0 - \log \frac{1}{2} = \log 2. \end{aligned}$$

The result for  $\rho_\varepsilon$  cannot be calculated so easily. We have that

$$\begin{aligned} \hat{H}_{\rho_\varepsilon}(A|B) &= -\frac{1}{2} \text{Tr} \left[ |0_B\rangle\langle 0_B| \log (|0_B\rangle\langle 0_B|^{1/2} (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1} |0_B\rangle\langle 0_B|^{1/2}) \right] \\ &\quad - \frac{1}{2} \text{Tr} \left[ |\varepsilon_B\rangle\langle \varepsilon_B| \log (|\varepsilon_B\rangle\langle \varepsilon_B|^{1/2} (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1} |\varepsilon_B\rangle\langle \varepsilon_B|^{1/2}) \right] \\ &= -\frac{1}{2} \log \text{Tr} [|0_B\rangle\langle 0_B| (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}] \\ &\quad - \frac{1}{2} \log \text{Tr} [|\varepsilon_B\rangle\langle \varepsilon_B| (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}], \end{aligned} \quad (32)$$

where in the first equality we used that  $|0_B\rangle\langle 0_B| |1_B\rangle\langle 1_B| = |1_B\rangle\langle 1_B| |0_B\rangle\langle 0_B| = 0$  and in the second equality that  $|\varepsilon_B\rangle\langle \varepsilon_B|$  and  $|0_B\rangle\langle 0_B|$  are rank-one projections. We find, using again the matrix representation in eq. (31), that

$$(|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1} \rightarrow \begin{pmatrix} 1 & \frac{\varepsilon-1}{\sqrt{\varepsilon}\sqrt{1-\varepsilon}} \\ \frac{\varepsilon-1}{\sqrt{\varepsilon}\sqrt{1-\varepsilon}} & \frac{2}{\varepsilon} - 1 \end{pmatrix}.$$

By forming matrix products and calculating the trace, we can immediately conclude that

$$\begin{aligned} \text{Tr} [|\varepsilon_B\rangle\langle \varepsilon_B| (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}] &= 1, \\ \text{Tr} [|0_B\rangle\langle 0_B| (|\varepsilon_B\rangle\langle \varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}] &= 1. \end{aligned}$$

If we insert this into eq. (32), we get  $\hat{H}_{\rho_\varepsilon}(A|B) = 0$ .  $\square$

This previous result shows in particular that we could only expect continuity and uniform continuity for the BS-conditional entropy on the set of full-rank states. The presence of the minimal eigenvalue of the states in the continuity bound provided below is thus not surprising.

**Corollary 6.8 (Uniform continuity of the BS-conditional entropy)**

The BS-conditional entropy over the bipartite Hilbert space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is for  $d_{\mathcal{H}}^{-1} > m > 0$  uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}_{\geq m}(\mathcal{H})$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$  it holds that

$$|\hat{H}_\rho(A|B) - \hat{H}_\sigma(A|B)| \leq 2l_m^{-1} \varepsilon \log d_A + \frac{l_m + \varepsilon}{l_m} (f_{m^{-1}, m^{-1}} + m^{-1}h) \left( \frac{\varepsilon}{l_m + \varepsilon} \right),$$

with  $l_m = 1 - d_{\mathcal{H}} m$ .

*Proof.* We find that  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant with  $s = md_{\mathcal{H}}$ . The justification of this choice is completely analogous to the reasoning in lemma 5.12 with  $\rho = d_{\mathcal{H}}^{-1} \mathbb{1}$ , i.e. the maximally mixed state. Furthermore,  $f(\cdot) = \widehat{H}(\cdot|B)$  is ALAFF with  $a_f = 0$  as  $\widehat{H}(\cdot|B)$  is concave, and  $b_f = m^{-1}h + f_{m^{-1}, m^{-1}}$  since the result in section 6.1 becomes independent of the states as we restrict to  $\widehat{H}(\cdot|B)$  using point 3 of proposition 6.5. We further find that

$$C_f^s \leq \sup_{\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})} |\widehat{H}_{\rho_1}(A|B) - \widehat{H}_{\rho_2}(A|B)| \leq 2 \log d_A,$$

using proposition 6.6. This allows us to apply theorem 4.6 where  $E_f^{\max}$  coincides with  $E_f$  as of point 4 in proposition 6.5. This concludes the claim.  $\square$

Even though a continuity bound for the BS-conditional entropy can only be proven for positive definite states, numerical simulations show us that we could expect a tighter bound on the previous proposition coinciding with that of corollary 5.5, i.e., without the dependence on the minimal eigenvalues of the states involved. One can find a visualisation of those numeric simulations that underlie the conjecture in fig. 5. The possibility of obtaining such a tighter bound is left for future work.

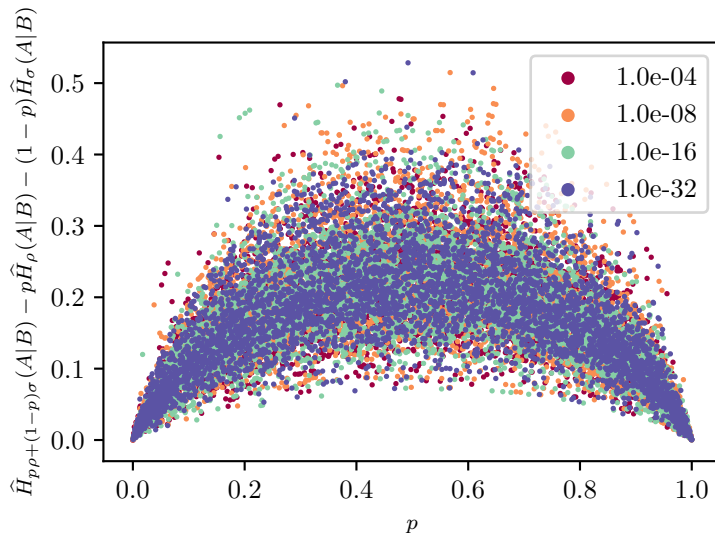


Figure 5: We investigate the dependence of the almost convex remainder term of the BS-conditional entropy on the minimal eigenvalue of the involved states. For the minimal eigenvalues  $10^{-4}, 10^{-8}, 10^{-16}, 10^{-32}$  we sampled five hundred pairs of qubits  $(\rho, \sigma)$  both of them with controlled eigenvalues. We then sampled for every state pair ten values of  $p$ , the convex interpolation parameter, and plotted the remainder. As can be seen from the plot, the remainder appears to be independent of the minimal eigenvalue and the shape suggests a binary entropy or Gini impurity. The result shows a similar pattern if the dimension is increased.

## 6.2.2 Uniform continuity for the BS-mutual information

Let us address now the case of the BS-mutual information. Since the BS-conditional entropy is a particular case of the latter (by assuming that one of the reduced states of  $\rho_{AB}$  is maximally mixed), the discontinuity issues presented in the previous subsection are expected to arise in the current

one as well. More specifically, the example of discontinuity of the BS-conditional entropy presented in proposition 6.7 also constitutes an example of discontinuity of the BS-mutual information. Thus, we can only expect to prove uniform continuity for the BS-mutual information for full-rank states

However, there is a subtle difference between the settings of the BS-conditional entropy and the BS-mutual information. As shown in proposition 6.6, the former is bounded between the same values as the (usual) conditional entropy, whereas the latter presents some pathological behaviour. Pathological in the sense that its (tight) upper bound depends on the minimal eigenvalues of the reduced state, as shown in eq. (29). For this reason, a continuity bound for the BS-mutual information will necessarily depend on the minimal eigenvalues of the states involved.

**Corollary 6.9 (Uniform continuity for the BS-mutual information)**

The BS-mutual information on a bipartite Hilbert space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  is for  $d_{\mathcal{H}}^{-1} > m > 0$  uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}_{\geq m}$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$  we find that

$$\begin{aligned} |\widehat{I}_\rho(A : B) - \widehat{I}_\sigma(A : B)| &\leq 2l_m^{-1}\varepsilon(\log \min\{d_A, d_B\} + \log m^{-1}) + \frac{l_m + \varepsilon}{l_m} z_m\left(\frac{\varepsilon}{l_m + \varepsilon}\right) \\ &\leq \frac{2 \log \min\{d_A, d_B\} + 4 \log m^{-1} + (\sqrt{2} + 2)m^{-1} + \sqrt{2}}{l_m} \sqrt{\varepsilon}, \end{aligned}$$

with  $l_m = 1 - md_{\mathcal{H}}$  and

$$z_m(p) = 2f_{m^{-1}, m^{-1}}(p) + (m^{-1} + 1)h(p).$$

For the second inequality, we used lemma 5.8,  $\log(1 + x) \leq x$  for  $0 \leq x$ ,  $\varepsilon \leq \sqrt{\varepsilon}$  for  $\varepsilon \in [0, 1]$  and  $l_m \leq l_m + \varepsilon$ .

*Proof.* As in the case of the BS-conditional entropy, we find that  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant with  $s = md_{\mathcal{H}}$ . To conclude that  $\widehat{I}(A : B)$  is ALAFF we first note that because of the convexity of  $\widehat{D}(\|\cdot\|)$ ,

$$\begin{aligned} \widehat{I}_{p\rho_1 + (1-p)\rho_2}(A : B) &\leq p\widehat{D}(\rho_1\|\rho_{1,A} \otimes (p\rho_{1,B} + (1-p)\rho_{2,B})) \\ &\quad + (1-p)\widehat{D}(\rho_2\|\rho_{2,A} \otimes (p\rho_{1,B} + (1-p)\rho_{2,B})) \\ &\leq p\widehat{I}_{\rho_1}(A : B) + (1-p)\widehat{I}_{\rho_2}(A : B) + h(p). \end{aligned}$$

In the last step, we used that  $\widehat{D}(\|\cdot\|)$  is monotone decreasing in its second argument, and  $p\rho_{1,B} \leq p\rho_{1,B} + (1-p)\rho_{2,B}$ ,  $(1-p)\rho_{2,B} \leq p\rho_{1,B} + (1-p)\rho_{2,B}$ , respectively. Hence  $b_f = h$ . We follow similar lines to obtain  $a_f$ . Starting with theorem 6.3 and point 2 in proposition 6.5 using that  $\|\rho_A^{-1}\|_\infty \leq \|\rho_{AB}^{-1}\|_\infty$ , and analogously for  $\rho_B$ , we find

$$\begin{aligned} \widehat{I}_{p\rho_1 + (1-p)\rho_2}(A : B) &\geq p\widehat{D}(\rho_1\|\rho_{1,A} \otimes (p\rho_{1,B} + (1-p)\rho_{2,B})) \\ &\quad + (1-p)\widehat{D}(\rho_2\|\rho_{2,A} \otimes (p\rho_{1,B} + (1-p)\rho_{2,B})) - m^{-1}h(p) - f_{m^{-1}, m^{-1}}(p) \\ &\geq p\widehat{I}_{\rho_1}(A : B) + (1-p)\widehat{I}_{\rho_2}(A : B) - m^{-1}h(p) - 2f_{m^{-1}, m^{-1}}(p). \end{aligned}$$

In the last step we used again that  $\widehat{D}(\|\cdot\|)$  is monotone decreasing in its second argument and that  $p\rho_{1,AB} + (1-p)\rho_{2,AB} \leq (p + (1-p)m^{-1})\rho_{1,AB}$  and  $p\rho_{1,AB} + (1-p)\rho_{2,AB} \leq (m^{-1}p + (1-p))\rho_{2,AB}$ , giving us another  $f_{m^{-1}, m^{-1}}(p)$ . Hence  $a_f = m^{-1}h + 2f_{m^{-1}, m^{-1}}$ . We conclude the proof by noticing again that  $\|\rho_A^{-1}\|_\infty \leq \|\rho_{AB}^{-1}\|_\infty \leq m^{-1}$ , yielding the upper bound

$$C_f^s \leq \sup_{\rho \in \mathcal{S}_0} \widehat{I}_\rho(A : B) \leq \log \min\{d_A, d_B\} + \log m^{-1}.$$

Finally, we apply theorem 4.6 and get the claimed bounds as  $E_f$  coincides with  $E_f^{\max}$ , due to point 4 in proposition 6.5.  $\square$

### 6.2.3 Uniform continuity for the BS-conditional mutual information

Next, we provide a result of uniform continuity for the BS-conditional mutual information, defined in eq. (4). As a difference between two BS-conditional entropies, it will not present the pathological behaviour from the BS-mutual information, as the BS-conditional entropies are bounded between the same limits as the (usual) conditional entropies. See proposition 6.6 for the specific bounds on all these BS-entropic quantities.

Nevertheless, the continuity bound we obtain below for the BS-conditional mutual information also depends on the minimal eigenvalues of the states involved, as happened in the case of the BS-conditional entropies.

#### Corollary 6.10 (Uniform continuity of the BS-conditional mutual information)

The BS-conditional mutual information over  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  is for  $d_{\mathcal{H}}^{-1} > m > 0$  uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}_{\geq m}(\mathcal{H})$  and for  $\rho, \sigma \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon \leq 1$  we find that

$$\begin{aligned} |\widehat{I}_\rho(A : B|C) - \widehat{I}_\sigma(A : B|C)| &\leq 2\varepsilon l_m^{-1} \log \min\{d_A, \sqrt{d_{ABC}}\} + 2g_m(\varepsilon) \\ &\leq \frac{2 \log \min\{d_A, \sqrt{d_{ABC}}\} + 2 \log m^{-1} + 2(\sqrt{2} + 1)m^{-1}}{l_m} \sqrt{\varepsilon}, \end{aligned}$$

with  $l_m = 1 - md_{\mathcal{H}}$  and

$$g_m(\varepsilon) = \frac{l_m + \varepsilon}{l_m} (f_{m^{-1}, m^{-1}} + m^{-1}h) \left( \frac{\varepsilon}{l_m + \varepsilon} \right).$$

For the second inequality, we used lemma 5.8,  $\log(1+x) \leq x$  for  $0 \leq x$ ,  $\varepsilon \leq \sqrt{\varepsilon}$  for  $\varepsilon \in [0, 1]$  and  $l_m \leq l_m + \varepsilon$ .

*Proof.* We have that  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant using the same reasoning as in the proof of corollary 6.8. Because of the representation  $\widehat{I}(A : B|C) = \widehat{H}(A|C) - \widehat{H}(A|BC)$  we can immediately conclude that  $\widehat{I}(A : B|C)$  is ALAFF with  $a_f = f_{m^{-1}, m^{-1}} + m^{-1}h$  and  $b_f = f_{m^{-1}, m^{-1}} + m^{-1}h$  arguing along the same lines as in corollary 6.8. Using proposition 6.6 we can conclude

$$C_f^s \leq \sup_{\rho \in \mathcal{S}(\mathcal{H})} \widehat{I}_\rho(A : B|C) \leq 2 \log \min\{d_A, \sqrt{d_{ABC}}\}.$$

Applying theorem 4.6 and using point 4 of proposition 6.5 we get that  $E_f = E_f^{\max}$  and thereby conclude the assertion.  $\square$

### 6.2.4 Divergence bound for the BS-entropy

We conclude this section by following the same lines as in the case of the relative entropy to provide a divergence bound for the BS-entropy. Firstly, we will prove the uniform continuity of the BS-entropy in the first argument and subsequently derive from that result the divergence bound. These results should be compared to their relative entropy analogues, namely corollary 5.9 and corollary 5.10, respectively.

#### Corollary 6.11 (Uniform continuity of the BS-entropy in the first argument)

Let  $\sigma \in \mathcal{S}_+(\mathcal{H})$  be fixed. Then  $\widehat{D}(\cdot \| \sigma)$  is uniformly continuous on  $\mathcal{S}_0 = \mathcal{S}(\mathcal{H})$ , and for  $\rho_1, \rho_2 \in \mathcal{S}_0$  with  $\frac{1}{2} \|\rho_1 - \rho_2\| \leq \varepsilon \leq 1$  we find that

$$|\widehat{D}(\rho_1 \| \sigma) - \widehat{D}(\rho_2 \| \sigma)| \leq \varepsilon \log(m_\sigma^{-1}) + (1 + \varepsilon)m_\sigma^{-1}h\left(\frac{\varepsilon}{1 + \varepsilon}\right),$$

with  $m_\sigma$  the minimal eigenvalue of  $\sigma$ .



*Proof.* The procedure is familiar. First,  $\mathcal{S}_0$  is 0-perturbed  $\Delta$ -invariant. Second  $f(\cdot) = \widehat{D}(\cdot\|\sigma)$  is ALAFF with  $a_f = m_\sigma^{-1}h$  and  $b_f = 0$  employing theorem 6.3 and point 1 of proposition 6.5. Further

$$C_f^\perp \leq \sup_{\rho \in \mathcal{S}(\mathcal{H})} \widehat{D}(\rho\|\sigma) \leq \log m_\sigma^{-1}$$

since  $\rho^{1/2}\sigma^{-1}\rho^{1/2} \leq \mathbb{1} m_\sigma^{-1}$ . Applying now remark 4.7 gives the claimed result.  $\square$

Utilizing the above result we obtain a divergence bound for the BS-entropy which constitutes the analogue to the one of the relative entropy in corollary 5.10. Note that even the divergence bounds obtained in both cases are similar, except for the presence of a factor  $m_\sigma^{-1}$  in the second term of the bound.

**Corollary 6.12 (Divergence bound for the BS-entropy)**

Let  $\rho \in \mathcal{S}(\mathcal{H})$  and  $\sigma \in \mathcal{S}_+(\mathcal{H})$ , then for  $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ , we have

$$\widehat{D}(\rho\|\sigma) \leq \varepsilon \log m_\sigma^{-1} + (1 + \varepsilon)m_\sigma^{-1}h\left(\frac{\varepsilon}{1 + \varepsilon}\right),$$

with  $m_\sigma$  the minimal eigenvalue of  $\sigma$ .

*Proof.* In the context of corollary 6.11, we just set  $\rho_1 = \rho$  and  $\rho_2 = \sigma$ , giving us that  $\frac{1}{2}\|\rho_1 - \rho_2\|_1 = \frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$ . Further  $\widehat{D}(\rho_2\|\sigma) = \widehat{D}(\sigma\|\sigma) = 0$  and  $|\widehat{D}(\rho_1\|\sigma)|$  loses the absolute value, as  $\widehat{D}(\cdot\|\cdot) \geq 0$ . The bound follows immediately.  $\square$

With this, we conclude our section on continuity bounds for entropic quantities derived from the BS-entropy. We have deliberately omitted the analogues of corollary 5.11 and theorem 5.13 for the BS-entropy, due to their high technicality and the complexity of the continuity bounds that we would obtain with our method. However, the same procedure as for the relative entropy would give analogous continuity bounds also in this setting.

## 7 Applications

In this section, we use some of the previously derived bounds to provide applications in various contexts within the field of quantum information.

### 7.1 Quantum hypothesis testing

In this section, we interpret our bounds in terms of hypothesis testing. Quantum state discrimination and quantum hypothesis testing are both well-studied tasks in quantum information theory.

In quantum state discrimination, you are given a source which prepares quantum states  $\rho_1$  and  $\rho_2$  with equal probability. The task is to perform a measurement in order to identify whether the state prepared by the source is  $\rho_1$  or  $\rho_2$ . In this setting, the optimal probability of successfully identifying the state is given in terms of the trace distance as

$$p_{\text{succ}} = \frac{1}{2} \left( 1 + \frac{1}{2} \|\rho_1 - \rho_2\|_1 \right) \tag{33}$$

using the Helstrom measurement (see textbooks such as [61]).

In quantum hypothesis testing, we consider an asymmetric setting with  $n$  copies and we are interested in the asymptotic performance. Again, the task is to discriminate between  $\rho$  and  $\sigma$ , using

a measurement  $\{E, \mathbb{1} - E\}$  where  $0 \leq E \leq \mathbb{1}$ . Upon the first outcome, the guess is  $\rho$ , and upon the second  $\sigma$ . Therefore, we define the errors of the first and second kind as

$$\alpha(E)_n = \text{Tr}[\rho^{\otimes n}(\mathbb{1} - E)]$$

and

$$\beta(E)_n = \text{Tr}[\sigma^{\otimes n}E].$$

We now want to fix the error of the first kind to be at most  $\varepsilon$  and define

$$\beta_\varepsilon(\rho^{\otimes n}||\sigma^{\otimes n}) := \min\{\beta(E)_n : \alpha(E)_n \leq \varepsilon\},$$

where the minimum runs over  $0 \leq E \leq \mathbb{1}$ . Then, the quantum Stein's lemma [45, 64] states that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log[\beta_\varepsilon(\rho^{\otimes n}||\sigma^{\otimes n})] = -D(\rho||\sigma).$$

Therefore, we can interpret the continuity bound in the way that two states that are hard to discriminate have almost the same performance in terms of hypothesis testing. We can illustrate this with corollary 5.9, just by taking  $1 + \varepsilon$  there to be  $2p_{\text{succ}}$  following eq. (33).

**Corollary 7.1** *Let  $\sigma \in \mathcal{S}(\mathcal{H})$  be fixed,  $0 < \varepsilon < 1$  and let us consider a source which produces  $\rho_1, \rho_2$  with equal probability. Moreover, let  $p$  be an upper bound on the probability  $p_{\text{succ}}$  of successfully identifying the state. Then, the difference in the asymptotic error exponent in hypothesis testing is bounded by*

$$\left| \lim_{n \rightarrow \infty} \frac{1}{n} \log[\beta_\varepsilon(\rho_1^{\otimes n}||\sigma^{\otimes n})] - \lim_{n \rightarrow \infty} \frac{1}{n} \log[\beta_\varepsilon(\rho_2^{\otimes n}||\sigma^{\otimes n})] \right| \leq (2p - 1) \log \tilde{m}_\sigma^{-1} + 2p h\left(\frac{2p - 1}{2p}\right),$$

with  $\tilde{m}_\sigma^{-1}$  the minimal non-zero eigenvalue of  $\sigma$ .

## 7.2 Free energy

In section 7.1, we already saw one interpretation of our results in terms of hypothesis testing. This section gives another interpretation using the language of quantum thermodynamics.

A ubiquitous quantity in quantum thermodynamics is free energy. To define it, we need to fix a Hamiltonian  $H \in \mathcal{B}(\mathcal{H})$ ,  $H = H^*$ , and some inverse temperature  $\beta > 0$ . The Gibbs state of this system, describing a quantum system in thermal equilibrium, is

$$\rho_\beta(H) = \frac{e^{-\beta H}}{\text{tr}[e^{-\beta H}]}.$$

Now, we can define the free energy as

$$F(\rho) = \text{tr}[H\rho] - \beta^{-1}S(\rho).$$

It can be related to the relative entropy as

$$D(\rho||\rho_\beta(H)) = \beta(F(\rho) - F(\rho_\beta(H))), \quad (34)$$

which can easily be verified by direct computation.

Inspired by quantum information theory, in particular entanglement theory, during the last years various descriptions of quantum thermodynamics as a resource theory have emerged. Resource

theories are described in terms of free states and free operations. In quantum thermodynamics, the free state is  $\rho_\beta(H)$ , whereas the choices of free operations can differ. Possible choices include the thermal operations (TO), their closure (CTO), and the Gibbs preserving covariant operations (GPC). Instead of giving a formal definition here, we refer the reader to [34, Section II.C]. In entanglement theory, we are interested in the distillation of EPR pairs from other states, possibly taking many copies. In the same spirit, in quantum thermodynamics, the corresponding task is the distillation of athermality. The asymptotic distillable athermality is quantified by the free energy difference in eq. (34) [20, 34]. Theorem V.1 of [34] states that for the asymptotic distillation rate of athermality

$$\text{Distill}_{\mathfrak{F}}(\rho, \rho_\beta(H)) = D(\rho || \rho_\beta(H)),$$

where  $\mathfrak{F} \in \{\text{TO}, \text{CTO}, \text{GPC}\}$ . Again, we refer the reader to [34] for the formal definitions. Thus, we can interpret corollary 5.9 as quantifying the continuity of distillable athermality.

**Corollary 7.2** *Let  $H$  be a fixed Hamiltonian with maximal eigenvalue  $\lambda_{\max}$ , minimal eigenvalue  $\lambda_{\min}$ , and  $\beta > 0$  an inverse temperature. Then, for  $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$  such that  $\frac{1}{2}\|\rho_1 - \rho_2\|_1 \leq \epsilon \leq 1$ , it holds that*

$$\begin{aligned} |\text{Distill}_{\mathfrak{F}}(\rho_1, \rho_\beta(H)) - \text{Distill}_{\mathfrak{F}}(\rho_2, \rho_\beta(H))| &\leq \epsilon \left( \beta \lambda_{\max} + \log(\text{Tr}[e^{-\beta H}]) \right) + (1 + \epsilon) h\left(\frac{\epsilon}{1 + \epsilon}\right) \\ &\leq \epsilon \beta (\lambda_{\max} - \lambda_{\min}) + \epsilon \log(d) + (1 + \epsilon) h\left(\frac{\epsilon}{1 + \epsilon}\right), \end{aligned}$$

where  $\mathfrak{F} \in \{\text{TO}, \text{CTO}, \text{GPC}\}$ .

### 7.3 Approximate Quantum Markov Chains

In this section, we consider a tripartite Hilbert space  $\mathcal{H}_{ABC} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  and  $\rho_{ABC} \in \mathcal{S}_+(\mathcal{H}_{ABC})$ . We further consider the conditional mutual information of  $\rho_{ABC}$  between  $A$  and  $C$  conditioned on  $B$ . The well-known property of strong subadditivity of the von Neumann entropy [53] is equivalent to the non-negativity of the conditional mutual information, which is furthermore known [42, 63] to vanish if, and only if,

$$\rho_{ABC} = \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2},$$

i.e., whenever  $\rho_{ABC}$  is a quantum Markov chain. In particular, if we denote  $\mathcal{P}_{B \rightarrow AB}(\rho_{BC}) = \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2}$ , we have

$$I_{\mathcal{P}_{B \rightarrow AB}(\rho_{BC})}(A : C|B) = 0.$$

Moreover, by the decomposition of the CMI of  $\rho_{ABC}$  in terms of a difference of conditional entropies, as well as the data processing inequality, we have

$$I_\rho(A : C|B) = H_\rho(C|B) - H_\rho(C|AB) \leq H_{\mathcal{P}_{B \rightarrow AB}(\rho_{BC})}(C|AB) - H_\rho(C|AB).$$

Here we w.l.o.g. assumed that  $d_A \leq d_C$  using the symmetry of the CMI in  $A$  and  $C$ . Therefore, we can apply our continuity bound for the CE from corollary 5.5 (which provides, in this case, a tighter result than corollary 5.7), cf. also [88], to obtain an upper bound on the CMI of  $\rho_{ABC}$  in terms of how far it is from being recovered with the Petz recovery map, i.e., in terms of

$$\left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1.$$

A similar direction was previously explored in [77, Eq. (26)]. Note that, as a direct consequence of corollary 5.5, we get the following bound for any state  $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_{ABC})$ :

$$I_\rho(A : C|B) \leq 2\varepsilon \log \min\{d_A, d_C\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right),$$

with

$$\varepsilon := \frac{1}{2} \left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1.$$

Moreover, we can use the following inequality

$$(1 + x)h\left(\frac{x}{1 + x}\right) \leq \sqrt{2x},$$

for every  $x \in [0, 1]$ , as well as the fact that, since  $\varepsilon \in [0, 1]$ , then  $\varepsilon \leq \sqrt{\varepsilon}$ , to upper bound the CMI of  $\rho_{ABC}$  by

$$I_\rho(A : C|B) \leq \left(\sqrt{2} \log \min\{d_A, d_C\} + 1\right) \left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1^{1/2}. \quad (35)$$

This bound should be compared to lower bounds for the conditional mutual information. On the one hand, Fawzi and Renner proved in [32] the following lower bound for such a quantity in terms of the fidelity  $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ :

$$I_\rho(A : C|B) \geq -\log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow AB}(\rho_{BC})),$$

where  $\mathcal{R}_{B \rightarrow AB}$  is another recovery map, the so-called *rotated Petz recovery map*, which was explicitly constructed in [46]. Several results have been provided in this line in the past decade. Here we specifically focus on [27], in which Carlen and Vershynina proved:

$$I_\rho(A : C|B) \geq \left(\frac{\pi}{8}\right)^4 \|\rho_B^{-1}\|_\infty^{-2} \|\rho_{ABC}^{-1}\|_\infty^{-2} \left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1^4, \quad (36)$$

Therefore, by combining eq. (35) with eq. (36) we obtain the following ‘‘sandwich’’ for the conditional mutual information of a tripartite density matrix  $\rho_{ABC}$  in terms of its trace distance to its Petz recovery map:

$$\begin{aligned} & \left(\frac{\pi}{8}\right)^4 \|\rho_B^{-1}\|_\infty^{-2} \|\rho_{ABC}^{-1}\|_\infty^{-2} \left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1^4 \\ & \leq I_\rho(A : C|B) \\ & \leq 2(\log \min\{d_A, d_C\} + 1) \left\| \rho_{ABC} - \rho_{AB}^{1/2} \rho_B^{-1/2} \rho_{BC} \rho_B^{-1/2} \rho_{AB}^{1/2} \right\|_1^{1/2}. \end{aligned}$$

In particular, this implies that a state  $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$  is an *approximate quantum Markov chain* [75] (i.e.  $I_\rho(A : C|B) < \epsilon$ ) if, and only if, it is close to its reconstructed state under the Petz recovery map. This idea was used in [48] to prove that a Gibbs state of a one-dimensional local Hamiltonian is an approximate quantum Markov chain, and subsequently, in [40] to provide an estimate on the time it takes for a Markovian evolution of a density matrix to become an approximate quantum Markov chain. Moreover, a similar inequality has recently been employed in [78] to study the decay of the CMI for purely generated finitely correlated states.

## 7.4 Difference between relative entropy and BS-entropy

It is well-known that the BS-entropy is an upper bound on the Umegaki relative entropy [43, 56, 62], i.e., that

$$D(\rho||\sigma) \leq \widehat{D}(\rho||\sigma),$$

and they coincide if and only if  $\rho$  and  $\sigma$  commute (see, e.g., [45] and [43, Proposition 4.7]). In this section, our aim is to quantify how large the difference between the two divergences can become. We start with two upper bounds on  $\widehat{D}(\rho||\sigma)$  in terms of  $D(\rho||\sigma)$ .

**Proposition 7.3** *Consider two positive definite states  $\rho, \sigma \in \mathcal{S}_+(\mathcal{H})$ . Then, the following inequality holds:*

$$\widehat{D}(\rho||\sigma) \leq D(\rho||\sigma) + m^{-1} \|\rho - \sigma\|_\infty,$$

where  $m$  is the minimal eigenvalue of  $\sigma$ .

*Proof.* We can upper bound the difference between the entropies by

$$\begin{aligned} \widehat{D}(\rho||\sigma) - D(\rho||\sigma) &= \text{Tr} \left[ \rho \left( \log(\rho^{1/2} \sigma^{-1} \rho^{1/2}) - \log \rho + \log \sigma \right) \right] \\ &= -D \left( \rho \middle| \middle| \exp \left\{ \log \sigma + \log(\rho^{1/2} \sigma^{-1} \rho^{1/2}) \right\} \right) \\ &\leq \log \text{Tr} \left[ \exp \left\{ \log \sigma + \log(\rho^{1/2} \sigma^{-1} \rho^{1/2}) \right\} \right] \\ &\leq \log \text{Tr} \left[ \sigma \rho^{1/2} \sigma^{-1} \rho^{1/2} \right], \end{aligned}$$

where we have used the non-negativity for the relative entropy of density matrices and Golden-Thompson inequality [33, 80]. Next, we can write

$$\text{Tr} \left[ \sigma \rho^{1/2} \sigma^{-1} \rho^{1/2} \right] = \text{Tr} \left[ \sigma \rho^{1/2} (\sigma^{-1} - \rho^{-1}) \rho^{1/2} \right] + 1.$$

Therefore, using  $\log(x+1) \leq x$ , we have

$$\widehat{D}(\rho||\sigma) - D(\rho||\sigma) \leq \text{Tr} \left[ \sigma \rho^{1/2} (\sigma^{-1} - \rho^{-1}) \rho^{1/2} \right].$$

Now, we can use the following expression for invertible matrices  $X$  and  $Y$ :

$$X^{-1} - Y^{-1} = Y^{-1}(Y - X)X^{-1}$$

Then,

$$\begin{aligned} \text{Tr} \left[ \sigma \rho^{1/2} (\sigma^{-1} - \rho^{-1}) \rho^{1/2} \right] &= \text{Tr} \left[ \sigma \rho^{-1/2} (\rho - \sigma) \sigma^{-1} \rho^{1/2} \right] \\ &\leq \|\rho^{-1/2} (\rho - \sigma) \sigma^{-1} \rho^{1/2}\|_\infty \\ &\leq \|\sigma^{-1}\|_\infty \|\rho - \sigma\|_\infty, \end{aligned}$$

by [15, Proposition IX.1.1] and Hölder's inequality.  $\square$

The previous proposition provides a general upper bound for the distance between both entropies in terms of the spectral norm and the minimal eigenvalue of the second input. This is valid for any pair of states but does not yield any further information on specific pairs with better conditions. Alternatively, we can prove the following bound, from which it is obvious that  $D(\rho||\sigma) = \widehat{D}(\rho||\sigma)$  if  $\rho$  and  $\sigma$  commute.

**Proposition 7.4** Consider two positive definite states  $\rho, \sigma \in \mathcal{S}_+(\mathcal{H})$ . Then, the following inequality holds:

$$\widehat{D}(\rho\|\sigma) \leq D(\rho\|\sigma) + f([\rho^{1/2}, \sigma^{-1/2}])$$

where the last term is given by

$$f([\rho^{1/2}, \sigma^{-1/2}]) := \left\| [\rho^{1/2}, \sigma^{-1/2}] \right\|_\infty^2 + 2 \left\| [\rho^{1/2}, \sigma^{-1/2}] \right\|_\infty.$$

In particular, whenever  $\rho$  and  $\sigma$  commute,  $f$  vanishes.

*Proof.* The proof proceeds in the same way as for proposition 7.3 until

$$\widehat{D}(\rho\|\sigma) - D(\rho\|\sigma) \leq \text{Tr} \left[ \sigma \rho^{1/2} (\sigma^{-1} - \rho^{-1}) \rho^{1/2} \right].$$

Let us define now

$$\eta := \sigma^{1/2} \rho^{1/2} \sigma^{-1} \rho^{1/2} \sigma^{1/2}.$$

Then,

$$\text{Tr} \left[ \sigma \rho^{1/2} (\sigma^{-1} - \rho^{-1}) \rho^{1/2} \right] = \text{Tr} [\eta - \sigma].$$

Introducing  $\rho$  gives

$$\text{Tr} [\eta - \sigma] = \text{Tr} [\eta - \sigma + \rho - \rho] = \text{Tr} [\eta - \rho] + \text{Tr} [\rho - \sigma] = \text{Tr} [\eta - \rho] \leq \|\eta - \rho\|_1.$$

Moreover, as appears in [18, Remark 2.2], the right-hand side above can be estimated by

$$\|\eta - \rho\|_1 \leq \left\| [\rho^{1/2}, \sigma^{-1/2}] \right\|_\infty^2 + 2 \left\| [\rho^{1/2}, \sigma^{-1/2}] \right\|_\infty.$$

This concludes the proof of the proposition.  $\square$

Finally, we want to compare our previous bounds, proven using inequalities such as Golden-Thompson or Hölder, with those we could obtain by means of our continuity bounds, as the BS-entropy can, in particular, be regarded as a relative entropy. For that, we can also apply the continuity bound we derived in theorem 5.13.

**Corollary 7.5** Let  $\rho \in \mathcal{S}(\mathcal{H})$ ,  $\sigma \in \mathcal{S}_+(\mathcal{H})$  and  $\tilde{m}$  such that  $d_{\mathcal{H}}^{-1} > 2\tilde{m} > 0$  and the minimal eigenvalue of  $\sigma$  is lower bounded by  $2\tilde{m}$ . Let

$$\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} = \sum_{i=1}^k \lambda_i P_i$$

be the spectral decomposition with eigenvalues  $\lambda_i$  and projections  $P_i$ . Define density matrices

$$p = \sum_{i=1}^k \lambda_i \frac{\text{tr}[\sigma P_i]}{\text{tr}[P_i]} P_i, \quad q = \sum_{i=1}^k \frac{\text{tr}[\sigma P_i]}{\text{tr}[P_i]} P_i.$$

Then, for  $\frac{1}{2} \|\rho - p\| \leq \varepsilon \leq 1$  and  $\frac{1}{2} \|\sigma - q\|_1 \leq \delta \leq 1$ , it holds that

$$|\widehat{D}(\rho\|\sigma) - D(\rho\|\sigma)| \leq \left( \varepsilon + \frac{\delta}{l_{\tilde{m}}} \right) \log \tilde{m}^{-1} + (1 + \varepsilon) h\left( \frac{\varepsilon}{1 + \varepsilon} \right) + 2 \frac{l_{\tilde{m}} + \delta}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left( \frac{\delta}{l_{\tilde{m}} + \delta} \right), \quad (37)$$

with  $l_{\tilde{m}} = 1 - \tilde{m}$ . In particular, if  $[\rho, \sigma] = 0$ ,  $\varepsilon$  and  $\delta$  can be taken as 0 such that the RHS of eq. (37) is zero.

Moreover, we can further simplify the previous bound to

$$|\widehat{D}(\rho\|\sigma) - D(\rho\|\sigma)| \leq (\sqrt{2} - \log \tilde{m}) \sqrt{\varepsilon} + 3 \frac{\log \tilde{m}^{-1}}{1 - \tilde{m}} \delta + 2 \log \left( 1 + \frac{\delta}{1 - \tilde{m} + \delta \tilde{m}} \right). \quad (38)$$

*Proof.* Our argument is a slight variation of Matsumoto’s minimal reverse test [56] (see also [43]). We can write the BS-entropy as the relative entropy of two commuting density matrices

$$\widehat{D}(\rho\|\sigma) = D(p\|q),$$

since we can verify with  $p_i = \lambda_i \text{tr}[\sigma P_i]$ ,  $q_i = \text{tr}[\sigma P_i]$  that

$$\begin{aligned} D(p\|q) &= \sum_{i=1}^k \text{Tr} \left[ \frac{P_i}{\text{Tr}[P_i]} p_i \left( \log \frac{p_i}{\text{Tr}[P_i]} - \log \frac{q_i}{\text{Tr}[P_i]} \right) \right] \\ &= \sum_{i=1}^k p_i (\log p_i - \log q_i) \\ &= \sum_{i=1}^k \lambda_i \text{tr}[\sigma P_i] \log \lambda_i \\ &= \text{Tr} \left[ \sigma \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}} \log(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}) \right] \\ &= \text{Tr} \left[ \rho \log(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}) \right]. \end{aligned}$$

Obviously, if  $m$  is the minimal eigenvalue of  $\sigma$ , then  $\frac{q_i}{\text{Tr}[P_i]} \geq m$  for all  $i \in \{1, \dots, k\}$ . Thus, the assertion follows from theorem 5.13. Moreover, it is clear that if  $[\rho, \sigma] = 0$  there is a unitary  $U$  which diagonalizes  $\rho$  and  $\sigma$  simultaneously such that  $\rho = p$  and  $\sigma = q$ .

Finally, the last simplification from eq. (38) is a direct consequence of theorem 5.13 and lemma 5.8.  $\square$

## 7.5 Weak quasi-factorization of the relative entropy

Results of quasi-factorization for a divergence allow us to split such a divergence in a bipartite space in terms of the sum of two “conditional” divergences on subsystems and a multiplicative error term that is related to the correlations between both subsystems on the second input of the divergences. A weak version of such a result presents instead an additive error term.

More specifically, it was proven in [23] that, given a bipartite space  $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$  and  $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$ , the following inequality holds:

$$D(\rho_{AB}\|\sigma_{AB}) \leq \frac{1}{1 - 2 \|h(\sigma_{AB})\|_\infty} [D_A(\rho_{AB}\|\sigma_{AB}) + D_B(\rho_{AB}\|\sigma_{AB})], \quad (39)$$

with

$$h(\sigma_{AB}) := \sigma_A^{-1/2} \otimes \sigma_B^{-1/2} \sigma_{AB} \sigma_A^{-1/2} \otimes \sigma_B^{-1/2} - \mathbf{1}_{AB},$$

and

$$D_X(\rho_{AB}\|\sigma_{AB}) := D(\rho_{AB}\|\sigma_{AB}) - D(\rho_{X^c}\|\sigma_{X^c}), \quad \text{for } X = A, B,$$

whenever  $\|h(\sigma_{AB})\|_\infty < 1/2$ . Note that the term  $\|h(\sigma_{AB})\|_\infty$  provides a measure of how far  $\sigma_{AB}$  is from being a tensor product between  $A$  and  $B$ . This result, and subsequent extensions with additional conditions on  $\sigma_{AB}$ , are expected to find applications on various tasks in quantum information theory, and in particular, have proven to be essential for some recent proofs of positivity of *modified logarithmic Sobolev inequalities* (MLSI) for quantum Markov semigroups modelling thermal dissipative evolutions on quantum spin systems [9–11, 24]. It is important to remark that eq. (39) is equivalent to a generalization of the property of superadditivity of the relative entropy, as shown in [22].

In [18], some authors of the current manuscript tried to extend the previous result for the Umegaki relative entropy to the BS-entropy framework. However, we showed that the BS-entropy cannot satisfy a property of superadditivity, which makes it impossible to obtain a quasi-factorization for the BS-entropy in the spirit of eq. (39) without an additive error term. Instead, we proved a result of *weak quasi-factorization*, from which we recovered eq. (39) if the marginals of  $\rho_{AB}$  and  $\sigma_{AB}$  commute. Here, we can prove another result along these lines as a consequence of our continuity bound for the relative entropy. Indeed, as a consequence of theorem 5.13, we obtain the following result of quasi-factorization for the relative entropy with an additive error term.

**Corollary 7.6 (Weak quasi-factorization for the relative entropy)**

Given  $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  such that  $\ker(\sigma_X) \subset \ker(\rho_X)$  for  $X = A, B, AB$ , we have:

$$D(\rho_{AB} \parallel \sigma_{AB}) \leq D_A(\rho_{AB} \parallel \sigma_{AB}) + D_B(\rho_{AB} \parallel \sigma_{AB}) + \xi_{RE}(\rho_{AB}, \sigma_{AB}),$$

with

$$\xi_{RE}(\rho_{AB}, \sigma_{AB}) := \left( \sqrt{2} - \log \tilde{m} \right) \sqrt{\varepsilon} + 3 \frac{\log \tilde{m}^{-1}}{l_{\tilde{m}}} \delta + 2 \log \left( 1 + \frac{\delta}{l_{\tilde{m}} + \delta \tilde{m}} \right),$$

where  $\tilde{m} = \frac{1}{2} \min \left\{ \left\| \sigma_A^{-1} \otimes \sigma_B^{-1} \right\|_{\infty}^{-1}, \left\| \sigma_{AB}^{-1} \right\|_{\infty}^{-1} \right\}$ ,  $\varepsilon = \varepsilon(\rho_{AB}) = \frac{1}{2} \|\rho_{AB} - \rho_A \otimes \rho_B\|_1$  and  $\delta = \delta(\sigma_{AB}) = \frac{1}{2} \|\sigma_{AB} - \sigma_A \otimes \sigma_B\|_1$ .

*Proof.* The difference between the relative and the two conditional entropies can be written as

$$D(\rho_{AB} \parallel \sigma_{AB}) - D_A(\rho_{AB} \parallel \sigma_{AB}) - D_B(\rho_{AB} \parallel \sigma_{AB}) = -D(\rho_{AB} \parallel \sigma_{AB}) + D(\rho_A \otimes \rho_B \parallel \sigma_A \otimes \sigma_B).$$

Therefore, we can apply theorem 5.13 to obtain a continuity bound for the difference between the last two relative entropies, obtaining

$$\begin{aligned} & |D(\rho_{AB} \parallel \sigma_{AB}) - D(\rho_A \otimes \rho_B \parallel \sigma_A \otimes \sigma_B)| \\ & \leq \left( \varepsilon + \frac{\delta}{l_{\tilde{m}}} \right) \log(\tilde{m}^{-1}) + (1 + \varepsilon) h\left(\frac{\varepsilon}{1 + \varepsilon}\right) + 2 \frac{l_{\tilde{m}} + \delta}{l_{\tilde{m}}} f_{\tilde{m}^{-1}, \tilde{m}^{-1}}\left(\frac{\delta}{l_{\tilde{m}} + \delta}\right), \end{aligned}$$

with

$$\varepsilon := \frac{1}{2} \|\rho_{AB} - \rho_A \otimes \rho_B\|_1, \quad \delta := \frac{1}{2} \|\sigma_{AB} - \sigma_A \otimes \sigma_B\|_1,$$

and  $l_{\tilde{m}} = 1 - \tilde{m}$ , for  $\tilde{m} = \frac{1}{2} \min \left\{ \left\| \sigma_A^{-1} \otimes \sigma_B^{-1} \right\|_{\infty}^{-1}, \left\| \sigma_{AB}^{-1} \right\|_{\infty}^{-1} \right\}$ . Moreover, we can apply the simplification of theorem 5.13 using lemma 5.8. We then have

$$|D(\rho_{AB} \parallel \sigma_{AB}) - D(\rho_A \otimes \rho_B \parallel \sigma_A \otimes \sigma_B)| \leq \left( \sqrt{2} - \log \tilde{m} \right) \sqrt{\varepsilon} + 3 \frac{\log \tilde{m}^{-1}}{l_{\tilde{m}}} \delta + 2 \log \left( 1 + \frac{\delta}{l_{\tilde{m}} + \delta \tilde{m}} \right),$$

concluding thus the proof.  $\square$

Note that, even though there is a caveat in this result in the form of an additive error term, which prevents it from being useful to prove the positivity of MLSIs, it presents the advantage with respect to eq. (39) that there is no multiplicative error term in this case, which might be of more interest for some other contexts, such as for entropy accumulation [58] or in the line of the applications given by the Brascamp-Lieb dualities [14].



## 7.6 Minimal distance to separable states

In this section, we show how to reprove the continuity bounds for the relative entropy of entanglement in [88] from the ALAFF method and how this strategy generalizes if we quantify the minimal distance to the set of separable states in terms of the BS-entropy instead.

Let  $\mathcal{C} \subset \mathcal{S}(\mathcal{H})$  be a compact convex subset of the set of quantum states with at least one positive definite state. We can define the minimal distance to  $\mathcal{C}$  in terms of the relative entropy as

$$D_{\mathcal{C}}(\rho) := \inf_{\gamma \in \mathcal{C}} D(\rho \| \gamma).$$

As explained in [88], the fact that  $\mathcal{C}$  contains a positive definite state guarantees that  $D_{\mathcal{C}}(\rho) < \infty$  for all  $\rho \in \mathcal{S}(\mathcal{H})$ . Moreover, the infimum is attained, as follows from the fact that the relative entropy is lower semi-continuous [62] and Weierstrass' theorem on extreme values of such functions [2, Theorem 2.43]. Examples of  $\mathcal{C}$  include  $\text{SEP}_{AB}$ , the set of separable states for systems  $A, B$ , and

$$\{d_A^{-1} \mathbb{1}_A \otimes \sigma_B : \sigma_B \in \mathcal{S}(\mathcal{H}_B)\},$$

which yields  $D_{\mathcal{C}}(\rho_{AB}) = -H_{\rho}(A|B) + \log d_A$ . The quantity  $D_{\text{SEP}_{AB}}$  is known as the relative entropy of entanglement [84, 85]. It constitutes a tight upper bound on the distillable entanglement [65, 85]. This is the quantity we focus on for now.

**Lemma 7.7** *Let  $\mathcal{C} \subset \mathcal{S}(\mathcal{H})$  be a compact convex set containing at least one positive definite state. Then,  $D_{\mathcal{C}}$  is convex on  $\mathcal{S}(\mathcal{H})$ .*

*Proof.* This follows directly from the joint convexity of the relative entropy.  $\square$

In order to apply the ALAFF method, we need to prove almost concavity next.

**Lemma 7.8** *Let  $\mathcal{C} \subset \mathcal{S}(\mathcal{H})$  be a compact convex set containing at least one positive definite state. Moreover, let  $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$  and  $p \in [0, 1]$ . Then,*

$$D_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2) \geq pD_{\mathcal{C}}(\rho_1) + (1-p)D_{\mathcal{C}}(\rho_2) - h(p).$$

*Proof.* We can use the almost concavity of the relative entropy. Let  $\tau$  the state that achieves the infimum in  $D_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2)$ . By theorem 5.1 and point 1 of proposition 5.2, we obtain that

$$\begin{aligned} D_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2) &\geq pD(\rho_1 \| \tau) + (1-p)D(\rho_2 \| \tau) - h(p) \\ &\geq pD_{\mathcal{C}}(\rho_1) + (1-p)D_{\mathcal{C}}(\rho_2) - h(p), \end{aligned}$$

which is the assertion.  $\square$

Finally, we need the following estimate:

**Lemma 7.9** *Let  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ . It holds that*

$$\sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1}} |D_{\text{SEP}_{AB}}(\rho) - D_{\text{SEP}_{AB}}(\sigma)| \leq \log \min\{d_A, d_B\}.$$

*Proof.* Without loss of generality, let  $d_A \leq d_B$ . For a pure state  $|\psi\rangle$  with Schmidt decomposition  $\sum_{i=1}^{d_A} \lambda_i |i_A\rangle \otimes |i_B\rangle$ , let

$$\tau_{\psi} = \frac{1}{d_A} \sum_{i=1}^{d_A} |i_A\rangle\langle i_A| \otimes |i_B\rangle\langle i_B|.$$

This state is manifestly separable. Then,

$$\begin{aligned} \sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1}} |D_{\text{SEP}_{AB}}(\rho) - D_{\text{SEP}_{AB}}(\sigma)| &\leq \sup_{|\psi\rangle\langle\psi| \in \mathcal{S}(\mathcal{H})} D(|\psi\rangle\langle\psi| \| \tau_\psi) \\ &= \log d_A. \end{aligned}$$

In the first inequality, we have used that  $D_{\text{SEP}_{AB}}$  is positive and convex.  $\square$

This allows us to prove via the ALAFF method a continuity bound for the relative entropy of entanglement:

**Theorem 7.10** *For  $\varepsilon \in [0, 1]$  and  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ , it holds that for  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$*

$$|D_{\text{SEP}_{AB}}(\rho) - D_{\text{SEP}_{AB}}(\sigma)| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

*Proof.* This follows from remark 4.7, using lemma 7.7, lemma 7.8, point 4 of proposition 5.2, and lemma 7.9.  $\square$

theorem 7.10 recovers the bound [88, Corollary 8], proven with very similar methods, which improved over the earlier bound in [28]. The interest of executing the proof here is that a similar strategy will give us bounds on a BS-entropy version of the relative entropy of entanglement, as we will show now. We define

$$\widehat{D}_{\mathcal{C}}(\rho) = \inf_{\gamma \in \mathcal{C}} \widehat{D}(\rho \| \gamma),$$

which measures how far  $\rho$  is from  $\mathcal{C}$  in terms of the BS-entropy. The infimum is attained as the BS-entropy is also lower semi-continuous [57, Section 10]. Convexity follows again from the joint-convexity of the BS-entropy.

**Lemma 7.11** *Let  $\mathcal{C} \subset \mathcal{S}(\mathcal{H})$  be a compact convex set containing at least one positive definite state. Then,  $\widehat{D}_{\mathcal{C}}$  is convex on  $\mathcal{S}(\mathcal{H})$ .*

Almost concavity requires more work in this case.

**Lemma 7.12** *Let  $\mathcal{C} \subset \mathcal{S}(\mathcal{H})$  be a compact convex set containing the maximally mixed state. Moreover, let  $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ ,  $p \in [0, 1]$ , and  $d \in \mathbb{N}$ ,  $d \geq 2$  the dimension of  $\mathcal{H}$ . Then,*

$$\widehat{D}_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2) \geq p\widehat{D}_{\mathcal{C}}(\rho_1) + (1-p)\widehat{D}_{\mathcal{C}}(\rho_2) - g_d(p).$$

Here,  $g_d(p) := \frac{d}{p^{1/d}}h(p) - \log(1 - p^{1/d})$  for  $p \in (0, 1)$  and  $g_d(0) := 0$ .

*Proof.* In order to apply the almost concavity of the BS-entropy, we need to control the minimal eigenvalue of  $\tau$ , the best approximation of  $\rho = p\rho_1 + (1-p)\rho_2$  in  $\mathcal{C}$ . To this end, we will use a strategy inspired by [28]. Let  $\tau_s$  be the state achieving the infimum in

$$\inf_{\tau \in \mathcal{C}} \widehat{D}\left(\rho \left\| s\tau + (1-s)\frac{\mathbb{1}}{d}\right.\right)$$

for some  $s \in (0, 1)$  which we will specify later. Clearly,

$$\widehat{D}_{\mathcal{C}}(\rho) \leq \widehat{D}\left(\rho \left\| s\tau_s + (1-s)\frac{\mathbb{1}}{d}\right.\right).$$

Furthermore, with  $\hat{\tau}$  a state such that  $\widehat{D}_C(\rho) = \widehat{D}(\rho \parallel \hat{\tau})$ ,

$$\begin{aligned} \widehat{D}\left(\rho \parallel s\tau_s + (1-s)\frac{\mathbb{1}}{d}\right) &\leq \widehat{D}\left(\rho \parallel s\hat{\tau} + (1-s)\frac{\mathbb{1}}{d}\right) \\ &\leq \widehat{D}_C(\rho) - \log s, \end{aligned}$$

as  $s\hat{\tau} + (1-s)\frac{\mathbb{1}}{d} \geq s\hat{\tau}$  and the logarithm is operator monotone. Since  $\widehat{D}_C(\rho) < \infty$  we have  $\ker \hat{\tau} \subseteq \ker \rho$ , thus, we can restrict  $\hat{\tau}$  to the support of  $\rho$ , where  $\hat{\tau}$  is positive definite. Combining this bound with theorem 6.3, we infer

$$\begin{aligned} \widehat{D}_C(p\rho_1 + (1-p)\rho_2) &\geq \widehat{D}\left(p\rho_1 + (1-p)\rho_2 \parallel s\tau_s + (1-s)\frac{\mathbb{1}}{d}\right) + \log s \\ &\geq p\widehat{D}_C(\rho_1) + (1-p)\widehat{D}_C(\rho_2) - \frac{d}{1-s}h(p) + \log s. \end{aligned}$$

Here, we have used point 1 of proposition 6.5. Finally, we have to choose  $s$  such that  $\frac{d}{1-s}h(p) - \log s$  goes to zero for  $p \rightarrow 0^+$  and is non-decreasing on  $p \in [0, 1/2]$ . It turns out that  $s = 1 - p^{1/d}$  is a convenient choice, see lemma G.1 and lemma G.2.  $\square$

**Remark 7.13** *Note that we could have substituted  $g_d$  in lemma 7.12 by a symmetrized version*

$$\tilde{g}_d(p) := \begin{cases} g_d(p) & p \in [0, 1/2] \\ g_d(1-p) & p \in [1/2, 1] \end{cases}$$

*in order to obtain*

$$\widehat{D}_C(p\rho_1 + (1-p)\rho_2) \geq p\widehat{D}_C(\rho_1) + (1-p)\widehat{D}_C(\rho_2) - \tilde{g}_d(p)$$

*for all  $p \in [0, 1]$  and  $\tilde{g}_d(0) = \tilde{g}_d(1) = 0$ . For the ALAFF method with  $s = 0$ , however, it is only relevant what happens on  $[0, 1/2]$ .*

The final estimate we need in order to apply the ALAFF method is proven in a very similar way as lemma 7.9.

**Lemma 7.14** *Let  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ . It holds that*

$$\sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2}\|\rho - \sigma\|_1 = 1}} |\widehat{D}_{\text{SEP}_{AB}}(\rho) - \widehat{D}_{\text{SEP}_{AB}}(\sigma)| \leq \log \min\{d_A, d_B\}.$$

*Proof.* Without loss of generality, let  $d_A \leq d_B$ . For a pure state  $|\psi\rangle$  with Schmidt decomposition  $\sum_{i=1}^{d_A} \lambda_i |i_A\rangle \otimes |i_B\rangle$ , let again

$$\tau_\psi = \frac{1}{d_A} \sum_{i=1}^{d_A} |i_A\rangle\langle i_A| \otimes |i_B\rangle\langle i_B|,$$

which is a separable state. Then,

$$\begin{aligned} \sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2}\|\rho - \sigma\|_1 = 1}} |\widehat{D}_{\text{SEP}_{AB}}(\rho) - \widehat{D}_{\text{SEP}_{AB}}(\sigma)| &\leq \sup_{|\psi\rangle \in \mathcal{S}(\mathcal{H})} \widehat{D}(|\psi\rangle\langle\psi| \parallel \tau_\psi) \\ &= \log d_A. \end{aligned}$$

In the above inequality, we have used that  $\widehat{D}_{\text{SEP}_{AB}}$  is positive and convex. Note that  $|\psi\rangle$  is in the support of  $\tau_\psi$ .  $\square$

**Theorem 7.15** For  $\varepsilon \in [0, 1]$ ,  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ , and  $d_{AB} \in \mathbb{N}$ ,  $d_{AB} \geq 2$ , it holds that for  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$

$$|\widehat{D}_{\text{SEP}_{AB}}(\rho) - \widehat{D}_{\text{SEP}_{AB}}(\sigma)| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)g_{d_{AB}} \left( \frac{\varepsilon}{1 + \varepsilon} \right).$$

Here,  $g_d(p) := \frac{d}{p^{1/d}} h(p) - \log(1 - p^{1/d})$  for  $p \in (0, 1)$  and  $g_d(0) = 0$ .

*Proof.* As shown in lemma G.3, it holds that  $g_d(p)/(1 - p)$  is non-decreasing on  $[0, 1]$  for all  $d \in \mathbb{N}$ ,  $d \geq 2$ . Thus, the assertion follows from remark 4.7 using lemma 7.11, lemma 7.12 with lemma G.1 and lemma G.2, and lemma 7.14.  $\square$

To end this section, let us investigate the choice

$$\mathcal{C}_0 := \{d_A^{-1} \mathbb{1}_A \otimes \sigma_B : \sigma_B \in \mathcal{S}(\mathcal{H}_B)\}.$$

From the discussion after eq. (5), we know that

$$\widehat{H}_\rho(A|B) \leq \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} -\widehat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) =: \widehat{H}_\rho^{\text{var}}(A|B),$$

but equality does not hold in general. This is different from the Umegaki relative entropy, where the conditional entropy coincides with its variational expression. Nonetheless, we obtain a continuity bound for  $\widehat{H}_\rho^{\text{var}}(A|B)$  from the approach in this section.

**Corollary 7.16** Let  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ . For  $\varepsilon \in [0, 1]$  and  $d_{AB} \in \mathbb{N}$ ,  $d_{AB} \geq 2$ , it holds that for  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$

$$|\widehat{H}_\rho^{\text{var}}(A|B) - \widehat{H}_\sigma^{\text{var}}(A|B)| \leq 2\varepsilon \log d_A + (1 + \varepsilon)g_{d_{AB}} \left( \frac{\varepsilon}{1 + \varepsilon} \right).$$

Here,  $g_d(p) := \frac{d}{p^{1/d}} h(p) - \log(1 - p^{1/d})$  for  $p \in (0, 1)$  and  $g_d(0) = 0$ .

*Proof.* It holds that for  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with  $\frac{1}{2} \|\rho - \sigma\|_1 \leq \varepsilon$

$$|\widehat{H}_\rho^{\text{var}}(A|B) - \widehat{H}_\sigma^{\text{var}}(A|B)| = |\widehat{D}_{\mathcal{C}_0}(\rho) - \widehat{D}_{\mathcal{C}_0}(\sigma)|,$$

since the normalization does not matter. Thus to apply ALAFF, we need to bound

$$\sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1}} |\widehat{D}_{\mathcal{C}_0}(\rho) - \widehat{D}_{\mathcal{C}_0}(\sigma)|.$$

Using eq. (28) and the fact that  $\widehat{D}_{\mathcal{C}_0}(\rho) \geq 0$  for all states  $\rho$ , we obtain

$$\begin{aligned} \sup_{\substack{\rho, \sigma \in \mathcal{S}(\mathcal{H}) \\ \frac{1}{2} \|\rho - \sigma\|_1 = 1}} |\widehat{D}_{\mathcal{C}_0}(\rho) - \widehat{D}_{\mathcal{C}_0}(\sigma)| &\leq \sup_{\rho \in \mathcal{S}(\mathcal{H})} -\widehat{H}_\rho^{\text{var}}(A|B) + \log d_A \\ &\leq 2 \log d_A. \end{aligned}$$

The assertion follows from combining the above with lemma 7.11, lemma 7.12 with lemma G.1 and lemma G.2, and lemma G.3 to apply remark 4.7.  $\square$

**Remark 7.17** Note that the findings of corollary 7.16 and proposition 6.7 provide a formal proof that  $\widehat{H}_\rho(A|B)$  and  $\widehat{H}_\rho^{\text{var}}(A|B)$  are different in general. Indeed, while we have just shown that the latter quantity is continuous on  $\mathcal{S}(\mathcal{H})$  as a consequence of the results of this section, in proposition 6.7 we showed that the former quantity is in general discontinuous on  $\mathcal{S}(\mathcal{H})$ .

## 7.7 Rains information

Inspired by the Rains bound from entanglement theory [66], for any divergence  $\mathbb{D}$ , the *generalized Rains bound* of a quantum state  $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  was defined in [82] by

$$\mathbb{R}(\rho_{AB}) := \min_{\sigma_{AB} \in \text{PPT}'(A:B)} \mathbb{D}(\rho_{AB} \parallel \sigma_{AB}),$$

where the minimization is taken over the Rains set

$$\text{PPT}'(A : B) := \left\{ \sigma_{AB} : \sigma_{AB} \geq 0, \left\| \sigma_{AB}^{T_B} \right\|_1 \leq 1 \right\}.$$

Where  $\cdot^{T_B}$  denotes the partial transpose in the  $B$ -system. This definition can be easily extended to channels in the following way. For a quantum channel  $T_{A' \rightarrow B} : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , we define

$$\mathbb{R}(T) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \mathbb{R}(T_{A' \rightarrow B}(\phi_{AA'})),$$

for  $\phi_{AA'}$  a purification of  $\rho_A$ . In particular, for the Umegaki relative entropy, we introduce the *Rains information* as

$$R(T) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \min_{\sigma_{AB} \in \text{PPT}'(A:B)} D(T_{A' \rightarrow B}(\phi_{AA'}) \parallel \sigma_{AB}),$$

as well as the BS-Rains information by

$$\widehat{R}(T) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \min_{\sigma_{AB} \in \text{PPT}'(A:B)} \widehat{D}(T_{A' \rightarrow B}(\phi_{AA'}) \parallel \sigma_{AB}).$$

In the rest of the subsection, we will drop the subindex from the channels whenever it is clear in which systems they act. In [29], it was proven that the latter two quantities constitute upper bounds to the quantum capacity of a quantum channel. Indeed, the following inequality holds for any channel  $T$ :

$$Q(T) \leq R(T) \leq \widehat{R}(T).$$

Moreover, the BS-Rains information is a limit of Rains informations induced by  $\alpha$ -geometric Rényi divergences, which can be written as single-letter formulas and computed via a semidefinite program (SDP), as shown in [29]. The study of these quantities is therefore of great interest for application in the context of strong converses of quantum capacities of channels.

Here, as a consequence of corollary 5.9 and corollary 6.11, respectively, we can provide continuity results for both the Rains information and the BS-Rains information, respectively, following the lines of theorem 7.10. Beforehand, we need to justify that both quantities are well-defined, i.e., that each of these quantities is attained at a certain  $\rho_A \in \mathcal{S}(\mathcal{H}_A)$  and  $\sigma_{AB} \in \text{PPT}'(A : B)$ , and thus the minimum and maximum in their definitions are properly written. For that, note that we are first taking an infimum on the second input over the compact set  $\text{PPT}'(A : B)$ . Then, the infimum is attained and the expression obtained is a continuous function, as we will show below in eq. (41). Next, we perform an optimization problem on the first input over another compact set, namely  $\mathcal{S}(\mathcal{H}_A)$ . Thus, that supremum is also attained and both Rains informations are well defined.

From now on, for simplicity and for similarity with the quantities introduced in the previous section, given  $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , let us define

$$D_{\text{PPT}'(A:B)}(\rho_{AB}) := \min_{\sigma_{AB} \in \text{PPT}'(A:B)} D(\rho_{AB} \parallel \sigma_{AB}).$$

Then, it is clear that we can rewrite, for a quantum channel  $T : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ ,

$$R(T) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} D_{\text{PPT}'(A:B)}(T(\phi_{AA'})),$$

for  $\phi_{AA'}$  a purification of  $\rho_A$ . The next step before applying the ALAFF method is bounding the difference between two Rains informations of two quantum channels. For that, we will use the  $1 \rightarrow 1$  norm of the difference between channels. Let us recall that for  $T : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  a quantum channel, its  $1 \rightarrow 1$  norm is given by

$$\|T\|_{1 \rightarrow 1} := \max_{\eta: \|\eta\|_1 \leq 1} \|T(\eta)\|_1.$$

For  $T_{A' \rightarrow B}$ , the  $1 \rightarrow 1$  norm coincides with the diamond norm. Now, as a consequence of lemma 7.9 and theorem 7.10 from the previous section, we can derive the following continuity bound for the Rains information.

**Theorem 7.18** *For  $\varepsilon \in [0, 1]$  and  $T_{A' \rightarrow B}^1, T_{A' \rightarrow B}^2 : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  two quantum channels with  $\frac{1}{2} \|T_{A' \rightarrow B}^1 - T_{A' \rightarrow B}^2\|_{1 \rightarrow 1} \leq \varepsilon$ , we have:*

$$|R(T_{A' \rightarrow B}^1) - R(T_{A' \rightarrow B}^2)| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right). \quad (40)$$

*Proof.* Let us drop the subscripts from the channels for ease of notation. Firstly, note that  $\text{SEP}_{AB} \subset \text{PPT}'(A : B)$ . Therefore,

$$R(T) = \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} D_{\text{PPT}'(A:B)}(T(\phi_{AA'})) \leq \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} D_{\text{SEP}_{AB}}(T(\phi_{AA'})).$$

Hence, in general

$$\begin{aligned} \max_{\substack{\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_{AB}) \\ \frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_1 = 1}} |D_{\text{PPT}'(A:B)}(\rho_{AB}) - D_{\text{PPT}'(A:B)}(\sigma_{AB})| &\leq \max_{\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})} D_{\text{PPT}'(A:B)}(\rho_{AB}) \\ &\leq \max_{\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})} D_{\text{SEP}_{AB}}(\rho_{AB}) \\ &\leq \log \min\{d_A, d_B\}, \end{aligned} \quad (41)$$

where in the last inequality we have used lemma 7.9. Following the lines of theorem 7.10, we have for  $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  with  $\frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_1 \leq \varepsilon$  the following continuity bound:

$$|D_{\text{PPT}'(A:B)}(\rho_{AB}) - D_{\text{PPT}'(A:B)}(\sigma_{AB})| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

Note that since  $\text{PPT}'(A : B)$  does not only contain states, but also subnormalized states, lemma 7.7 and lemma 7.8 are not directly applicable. One can however verify that the corresponding statements for  $\text{PPT}'(A : B)$  still hold using the same arguments. For simplicity, let us denote

$$b(\varepsilon) := \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

To estimate an upper bound on the difference that appears in eq. (40), first note that, given  $T^1, T^2 : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  two quantum channels with  $\frac{1}{2} \|T^1 - T^2\|_{1 \rightarrow 1} \leq \varepsilon$ , and  $\rho_A \in \mathcal{S}(\mathcal{H}_A)$  with  $\phi_{AA'}$  a purification of it, we have

$$\frac{1}{2} \|T^1(\phi_{AA'}) - T^2(\phi_{AA'})\|_1 \leq \frac{1}{2} \|T^1 - T^2\|_{1 \rightarrow 1} \leq \varepsilon.$$

Consider now  $\rho^1, \rho^2 \in \mathcal{S}(\mathcal{H}_A)$  with respective purifications  $\phi_{AA'}^1, \phi_{AA'}^2$ , the states in which the respective maxima of  $R(T^1)$  and  $R(T^2)$  are attained. Then, we clearly have, for  $i, j = 1, 2$  and  $i \neq j$ ,

$$|R(T^j) - D_{\text{PPT}'(A:B)}(T^i(\phi_{AA'}^j))| = |D_{\text{PPT}'(A:B)}(T^j(\phi_{AA'}^j)) - D_{\text{PPT}'(A:B)}(T^i(\phi_{AA'}^j))| \leq b(\varepsilon),$$

and thus,

$$R(T^i) \geq D_{\text{PPT}'(A:B)}(T^i(\phi_{AA'}^j)) \geq R(T^j) - b(\varepsilon).$$

Therefore, we can conclude

$$|R(T^1) - R(T^2)| \leq b(\varepsilon),$$

and consequently

$$|R(T^1) - R(T^2)| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)h\left(\frac{\varepsilon}{1 + \varepsilon}\right).$$

□

In a similar way, we can also prove uniform continuity and provide explicit continuity bounds for the BS-Rains information. Analogously to what we have done above for the Rains information, we can define for  $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  the following quantity:

$$\widehat{D}_{\text{PPT}'(A:B)}(\rho_{AB}) := \min_{\sigma_{AB} \in \text{PPT}'(A:B)} \widehat{D}(\rho_{AB} \| \sigma_{AB}),$$

and thus, we can rewrite, for a quantum channel  $T : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ ,

$$\widehat{R}(T) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \widehat{D}_{\text{PPT}'(A:B)}(T(\phi_{AA'})),$$

for  $\phi_{AA'}$  a purification of  $\rho_A$ . We can finally use lemma 7.14 and theorem 7.15 from the previous section, for the BS-entropy, to obtain a continuity bound for the BS-Rains information. However, the bound obtained, as well as the procedure employed to derive it, are a straightforward combination of the strategies of the continuity bound for the Rains information theorem 7.18 and the continuity bound for the BS-entropy of entanglement from theorem 7.15. Therefore, we omit it, to avoid unnecessary repetitions.

**Theorem 7.19** *For  $\varepsilon \in [0, 1]$  and  $T_{A' \rightarrow B}^1, T_{A' \rightarrow B}^2 : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \rightarrow \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  two quantum channels with  $\frac{1}{2}\|T^1 - T^2\|_{1 \rightarrow 1} \leq \varepsilon$ , we have:*

$$|\widehat{R}(T^1) - \widehat{R}(T^2)| \leq \varepsilon \log \min\{d_A, d_B\} + (1 + \varepsilon)g_{d_{AB}}\left(\frac{\varepsilon}{1 + \varepsilon}\right),$$

where  $g_d(t) := \frac{d}{t^{1/d}}h(t) - \log(1 - t^{1/d})$ .

## 8 Outlook

In this paper, we have introduced a generalisation of the Alicki-Fannes-Winter method by Shirokov and applied it to derive results of uniform continuity and explicit continuity bounds for divergences. We gave this generalisation the name ALAFF (cf. theorem 4.6) after the functions to which it applies (almost locally affine functions). The method allows deriving various continuity bounds for entropic quantities, by simply proofing (joint) convexity and almost (joint) concavity of the underlying divergence.

In particular, in the current paper, we have applied our ALAFF method to the specific cases of the Umegaki and the Belavkin-Staszewski relative entropies. For both of them, we have proven results of almost concavity (for the Umegaki case, our result is shown to be tight), and these, together with the well-known results of convexity for these quantities, have yielded a plethora of results of continuity bounds for both the Umegaki and BS-entropies, as well as for many other quantities derived from them. In particular, our results recover the previously known almost tight continuity bounds for the conditional entropy and the (conditional) mutual information.

A natural question arises from the findings of this paper: Is our method applicable to any other family of divergences? We expect this to be the case, since, as shown in section 2, our method only requires almost concavity and convexity (already known for divergences) in order to work. Therefore, a result of almost concavity with a “well-behaved” correction factor would be enough for the ALAFF method and is expected to exist, for families such as the  $\alpha$ -sandwiched Rényi divergences or the  $\alpha$ -geometric Rényi divergences, as they converge to the quantities studied in this paper. This possibility will be explored in a future manuscript.

Let us conclude this section, and our paper, with some analysis of the results obtained here. For both the Umegaki and the BS-entropies, we have presented results of almost concavity in order to provide some continuity bounds. However, while for the former (cf. theorem 5.1) we have shown that the result is tight, for the latter (cf. theorem 6.3) we are certain that there is room for improvement. Indeed, our almost concavity bound for the BS-entropy depends on the minimal eigenvalues of some of the states involved even in the simplified case of the BS-conditional entropy. In such a case, numerical simulations, as well as analytical proof, have shown us that there is a universal bound for the BS-conditional entropy of a state which is independent of the state involved. Therefore, we would expect an almost convexity result for the BS-conditional entropy being independent of the states involved, and this is clearly not the case at the moment. Nevertheless, there is no doubt that the BS-entropy, and quantities derived from it, are “pathological” in some sense. First of all, we have shown that the BS-conditional entropy exhibits discontinuities in the presence of vanishing eigenvalues (cf. proposition 6.7), as opposed to the conditional entropy, which behaves well in that setting. This motivates the idea that the minimal eigenvalue of the involved states should appear in the most general bounds of almost concavity and continuity. Additionally, we can compare some upper bounds of some entropic quantities derived from the Umegaki and the BS-entropy:

- For the relative entropy, we have the following 3 bounds:

$$-H_\rho(A|B) \leq \log d_A, \quad I_\rho(A : B) \leq 2 \log \min\{d_A, d_B\}, \quad D(\rho||\sigma) \leq \log \tilde{m}_\sigma^{-1}.$$

- For the BS-entropy, we have the following 3 bounds (cf. proposition 6.6):

$$-\hat{H}_\rho(A|B) \leq \log d_A, \quad \hat{I}_\rho(A : B) \leq \log d_A \tilde{m}_{(\rho_A)}^{-1}, \quad \hat{D}(\rho||\sigma) \leq \log m_\sigma^{-1}.$$

In the above  $m$ .,  $\tilde{m}$ . denote the minimal respectively minimal non-zero eigenvalue of the state in the index. It is remarkable that for the conditional and BS-conditional entropy and the mutual information, there appears no dependence on the minimal eigenvalue of the argument, whilst for the BS-mutual information this is the case.

Moreover, let us recall that, from the discussion in remark 7.17, we know that the conditional BS-entropy and its variational counterpart are different because the latter is continuous on  $\mathcal{S}(\mathcal{H})$  and the former is not. One could wonder whether the same difference appears for the BS-mutual information. Analogously to the case of the (Umegaki) mutual information, we could define four possible versions of such a notion by optimizing over one marginal, both or none. Remarkably, we



find that, when optimizing over both marginals, we have, assuming w.l.o.g.  $d_A \leq d_B$ ,

$$\widehat{I}_\rho^{\text{var}}(A : B) := \inf_{\sigma_A, \sigma_B} \widehat{D}(\rho_{AB} \| \sigma_A \otimes \sigma_B) \leq \log d_A - \widehat{H}_\rho^{\text{var}}(A|B) \leq 2 \log \min\{d_A, d_B\}.$$

Comparing this bound to the one shown above for  $\widehat{I}_\rho(A : B)$ , which we prove to be tight in proposition 6.6, we realize that the BS-mutual information and its variational counterpart (with optimization over both marginals) are also different in general.

To conclude, the literature concerning continuity bounds for entropic quantities is much broader than the results collected here. For Rényi and Tsallis entropies, many results concerning their continuity can be derived from other techniques, such as majorization flows, and can be found in texts such as [37–39]. Additionally, some of these results for the von Neumann entropy, Rényi and Tsallis entropies, as well as their classical counterparts, can be extended to energy-constrained systems in infinite dimensions, as shown in [12], [88] (see also the recent [73]). We leave for future work the possibility of extending the results presented here to a similar framework.

**Acknowledgements:** The authors are grateful to Li Gao for the interesting discussions, to Peter Brown for the example used in proposition 6.7, and to Mark Wilde for the discussion on the difference between BS-mutual information and its variational counterpart. Moreover, the authors would like to thank Álvaro Alhambra for pointing them to [34]. The authors also express their thanks to Ludovico Lami and Marco Tomamichel for spotting an error in Lemma 5.8 of an earlier version of the paper. A.B. acknowledges financial support from the European Research Council (ERC Grant Agreement No. 81876) and VILLUM FONDEN via the QMATH Centre of Excellence (Grant No.10059). A.P.H acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (grant PID2020- 113523GB-I00) and Comunidad de Madrid (grant QUITEMAD-CMS2018/TCS-4342). This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 470903074 – TRR 352.

## References

- [1] R. Alicki and M. Fannes. Continuity of quantum conditional information. *J. Phys. A: Math. Gen.*, 37(5):L55–L57, 2004. [3](#), [8](#)
- [2] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker’s Guide*. Springer Berlin, Heidelberg, third edition, 2006. [40](#)
- [3] H. Araki. On an inequality of Lieb and Thirring. *Lett. Math. Phys.*, 19:167–170, 1990. [13](#)
- [4] K. M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *J. Phys. A: Math. Theor.*, 40(28):8127, 2007. [3](#), [15](#)
- [5] K. M. R. Audenaert. Quantum skew divergence. *J. Math. Phys.*, 55:112202, 2014. [3](#), [12](#)
- [6] K. M. R. Audenaert and N. Datta.  $\alpha$ - $z$ -Rényi relative entropies. *J. Math. Phys.*, 56:022202, 2015. [17](#)
- [7] K. M. R. Audenaert and J. Eisert. Continuity bounds on the quantum relative entropy. *J. Math. Phys.*, 46(10):102104, 2005. [5](#), [19](#)
- [8] K. M. R. Audenaert and J. Eisert. Continuity bounds on the quantum relative entropy — II. *J. Math. Phys.*, 52:112201, 2011. [5](#), [18](#), [19](#)

- [9] I. Bardet, Á. Capel, L. Gao, A. Lucia, D. Pérez-García, and C. Rouzé. Entropy decay for Davies semigroups of a one dimensional quantum lattice. *arXiv preprint arXiv:2112.00601*, 2021. [3](#), [38](#)
- [10] I. Bardet, A. Capel, L. Gao, A. Lucia, D. Pérez-García, and C. Rouzé. Rapid thermalization of spin chain commuting Hamiltonians. *Phys. Rev. Lett.*, 130:060401, 2023. [3](#), [38](#)
- [11] I. Bardet, Á. Capel, A. Lucia, D. Pérez-García, and C. Rouzé. On the modified logarithmic Sobolev inequality for the heat-bath dynamics for 1D systems. *J. Math. Phys.*, 62:061901, 2021. [38](#)
- [12] S. Becker, N. Datta, and M. G. Jabbour. From classical to quantum: Uniform continuity bounds on entropies in infinite dimensions. *IEEE Trans. Inf. Theory*, 2023. [48](#)
- [13] V. P. Belavkin and P. Staszewski.  $C^*$ -algebraic generalization of relative entropy and entropy. *Ann. Inst. Henri Poincaré, section A*, 37(1):51–58, 1982. [3](#), [4](#), [6](#)
- [14] M. Berta, D. Sutter, and M. Walter. Quantum Brascamp-Lieb dualities. *Commun. Math. Phys.*, 2023. [39](#)
- [15] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate texts in mathematics*. Springer New York, NY, 1997. [36](#)
- [16] A. Bluhm and Á. Capel. A strengthened data processing inequality for the Belavkin-Staszewski relative entropy. *Rev. Math. Phys.*, 32(2):2050005, 2020. [3](#)
- [17] A. Bluhm, Á. Capel, and A. Pérez-Hernández. Exponential decay of mutual information for Gibbs states of local Hamiltonians. *Quantum*, 6:650, 2022. [3](#)
- [18] A. Bluhm, A. Capel, and A. Pérez-Hernández. Weak Quasi-Factorization for the Belavkin-Staszewski relative entropy. *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 118–123, 2021. [3](#), [7](#), [37](#), [39](#)
- [19] A. Bluhm, P. Capel, A. Gondolf, and A. Pérez-Hernández. General Continuity Bounds for Quantum Relative Entropies. *2023 IEEE International Symposium on Information Theory (ISIT) (to appear)*, 2023. [1](#), [3](#)
- [20] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens. Resource theory of quantum states out of thermal equilibrium. *Phys. Rev. Lett.*, 111(25):250404, 2013. [9](#), [13](#), [34](#)
- [21] O. Bratteli and D. W. Robinson. *Operator algebras and quantum-statistical mechanics II. Equilibrium states. Models in quantum statistical mechanics*. Texts and Monographs in Physics. Springer Berlin, Heidelberg, 1981. [5](#), [19](#)
- [22] Á. Capel, A. Lucia, and D. Pérez-García. Superadditivity of quantum relative entropy for general states. *IEEE Trans. Inf. Theory*, 64(7):4758–4765, 2017. [38](#)
- [23] Á. Capel, A. Lucia, and D. Pérez-García. Quantum conditional relative entropy and quasi-factorization of the relative entropy. *J. Phys. A: Math. Theor.*, 51(48):484001, 2018. [38](#)
- [24] Á. Capel, C. Rouzé, and D. Stilck França. The modified logarithmic Sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions. *Commun. Math. Phys. (to appear)*, 2023. [38](#)

- [25] E. A. Carlen. Trace inequalities and quantum entropy: An introductory course. In R. Sims and D. Ueltschi, editors, *Entropy and the Quantum*, volume 529 of *Contemporary Mathematics*, pages 73–140. American Mathematical Society, 2009. [22](#), [25](#), [59](#)
- [26] E. A. Carlen and E. H. Lieb. Remainder terms for some quantum entropy inequalities. *J. Math. Phys.*, 55:042201, 2014. [3](#)
- [27] E. A. Carlen and A. Vershynina. Recovery map stability for the data processing inequality. *J. Phys. A: Math. Theor.*, 53(3):035204, 2020. [35](#)
- [28] M. J. Donald and M. Horodecki. Continuity of relative entropy of entanglement. *Phys. Lett. A*, 264:257–260, 1999. [41](#)
- [29] K. Fang and H. Fawzi. Geometric Rényi divergence and its applications in quantum channel capacities. *Commun. Math. Phys.*, 384:1615–1677, 2021. [3](#), [44](#)
- [30] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Commun. Math. Phys.*, 31:291–294, 1973. [3](#)
- [31] H. Fawzi and O. Fawzi. Defining quantum divergences via convex optimization. *Quantum*, 5:387, 2021. [27](#)
- [32] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate Markov chains. *Commun. Math. Phys.*, 340:575–611, 2015. [35](#)
- [33] S. Golden. Lower bounds for the Helmholtz function. *Phys. Rev.*, 137(4B):B1127–B1128, 1965. [36](#)
- [34] G. Gour. Role of quantum coherence in thermodynamics. *PRX Quantum*, 3:040323, 2022. [34](#), [48](#)
- [35] G. Gour and M. Tomamichel. Optimal extensions of resource measures and their applications. *Phys. Rev. A*, 102:062401, Dec 2020. [18](#), [20](#)
- [36] G. Gour and M. Tomamichel. Entropy and relative entropy from information-theoretic principles. *IEEE Trans. Inf. Theory*, 67(10):6313–6327, 2021. [18](#), [20](#)
- [37] E. P. Hanson. Entropic continuity bounds & eventually entanglement-breaking channels. *arXiv preprint, arXiv:2010.02408*, 2020. [19](#), [48](#)
- [38] E. P. Hanson and N. Datta. Maximum and minimum entropy states yielding local continuity bounds. *J. Math. Phys.*, 59:042204, 2018. [48](#)
- [39] E. P. Hanson and N. Datta. Universal proofs of entropic continuity bounds via majorization flow. *arXiv preprint, arXiv:1909.06981*, 2019. [48](#)
- [40] E. P. Hanson, C. Rouzé, and D. Stileck França. Eventually entanglement breaking Markovian dynamics: Structure and characteristic times. *Ann. Henri Poincaré*, 21:1517–1571, 2020. [35](#)
- [41] M. Hayashi. *Quantum information: An introduction*. Springer Berlin, Heidelberg, 2006. [15](#), [16](#)
- [42] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Commun. Math. Phys.*, 246:359–374, 2004. [34](#)

- [43] F. Hiai and M. Mosonyi. Different quantum  $f$ -divergencies and the reversibility of quantum operations. *Rev. Math. Phys.*, 29(7):1750023, 2017. [3](#), [5](#), [6](#), [36](#), [38](#)
- [44] F. Hiai, M. Mosonyi, D. Petz, and C. Bény. Quantum  $f$ -divergences and error correction. *Rev. Mat. Phys.*, 23(07):691–747, 2011. [8](#), [22](#), [23](#), [60](#)
- [45] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.*, 143:99–114, 1991. [33](#), [36](#)
- [46] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. Universal recovery from a decrease of quantum relative entropy. *Ann. Henri Poincaré*, 19:2955, 2018. [35](#)
- [47] V. Katariya and M. M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. *Quantum Inf. Process.*, 20:78, 2021. [24](#)
- [48] K. Kato and F. G. S. L. Brandão. Quantum approximate Markov chains are thermal. *Commun. Math. Phys.*, 370:117–149, 2019. [35](#)
- [49] I. Kim and M. Ruskai. Bounds on the concavity of quantum entropy. *J. Math. Phys.*, 55:092201, 2014. [3](#)
- [50] I. H. Kim. Modulus of convexity for operator convex functions. *J. Math. Phys.*, 55:082201, 2014. [3](#)
- [51] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Math. Stat.*, 22(1):79–86, 1951. [3](#)
- [52] D. Leung and G. Smith. Continuity of quantum channel capacities. *Comm. Math. Phys.*, 292:201–215, 2009. [3](#)
- [53] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum mechanical entropy. *J. Math. Phys.*, 14:1938–1941, 1973. [34](#)
- [54] E. H. Lieb and W. E. Thirring. Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In E. H. Lieb, editor, *Studies in Mathematical Physics*, pages 269–304. Princeton University Press, 1976. [13](#)
- [55] G. Lindblad. Expectations and entropy inequalities for finite quantum systems. *Comm. Math. Phys.*, 39:111–119, 1974. [4](#), [11](#)
- [56] K. Matsumoto. Reverse test and characterization of quantum relative entropy. *arXiv preprint arXiv:1010.1030*, 2010. [3](#), [5](#), [22](#), [23](#), [36](#), [38](#)
- [57] K. Matsumoto. A new quantum version of  $f$ -divergence. In *Reality and Measurement in Algebraic Quantum Theory*, pages 229–273. Springer Singapore, 2018. [3](#), [41](#)
- [58] T. Metger and R. Renner. Security of quantum key distribution from generalised entropy accumulation. *arXiv preprint, arXiv:2203.04989*, 2022. [39](#)
- [59] M. Mosonyi and F. Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Trans. Inf. Theory*, 57(4):2474–2487, 2011. [3](#), [8](#)
- [60] M. A. Nielsen. Continuity bounds for entanglement. *Phys. Rev. A*, 61:064301, 2000. [3](#)

- [61] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 10th anniversary edition, 2010. [32](#)
- [62] M. Ohya and D. Petz. *Quantum Entropy and Its Use*. Texts and Monographs in Physics. Springer Berlin, Heidelberg, 1993. [13](#), [23](#), [25](#), [36](#), [40](#), [55](#)
- [63] D. Petz. Monotonicity of quantum relative entropy revisited. *Rev. Math. Phys.*, 15(1):79–91, 2003. [34](#)
- [64] D. Petz. *Quantum Information Theory and Quantum Statistics*. Springer Berlin, Heidelberg, 2008. [3](#), [33](#)
- [65] E. M. Rains. Bound on distillable entanglement. *Phys. Rev. A*, 60:179–184, 1999. [40](#)
- [66] E. M. Rains. A semidefinite program for distillable entanglement. *IEEE Trans. Inf. Theory*, 47(7):2921–2933, 2001. [44](#)
- [67] A. E. Rastegin. Upper continuity bounds on the relative  $q$ -entropy for  $q > 1$ . *J. Math. Phys.*, 52:062203, 2011. [17](#)
- [68] S. O. Scalet, A. M. Alhambra, G. Styliaris, and J. I. Cirac. Computable Rényi mutual information: Area laws and correlations. *Quantum*, 5:541, 2021. [3](#), [7](#)
- [69] M. E. Shirokov. Tight uniform continuity bounds for the quantum conditional mutual information, for the Holevo quantity, and for capacities of quantum channels. *J. Math. Phys.*, 58(10):102202, 2017. [3](#)
- [70] M. E. Shirokov. Adaptation of the Alicki-Fannes-Winter method for the set of states with bounded energy and its use. *Reports on Mathematical Physics*, 81(1):81–104, 2018. [15](#)
- [71] M. E. Shirokov. Uniform continuity bounds for information characteristics of quantum channels depending on input dimension and on input energy. *J. Phys. A: Math. Theor.*, 52(1):014001, 2019. [16](#)
- [72] M. E. Shirokov. Advanced Alicki-Fannes-Winter method for energy-constrained quantum systems and its use. *Quantum Inf. Process.*, 19:164, 2020. [3](#), [8](#), [15](#)
- [73] M. E. Shirokov. Quantum relative entropy: general convergence criterion and preservation of convergence under completely positive linear maps. *arXiv preprint: arXiv 2205.10341*, 2022. [48](#)
- [74] M. E. Shirokov. Quantifying continuity of characteristics of composite quantum systems. *Physica Scripta*, 98(4):042002, mar 2023. [3](#), [8](#), [9](#)
- [75] D. Sutter. *Approximate Quantum Markov Chains*. SpringerBriefs in Mathematical Physics. Springer Cham, 2018. [17](#), [35](#)
- [76] D. Sutter, M. Berta, and M. Tomamichel. Multivariate trace inequalities. *Commun. Math. Phys.*, 352:37–58, 2017. [13](#), [25](#), [55](#)
- [77] D. Sutter and R. Renner. Necessary criterion for approximate recoverability. *Ann. Henri Poincaré*, 19(10):3007–3029, 2018. [16](#), [35](#)

- [78] P. Svetlichnyy and T. A. B. Kennedy. Decay of quantum conditional mutual information for purely generated finitely correlated states. *J. Math. Phys.*, 63:072201, 2022. [35](#)
- [79] B. Synak-Radtke and M. Horodecki. On asymptotic continuity of functions of quantum states. *J. Phys. A: Math. Gen.*, 39(26):L423, 2006. [3](#)
- [80] C. J. Thompson. Inequality with applications in statistical mechanics. *J. Math. Phys.*, 6:1812–1813, 1965. [36](#)
- [81] M. Tomamichel. *Quantum Information Processing with Finite Resources*. Springer Cham, 2016. [8](#)
- [82] M. Tomamichel, M. M. Wilde, and A. Winter. Strong converse rates for quantum communication. *IEEE Trans. Inf. Theory*, 63(1):715–727, 2017. [44](#)
- [83] H. Umegaki. Conditional expectation in an operator algebra IV. Entropy and information. *Kodai Math. Sem. Rep.*, 14:59–85, 1962. [3](#), [4](#), [6](#)
- [84] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619–1633, 1998. [40](#)
- [85] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. Quantifying entanglement. *Phys. Rev. Lett.*, 78:2275–2279, 1997. [40](#)
- [86] A. Vershynina. Upper continuity bound on the quantum quasi-relative entropy. *J. Math. Phys.*, 60:102201, 2019. [5](#), [17](#), [19](#)
- [87] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2016. [3](#), [5](#), [12](#), [15](#), [16](#)
- [88] A. Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Commun. Math. Phys.*, 347:291–313, 2016. [3](#), [5](#), [8](#), [15](#), [34](#), [40](#), [41](#), [48](#)
- [89] Y. Zhai, B. Yang, and Z. Xi. Belavkin–Staszewski relative entropy, conditional entropy, and mutual information. *Entropy*, 24:837, 2022. [3](#), [7](#)

## A Numerical investigation of the variational definition of the BS-conditional entropy

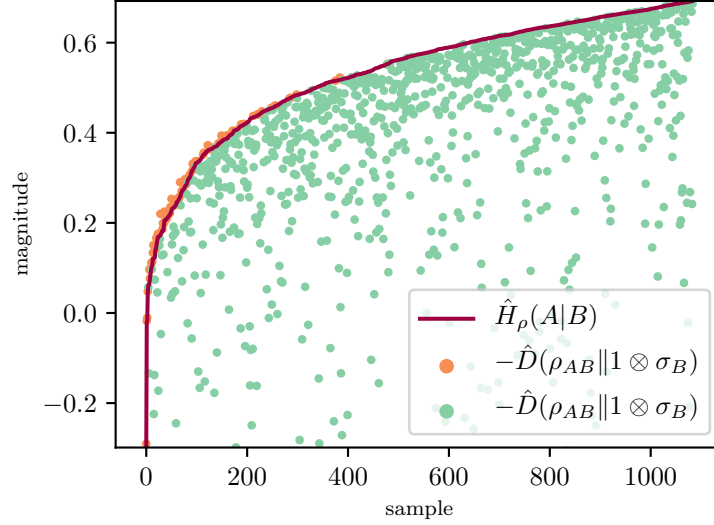


Figure 6: The red line is the BS-conditional entropy defined via the partial trace evaluated at  $\rho_{AB}$ . The dots are the BS-entropy between the state  $\rho_{AB}$  and  $\mathbb{1}_A \otimes \sigma_B$  with  $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$  sampled at random. The orange dots are the cases when the  $-\hat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$  exceeds  $\hat{H}(A|B)_\rho$ . We sampled a total of 100.000 pairs of  $\rho_{AB}$  and  $\sigma_B$  and evaluated both  $\hat{H}(A|B)_\rho$  and  $-\hat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)$ . Only a tenth of all samples were kept in addition to the ones that violated the bound. Those were then plotted in ascending order w.r.t the magnitude of their BS-conditional entropy. We further controlled the minimal eigenvalue and set  $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^2 \otimes \mathbb{C}^2$  to reduce the risk of numerical flaws.

## B Supplements to the proof of theorem 5.1

We will now show that the result of the inequality in eq. (12) is still true, even if  $\rho_1, \rho_2, \sigma_1, \sigma_2$  are not full rank. We have that

$$\ker \sigma \subseteq \ker \sigma_1 \subseteq \ker \rho_1.$$

If  $\ker \sigma \subsetneq \ker \rho_1$  we set

$$\tilde{\Pi}_{\rho_1} := P_{\ker \rho_1 \cap (\ker \sigma)^\perp} \quad , \quad \Pi_{\rho_1} := \left\| \tilde{\Pi}_{\rho_1} \right\|_1^{-1} \tilde{\Pi}_{\rho_1} ,$$

and if  $\ker \sigma \subsetneq \ker \sigma_1$ ,

$$\tilde{\Pi}_{\sigma_1} := P_{\ker \sigma_1 \cap (\ker \sigma)^\perp} \quad , \quad \Pi_{\sigma_1} := \left\| \tilde{\Pi}_{\sigma_1} \right\|_1^{-1} \tilde{\Pi}_{\sigma_1} ,$$

normalised projections on the spaces in the index. Both of the latter are quantum states and fulfil

$$\Pi_{\rho_1} \rho_1 = \rho_1 \Pi_{\rho_1} = 0, \quad \Pi_{\sigma_1} \sigma_1 = \sigma_1 \Pi_{\sigma_1} = 0, \quad \Pi_{\sigma_1} \rho_1 = \rho_1 \Pi_{\sigma_1} = 0. \quad (42)$$

For  $1 > \varepsilon > 0$  and  $1 > \delta > 0$ , let

$$\rho_{1,\varepsilon} = \begin{cases} \varepsilon \Pi_{\rho_1} + (1 - \varepsilon) \rho_1 & \text{if } \ker \sigma \subsetneq \ker \rho_1 \\ \rho_1 & \text{if } \ker \sigma = \ker \rho_1 \end{cases},$$

$$\sigma_{1,\delta} = \begin{cases} \delta \Pi_{\sigma_1} + (1 - \delta) \sigma_1 & \text{if } \ker \sigma \subsetneq \ker \sigma_1 \\ \sigma_1 & \text{if } \ker \sigma = \ker \sigma_1 \end{cases}.$$

We then have that  $\ker \rho_{1,\varepsilon} = \ker \sigma_{1,\delta} = \ker \sigma$ . This means, however, considering  $\text{Tr} [\rho_{1,\varepsilon}(\log \sigma - \log \sigma_{1,\delta})]$  we can reduce to the subspace where they are all full rank. We then apply the Peierls-Bogoliubov inequality [62] and the multivariate trace inequality by Sutter et al. [76, Corollary 3.3]

$$\begin{aligned} \text{Tr} [\rho_{1,\varepsilon}(\log \sigma - \log \sigma_{1,\delta})] &\leq \log \text{Tr} [\exp(\log(\rho_{1,\varepsilon}) + \log(\sigma) - \log(\sigma_{1,\delta}))] \\ &\leq \log \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_{1,\varepsilon} \sigma_{1,\delta}^{\frac{it-1}{2}} \sigma \sigma_{1,\delta}^{\frac{-it-1}{2}} \right]. \end{aligned} \quad (43)$$

Both of the traces on the LHS and RHS of eq. (43) can without change be extended to the full Hilbert space again. Next, we take limits on both sides of the inequality and in doing so recover the claim. We first note that the limit  $\varepsilon \rightarrow 0$  requires no more argument as both sides are linear in  $\varepsilon$ . Hence, we get

$$\text{Tr} [\rho_1(\log \sigma - \log \sigma_{1,\delta})] \leq \log \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_1 \sigma_{1,\delta}^{\frac{it-1}{2}} \sigma \sigma_{1,\delta}^{\frac{-it-1}{2}} \right]. \quad (44)$$

The limit  $\delta \rightarrow 0$  on the other hand is, in the case of  $\ker \sigma \subsetneq \ker \sigma_1$ , a little more involved. Due to the orthogonality in eq. (42) we cannot only split up the logarithm but also eliminate terms. More specifically, we have

$$\log \sigma_{1,\delta} = \log(\delta \Pi_{\sigma_1}) + \log((1 - \delta) \sigma_1),$$

where the logarithms in the RHS have to be understood as living in the support of the respective argument (and complemented with zeros in the rest). Hence, we obtain for the LHS of eq. (44)

$$\begin{aligned} \text{Tr} [\rho_1(\log \sigma - \log \sigma_{1,\delta})] &= \text{Tr} [\rho_1(\log \sigma - \log(\delta \Pi_{\sigma_1} + (1 - \delta) \sigma_1))] \\ &= \text{Tr} [\rho_1(\log \sigma - \log((1 - \delta) \sigma_1))] + \text{Tr} [\rho_1 \log(\delta \Pi_{\sigma_1})] \\ &= \text{Tr} [\rho_1(\log \sigma - \log((1 - \delta) \sigma_1))] \\ &= \text{Tr} [\rho_1(\log \sigma - \log \sigma_1)] + \log(1 - \delta). \end{aligned}$$

Moreover, for the RHS of eq. (44) we use that

$$\sigma_{1,\delta}^z = \delta^z \Pi_{\sigma_1}^z + (1 - \delta)^z \sigma_1^z,$$



for any  $z \in \mathbb{C}$ , where the last exponential has to be understood again in the support of the respective argument. Thus, we obtain

$$\begin{aligned}
\mathrm{Tr} \left[ \rho_1 \sigma_{1,\delta}^{\frac{it-1}{2}} \sigma \sigma_{1,\delta}^{\frac{-it-1}{2}} \right] &= (1-\delta)^{-1} \mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right] \\
&+ (1-\delta)^{\frac{it-1}{2}} \delta^{\frac{-it-1}{2}} \mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \Pi_{\sigma_1}^{\frac{-it-1}{2}} \right] \\
&+ \delta^{\frac{it-1}{2}} (1-\delta)^{\frac{-it-1}{2}} \mathrm{Tr} \left[ \rho_1 \Pi_{\sigma_1}^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right] \\
&+ \delta^{-1} \mathrm{Tr} \left[ \rho_1 \Pi_{\sigma_1}^{\frac{it-1}{2}} \sigma \Pi_{\sigma_1}^{\frac{-it-1}{2}} \right] \\
&= (1-\delta)^{-1} \mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right].
\end{aligned}$$

Taking the limit  $\delta \rightarrow 0$  now directly follows from the continuity of the logarithm. We thereby conclude

$$p \mathrm{Tr} [\rho_1 (\log(\sigma) - \log(\sigma_1))] \leq p \log \int_{-\infty}^{\infty} dt \beta_0(t) \mathrm{Tr} \left[ \rho_1 \sigma_1^{\frac{it-1}{2}} \sigma \sigma_1^{\frac{-it-1}{2}} \right],$$

for  $\sigma_1, \sigma_2, \rho_1$  not full rank.

## C Proof of proposition 5.2

We first of all note that for all  $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$  we have  $\frac{1}{2} \|\rho_1 - \rho_2\|_1 \leq 1$ , hence as a direct consequence  $f_{c_1, c_2} + \frac{1}{2} \|\rho_1 - \rho_2\|_1 h \leq f_{c_1, c_2} + h$ . We therefore will drop the  $\frac{1}{2} \|\rho_1 - \rho_2\|_1$  in front of the  $h$  here already.

1. If  $\sigma_1 = \sigma_2 =: \sigma$ , we find for  $j = 1, 2$  that

$$c_j = \int_{-\infty}^{\infty} dt \beta_0(t) \mathrm{Tr} \left[ \rho_j \sigma^{\frac{it-1}{2}} \sigma \sigma^{\frac{-it-1}{2}} \right] = \int_{-\infty}^{\infty} dt \beta_0(t) \mathrm{Tr} [\rho_j] = 1.$$

The reduction of  $f_{c_1, c_2} + h$  to  $h$  then happens because  $\log(p + (1-p)) = \log(1) = 0$  gives  $f_{c_1, c_2} = 0$ .

2. With  $j, k = 1, 2, j \neq k$  and  $\tilde{m} \rho_j \leq \sigma_j$ , we find

$$\sigma_j^{\frac{it-1}{2}} \rho_j \sigma_j^{\frac{-it-1}{2}} \leq \sigma_j^{\frac{it-1}{2}} \tilde{m}^{-1} \sigma_j \sigma_j^{\frac{-it-1}{2}} \leq \tilde{m}^{-1} P_{\sigma_j} \leq \tilde{m}^{-1} \mathbf{1}$$

where  $P_{\sigma_j}$  is the projection onto the support of  $\sigma_j$ . We therefore find

$$c_j \leq \int_{-\infty}^{\infty} dt \beta_0(t) \tilde{m}^{-1} \mathrm{Tr} [\sigma_k] = \tilde{m}^{-1}.$$

By the monotonicity of the logarithm, we obtain  $f_{c_1, c_2} \leq f_{\tilde{m}^{-1}, \tilde{m}^{-1}}$  and hence  $f_{c_1, c_2} + h \leq f_{\tilde{m}^{-1}, \tilde{m}^{-1}} + h$ .

3. For  $j, k = 1, 2, j \neq k$  we have

$$\begin{aligned}
c_j &= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_{j,AB} (\mathbb{1}_A \otimes \rho_{j,B})^{\frac{it-1}{2}} \mathbb{1}_A \otimes \rho_{k,B} (\mathbb{1}_A \otimes \rho_{j,B})^{-\frac{it-1}{2}} \right] \\
&= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_{j,AB} \mathbb{1}_A \otimes \left( \rho_{j,B}^{\frac{it-1}{2}} \rho_{k,B} \rho_{j,B}^{-\frac{it-1}{2}} \right) \right] \\
&= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_{j,B} \left( \rho_{j,B}^{\frac{it-1}{2}} \rho_{k,B} \rho_{j,B}^{-\frac{it-1}{2}} \right) \right] \\
&= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} [\rho_{k,B}] = 1.
\end{aligned}$$

We used that the functional calculus has the property that  $f(A \otimes B) = f(A) \otimes f(B)$  for  $A, B$  self-adjoint, as can easily be verified by direct computation, and that the trace is cyclic. This gives us  $f_{c_1, c_2} = f_{1,1} = 0$  which concludes the claim.

4. The derivative of  $p \mapsto \frac{1}{1-p} h(p)$  at  $p \in (0, 1)$  is  $-\frac{\log(p)}{(1-p)^2} \geq 0$ , which proves the second assertion. For  $p \mapsto \frac{1}{1-p} f_{m_1, m_2}(p) = \frac{p}{1-p} \log(p + m_1(1-p)) + \log(1-p + m_2 p)$  we use similar reasoning. First we use that  $m_2 \geq 1$  hence  $\log(1-p + m_2 p) = \log(1 + (m_2 - 1)p)$  is monotone in  $p$ , i.e. in particular non-decreasing. Second we note that  $p \mapsto \frac{p}{1-p} \log(p + m_1(1-p))$  is monotone in  $p$ , because forming the derivative at  $p \in (0, 1)$ , we get

$$\begin{aligned}
&\frac{1}{(1-p)^2} \left( \frac{p}{p + (1-p)m_1} + \log(p + m_1(1-p)) - p \right) \\
&\geq \frac{1}{(1-p)^2} \left( \frac{p}{p + (1-p)m_1} + \frac{p + (1-p)m_1 - 1}{p + m_1(1-p)} - p \right) \\
&= \frac{1}{(1-p)^2} \left( \frac{m_1(1-p) + 2p - 1}{p + (1-p)m_1} - p \right) \\
&= \frac{1}{(1-p)^2} \left( \frac{m_1(1-p) + 2p - 1 - p(p + (1-p)m_1)}{p + (1-p)m_1} \right) \\
&= \frac{1}{(1-p)^2} \left( \frac{(m_1 - 1)(p - 1)^2}{p + (1-p)m_1} \right) \\
&\geq 0.
\end{aligned}$$

We used that for  $x \geq 1$ ,  $\log(x) \geq \frac{x-1}{x}$  (this can be seen by taking the derivative and realizing that both sides coincide for  $x = 1$ ) and  $m_1 \geq 1$ . This concludes the claim.

## D Proof of lemma 5.12

We first show that for  $s \geq \tilde{m}$ ,  $\mathcal{S}_0$  is  $s$ -perturbed  $\Delta$ -invariant. For that purpose let  $\sigma_1, \sigma_2 \in \mathcal{S}_0$ , then we find

$$\Delta^\pm(\sigma_1, \sigma_2, \rho) = s\rho + (1-s)[\sigma_1 - \sigma_2]_\pm \geq \tilde{m}\rho,$$

which immediately gives the kernel inclusion as well as the condition to be lower bounded by  $\tilde{m}\rho$ . Therefore,  $\Delta^\pm(\sigma_1, \sigma_2, \tau) \in \mathcal{S}_0$  which makes  $\mathcal{S}_0$  an  $s$ -perturbed  $\Delta$ -invariant set. We show the other

direction by contrapositive. Let  $s < \tilde{m}$ . Since  $\tilde{m} < 1$  and  $\text{rank } \rho \geq 2$  we find an  $\varepsilon > 0$  and two orthonormal  $|0\rangle, |1\rangle \in \text{supp } \rho$ , such that  $\tilde{m}\rho < \rho - \frac{\varepsilon}{2} |i\rangle\langle i|$  for  $i = 0, 1$ . We then have that

$$\begin{aligned}\sigma_1 &= \rho + \frac{\varepsilon}{2} |0\rangle\langle 0| - \frac{\varepsilon}{2} |1\rangle\langle 1| \\ \sigma_2 &= \rho - \frac{\varepsilon}{2} |0\rangle\langle 0| + \frac{\varepsilon}{2} |1\rangle\langle 1|\end{aligned}$$

manifestly are contained in  $\mathcal{S}_0$ . Furthermore,  $\frac{1}{2} \|\sigma_1 - \sigma_2\|_1 = \varepsilon$  and

$$\begin{aligned}\varepsilon^{-1}[\sigma_1 - \sigma_2]_+ &= |0\rangle\langle 0| \\ \varepsilon^{-1}[\sigma_1 - \sigma_2]_- &= |1\rangle\langle 1|\end{aligned}$$

We will now show that there exists no  $\tau \in \mathcal{S}(\mathcal{H})$  such that  $\Delta^\pm(\sigma_1, \sigma_2, \tau) \in \mathcal{S}_0$  again, meaning  $\mathcal{S}_0$  is not  $s$ -perturbed  $\Delta$ -invariant. Assume there is an operator  $\tau \geq 0$  such that  $\Delta^\pm(\sigma_1, \sigma_2, \tau) \in \mathcal{S}_0$  we then would have

$$\begin{aligned}|0\rangle\langle 0|^\perp \Delta^+(\sigma_1, \sigma_2, \tau) |0\rangle\langle 0|^\perp &= |0\rangle\langle 0|^\perp s\tau |0\rangle\langle 0|^\perp \geq \tilde{m} |0\rangle\langle 0|^\perp \rho |0\rangle\langle 0|^\perp \\ |1\rangle\langle 1|^\perp \Delta^-(\sigma_1, \sigma_2, \tau) |1\rangle\langle 1|^\perp &= |1\rangle\langle 1|^\perp s\tau |1\rangle\langle 1|^\perp \geq \tilde{m} |1\rangle\langle 1|^\perp \rho |1\rangle\langle 1|^\perp\end{aligned}\tag{45}$$

where  $|i\rangle\langle i|^\perp := P_\rho - |i\rangle\langle i|$  for  $i = 0, 1$ . Here  $P_\rho$  is the projection on the support of  $\rho$ . We further used  $\Delta^\pm(\sigma_1, \sigma_2, \tau) \geq \tilde{m}\rho$  as  $\Delta^\pm(\sigma_1, \sigma_2, \tau)$  are in  $\mathcal{S}_0$  by assumption. To fulfil eq. (45) we clearly need to choose  $s > 0$  and since  $s < \tilde{m}$  we directly obtain the conditions

$$|0\rangle\langle 0|^\perp \tau |0\rangle\langle 0|^\perp \gneq |0\rangle\langle 0|^\perp \rho |0\rangle\langle 0|^\perp \quad \text{and} \quad |1\rangle\langle 1|^\perp \tau |1\rangle\langle 1|^\perp \gneq |1\rangle\langle 1|^\perp \rho |1\rangle\langle 1|^\perp.$$

This gives us,

$$\begin{aligned}\text{Tr } [\tau] &\geq \text{Tr} \left[ |0\rangle\langle 0|^\perp \tau |0\rangle\langle 0|^\perp + |0\rangle\langle 0| \tau |0\rangle\langle 0| \right] = \text{Tr} \left[ |0\rangle\langle 0|^\perp \tau |0\rangle\langle 0|^\perp + |0\rangle\langle 0| |1\rangle\langle 1|^\perp \tau |1\rangle\langle 1|^\perp |0\rangle\langle 0| \right] \\ &> \text{Tr} \left[ |0\rangle\langle 0|^\perp \rho |0\rangle\langle 0|^\perp + |0\rangle\langle 0| |1\rangle\langle 1|^\perp \rho |1\rangle\langle 1|^\perp |0\rangle\langle 0| \right] = \text{Tr} \left[ |0\rangle\langle 0|^\perp \rho |0\rangle\langle 0|^\perp + |0\rangle\langle 0| \rho |0\rangle\langle 0| \right] \\ &= \text{Tr} [P_\rho \rho] = \text{Tr} [\rho] = 1,\end{aligned}$$

where we used that  $|0\rangle$  and  $|1\rangle$  are orthogonal, hence  $|0\rangle\langle 0| |1\rangle\langle 1|^\perp = |1\rangle\langle 1|^\perp |0\rangle\langle 0| = |0\rangle\langle 0|$  and  $|0\rangle\langle 0|^2 = |0\rangle\langle 0|, (|0\rangle\langle 0|^\perp)^2 = |0\rangle\langle 0|^\perp$ . We thus conclude  $\tau \notin \mathcal{S}(\mathcal{H})$  proving the claim.

## E Proof of proposition 6.5

1. If  $\sigma_1 = \sigma_2 = \sigma$ , then for  $j = 1, 2$

$$\begin{aligned}\hat{c}_j &= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_j (\rho_j^{1/2} \sigma^{-1} \rho_j^{1/2})^{\frac{it+1}{2}} \rho_j^{-1/2} \sigma \rho_j^{-1/2} (\rho_j^{1/2} \sigma^{-1} \rho_j^{1/2})^{-\frac{it+1}{2}} \right] \\ &= \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} [\rho_j] = \int_{-\infty}^{\infty} dt \beta_0(t) = 1\end{aligned}$$

which gives us immediately  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h = \hat{c}_0 h$ .

2. For  $j, k = 1, 2$  with  $j \neq k$  we first have  $\sigma_k \leq m^{-1}\sigma_j$  giving us

$$\begin{aligned}\hat{c}_j &\leq \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} \left[ \rho_j (\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2})^{\frac{it+1}{2}} \rho_j^{-1/2} m^{-1} \sigma_j \rho_j^{-1/2} (\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2})^{\frac{-it+1}{2}} \right] \\ &= m^{-1} \int_{-\infty}^{\infty} dt \beta_0(t) \text{Tr} [\rho_j] = m^{-1}.\end{aligned}$$

Since  $\hat{c}_0 \leq m^{-1}$  and because the logarithm is monotone this immediately gives  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h \leq f_{m^{-1}, m^{-1}} + m^{-1} h$ .

3. The proof is along the same lines as the one for 2., however with  $\sigma_j = d_A^{-1} \mathbb{1}_A \otimes \rho_{j,B}$ . We just have to show that the minimal eigenvalue of  $\sigma_j$  is bounded from below by  $m$ . We use that  $T_A : \tau \mapsto d_A^{-1} \mathbb{1}_A \otimes \tau_B$  is a conditional expectation and that  $d_A^{-1} \mathbb{1}_A \otimes \tau_B$  is full rank if  $\tau$  was full rank [25, Theorem 4.13]. This means, however,

$$(d_A^{-1} \mathbb{1}_A \otimes \rho_{j,B})^{-1} = T_A(\rho_j)^{-1} \leq T_A(\rho_j^{-1}),$$

where we used [25, Theorem 4.16]. This gives us

$$\|(d_A^{-1} \mathbb{1}_A \otimes \rho_{j,B})^{-1}\|_{\infty} \leq \|T_A(\rho_j^{-1})\|_{\infty} \leq \|\rho_j^{-1}\|_{\infty} \leq m^{-1}. \quad (46)$$

Hence, we have that  $\|(d_A^{-1} \otimes \rho_{j,B})^{-1}\|_{\infty}^{-1}$  the minimal eigenvalue of  $d_A^{-1} \otimes \rho_{j,B}$  is bounded from below by  $m$ . From here on the proof is analogous to the one in 2. We obtain  $f_{\hat{c}_1, \hat{c}_2} + \hat{c}_0 h \leq f_{m^{-1}, m^{-1}} + \hat{c}_0 h$  and again use eq. (46) to get  $f_{m^{-1}, m^{-1}} + \hat{c}_0 h \leq f_{m^{-1}, m^{-1}} + m^{-1} h$ .

4. The proof is completely analogous to the one in 4. of appendix C.

## F Proof of proposition 6.6

1. We begin with the BS-conditional information. The upper bound on  $\hat{H}(A|B)$  can be obtained by

$$\hat{H}_{\rho}(A|B) = -\hat{D}(\rho_{AB} \| d_A^{-1} \mathbb{1}_A \otimes \rho_B) + \log d_A \leq \log d_A.$$

where we used the non-negativity of  $\hat{D}(\cdot \| \cdot)$  on quantum states. The bound is attained if one inserts the maximally mixed state, i.e.,  $\rho_{AB} = d_{AB}^{-1} \mathbb{1}_{AB}$ . For the lower bound we use that  $-\hat{D}(\cdot \| \cdot)$  is jointly concave and  $\text{Tr}_A[\cdot]$  linear which means without loss of generality one can assume  $\rho$  to be pure, i.e., a rank one projection. Then

$$\begin{aligned}\hat{H}_{|\psi\rangle\langle\psi|}(A|B) &= -\hat{D}(|\psi\rangle\langle\psi| \| \mathbb{1}_A \otimes P_B) = -\text{Tr} \left[ |\psi\rangle\langle\psi| \log |\psi\rangle\langle\psi|^{1/2} (\mathbb{1}_A \otimes P_B^{-1}) |\psi\rangle\langle\psi|^{1/2} \right] \\ &= -\log \text{Tr} [|\psi\rangle\langle\psi| (\mathbb{1}_A \otimes P_B^{-1})] = -\log \text{Tr} [P_B P_B^{-1}],\end{aligned}$$

with  $P_B = \text{Tr}_A[|\psi\rangle\langle\psi|]$ . Employing the Schmidt decomposition to  $|\psi\rangle\langle\psi|$  we find that

$$P_B = \sum_{i=1}^d \lambda_i^2 P_i$$

with  $P_i$  orthogonal rank one projections on  $\mathcal{H}_B$ ,  $\lambda_i^2 > 0$  and  $\sum_{i=1}^d \lambda_i^2 = 1$ . Further  $d \leq \min\{d_A, d_B\}$  the Schmidt rank. This gives us that

$$\text{Tr} [P_B P_B^{-1}] = \sum_{i=1}^d \lambda_i^2 \lambda_i^{-2} = d \leq \min\{d_A, d_B\}.$$

Through monotonicity of the logarithm, we obtain the lower bound, i.e.,

$$\widehat{H}_\rho(A|B) \geq -\log \min\{d_A, d_B\}.$$

This bound is attained for  $\rho$  a pure state with full Schmidt rank, which can directly be seen from the above calculations.

2. We now tackle the BS-mutual information. The lower bound, i.e.  $\widehat{I}_\rho(A : B) \geq 0$ , is a direct consequence of the data processing inequality [44]. Applying  $\text{Tr}_A[\cdot]$ , we find

$$\widehat{I}_\rho(A : B) = \widehat{D}(\rho_{AB} \| \rho_A \otimes \rho_B) \geq \widehat{D}(\rho_B \| \rho_B) = 0.$$

To proof the upper bound, we w.l.o.g assume that  $\|\rho_A^{-1}\|_\infty \leq \|\rho_B^{-1}\|_\infty$ . We then use that  $\rho_A \otimes \rho_B \geq \|\rho_A^{-1}\|_\infty^{-1} P_{\rho_A} \otimes \rho_B$ , where  $P_{\rho_A}$  is the projection to the support of  $\rho_A$ . This gives us

$$\begin{aligned} \widehat{I}_\rho(A : B) &= \widehat{D}(\rho_{AB} \| \rho_A \otimes \rho_B) \leq \widehat{D}(\rho_{AB} \| P_{\rho_A} \otimes \rho_B) + \log \|\rho_A^{-1}\|_\infty \\ &= \widehat{D}(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B) + \log \|\rho_A^{-1}\|_\infty = -\widehat{H}_\rho(A|B) + \log \|\rho_A^{-1}\|_\infty \\ &\leq \log \min\{d_A, d_B\} + \log \|\rho_A^{-1}\|_\infty \\ &\leq \log \min\{d_A, d_B\} + \log \min\{\|\rho_A^{-1}\|_\infty, \|\rho_B^{-1}\|_\infty\} \end{aligned}$$

In the second equality we used that  $(\ker \rho_A) \otimes \mathcal{H}_B \subseteq \ker \rho_{AB}$ , so extending  $P_{\rho_A}$  to  $\mathbb{1}_A$  has no effect. With the next example, we will see that the bound is tight and scales with  $\log \max\{\|\rho_A^{-1}\|_\infty, \|\rho_B^{-1}\|_\infty\}$  in some cases. For that purpose let  $d_A \in \mathbb{N}$ ,  $d_A \geq 2$  and a bipartite space  $\mathcal{H}_A \otimes \mathcal{H}_B$  with  $\mathcal{H}_A$  having dimension  $d_A$  and  $\mathcal{H}_B$  dimension  $d_B = d_A + 1$ . Furthermore, let  $\varepsilon \in (0, 1)$ . We then consider sets of orthonormal vectors  $\{|i_A\rangle\}_{i=1}^{d_A} \subset \mathcal{H}_A$ ,  $\{|i_B\rangle\}_{i=1}^{d_A} \subset \mathcal{H}_B$  and define

$$|\psi\rangle := \sum_{i=1}^{d_A-1} \sqrt{\frac{\varepsilon}{d_A-1}} |i_A\rangle \otimes |i_B\rangle + \sqrt{1-\varepsilon} |(d_A)_A\rangle \otimes |(d_A)_B\rangle = \sum_{i=1}^{d_A} \sqrt{\lambda_i} |i_A\rangle \otimes |i_B\rangle.$$

with the  $\lambda_i$  defined accordingly. We find that

$$\begin{aligned} \rho_A &:= \text{Tr}_B[|\psi\rangle\langle\psi|] = \sum_{i=1}^{d_A} \lambda_i |i_A\rangle\langle i_A|, \\ \rho_B &:= \text{Tr}_A[|\psi\rangle\langle\psi|] = \sum_{i=1}^{d_A} \lambda_i |i_B\rangle\langle i_B|, \end{aligned}$$

and the Moore-Penrose pseudoinverse (in the case of  $\rho_A$  it is an inverse)

$$\begin{aligned} \rho_A^{-1} &= \sum_{i=1}^{d_A} \lambda_i^{-1} |i_A\rangle\langle i_A|, \\ \rho_B^{-1} &= \sum_{i=1}^{d_A} \lambda_i^{-1} |i_B\rangle\langle i_B|. \end{aligned}$$

We find

$$\begin{aligned}
\text{Tr} [|\psi\rangle\langle\psi| \rho_A^{-1} \otimes \rho_B^{-1}] &= \sum_{i,j,k,l} \frac{\sqrt{\lambda_i} \sqrt{\lambda_j}}{\lambda_k \lambda_l} \langle i_A | k_A \rangle \langle k_A | j_A \rangle \langle i_B | l_B \rangle \langle l_B | j_B \rangle \\
&= \sum_{i,j,k,l} \frac{\sqrt{\lambda_i} \sqrt{\lambda_j}}{\lambda_k \lambda_l} \delta_{ik} \delta_{kj} \delta_{il} \delta_{lj} \\
&= \sum_i \frac{1}{\lambda_i} = \frac{(d_A - 1)^2}{\varepsilon} + \frac{1}{1 - \varepsilon},
\end{aligned}$$

with which, as  $|\psi\rangle\langle\psi|$  is a rank one projection

$$\begin{aligned}
\widehat{I}_{|\psi\rangle\langle\psi|}(A : B) &= \text{Tr} \left[ |\psi\rangle\langle\psi| \log(|\psi\rangle\langle\psi|^{1/2} \rho_A^{-1} \otimes \rho_B^{-1} |\psi\rangle\langle\psi|^{1/2}) \right] \\
&= \log \text{Tr} [|\psi\rangle\langle\psi| \rho_A^{-1} \otimes \rho_B^{-1}] \\
&= \log \left( \frac{(d_A - 1)^2}{\varepsilon} + \frac{1}{1 - \varepsilon} \right) \geq \log \left( \frac{(d_A - 1)^2}{\varepsilon} \right).
\end{aligned}$$

We directly obtain  $\|\rho_A^{-1}\|_\infty = \|\rho_B^{-1}\|_\infty = \frac{d_A - 1}{\varepsilon}$  and by construction  $d_A < d_B$ , hence the bound in eq. (29) gives

$$\widehat{I}_{|\psi\rangle\langle\psi|}(A : B) \leq \log \left( \frac{d_A (d_A - 1)}{\varepsilon} \right). \quad (47)$$

We first note that for  $\varepsilon = 1 - \frac{1}{d_A}$  we get equality in eq. (47). What is, however, more interesting is the fact that

$$\log \left( \frac{(d_A - 1)^2}{\varepsilon} \right) \leq \widehat{I}_{|\psi\rangle\langle\psi|}(A : B) \leq \log \left( \frac{d_A (d_A - 1)}{\varepsilon} \right),$$

with

$$\left| \log \left( \frac{d_A (d_A - 1)}{\varepsilon} \right) - \log \left( \frac{(d_A - 1)^2}{\varepsilon} \right) \right| = \log \left( \frac{d_A}{d_A - 1} \right).$$

I.e., the error of the bound is of order  $\log \left( \frac{d_A}{d_A - 1} \right)$  independent of the  $\varepsilon$ . This means, that the scaling behaviour of the bound, in terms of the minimal non-zero eigenvalue of  $\rho_A$  and  $\rho_B$  respectively is the best one can do.

3. The lower bound of the BS-CMI is again a consequence of the data processing inequality. The upper bound is a direct consequence of the bounds obtained for the BS-conditional information due to the definition of the conditional mutual information in eq. (4)

$$\begin{aligned}
\widehat{I}_\rho(A : B|C) &= \widehat{H}_\rho(A|C) - \widehat{H}_\rho(A|BC) \\
&\leq \log d_A + \log \min\{d_A, d_{BC}\} \\
&= \log \min\{d_A^2, d_{ABC}\}.
\end{aligned}$$

We expect that the tightness of such a bound can be proven in a similar way to the one for the BS-mutual information.

## G Behavior of $g_d$

In this section, we study the function  $g_d(p) := \frac{d}{p^{1/d}} h(p) - \log(1 - p^{1/d})$  for  $p \in (0, 1)$  and a fixed  $d \in \mathbb{N}$ ,  $d \geq 2$ . This function appears in some of the continuity bounds in section 7.6.

**Lemma G.1** *Let  $d \in \mathbb{N}$ ,  $d \geq 2$ . Then,  $\lim_{p \rightarrow 0^+} g_d(p) = 0$ . In particular,  $g_d$  is continuous on  $p \in [0, 1)$ .*

*Proof.* Since  $\lim_{p \rightarrow 0^+} \log(1 - p^{1/d}) = 0$ , we can focus on  $\frac{d}{p^{1/d}} h(p)$ . The assertion follows from applying L'Hospital's rule twice. Indeed,

$$\begin{aligned} \lim_{p \rightarrow 0^+} \frac{d}{p^{1/d}} h(p) &= \lim_{p \rightarrow 0^+} \frac{d(\log(1-p) - \log(p))}{p^{1/d-1}/d} \\ &= \lim_{p \rightarrow 0^+} \frac{d(-(1-p)^{-1} - p^{-1})}{(1-d)p^{1/d-2}/d^2} \\ &= \lim_{p \rightarrow 0^+} \frac{d^3}{d-1} \left( \frac{p^{2-1/d}}{1-p} + p^{1-1/d} \right) \\ &= 0. \end{aligned}$$

Continuity, therefore, follows from the definition of the function.  $\square$

**Lemma G.2** *Let  $d \in \mathbb{N}$ ,  $d \geq 2$ . Then, the function  $g_d$  is non-decreasing on  $[0, 1/2]$ .*

*Proof.* We can differentiate  $g_d(p)$  on  $(0, 1/2)$ . This yields

$$\begin{aligned} \frac{\partial}{\partial p} g_d(p) &= \frac{1}{p^{1/d}} \left( \frac{p^{2/d-1}}{d(1-p^{1/d})} + (d-1+p^{-1}) \log(1-p) - (d-1) \log(p) \right) \\ &=: \frac{1}{p^{1/d}} g'_d(p). \end{aligned} \tag{48}$$

We will now show monotonicity in  $d$  of  $g'_d(p)$  for all  $p \in (0, 1/2)$ . This will allow us to show the non-negativity of eq. (48) on  $(0, 1/2)$  only for  $d = 2$  and conclude it for all  $d \geq 2$ . We have

$$\frac{\partial}{\partial d} g'_d(p) = \frac{p^{2/d-1} (d(p^{1/d} - 1) + (p^{1/d} - 2) \log(p))}{d^3(p^{1/d} - 1)^2} + \log(1-p) - \log(p).$$

The above is non-negative for  $p \in (0, 1/2)$ , if

$$(2 - p^{1/d}) \log \frac{1}{p} \geq d(1 - p^{1/d}) \quad \Leftrightarrow \quad \left(1 + \frac{1}{1 - p^{1/d}}\right) \log \frac{1}{p} \geq d$$

One obtains the last inequality by substitution of  $p = e^{dt}$  with  $t \in (-\infty, -\frac{\log(2)}{d})$  giving us

$$-dt \left(1 + \frac{1}{1 - e^t}\right) \geq d \quad \Leftrightarrow \quad -t \left(1 + \frac{1}{1 - e^t}\right) \geq 1$$

which is true for  $t \in (-\infty, 0)$  hence in particular on  $(-\infty, -\frac{\log(2)}{d})$ . We thereby have that for  $d \geq 2$   $p \in (0, 1/2)$   $g'_d(p) \geq g'_2(p)$ . It is straightforward to see that  $g'_2(p) \geq 0$  on  $p \in (0, 1/2)$ . This finally lets us conclude the claim that  $g_d(p)$  is non-decreasing on  $p \in [0, 1/2]$  as  $g_d(p)$  is continuous on  $[0, 1/2]$  by lemma G.1.  $\square$

**Lemma G.3** *Let  $d \in \mathbb{N}$ ,  $d \geq 2$ . Then, the function  $p \mapsto g_d(p)/(1-p)$  is non-decreasing on  $[0, 1)$ .*

*Proof.* The argument follows similar lines as the one in lemma G.2. We first note that  $p \mapsto \frac{1}{1-p}$  is non decreasing on  $[0, 1/2)$  and  $p \mapsto g_d(p)$  is as well, as proven in lemma G.2. Hence  $p \mapsto \frac{1}{1-p}g_d(p)$  is non decreasing on  $[0, 1/2]$ . What now remains to show is that it is non-decreasing on  $[1/2, 1)$ . We can differentiate the function on the interval  $[1/2, 1)$  and obtain

$$\begin{aligned} \left. \frac{\partial}{\partial x} \frac{g_d(x)}{1-x} \right|_{x=p} &= \frac{1}{1-p} \left( \frac{dp^{-1/d}h(p) - \log(1-p^{1/d})}{1-p} \right. \\ &\quad \left. + \frac{p^{1/d-1}}{d(1-p^{1/d})} - p^{-1/d}p^{-1}h(p) + dp^{-1/d}(\log(1-p) - \log(p)) \right). \\ &\geq \frac{1}{1-p} \left( p^{-1/d}h(p) \left( \frac{1}{1-p} - \frac{1}{p} \right) + (d-1)p^{-1/d} \left( \frac{h(p)}{1-p} + \log(1-p) \right) \right. \\ &\quad \left. + p^{-1/d} \log(1-p) - \frac{\log(1-p^{1/d})}{1-p} \right) \\ &\geq 0. \end{aligned}$$

The last inequality holds since  $p \geq \frac{1}{2}$  and  $d \geq 2$  hence

$$\begin{aligned} \frac{1}{1-p} - \frac{1}{p} &\geq 0, \\ \frac{h(p)}{1-p} + \log(1-p) &\geq 0, \\ p^{-1/d} \log(1-p) - \frac{\log(1-p^{1/d})}{1-p} &\geq 0. \end{aligned}$$

To see the last inequality, one can verify that  $p^{-1/d} \leq 1/(1-p)$  in this regime and that  $\log(1-p) \geq \log(1-p^{1/d})$ . Thus  $p \mapsto \frac{g_d(p)}{1-p}$  is non-decreasing on  $[1/2, 1)$ , which concludes the proof.  $\square$