# Continuity of quantum entropic quantities via almost convexity

Master Thesis

Paul Elias Gondolf

paul.gondolf@student.uni-tuebingen.de

Fachbereich Mathematik
Universität Tübingen

**Supervisor:**
jProf. Dr. Ángela Capel-Cuevas

August 15, 2025

# Acknowledgements

[excluded]

# Abstract

Based on the proofs of the continuity of the conditional entropy by Alicki, Fannes, and Winter, we introduce in this thesis 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 new continuity bounds for the relative entropy itself. 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 and give an outlook on future lines of work.

**Keywords:** Continuity bounds, Umegaki relative entropy, Belavkin-Staszewski relative entropy.

# Contents

Chapter 1

# Preface

In information theory and in particular, quantum information theory one major task is to evaluate the difference between quantum states. Although the family of operator norms seem to be a natural choice from a mathematical perspective, problems from physics naturally bring up so-called divergence measures. A wide variety of these quantities exist and their mathematical understanding is crucial if one wants to work with them in both the math and physics context.

In this work, we focus our attention on continuity bounds for those quantities and for related functions such as conditional entropy, (conditional) mutual information and divergence bounds for two particular divergences. The first one is the Umegaki and the second one is the Belavkin-Staszewski relative entropy.

The thesis is structured as follows. We will first give an introduction to the mathematical framework we work with. Starting with operator theory and convex analysis in Section 2.1, we go to norms and trace inequalities in Section 2.2 and finally talk about divergences and the definitions of quantities like the conditional entropy and (conditional) mutual information for a general divergence. This final section of the chapter also includes a discussion of the ambiguities that arise in the case of the BS-entropy.[1]

Building on that, Chapter 3 introduces the almost locally affine (ALAFF) method which translates joint convexity and almost joint concavity into a plethora of results. Amongst them are continuity bounds for entropic quantities derived from the divergence and continuity bounds for the divergence itself on a suitably defined set.

We then go on and prove this almost joint concavity for the relative entropy in Chapter 4 and demonstrate how the ALAFF method works. In the course of that, we recover the well-established (almost) tight bounds for the conditional entropy by Winter [2] and also the ones for the (conditional) mutual information. We further manage to derive something that we call a divergence bound which

[1] Let it be noted that we will not give an introduction to continuity bounds in this work and just discuss related work right when we derive our result. The interested reader is referred to [1].

performs slightly worse than the preexisting bound but has the advantage that none of the involved states has to be full rank. Finally, we construct a continuity bound for the relative entropy itself.

The subsequent chapter, Chapter 5, follows a similar structure. We again prove almost joint concavity, but this time for the Belavkin-Staszewski or BS-entropy for short. Obtaining similar results for the BS-conditional, (conditional) mutual information and divergence bounds is slightly more involved than in the case of the relative entropy. This is due to some pathologies, i.e. discontinuities that these quantities exhibit, and that we discuss in that chapter as well.

We conclude this work with Chapter 6, revolving around applications of the obtained results in various contexts of the field of quantum information. They allow us to obtain bounds on the conditional mutual information of a state in terms of its norm distance to its Petz recovery. We further obtain a bound on the difference between the relative and the BS-entropy and establish continuity bounds for the Rains and the BS-Rains information. At last, we give an outlook on open problems and future lines of work.

This whole thesis has been adapted into a paper [1] together with Andreas Bluhm, Ángela Capel-Cuevas and Antonio Pérez-Hernández.

CHAPTER 2

# Introduction

In this chapter, we introduce the mathematical framework. We start with the fundamentals, like introducing a Hilbert space, and its operator space, and very roughly describe concepts like adjoint operators, functional calculus and operator monotonicity, convexity and concavity respectively. We build upon these basics when introducing norms on the operator space and further state important trace inequalities that are essential when working with entropic functions and functions of operators in general. We wrap up with the discussion of divergences and how one builds all the derived quantities such as for example the conditional entropy from those. In the course of this, we will come across ambiguities that lead to open questions we leave for future work (see Section 6.2).

## 2.1 Operator theory and convex analysis

Our starting point is a Hilbert space of finite dimension $d$, which can be isomorphically identified with $\mathbb{C}^d$. We will not make the identification explicit every time and just write abstractly $\mathcal{H}$. Elements of this Hilbert space will be denoted by $|\psi\rangle$, $|\varphi\rangle$ or $|i\rangle$ with $i \in \mathbb{N}$. A Hilbert space carries by definition an inner product, which we will denote by $\langle \cdot | \cdot \rangle$. It further, as every other vector space, admits an algebra of (bounded[1]) linear operators $\mathcal{B}(\mathcal{H})$, which can be identified with the complex $n \times n$ matrices. We denote elements of $\mathcal{B}(\mathcal{H})$ using capital letters and often call them just operators. This latter space, or in particular the subset of normalised positive semi-definite operators $\mathcal{S}(\mathcal{H})$, is the fundamental space we are interested in.

To get an understanding of the defining property of a quantum state, i.e. an element of $\mathcal{S}(\mathcal{H})$, we need the concept of normalisation as well as the Löwner order on the set of self-adjoint operators. First, note that for every operator $A$ in $\mathcal{B}(\mathcal{H})$, there exists an adjoint operator $A^*$, implicitly defined via the inner

[1] Since the underlying Hilbert space is finite, linearity is equivalent to boundedness as well as continuity.

product:

$$\langle\psi|A\phi\rangle = \langle A^*\psi|\phi\rangle \qquad \forall\,|\psi\rangle\,,|\phi\rangle \in \mathcal{H}\ .$$

An operator that coincides with its adjoint is called *self-adjoint* or *Hermitian.* Using again the structure given by the inner product, we can establish a *partial order* on the set of self-adjoint operators, where for $A, B \in \mathcal{B}(\mathcal{H})$ self-adjoint

$$A \geq B \quad :\Leftrightarrow \quad \langle\psi|(A-B)\psi\rangle \geq 0 \quad \forall\,|\psi\rangle \in \mathcal{H}\ . \tag{2.1}$$

The inequality can be made strict, by just replacing $\geq$ with $>$ on both sides of Eq. (2.1). If for $A \in \mathcal{B}(\mathcal{H})$ with $A$ self adjoint, $A \geq 0$ we call it positive semi-definite and further positive definite if $A > 0$. If such a self-adjoint operator further has trace one, i.e. the function

$$\mathrm{tr} : \mathcal{B}(\mathcal{H}) \to \mathbb{C}, \qquad A \mapsto \sum_{i=1}^{d} \langle i|Ai\rangle$$

for $(|i\rangle)_{i=1}^{d}$ an orthonormal basis, evaluates to one, we call it a *quantum state, density matrix,* or just *state.* The set of such operators, as we have already mentioned, is denoted by $\mathcal{S}(\mathcal{H})$ and we use $\rho, \sigma, \tau$ and $\gamma$ to identify its elements. 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 moreover only want to admit states with a minimal eigenvalue greater than $m$, we write $\mathcal{S}_{\geq m}(\mathcal{H})$.

To define divergences, entropies or just general functions of operators, we need to dive a little into spectral theory. We first note that a self-adjoint operator $A$ has a spectral decomposition

$$A = \sum_{i=1}^{n} \lambda_i P_i\,,$$

where $\lambda_i$ and $P_i$ are, respectively, its eigenvalues and spectral projections. For any function $f : \mathbb{R} \to \mathbb{R}$, one then defines the operator $f(A) \in \mathcal{B}(\mathcal{H})$ as

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) P_i\,.$$

With the partial order defined on the self-adjoint operators, one can extend the concept of monotonicity and convexity (concavity) to functions on operators. We call a function $f : I \to \mathbb{R}$ on an interval $I \subseteq \mathbb{R}$ *operator montone* if for all finite-dimensional Hilbert spaces, all self-adjoint operators $A, B \in \mathcal{B}(\mathcal{H})$ with spectrum contained in $I$ and $A \leq B$,

$$f(A) \leq f(B)$$

holds true. We further call $f$ *operator convex* if for $\lambda \in [0, 1]$

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B)\,.$$

Following from the above we call $f$ operator concave if $-f$ is operator convex. Operator convex functions include $x \mapsto x\log(x)$ and $x \mapsto x^t$ for $t \in [1,2]$ and operator concave and operator monotone functions are for example $x \mapsto \log(x)$ and $x \to -x^t$ for $t \in [-1,0]$ all on the domain $(0,\infty)$. The theorem from which all these properties follow is the famous Löwner-Heinz theorem.

From the operator monotonicity and convexity (concavity) follow several properties such as the operator Jensen inequality, the Sherman-Davis inequality and more. We will only state the Sherman-Davis inequality to give the reader an intuition for such kinds of inequalities and because it is used later.

**Theorem 2.1** (Sherman-Davis inequality).
*Let $f : I \to \mathbb{R}$ an operator convex function, $A \in \mathcal{B}(\mathcal{H})$ self-adjoint and with spectrum in $I$ and $P$ an orthogonal projection, then*

$$Pf(PAP)P \leq Pf(A)P$$

The proofs of the above theorem as well as of the other claims made here can be found in [3], which together with [4] gave the inspiration for this section.

## 2.2 Norms and trace inequalities

We have already talked about normalisation and while doing so swept under the rock that the trace that we use in this context can also be written as a norm, the so-called one norm. The fact that it reduces just to the trace in the case of positive semi-definite operators is due to the exact same property, i.e. positive semi-definiteness. For general operators, one defines the *one norm* as

$$\|\cdot\|_1 : \mathcal{B}(\mathcal{H}) \to [0,\infty), \qquad A \mapsto \operatorname{tr}\left[\sqrt{A^*A}\right] = \operatorname{tr}[|A|]\,.$$

Naturally $|A| := \sqrt{A^*A}$. In the case of positive semi-definite operators, all eigenvalues are positive, hence the one norm reduces just to the trace.

The above is a representative of the Schatten $p$-norms, which are defined for every $p \geq 1$ as

$$\|\cdot\|_p : \mathcal{B}(\mathcal{H}) \to [0,\infty), \qquad A \mapsto \operatorname{tr}[|A|^p]^{\frac{1}{p}}\,.$$

For $p = 1$ we clearly recover the one norm whilst for $p \to \infty$ one obtains the *spectral norm* or *operator norm*. Besides being norms those functions have several more useful properties, for example satisfy for $A \in \mathcal{B}(\mathcal{H})$, $\|A\|_p = \|A^*\|_p$ and are submultiplicative, meaning for $A, B \in \mathcal{B}(\mathcal{H})$, we find $\|AB\|_p \leq \|A\|_p \|B\|_p$. Similarly to $L_p$-norms we also have a Hölder inequality, i.e. for $p \in [1,\infty)$ and $q$ such that $\frac{1}{p} + \frac{1}{q} = 1$[2], we find

$$\|AB\|_1 \leq \|A\|_p \|B\|_q\,,$$

[2] The case $p = 1, q = \infty$ is the limiting case inequalities with $p$ and $q$ finite.

with $A, B \in \mathcal{B}(\mathcal{H})$. To find a more thorough discussion we again refer to [3] and [5].

Closely related to those norm inequalities are the trace inequalities that have shown to be fundamental when looking at properties of entropic quantities. We will only introduce the Peierls-Bogolubov inequality as well as a Corollary that is a consequence of one of the so-called multivariant trace inequalities [6]. Both of them are substantial when proving results for the relative as well as the BS-entropy. We begin with the Peierls-Bogolubov inequality.

**Theorem 2.2** (Peierls-Bogolubov, [7])**.**
*For $A, B \in \mathcal{B}(\mathcal{H})$ self-adjoint and $\operatorname{tr}\left[e^A\right] = 1$, we find*

$$\log \operatorname{tr}\left[e^A e^B\right] \geq \log \operatorname{tr}\left[e^{A+B}\right] \geq \operatorname{tr}\left[B e^A\right]$$

The following Corollary can be found as Corollary 3.3. in [6] and is the limiting case of the generalised Araki-Lieb-Thirring inequality presented in the very same paper.

**Corollary 2.3.** *Let $p \geq 1$, $\beta_0(t) = \frac{\pi}{2}(\cosh(\pi t) + 1)^{-1}$, $n \in \mathbb{N}$ and consider a collection $\{H_k\}_{k=1}^n \subset \mathcal{B}(\mathcal{H})$ of self-adjoint operators. Then*

$$\log \left\| \exp\left( \sum_{k=1}^n H_k \right) \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_0(t) \log \left\| \prod_{k=1}^n \exp((1+it)H_k) \right\|_p .$$

*For the case $n = 3$ and $p = 2$ substituting $H_k \leftarrow \frac{1}{2} H_k$ and using the concavity of the logarithm as well as Jensen's inequality, translates the above into*

$$\begin{aligned} &\operatorname{tr}[\exp(H_1 + H_2 + H_3)] \\ &\quad \leq \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\left[ \exp(H_1) \exp\left( \frac{1+it}{2} H_2 \right) \exp(H_3) \exp\left( \frac{1+it}{2} H_2 \right) \right] . \end{aligned} \tag{2.2}$$

## 2.3 Divergences

We will now come to the quantities we are investigating in this thesis, namely quantum divergences and their derivates. As there exists an axiomatic definition of the latter, we want to give a little context on how those came along, starting with their classical counterparts. Note, that there exists an embedding of the classical setting into the quantum setting just by requiring the involved states to commute. This allows for simultaneous diagonalisation meaning they can be identified with just diagonal matrices with non-negative entries that sum to one.

In 1961 Alfréd Rényi took an axiomatic approach [8] to define the well known Shannon entropy [9] and the Kullback-Leibler divergence [10]. Indeed he stated five axioms that describe desirable properties of uncertainty measures and six axioms accomplishing the same for measures of distinguishability respectively. Those axioms only allowed for the two named quantities as solutions. By relaxing one of the five (six) axioms he further managed to derive a whole family of entropies and divergences, controlled by a parameter $\alpha \in (0,1) \cup (1,\infty)$. Today those families are called, respectively, $\alpha$-Rényi entropies and $\alpha$-Rényi divergences after him.

Following a similar approach one can axiomatically define quantum divergences, as it is done in for example [11]. It turns out that a generalisation[3] of the original axioms is not enough and one has to add an additional one to reach a sensible result. This complementary axiom requires a divergence $\mathbb{D}$ to fulfil the so-called data-processing inequality (DPI). Meaning for every completely positive, trace-preserving (CPTP) map $T$ and quantum states $\rho, \sigma$:

$$\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}(T(\rho)\|T(\sigma))\,.$$

In the classical setting, the DPI is a consequence of the six initial axioms, while in the quantum setting, this is still an open question.

It is also remarkable that there is not a unique family of functionals admitted by the axioms, as in the classical case, but several so-called quantum Rényi divergences. This is due to the non-commutative nature of quantum mechanics. Indeed, as soon as one assumes that the involved states (operators) commute, all the quantities reduce to the classical unique counterpart.

We will not dive into this jungle of divergences nor discuss any family in particular but only focus on the two divergences that are relevant for this work, the Umegaki [12] and the Belavkin-Staszewski [13] relative entropy. For an extensive discussion of quantum divergences, definitions and proofs of claims made here we refer the interested reader to [11]. The first quantity we want to introduce is the Umegaki relative entropy.

**Definition 2.4** (Umegaki relative entropy, [12])**.**
For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, we define the Umegaki relative entropy, or just relative entropy as

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

The second one is the Belavkin-Staszewski relative entropy.

**Definition 2.5** (Belavkin-Staszewski relative entropy, [13])**.**
For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, we define the Belavkin-Staszewski entropy or BS-entropy for

[3] As already mentioned classical probability distributions can be represented by diagonal matrices.

short, as

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

Note that in both definitions we used the convention $\frac{0}{0} = 0$ and $0 \log 0 = 0$. It holds in general that for $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, $D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma)$ [11, Eq. (4.35)] with equality iff $\rho$ and $\sigma$ commute. We further note that both, the relative and the BS-entropy, can also be defined in terms of positive semi-definite operators $A, B$ (not necessarily normalised), by replacing $\rho$ with $A$, $\sigma$ with $B$ and dividing by $\operatorname{tr}[A]$.

Before we can define the entropic quantities derived from a divergence, we have to first introduce bipartite Hilbert spaces and the partial trace. A bipartite Hilbert space is given as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ with $\mathcal{H}_A$ and $\mathcal{H}_B$ both Hilbert spaces themselves and $\otimes$ denoting the tensor product. The explicit construction can be found in for example [3]. We will write an index to all quantities that are associated with either one of the spaces. For example, $d_B$ is the dimension of the space $\mathcal{H}_B$ and $\rho_B \in \mathcal{S}(\mathcal{H}_B)$ is the image of $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ under the CPTP map that is the partial trace $\operatorname{tr}_A[\cdot]$. Note that the index of the partial trace indicates the system we trace out while the index of the state is the system the state lives on. Further note that sometimes, under slight misuse of notation, we write the system in the index of the state without the state being an image under a partial trace, i.e. for $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we write $\rho_{AB}$ or for $\sigma \in \mathcal{S}(\mathcal{H}_B)$, $\sigma_B$. This is done to put emphasis on the system the state stems from. As already mentioned the *partial trace* is a CPTP map

$$\operatorname{tr}_B[\cdot] : \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B), \qquad N \mapsto \operatorname{tr}_A[N]\,,$$

where $\operatorname{tr}_A[N]$ is the unique element of $\mathcal{B}(\mathcal{H}_B)$ such that the equality

$$\operatorname{tr}[M \operatorname{tr}_A[N]] = \operatorname{tr}[\mathbb{1}_A \otimes M N]$$

holds for all $M \in \mathcal{B}(\mathcal{H}_B)$. Analogously one can define $\operatorname{tr}_B[\cdot]$ and further extend those concepts also to $n$-partite spaces.

Now that we know what $n$-partite spaces and the partial trace are, we can come to define conditional entropy, mutual information and conditional mutual information for a general divergence. The course we will take is starting with the representations of those quantities in terms of the *von Neuman entropy* [14] given as

$$S(\rho) = -\operatorname{tr}[\rho \log \rho]$$

for a $\rho \in \mathcal{S}(\mathcal{H})$[4], rewrite them in terms of the relative entropy and then abstract from those to definitions. We summarise those steps in the following two definitions, beginning with the original quantities.

[4]We again used the convention $0 \log 0 = 0$.

**Definition 2.6** (Conditional entropy, mutual and conditional mutual information)**.** Let $\mathcal{H}_A \otimes \mathcal{H}_B$ a bipartite Hilbert space, then the *conditional entropy* for $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ conditioned on the system $B$ is given as

$$\begin{aligned} H_\rho(A|B) &:= S(\rho_{AB}) - S(\rho_B) \\ &= -D(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B) \\ &= \max_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} -D(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B)\,, \end{aligned}$$

and the *mutual information* is given by

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

Further for a tripartite system $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ and $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, we can define the *conditional mutual information* as

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

All the above equalities are well known and most of them are obtained directly by inserting definitions. Note that both, the mutual information and the conditional mutual information, are symmetric in the $A$ and $B$ systems.

Based on the representations involving the relative entropy we now proceed and define the conditional divergence (analogue of the conditional entropy), mutual information and conditional mutual information for an arbitrary quantum divergence.

**Definition 2.7** (Conditional divergence, mutual and conditional mutual information for a quantum divergence)**.** Let $\mathbb{D}$ be a quantum divergence. Let $\mathcal{H}_A \otimes \mathcal{H}_B$ a bipartite Hilbert space, then the *conditional divergence* for $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ conditioned on the system $B$ is given by

$$\mathbb{H}_\rho(A|B) := -\,\mathbb{D}(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B)\,, \tag{2.3}$$

the *variational conditional divergence* by

$$\mathbb{H}_\rho^{\mathrm{var}}(A|B) := \sup_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} -\,\mathbb{D}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) \tag{2.4}$$

and the *mutual information* as

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

Further for a tripartite system $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ and $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ we can define the *(one sided) conditional mutual information* as

$$\mathbb{I}_\rho^{\mathrm{os}}(A:B|C) := \mathbb{H}_\rho(A|C) - \mathbb{H}_\rho(A|BC)\,, \tag{2.5}$$

and the *(two sided) conditional mutual information* as

$$\mathbb{I}_\rho^{\mathrm{ts}}(A:B|C) := \mathbb{I}_\rho(A:BC) - \mathbb{I}_\rho(A:C)\,.$$

In the case of the relative entropy (and possibly some others) the different definitions for conditional divergence, mutual information and conditional mutual information agree respectively. This is, however, not the case in general. For the BS-conditional entropy, for instance, which we denote by $\widehat{H}$, the general inequality

$$\mathbb{H}_\rho(A|B) \leq \mathbb{H}_\rho^{\text{var}}(A|B)$$

is strict in some cases. The numerics to support this claim are visualised in

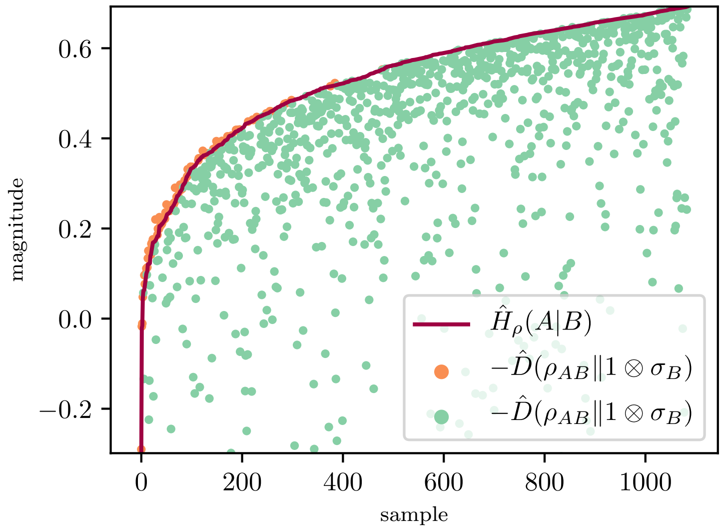



Figure 2.1: The red line is the BS-conditional entropy defined via the partial trace evaluated at $\rho_{AB}$. The dots are the BS-entropy of the states $\rho_{AB}$ and $\mathbb{1}_A \otimes \sigma_B$ with a state $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$. The orange dots are the cases when the $-\widehat{D}(\rho_{AB}\| \mathbb{1}_A \otimes \sigma_B)$ exceeds $\widehat{H}(A|B)_\rho$. We sampled a total of 100.000 pairs of $\rho_{AB}$ and $\sigma_B$ and evaluated both $\widehat{H}(A|B)_\rho$ and $-\widehat{D}(\rho_{AB}\| \mathbb{1}_A \otimes \sigma_B)$. Only a tenth of all samples have been 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 to reduce the risk of numerical flaws. The numerical simulation was conducted on $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^2 \otimes \mathbb{C}^2$.

Fig. 2.1. Not only that but we further find that while $\widehat{H}$ is discontinuous on $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ (cf. Proposition 5.6), $\widehat{H}^{\text{var}}$ is uniformly continuous (cf. Corollary 6.11). Note that in the rest of this work when talking about the BS-conditional entropy we mean the definition in Eq. (2.3). Further we mean Eq. (2.5) when talking about the BS-conditional mutual information and denote it by $\widehat{I}(A:B|C)$ and analogously we use $\widehat{I}(A:B)$ for the BS-mutual information. The reason for choosing the one-sided over the two-sided is that the definition of the latter involves the BS-mutual information which has no constant upper

bound and therefore is the more pathological quantity (cf. Proposition 5.5).

The definitions of conditional divergence, mutual information and conditional mutual information are not new but have already appeared in some texts, such as [15, 16, 17].

Chapter 3

# The ALAFF method

In this chapter, we will introduce the method that lies at the heart of our approach: The locally almost affine (ALAFF) method. It translates the joint convexity and almost joint concavity of a divergence into continuity bounds for derived entropic quantities of that divergence. The flow chart in Fig. 3.1, visualises the procedure and lists entropic quantities (see Definition 2.7) for which uniform continuity (on properly constructed sets) can be obtained. We will drop the joint in front of convex and almost concave, for better readability respectively.

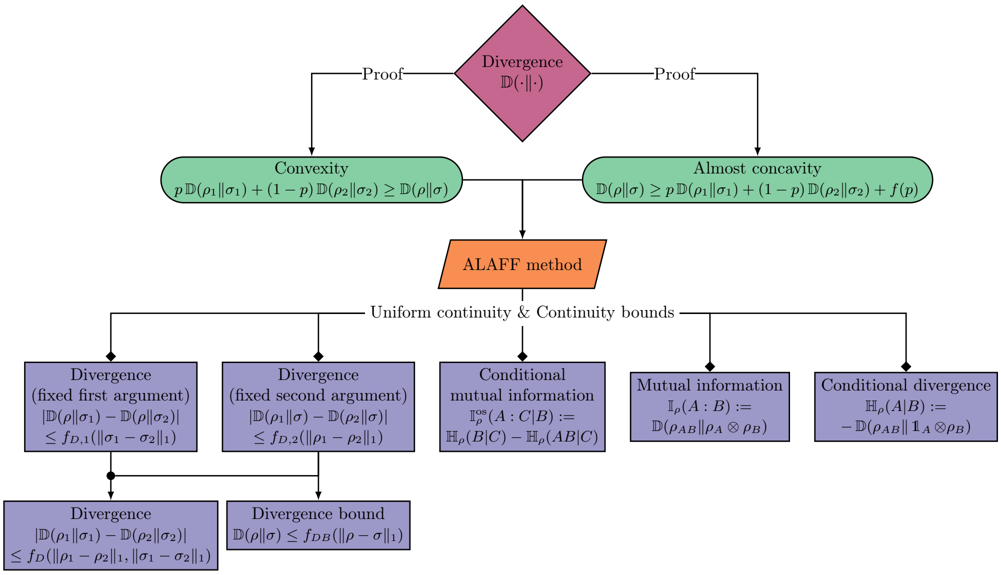



Figure 3.1: The flow chart demonstrates how convexity and almost concavity of a divergence can be used to obtain uniform continuity and explicit continuity bounds on entropic quantities derived from that divergence.

It is immediately clear what is meant by convexity which is often even a defining property of a divergence [18] or a direct consequence thereof[1] [11, Proposition 4.2]. The almost (joint) concavity, however, needs yet to be defined.

[1]Some authors define divergences as function on two density operators fulfilling a data

**Definition 3.1** (Almost (joint) concavity of a divergence)**.** A divergence $\mathbb{D}$ 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 $f : [0,1] \to \mathbb{R}$ with $f(0) = f(1) = 0$ such that, for all $p \in [0,1]$,

$$\mathbb{D}(\rho\|\sigma) \geq p\,\mathbb{D}(\rho_1\|\sigma_1) + (1-p)\,\mathbb{D}(\rho_2\|\sigma_2) - f(p) \tag{3.1}$$

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* 3.2. 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. (3.1). 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.

In its earliest form, it was developed and used by Alicki and Fannes [19], as well as Winter [2], to prove uniform continuity and give an explicit continuity bound for the conditional entropy. Shirokov noticed its potential beyond this specific application and moulded it into a method that can be applied to functions defined on convex and $\Delta$-invariant subsets of $\mathcal{S}(\mathcal{H})$ [20, 21]. In short, $\Delta$-invariance means that for two elements their normalised positive and negative part again lies in the set (see also Definition 3.3). 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 conditional BS-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 $s$.

**Definition 3.3** (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

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 [18, Corollary 4.7].

$\tau \in \mathcal{S}(\mathcal{H})$ such that the two states

$$\Delta^{\pm}(\rho, \sigma, \tau) = s\tau + (1-s)\varepsilon^{-1}[\rho - \sigma]_{\pm} \tag{3.2}$$

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 [21].

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

*Remark* 3.4.

1. Let $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$ $s$-perturbed $\Delta$-invariant. 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.

It has already been mentioned that almost concavity is not enough but we need a well-behaved remainder function that becomes uniform in case the states fulfil certain structural requirements (e.g. one being a marginal of the other). These structural restrictions lead to functions that now only take one state as an argument while still being convex and almost concave. However, due to the uniformity of the remainder function, the almost concavity constitutes a stronger property. Namely almost local affinity.

**Definition 3.5** (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

$$-a_f(p) \leq f(p\rho + (1-p)\sigma) - pf(\rho) - (1-p)f(\sigma) \leq b_f(p) \tag{3.3}$$

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

The notion of almost locally affine functions as above has appeared previously in the literature, also under the name “approximate affinity" (see e.g. [22]). We can now formulate the following theorem, whose proof is inspired by Shirokov [20].

**Theorem 3.6** (Almost locally affine (ALAFF) method)**.** *Let $s \in [0, 1)$ and $\mathcal{S}_0 \subseteq \mathcal{S}(\mathcal{H})$ be a $s$-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 $\varepsilon \in (0,1]$*

$$\sup_{\substack{\rho,\sigma\in\mathcal{S}_0 \\ \frac{1}{2}\|\rho-\sigma\|_1\leq\varepsilon}} |f(\rho)-f(\sigma)| \leq C_f^s \frac{\varepsilon}{1-s} + \frac{1-s+\varepsilon}{1-s} E_f^{\max}\Big(\frac{\varepsilon}{1-s+\varepsilon}\Big)\,, \tag{3.4}$$

*with*

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

*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. (3.2). 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. (3.3) for both of the representations we have for the state in question. This gives us

$$-a_f(p) \leq f(\omega) - (1-p)f(\rho) - pf(\gamma_-) \leq b_f(p)\,,$$
$$-a_f(p) \leq f(\omega) - (1-p)f(\sigma) - pf(\gamma_+) \leq b_f(p)\,,$$

where we set $p = \frac{\varepsilon}{1-s+\varepsilon}$ 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

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

Those two inequalities immediately give us that

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

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

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

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 then 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. (3.4). The reduction of $E_f^{\max}$ to $E_f$ on $\varepsilon \in (0, 1-s]$ $E_f^{\max}$ is due to $E_f$ being non-decreasing on $[0, \frac{1}{2}]$. This means, however, that $E_f^{\max}$ inherits the vanishing property as $p \to +0$, which translates to $E_f^{\max}(p(\varepsilon)) \to 0$ if $\varepsilon \to +0$. Thus we conclude uniform continuity. □

*Remark* 3.7. We have restricted to $\varepsilon \in (0,1]$ as the maximal one norm distance of two quantum states is bounded by 2, hence there is no need to cover the case $\varepsilon > 1$.

*Remark* 3.8. 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. (3.4) 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

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

as states with maximal trace distance have orthogonal support.

In the subsequent chapters, we will use Theorem 3.6 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 related entropic quantities. Depending on the case, we will sometimes have to employ the whole machinery devised in Theorem 3.6, whereas at other times the simplification provided in Remark 3.8 will be enough.

Chapter 4

# The Umegaki relative entropy

In this chapter, we apply the ALAFF method introduced in Chapter 3 for the particular case of the relative entropy, as well as some other entropic quantities derived from it. Since the relative entropy is in particular a divergence, it is (jointly) convex. Thus it remains to show that this quantity satisfies proper almost concavity. The proof of that feature, as well as the tightness of the result obtained, is presented in Section 4.1.

The ALAFF method then yields a plethora of results of uniform continuity for entropic quantities derived from the relative entropy. These are all presented in Section 4.2. In particular, we recover the well-known (and almost tight) continuity bound for the conditional entropy by Winter [2].

All the results provided in this chapter are summarized in Fig. 4.1.

## 4.1 Almost concavity for the relative entropy

The (joint) convexity of the relative entropy is a well-established result with proofs found for example in [24]. In this section, we complement this result with almost concavity and further prove that the bound we obtain is tight.

**Theorem 4.1** (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* $\sigma = p\sigma_1 + (1-p)\sigma_2$,

$$D(\rho\|\sigma) \geq pD(\rho_1\|\sigma_1) + (1-p)D(\rho_2\|\sigma_2) - h(p)\frac{1}{2}\|\rho_1 - \rho_2\|_1 - f_{c_1,c_2}(p)\,. \quad (4.1)$$

*Here,*

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

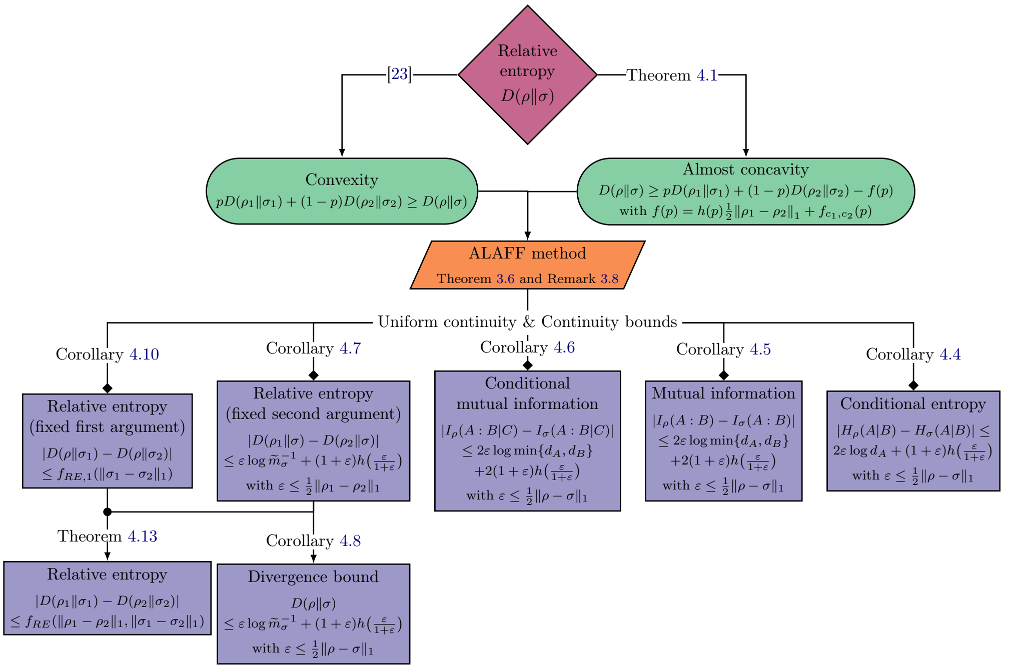



Figure 4.1: In this flow chart we collect the main results from this chapter, starting with the almost concavity of the relative entropy, which together with the ALAFF method outputs a collection of continuity bounds for related entropic quantities. For the convexity and almost concavity, we are setting $\rho = p\rho_1 + (1-p)\rho_2$ and $\sigma = p\sigma_1 + (1-p)\sigma_2$, with $p \in [0,1]$. We denote by $\widetilde{m}_\sigma$ 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.

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

$$c_1 := \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\left[\rho_1 \sigma_1^{\frac{it-1}{2}} \sigma_2 \sigma_1^{\frac{-it-1}{2}}\right] < \infty,$$

$$c_2 := \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\left[\rho_2 \sigma_2^{\frac{it-1}{2}} \sigma_1 \sigma_2^{\frac{-it-1}{2}}\right] < \infty.$$

*Here, $\beta_0$ is a probability density on $\mathbb{R}$ (see Eq. (4.4) 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 $\mathcal{S}_{\text{ker}}$ is a convex set and that the bound holds trivially for $p = 0$ and $p = 1$. Hence let $p \in (0,1)$ in the following and $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in \mathcal{S}_{\text{ker}}$.

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 [25, Theorem 14]

$$S(\rho) \leq \frac{1}{2}\|\rho_1 - \rho_2\|_1 h(p) + pS(\rho_1) + (1-p)S(\rho_2)\,. \tag{4.2}$$

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

$$\begin{aligned} p\operatorname{tr}[\rho_1(\log(\sigma) - \log(\sigma_1))] &= p\operatorname{tr}\left[\exp(\log(\rho_1))(\log(\sigma) - \log(\sigma_1))\right] \\ &\leq p\log\operatorname{tr}\left[\exp\left(\log(\rho_1) + \log(\sigma) - \log(\sigma_1)\right)\right] \\ &\leq p\log\int_{-\infty}^{\infty} dt\,\beta_0(t)\ \operatorname{tr}\left[\rho_1\sigma_1^{\frac{it-1}{2}}\sigma\sigma_1^{\frac{-it-1}{2}}\right]\,. \end{aligned} \tag{4.3}$$

The first estimate follows immediately using the well-known Peierls-Bogolubov inequality [7]. The second one is the inequality in Eq. (2.2) from Corollary 2.3 with $H_1 = \log\rho_1$, $H_2 = -\log\sigma_1$ and $H_3 = \log\sigma$. Note that

$$\beta_0(t) = \frac{\pi}{2}\frac{1}{\cosh(\pi t) + 1}\,, \tag{4.4}$$

appearing in this inequality is a probability density on $\mathbb{R}$. In the above steps, i.e. Eq. (4.3), 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 A.1. Note here that in the most general case $\cdot^{-1}$ in the RHS of Eq. (4.3) is the Moore-Penrose pseudoinverse. The trace in the integral can now be estimated for each $t$ by

$$\operatorname{tr}\left[\rho_1\sigma_1^{\frac{it-1}{2}}\sigma\sigma_1^{\frac{-it-1}{2}}\right] = p + (1-p)\operatorname{tr}\left[\rho_1\sigma_1^{\frac{it-1}{2}}\sigma_2\sigma_1^{\frac{-it-1}{2}}\right]. \tag{4.5}$$

Here, we just split $\sigma$ and used the cyclicity of the trace to get rid of the unitary. To see that $c_1 < \infty$, we upper bound $\sigma_2$ by $\mathbb{1}$ and $\sigma_1^{-1}$ by $\widetilde{m}_{\sigma_1}^{-1}\mathbb{1}$ where $\widetilde{m}_{\sigma_1}$ is the smallest non-zero eigenvalue of $\sigma_1$. This is valid, since $\ker\sigma_1 \subseteq \ker\rho_1$. We end up with $c_1 \leq \widetilde{m}_{\sigma_1}^{-1} < \infty$. Inserting Eq. (4.5) into Eq. (4.3), we obtain the first part of $f_{c_1,c_2}(p)$ and repeating the steps for $(1-p)\operatorname{tr}[\rho_2(\log(\sigma) - \log(\sigma_2))]$ the second one as well. This concludes the proof. □

We remark that Eq. (4.1) provides a result of almost concavity for the relative entropy in the sense of Definition 3.1. Indeed, the additive "correction" term obtained 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 4.2** (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 4.1 is well behaved in the following sense: Let $j = 1, 2$ and $(\rho_j, \sigma_j) \in \mathcal{S}_{\mathrm{ker}}$. 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 each $\sigma_j$, has a minimal non-zero eigenvalue that is bounded from below by some $\widetilde{m} > 0$, then $f_{c_1,c_2} + h\frac{1}{2}\|\rho_1 - \rho_2\|_1 \leq f_{\widetilde{m}^{-1},\widetilde{m}^{-1}} + h$.*
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 $[0, 1)$.*

Finally, before using these results of almost concavity for the relative entropy jointly with the ALAFF method to provide some continuity bounds for the relative entropy and derived quantities, we conclude this section with some discussion of our almost concave bound. In Proposition 4.2, we have shown that our almost concave bound is well behaved in the sense that, in some specific cases, it is independent of the states. However, we can additionally show that it is tight, meaning there exist states that saturate the inequality in Eq. (4.1).

**Proposition 4.3** (Almost concavity estimate of the relative entropy is tight)**.**
*The bound presented in Theorem 4.1 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. (4.1).*

*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 evaluates to

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

and

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

This gives us

$$\begin{aligned} &pD(\rho_1\|\sigma_1) + (1-p)D(\rho_2\|\sigma_2) - D(\rho\|\sigma) \\ &\qquad = h(p) + p\log\left(p + (1-p)\frac{1-t}{t}\right) + (1-p)\log\left((1-p) + p\frac{1-t}{t}\right). \end{aligned} \tag{4.6}$$

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 4.1 are given by

$$c_i = \mathrm{tr}\left[\rho_i\sigma_i^{-1}\sigma_j\right] = \frac{1-t}{t}\,.$$

for $i, j = 1, 2$ and $i \neq j$. Finally 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. (4.6) from the almost concavity estimate in Eq. (4.1). This concludes the claim. □

## 4.2 Continuity bounds for the relative entropy

In this section we will harvest the fruits of our work and prove a number of corollaries that are direct consequences of the results of almost concavity in Theorem 4.1 and Proposition 4.2 in combination with the results in Theorem 3.6 and Remark 3.8. All of them concern quantities which are derived from the relative entropy.

### 4.2.1 Uniform continuity for the conditional entropy

Let us first consider a bipartite space and the conditional entropy of a state 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 state, but we do not require positive definiteness. This should be compared to the findings of the subsequent chapter for the BS-entropy and derived quantities, where discontinuities appear with vanishing eigenvalues.

**Corollary 4.4** (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*

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

*Proof.* First of all, $\mathcal{S}_0$ is clearly 0-perturbed $\Delta$-invariant. Setting $f(\cdot) = H_.(A|B)$, we find that it is ALAFF with $a_f = 0$ as $H_.(A|B)$ is concave, and $b_f = h$ since the result in Theorem 4.1 becomes independent of the states as we go to $H_.(A|B)$ using point 3 of Proposition 4.2. Finally, we find that

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

where we used $-\log d_X \leq H_.(X|Y) \leq \log d_X$ shown, for example, in [24]. Using Theorem 3.6 in the form of Remark 3.8, we can infer the claimed continuity bound. □

As we have already mentioned, this coincides with the result of Winter [2], which proved to be almost tight.

### 4.2.2 Uniform continuity for the mutual information

The previous result can now be easily adapted to the mutual information. Indeed, as the mutual information can be obtained from a conditional entropy and a von Neumann entropy, a continuity bound for the former in terms of those for the latter quantities is a direct consequence.

**Corollary 4.5** (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*

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

*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. (4.2) and $-H_.(A|B)$ is almost locally affine with $a_{-H_.(A|B)} = 0$ and $b_{-H_.(A|B)} = h$. Combined we get for $f(\cdot) = I_.(A:B)$, $a_f = h$ and $b_f = h$. We further have that

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

where we used that $0 \leq I.(A : B)$ and $I.(A : B) \leq 2 \log \min\{d_A, d_B\}$ [24]. Applying Theorem 3.6 in the form of Remark 3.8, we can conclude the claim and obtain the given continuity bound. $\square$

This bound also coincides with the tightest previously-known continuity bound for the mutual information (see e.g. [20]).

### 4.2.3 Uniform continuity for the conditional mutual information

Next, we use again a similar approach to derive the result for the conditional mutual information. Note that this can be done by viewing the conditional mutual information as the difference between two mutual informations.

**Corollary 4.6** (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*

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

*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 $f(\cdot) = I.(A : B|C) = H.(A|BC) - H.(A|C)$. 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

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

as the conditional mutual information is non-negative and again $-\log d_X \leq H.(X|Y) \leq \log d_X$. Using Theorem 3.6 in the form of Remark 3.8, we can conclude the claim and obtain the given continuity bound. $\square$

This continuity bound for the conditional mutual information also coincides with the best previously-known continuity bound for the named quantity (see e.g. [26, Lemma 4]).

### 4.2.4 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 [27, 28, 29]; however, we consider this name to be a bit misleading since the bound involves $\rho$ and $\sigma$. For a continuity bound, 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, to prevent confusion with the result in Section 4.2.5. This name is fitting, since we are relating the strength of divergence (between $\rho$ and $\sigma$) to a fixed norm distance (the one norm).

We now give the divergence bound we obtain when using the convexity and almost concavity of $D(\rho\|\sigma)$ together with Theorem 3.6 by going through uniform continuity of the relative entropy in its first argument.

**Corollary 4.7** (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*

$$|D(\rho_1\|\sigma) - D(\rho_2\|\sigma)| \leq \varepsilon \log \widetilde{m}_\sigma^{-1} + (1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big),$$

*with $\widetilde{m}_\sigma$ 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 4.1 and point 1 of Proposition 4.2, we find that $D(\cdot\|\sigma)$ is ALAFF with $a_f = h$ and $b_f = 0$. At last, we have that

$$C_f^\perp = \sup_{\substack{\rho_1,\rho_2\in\mathcal{S}_0 \\ \frac{1}{2}\|\rho_1-\rho_2\|=0}} |D(\rho_1\|\sigma) - D(\rho_2\|\sigma)| \leq \sup_{\rho\in\mathcal{S}(\mathcal{H})} D(\rho\|\sigma) \leq \log \widetilde{m}_\sigma^{-1}.$$

In the first inequality, we used that $D(\rho\|\sigma) \geq 0$, and in the second one that $\widetilde{m}_\sigma \rho \leq \sigma$ hence $D(\rho\|\sigma) \leq \log \widetilde{m}_\sigma^{-1}$. Using Theorem 3.6 in the form of Remark 3.8 concludes the claim. □

We can subsequently use the Corollary 4.7 to prove a divergence bound for the relative entropy.

**Corollary 4.8** (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*

$$D(\rho\|\sigma) \leq \varepsilon \log \widetilde{m}_\sigma^{-1} + (1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big).$$

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

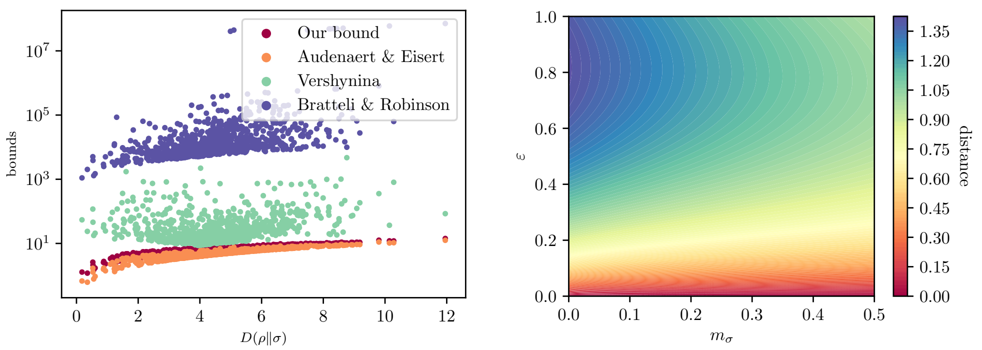
*Proof.* In the context of Corollary 4.7, we just set $\rho_1 = \rho$ and $\rho_2 = \sigma$, giving us that $\frac{1}{2}\|\rho_1 - \rho_2\|_1 = \frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$. Furthermore, $D(\rho_2\|\sigma) = D(\sigma\|\sigma) = 0$ and $|D(\rho_1\|\sigma)|$ loses the absolute value, as $D(\cdot\|\cdot) \geq 0$. The bound follows immediately. □

*Remark* 4.9. For a better understanding of the dependence of the previous divergence bound in terms of $\varepsilon$, we can use the following inequality:

$$(1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big) \leq \sqrt{2\varepsilon}\,,$$

jointly with the fact that $\varepsilon \leq \sqrt{\varepsilon}$ for any $\varepsilon \in [0,1]$. Therefore, we obtain

$$D(\rho\|\sigma) \leq \Big(1 + \frac{\log \widetilde{m}_\sigma^{-1}}{\sqrt{2}}\Big)\|\rho - \sigma\|_1^{1/2}\,.$$




(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 4.1.

(b) The difference between the bound from Corollary 4.8 and the one of Audenaert & Eisert [30, 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 4.2: Two plots comparing the divergence bounds from Table 4.1.

Some bounds for the relative entropy between two density operators in a similar direction as ours have previously appeared in the literature. In particular, in [31, 29], 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 [32] a similar bound is provided in terms of the operator norm of the difference between the states. One of the bounds in [31] is further generalised in [30] 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 [32, 29]. In Table 4.1 and Fig. 4.2 we compare the aforementioned bounds from [30, 32, 29]. From Fig. 4.2a 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. 4.2a and second plot Fig. 4.2b 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.

| Bound by | not full rank $\rho$ | not full rank $\sigma$ | Bound on $D(\rho\\|\sigma)$ |
|---|---|---|---|
| Corollary 4.8 | ✓ | ✓ | $\varepsilon \log \widetilde{m}_\sigma^{-1} + (1+\varepsilon) h\left(\frac{\varepsilon}{1+\varepsilon}\right)$ |
| Audenaert & Eisert [30, 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 [29] | x | x | $2\varepsilon\lambda_\rho \frac{\log m_\rho - \log m_\sigma}{m_\rho - m_\sigma}$ |
| Bratteli & Robinson [32] | x | x | $m_\sigma^{-1}\\|\rho-\sigma\\|_\infty$ |

Table 4.1: A comparison of different divergence bounds. Here $\varepsilon = \frac{1}{2}\|\rho - \sigma\|_1$ and $m_.$ and $\widetilde{m}_.$ 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 $m_\rho \to +0$.

### 4.2.5 Continuity bounds for the relative entropy

We conclude this chapter 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 of the relative entropy simultaneously. This presents some challenges that need to be dealt with, as the relative entropy presents problems of 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 4.10** (Uniform continuity of the relative entropy in the second argument)**.**
*Let $\rho \in \mathcal{S}(\mathcal{H})$ be fixed and $1 > \widetilde{m} > 0$. Then, $D(\rho\|\cdot)$ is uniformly continuous on*

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

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

$$|D(\rho\|\sigma_1) - D(\rho\|\sigma_2)| \leq \frac{\varepsilon}{l_{\widetilde{m}}} \log\big(\widetilde{m}^{-1}\big) + \frac{l_{\widetilde{m}} + \varepsilon}{l_{\widetilde{m}}} f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{\varepsilon}{l_{\widetilde{m}} + \varepsilon}\Big) , \tag{4.7}$$

*where $l_{\widetilde{m}} = 1 - \widetilde{m}$.*

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

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

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 $s = \widetilde{m}$. This is because one can just perturb with $\tau = \rho$ and get the kernel inclusion as well as the minorization by $\widetilde{m}\rho$. Employing point 2 of Proposition 4.2 we further get that $f(\cdot) = D(\rho\|\cdot)$ satisfies Eq. (3.3) with $b_f = 0$ and $a_f = f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}$, hence $E_f = f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}$ , and using again Proposition 4.2 (point 4, since $\tilde{m} \leq 1$) we find $E_f^{\max} = f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}$. At last, we have that

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

where we used that $D(\rho\|\cdot) \geq 0$ and for the last inequality that $\widetilde{m}\rho \leq \sigma$ for all $\sigma \in \mathcal{S}_0$. Employing now Theorem 3.6 we obtain uniform continuity and the claimed continuity bound. □

*Remark* 4.11. The continuity bound obtained in the previous corollary is relatively involved. For a better understanding of its behaviour, let us remark that we can bound the last term of Eq. (4.7) in the following form:[1]

$$\frac{l_{\widetilde{m}}+\varepsilon}{l_{\widetilde{m}}} f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{\varepsilon}{l_{\widetilde{m}}+\varepsilon}\Big) \leq \frac{2\log^2 \widetilde{m}^{-1}}{l_{\widetilde{m}}}\sqrt{\varepsilon}\,,$$

if $\widetilde{m} \leq \frac{1}{2}$. Noticing now that $\varepsilon \leq \sqrt{\varepsilon}$ for any $\varepsilon \in [0,1]$, and replacing $l_{\widetilde{m}} = 1-\widetilde{m}$, we obtain the following modified continuity bound for the relative entropy in the second argument:

$$|D(\rho\|\sigma_1) - D(\rho\|\sigma_2)| \leq \frac{3\log^2 \widetilde{m}^{-1}}{1-\widetilde{m}}\sqrt{\varepsilon}\,.$$

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

**Lemma 4.12.** *Let $\rho \in \mathcal{S}(\mathcal{H})$ with* $\operatorname{rank}\rho \geq 2$*, further $\widetilde{m} \in (0,\infty)$ and*

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

*Then, the following is true:*

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

We will only give proof for the first one in Appendix A.3 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 for the relative entropy on suitable pairs of states. Since we have already explored the cases in which we either fix the second (Corollary 4.10) or first (Corollary 4.7) density operator, we now combine both results in the proof of the next theorem.

**Theorem 4.13** (Uniform continuity of the relative entropy)**.**
*Let $1 > 2\widetilde{m} > 0$ and*

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

[1]This bound can be easily checked by noticing that the function $g : (0,1)\times(0,1) \to \mathbb{R}$ given by

$$g(m,\varepsilon) := 2\log^2 m^{-1}\sqrt{\varepsilon} - \varepsilon\log\Big(\varepsilon + \frac{1-m}{m}\Big) - (1-m)\log\Big(1-m+\frac{\varepsilon}{m}\Big) + (1-m+\varepsilon)\log(1-m+\varepsilon)$$

is monotonically increasing in $\varepsilon$ and decreasing in $m$. As we are interested in $m < 1/2$ it is enough to study the case $g(1/2,\varepsilon)$, for $\varepsilon \in (0,1)$. It is not difficult to check that $\lim_{\varepsilon\to 0} g(1/2,\varepsilon) \geq 0$.

*with $\widetilde{m}_\sigma$ 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 $\frac{1}{2}\|\sigma_1-\sigma_2\|_1\leq\delta\leq 1$*

$$\begin{aligned}|D(\rho_1\|\sigma_1)-D(\rho_2\|\sigma_2)|\leq\Big(\varepsilon+\frac{\delta}{l_{\widetilde{m}}}\Big)\log\big(\widetilde{m}^{-1}\big)+(1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big)\\ +2\frac{l_{\widetilde{m}}+\delta}{l_{\widetilde{m}}}f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{\delta}{l_{\widetilde{m}}+\delta}\Big)\,,\end{aligned}\tag{4.8}$$

*with $l_{\widetilde{m}}=1-\widetilde{m}$.*

*Proof.* We will prove the uniform continuity by proving that the bound Eq. (4.8) 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

$$\overline{\sigma}=\frac{1}{2}\sigma_1+\frac{1}{2}\sigma_2\,,\tag{4.9}$$

and obtain

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

Using this, we get

$$\begin{aligned}|D(\rho_1\|\sigma_1)-D(\rho_2\|\sigma_2)|\leq|D(\rho_1\|\sigma_1)-D(\rho_1\|\overline{\sigma})|+|D(\rho_1\|\overline{\sigma})-D(\rho_2\|\overline{\sigma})|\\ +|D(\rho_2\|\overline{\sigma})-D(\rho_2\|\sigma_2)|\,.\end{aligned}$$

The middle term can be bounded using Corollary 4.7 and the fact that

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

One obtains

$$|D(\rho_1\|\overline{\sigma})-D(\rho_2\|\overline{\sigma})|\leq\varepsilon\log\widetilde{m}^{-1}+(1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big).$$

The other two terms are bounded using Corollary 4.10 and the fact that $\widetilde{m}\rho_1\leq\frac{1}{2}\sigma_1\leq\overline{\sigma}$ and $\widetilde{m}\rho_2\leq\frac{1}{2}\sigma_2\leq\overline{\sigma}$ by construction of $\mathcal{S}_0$ and the definition of $\overline{\sigma}$, respectively. We therefore obtain

$$\begin{aligned}|D(\rho_1\|\sigma_1)-D(\rho_1\|\overline{\sigma})|\leq\frac{\delta}{2l_{\widetilde{m}}}\log\big(\widetilde{m}^{-1}\big)+\frac{l_{\widetilde{m}}+2^{-1}\delta}{l_{\widetilde{m}}}f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{2^{-1}\delta}{l_{\widetilde{m}}+2^{-1}\delta}\Big)\,,\\ |D(\rho_2\|\overline{\sigma})-D(\rho_2\|\sigma_2)|\leq\frac{\delta}{2l_{\widetilde{m}}}\log\big(\widetilde{m}^{-1}\big)+\frac{l_{\widetilde{m}}+2^{-1}\delta}{l_{\widetilde{m}}}f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{2^{-1}\delta}{l_{\widetilde{m}}+2^{-1}\delta}\Big)\,.\end{aligned}$$

Combining the bounds and point 2 of Proposition 4.2, as $\tilde{m}\leq 1$ and further using that

$$\frac{l_{\widetilde{m}}+2^{-1}\delta}{l_{\widetilde{m}}}f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{2^{-1}\delta}{l_{\widetilde{m}}+2^{-1}\delta}\Big)\leq\frac{l_{\widetilde{m}}+\delta}{l_{\widetilde{m}}}f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{\delta}{l_{\widetilde{m}}+\delta}\Big)\,,$$

we obtain the claimed bound, and thereby also uniform continuity. □

Similarly to the discussion of Remark 4.11, we can further simplify the latter continuity bound as we show below.

*Remark* 4.14. The continuity bound for the relative entropy from Theorem 4.13 can be simplified by bounding the terms involved. For $\widetilde{m} \leq \frac{1}{2}$ on the one side, we have

$$2\frac{l_{\widetilde{m}} + \delta}{l_{\widetilde{m}}} f_{\widetilde{m}^{-1},\widetilde{m}^{-1}}\Big(\frac{\delta}{l_{\widetilde{m}} + \delta}\Big) \leq \frac{4\log^2 \widetilde{m}^{-1}}{l_{\widetilde{m}}}\sqrt{\delta}\,.$$

whereas, on the other side, we can bound the binary entropy term using

$$(1+\varepsilon)h\Big(\frac{\varepsilon}{1+\varepsilon}\Big) \leq \sqrt{2\varepsilon}\,.$$

Using these inequalities, jointly with the fact that $\varepsilon, \delta \in (0,1]$, and thus $\varepsilon \leq \sqrt{\varepsilon}$ and $\delta \leq \sqrt{\delta}$, as well as replacing $l_{\widetilde{m}} = 1 - \widetilde{m}$, we obtain the following simplified continuity bound for the relative entropy:

$$|D(\rho_1\|\sigma_1) - D(\rho_2\|\sigma_2)| \leq \Big(1 + \frac{\log \widetilde{m}^{-1}}{\sqrt{2}}\Big)\|\rho_1 - \rho_2\|_1^{1/2} + \frac{5\log^2 \widetilde{m}^{-1}}{\sqrt{2}(1-\widetilde{m})}\|\sigma_1 - \sigma_2\|_1^{1/2}\,. \tag{4.10}$$

Let us conclude this section by emphasizing that there might be some room for improvement in the previous result. For instance, it might be possible to improve the interpolation between $\sigma_1$ and $\sigma_2$ considered in Eq. (4.9) by optimizing over some probabilities $p$ and $1-p$ associated to $\sigma_1$ and $\sigma_2$, respectively, instead of just assigning both probability $1/2$. However, we believe this would not change the appearance of the bound drastically and thus the reason for not performing this optimization.

Chapter 5

# The Belavkin-Staszewski entropy

Following the same lines as in the previous chapter, we deploy the ALAFF method introduced in Chapter 3 to the particular case of the BS-entropy. For that, we need to prove a result of almost concavity, which is presented in Section 5.1. However, in contrast to the case of the relative entropy, our result for the BS-entropy is not tight. We leave the discussion of the bound and the difficulties that arise in the BS-entropy case for the next section.

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 constructed from the BS-entropy in Definition 2.7. All the continuity bounds obtained in this section are summarized in Fig. 5.1.

## 5.1 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 [33, Theorem 4.4], [18, Corollary 4.7]. We first need to give some auxiliary results that are necessary to proceed. The first of these auxiliary results is an operator inequality for the term inside the trace in the definition of the BS-entropy.

**Lemma 5.1.** *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 $h_{A_1,A_2}(p) = -p\log(p)\operatorname{tr}[A_1] - (1-p)\log(1-p)\operatorname{tr}[A_2]$ the distorted binary entropy.*

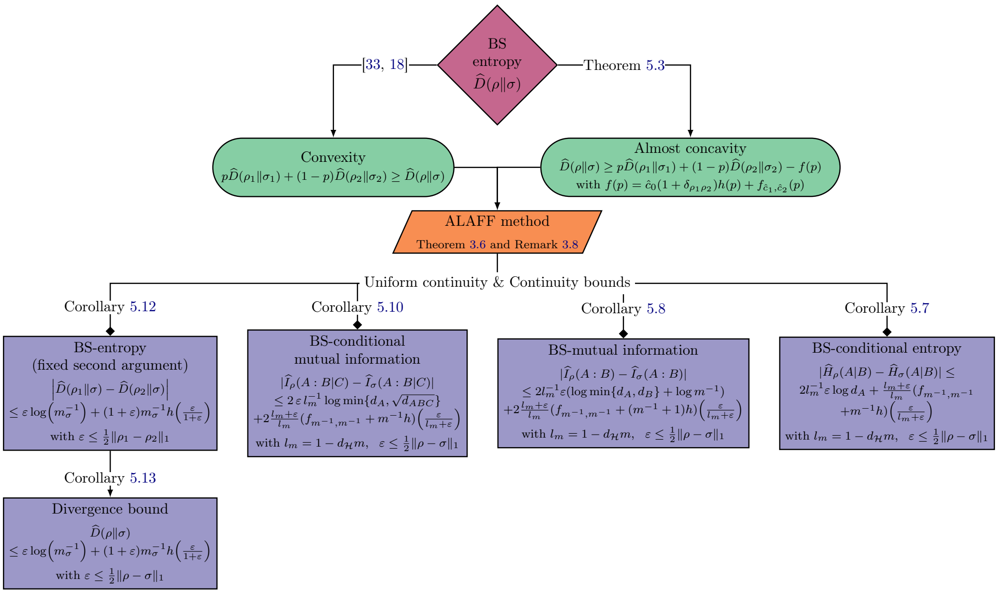



Figure 5.1: In this flow chart we collect the main results from this chapter, starting with the almost concavity for the BS-entropy, which together with the ALAFF method outputs a plethora of continuity bounds for entropic quantities. For the convexity and almost concavity of the BS-entropy we are setting $\rho = p\rho_1 + (1-p)\rho_2$ 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 $\left\|\sigma_1^{-1}\right\|_\infty$ and $\left\|\sigma_2^{-1}\right\|_\infty$. Additionally, we assume in all the continuity bounds that $m \leq \left\|\eta^{-1}\right\|_\infty$, for $\eta = \sigma, \rho$.

*Proof.* It holds that

$$\begin{aligned} &-A\log(A) + pA_1\log(A_1) + (1-p)A_2\log(A_2) \\ &\qquad \leq \|-A\log(A) + pA_1\log(A_1) + (1-p)A_2\log(A_2)\|_1 \mathbb{1} \,. \end{aligned} \tag{5.1}$$

Now, since $x \mapsto -x\log(x)$ is operator concave [3, Theorem 2.6], we have

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

giving us that

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

and hence

$$\begin{aligned} &\|-A\log(A) + pA_1\log(A_1) + (1-p)A_2\log(A_2)\|_1 \\ &\qquad = \operatorname{tr}[-A\log(A) + pA_1\log(A_1) + (1-p)A_2\log(A_2)]\,. \end{aligned} \tag{5.2}$$

We now use operator monotonicity of the logarithm to find

$$\begin{aligned}
-\operatorname{tr}[A\log(A)] &= -p\operatorname{tr}[A_1\log(A)] - (1-p)\operatorname{tr}[A_2\log(A)]\\
&\le -p\operatorname{tr}[A_1\log(pA_1)] - (1-p)\operatorname{tr}[A_2\log((1-p)A_2)]\\
&= -p\operatorname{tr}[A_1\log(A_1)] - (1-p)\operatorname{tr}[A_2\log(A_2)] + h_{A_1,A_2}(p)\,.
\end{aligned}$$

Inserting this into Eq. (5.2) and then into Eq. (5.1) yields the claimed result. □

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. [34, Eq. (7.35)]). We include here a short proof of this result for completeness.

**Lemma 5.2.** *Let $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{S}_+(\mathcal{H})$, then*

$$\widehat{D}(\rho\|\sigma) = \operatorname{tr}\Big[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})\log\Big(\sigma^{-1/2}\rho\sigma^{-1/2}\Big)\Big]\,.$$

*Proof.*

$$\begin{aligned}
\widehat{D}(\rho\|\sigma) &= \operatorname{tr}\Big[\rho\log\Big(\rho^{1/2}\sigma^{-1}\rho^{1/2}\Big)\Big]\\
&= \operatorname{tr}\Big[\log\Big(\rho^{1/2}\sigma^{-1/2}\sigma^{-1/2}\rho^{1/2}\Big)\rho^{1/2}\sigma^{-1/2}\sigma^{1/2}\rho^{1/2}\Big]\\
&= \operatorname{tr}\Big[\rho^{1/2}\sigma^{-1/2}\log\Big(\sigma^{-1/2}\rho\sigma^{-1/2}\Big)\sigma^{1/2}\rho^{1/2})\Big]\\
&= \operatorname{tr}\Big[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})\log\Big(\sigma^{-1/2}\rho\sigma^{-1/2}\Big)\Big]\,.
\end{aligned}$$

We used the cyclicity of the trace several times, and further, the 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 $f$ [35, Lemma 61.]. □

Building on the previous results from this section, we proceed to prove now the main result, namely the almost concavity for the BS-entropy, in the line of results of almost concavity discussed in Definition 3.1.

**Theorem 5.3** (Almost concavity of the BS-entropy)**.**
*Let $(\rho_1,\sigma_1),(\rho_2,\sigma_2) \in \mathcal{S}_{\ker,+}$ with*

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

*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*

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

*with*

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

*and the constants*

$$
\begin{aligned}
\hat{c}_0 &:= \max\{\left\|\sigma_1^{-1}\right\|_\infty, \left\|\sigma_2^{-1}\right\|_\infty\}\,,\\
\hat{c}_1 &:= \int_{-\infty}^{\infty} dt\beta_0(t)\,\mathrm{tr}\Big[\rho_1(\rho_1^{1/2}\sigma_1^{-1}\rho_1^{1/2})^{\frac{it+1}{2}}\rho_1^{-1/2}\sigma_2\rho_1^{-1/2}(\rho_1^{1/2}\sigma_1^{-1}\rho_1^{1/2})^{\frac{-it+1}{2}}\Big]\,,\\
\hat{c}_2 &:= \int_{-\infty}^{\infty} dt\beta_0(t)\,\mathrm{tr}\Big[\rho_2(\rho_2^{1/2}\sigma_2^{-1}\rho_2^{1/2})^{\frac{it+1}{2}}\rho_2^{-1/2}\sigma_1\rho_2^{-1/2}(\rho_2^{1/2}\sigma_2^{-1}\rho_2^{1/2})^{\frac{-it+1}{2}}\Big]\,,
\end{aligned}
\tag{5.3}
$$

*with the probability density* $\beta_0$ *defined as in Eq.* (4.4).

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

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

Indeed, as of Lemma 5.2 and then Lemma 5.1 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

$$
\begin{aligned}
-\widehat{D}(\rho\|\sigma) &= \mathrm{tr}\Big[\sigma\left(-\sigma^{-1/2}\rho\sigma^{-1/2}\log\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\right)\Big]\\
&\le p\,\mathrm{tr}\Big[\sigma\left(-\sigma^{-1/2}\rho_1\sigma^{-1/2}\log\left(\sigma^{-1/2}\rho_1\sigma^{-1/2}\right)\right)\Big]\\
&\qquad + (1-p)\,\mathrm{tr}\Big[\sigma\left(-\sigma^{-1/2}\rho_2\sigma^{-1/2}\log\left(\sigma^{-1/2}\rho_2\sigma^{-1/2}\right)\right)\Big] + h_{A_1,A_2}(p)\\
&= -p\widehat{D}(\rho_1\|\sigma) - (1-p)\widehat{D}(\rho_2\|\sigma) + h_{A_1,A_2}(p)\,.
\end{aligned}
$$

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

Now we are left with terms of the form $\widehat{D}(\rho_j\|\sigma_j) - \widehat{D}(\rho_j\|\sigma)$ for $j = 1, 2$. To estimate those we use the Peierls-Bogolubov inequality [7] and Eq. (2.2) from

Corollary 2.3 analagous to the case of the relative entropy:

$$\begin{aligned}\widehat{D}(\rho_j\|\sigma_j) - \widehat{D}(\rho_j\|\sigma) &= \operatorname{tr}\Big[\rho_j\,\Big(\log\Big(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}\Big) - \log\Big(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}\Big)\Big)\Big]\\ &\le \operatorname{tr}\Big[\exp\Big(\log(\rho_j) + \log\Big(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}\Big) - \log\Big(\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}\Big)\Big)\Big]\\ &\le \operatorname{tr}\Big[\exp\Big(\log(\rho_j) + \log\Big(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}\Big) + \log\Big(\rho_j^{-1/2}\sigma\rho_j^{-1/2}\Big)\Big)\Big]\\ \le \log\Big(\int_{-\infty}^{\infty} dt\beta_0(t)\operatorname{tr}\Big[\rho_j(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2})^{\frac{it+1}{2}}\rho_j^{-1/2}\sigma\rho_j^{-1/2}(\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2})^{\frac{-it+1}{2}}\Big]\Big)&\\ &= \begin{cases}\log(p+(1-p)\hat{c}_1) & j=1\\ \log((1-p)+p\hat{c}_2) & j=2\end{cases}.\end{aligned} \tag{5.4}$$

In the third line, we use that

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

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

$$P_\rho(P_\rho\sigma P_\rho)^{-1}P_\rho \le P_\rho\sigma^{-1}P_\rho\,,$$

as $x \to x^{-1}$ is operator convex and hence fulfills the Sherman-Davis inequality [3, Theorem 4.19]. Note that $\sigma$ is invertible and that by $(P_\rho\sigma P_\rho)^{-1}$ we mean the Moore-Penrose pseudoinverse. We find

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

The argument why Eq. (5.4) holds 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 4.1 and Appendix A.1). For the BS-entropy, we can already restrict Eq. (5.4) to the support of $\rho_j$ as all operators involved, $\rho_j$, $\rho_j^{1/2}\sigma_j^{-1}\rho_j^{1/2}$ and $\rho_j^{1/2}\sigma^{-1}\rho_j^{1/2}$, commute with the projection onto the support of $\rho_j$.

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. □

We strongly suspect that Theorem 5.3 can be improved because of two arguments. The first one is that we would expect the results of almost concavity of the relative and the BS-entropy to coincide if the involved states commute. That is because then the involved states are classical and both quantities coincide with the Kullback-Leibler divergence. A straightforward calculation shows that in 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 is generally not the case.

The other reason is given by the bounds we obtain for the BS-conditional entropy in Proposition 5.5, which show that there is no dependence on the minimal eigenvalue if the state $\sigma$ is full rank (The full rank requirement is necessary, however, as we will show in Proposition 5.6). Hence we would also suspect that an optimal bound would be eigenvalue independent in the case of the BS-conditional entropy.

As in 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 we make structural restrictions to the set of allowed states.

**Proposition 5.4** (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 5.3 is well behaved in the following sense: Let $j = 1, 2$ and $(\rho_j, \sigma_j) \in \mathcal{S}_{\mathrm{ker},+}$. 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 that 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 4.2, its analogue for the relative entropy. The proof can be found in Appendix B.1. We will use the reductions from Proposition 5.4 to simplify the terms in Theorem 5.3 for the various applications presented in the subsequent section.

## 5.2 Continuity bounds for the BS-entropy

In this section, we will use the almost concavity for the BS-entropy from Theorem 5.3 together with the ALAFF method in its full generality, Theorem 3.6, as well as the reduction of the correction terms from Proposition 5.4, to prove a collection of results of continuity bounds for entropic quantities derived from the BS-entropy. However, as we will show in the next pages, these bounds are generally more involved than in the analogous case for the relative entropy, both in their forms as well as in their proofs. In particular, all of them depend in one way or another on the minimal eigenvalue of the second input in the BS-entropy. The reason for this apparent caveat will become clear in the next few subsections, where we discuss the discontinuities present in the BS-entropy.

Beforehand, we need to collect some lower and upper estimates of certain entropic quantities derived from the BS-entropy (see Definition 2.7 and the following for the definitions).

**Proposition 5.5** (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:*
$$-\log \min\{d_A, d_B\} \leq \widehat{H}_\rho(A|B) \leq \log d_A \,. \tag{5.5}$$
2. *For the BS-mutual information:*
$$0 \leq \widehat{I}_\rho(A:B) \leq \log \min\{d_A, d_B\} + \log \min\{\left\|\rho_A^{-1}\right\|_\infty, \left\|\rho_B^{-1}\right\|_\infty\}\,, \tag{5.6}$$
*with $\cdot^{-1}$ the Moore-Penrose pseudoinverse.*
3. *For $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, for the BS-conditional mutual information:*
$$0 \leq \widehat{I}_\rho(A:B|C) \leq \min\{\log d_A^2, \log d_{ABC}\}\,.$$

*The two first 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 B.2. We further want to remark that the scaling of the bound with respect to the minimal non-zero eigenvalue of $\rho_A$ or $\rho_B$ is justified. The reasoning can be found in Appendix B.2 as well.

### 5.2.1 Uniform continuity for the BS-conditional entropy

The case of the BS-conditional entropy is more involved than the one of the conditional entropy that we have covered in Corollary 4.4. This is because the almost concave bound of the BS-entropy depends on the minimal eigenvalue of the second argument (see Eq. (5.3)), hence we need to require the second argument to be full rank, which in the case of the BS-conditional entropy means we have to require the argument 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 [36, Remark 3.3] for a similar behaviour of the sharp quantum Rényi divergences.

**Proposition 5.6** (Discontinuity of the BS-conditional entropy on positive semi-definite states)**.**
*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 further fulfil

$$\begin{aligned}\|\rho_0 - \rho_\varepsilon\|_1 &= \frac{1}{2}\||1_A\rangle\langle 1_A| \otimes (|0_B\rangle\langle 0_B| - |\varepsilon_B\rangle\langle\varepsilon_B|)\|_1\\ &= \frac{1}{2}\||0_B\rangle\langle 0_B| - |\varepsilon_B\rangle\langle\varepsilon_B|\|_1 = \sqrt{\varepsilon}\,.\end{aligned} \tag{5.7}$$

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

$$|0_B\rangle\langle 0_B| \to \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \quad \text{and} \quad |\varepsilon_B\rangle\langle\varepsilon_B| \to \begin{pmatrix}1-\varepsilon & \sqrt{\varepsilon}\sqrt{1-\varepsilon}\\ \sqrt{\varepsilon}\sqrt{1-\varepsilon} & \varepsilon\end{pmatrix}. \tag{5.8}$$

Calculating the eigenvalues of the difference and taking the sum of their absolute value gives $2\sqrt{\varepsilon}$ and thereby Eq. (5.7). Since clearly $[\rho_1, \mathbb{1} \otimes \mathrm{tr}_A[\rho_1]] = 0$, the BS-conditional and conditional entropy coincide and we find

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

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

$$\begin{aligned}\widehat{H}_{\rho_\varepsilon}(A|B) &= \frac{1}{2}\,\mathrm{tr}\Big[|0_B\rangle\langle 0_B| \log\Big(|0_B\rangle\langle 0_B|^{1/2}\,(|\varepsilon_B\rangle\langle\varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}\,|0_B\rangle\langle 0_B|^{1/2}\Big)\Big]\\ &\quad + \frac{1}{2}\,\mathrm{tr}\Big[|\varepsilon_B\rangle\langle\varepsilon_B| \log\Big(|\varepsilon_B\rangle\langle\varepsilon_B|^{1/2}\,(|\varepsilon_B\rangle\langle\varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}\,|\varepsilon_B\rangle\langle\varepsilon_B|^{1/2}\Big)\Big]\\ &= \frac{1}{2}\log\mathrm{tr}\big[|0_B\rangle\langle 0_B|\,(|\varepsilon_B\rangle\langle\varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}\big]\\ &\qquad + \frac{1}{2}\log\mathrm{tr}\big[|\varepsilon_B\rangle\langle\varepsilon_B|\,(|\varepsilon_B\rangle\langle\varepsilon_B| + |0_B\rangle\langle 0_B|)^{-1}\big]\,,\end{aligned} \tag{5.9}$$

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. (5.8), that

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

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

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

If we insert this into Eq. (5.9), we find $\widehat{H}_{\rho_\varepsilon}(A|B) = 0$. □

The author would like to thank Peter Brown for the above counterexample. It shows in particular that we could only expect to be able to prove uniform continuity for the BS-conditional entropy for full-rank states. The presence of the minimal eigenvalue of the states in the continuity bound provided below for the BS-conditional entropy is thus not surprising.

**Corollary 5.7** (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*

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

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

*Proof.* We find that $\mathcal{S}_0$ is $s$-perturbed $\Delta$-invariant with $s = m d_{\mathcal{H}}$. The justification of this choice is completely analogous to the reasoning in Lemma 4.12 with $\rho = d_{\mathcal{H}}^{-1}\mathbb{1}$, i.e. the maximally mixed state. Furthermore, $f(\cdot) = \widehat{H}_\cdot(A|B)$ is ALAFF with $a_f = 0$ as $\widehat{H}_\cdot(A|B)$ is concave, and $b_f = m^{-1}h + f_{m^{-1},m^{-1}}$ since the result in Section 5.1 becomes independent of the states as we go to $\widehat{H}_\cdot(A|B)$ using point 3 of Proposition 5.4. 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 5.5. This allows us to apply Theorem 3.6 where $E_f^{\max}$ coincides with $E_f$ as of point 4 in Proposition 5.4. This concludes the claim. □

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 4.4, 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. 5.2. The possibility of obtaining such a tighter bound is left for future work.

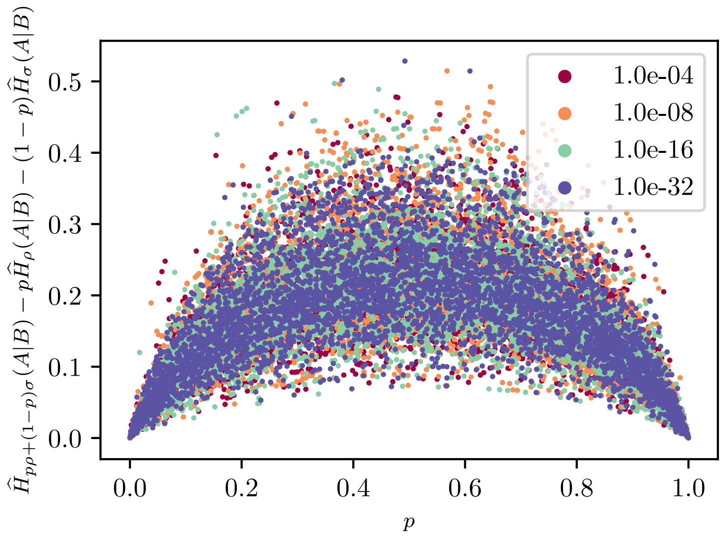



Figure 5.2: We investigate the dependence of the almost convex remainder of the BS-conditional entropy on the minimal eigenvalue of the involved states. For the minimal eigenvalues $10^{-4}, 10^{-8}, 10^{-16}, 10^{-32}$ we sampled five hundred pairs of qubits $(\rho, \sigma)$ both of them with controlled 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 as an upper bound. The plots show a similar pattern if the dimension is increased.

### 5.2.2 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 5.6 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 5.5, 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. (5.6). For this reason, a continuity bound for the BS-mutual information necessarily will depend on the minimal eigenvalues of the states involved.

**Corollary 5.8** (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*

$$|\widehat{I}_\rho(A:B) - \widehat{I}_\sigma(A:B)| \leq 2l_m^{-1}\varepsilon(\log\min\{d_A, d_B\} + \log m^{-1}) + \frac{l_m + \varepsilon}{l_m} z_m\Big(\frac{\varepsilon}{l_m + \varepsilon}\Big),$$

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

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

*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 $I.(A:B)$ is ALAFF we first note that because of the convexity of $\widehat{D}(\cdot\|\cdot)$

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

In the last step, we further used that $\widehat{D}(\cdot\|\cdot)$ 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 $a_f = h$. We follow similar lines to obtain $b_f$. Starting with Theorem 5.3 and point 2 in Proposition 5.4 using that $\|\rho_A^{-1}\|_\infty \leq \|\rho_{AB}^{-1}\|_\infty$, and analogously for $\rho_B$, we find

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

In the last step we used again that $\widehat{D}(\cdot\|\cdot)$ is monotone decreasing in its second argument and that $p\rho_{1,A} + (1-p)\rho_{2,A} \leq (p + (1-p)m^{-1})\rho_{1,A}$ and $p\rho_{1,A} + (1-p)\rho_{2,A} \leq (m^{-1}p + (1-p))\rho_{2,A}$, giving us another $f_{m^{-1},m^{-1}}(p)$. Hence $b_f = m^{-1}h + 2f_{m^{-1},m^{-1}}$. 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 3.6 and get the claimed bounds as $E_f$ coincides with $E_f^{\max}$, due to point 4 in Proposition 5.4. □

Again we further simplify the continuity bound from the previous result, as we did in Remark 4.11 and Remark 4.14.

*Remark* 5.9. To simplify the continuity bound from Corollary 5.8, let us upper bound each of the terms $z_m$. Firstly, for $f$, we have

$$2\frac{l_m+\varepsilon}{l_m}f_{m^{-1},m^{-1}}\Big(\frac{\varepsilon}{l_m+\varepsilon}\Big) \le \frac{4\log^2 m^{-1}}{l_m}\sqrt{\varepsilon}\,.$$

Then, we can bound the binary entropy term similarly, as the following holds for any $a \in (0,1]$:

$$(a+x)h\left(\frac{x}{a+x}\right) \le \sqrt{2x}\,,$$

since an inspection of the derivative shows that the function is non-decreasing in $a$. Then, we obtain:

$$(m^{-1}+1)\frac{l_m+\varepsilon}{l_m}h\Big(\frac{\varepsilon}{l_m+\varepsilon}\Big) \le \frac{m^{-1}+1}{l_m}\sqrt{2\varepsilon}\,.$$

Therefore, combining these inequalities with the fact that $\varepsilon \le \sqrt{\varepsilon}$ for $\varepsilon \le 1$, and replacing $l_m = 1 - md_{\mathcal{H}}$, we have

$$|\widehat{I}_\rho(A:B) - \widehat{I}_\sigma(A:B)| \le \frac{2\log\min\{d_A,d_B\} + 6\log^2 m^{-1} + \sqrt{2}(m^{-1}+1)}{1-md_{\mathcal{H}}}\sqrt{\varepsilon}\,.$$

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

Next, we provide a result of uniform continuity for the BS-conditional mutual information, as defined in Eq. (2.5). This constitutes the analogue of its relative entropy counterpart, presented in Corollary 4.6. Since the BS-conditional mutual information considered in this manuscript is the difference between two BS-conditional entropies, a continuity bound for the former can be directly obtained from a continuity bound for the latter. Moreover, 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 5.5 for the specific bounds on all these BS-entropic quantities.

Nevertheless, the continuity bound we obtain for the BS-conditional mutual information also depends on the states' minimal eigenvalues, as in the case of the BS-conditional entropies. We again believe that our bound should be improbable. However, there is no uniform continuity for positive semi-definite states due to the discontinuities of the BS-conditional entropy (Proposition 5.6).

**Corollary 5.10** (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*

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

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

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

*Proof.* We have that $\mathcal{S}_0$ is $s$-perturbed $\Delta$-invariant using the same reasoning as in the proof of Corollary 5.7. 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 5.7. Using Proposition 5.5 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 3.6 and using point 4 of Proposition 5.4 we get that $E_f = E_f^{\max}$ and thereby conclude the claim. □

Similarly to Remark 5.9, we can write the previous continuity bound in a more straightforward form.

*Remark* 5.11. We can further simplify the continuity bound from Corollary 5.10 by bounding the $g_m$ term as in the case of Corollary 5.8. Firstly, for $f$, we have

$$2\frac{l_m + \varepsilon}{l_m} f_{m^{-1},m^{-1}}\Big(\frac{\varepsilon}{l_m + \varepsilon}\Big) \leq \frac{4\log^2 m^{-1}}{l_m}\sqrt{\varepsilon}\,.$$

Then, we can bound the binary entropy term similarly:

$$2m^{-1}\frac{l_m + \varepsilon}{l_m} h\Big(\frac{\varepsilon}{l_m + \varepsilon}\Big) \leq \frac{2m^{-1}}{l_m}\sqrt{2\varepsilon}\,.$$

Combining these inequalities with the fact that $\varepsilon \leq \sqrt{\varepsilon}$ for $\varepsilon \leq 1$, and replacing $l_m = 1 - md_{\mathcal{H}}$, we have

$$|\widehat{I}_\rho(A:B|C) - \widehat{I}_\sigma(A:B|C)| \leq \frac{2\log\min\{d_A, \sqrt{d_{ABC}}\} + 4\log^2 m^{-1} + 2\sqrt{2}m^{-1}}{1 - md_{\mathcal{H}}}\sqrt{\varepsilon}\,.$$

### 5.2.4 Divergence bound for the BS-entropy

Finally, we conclude this chapter by providing a divergence bound for the BS-entropy following the same lines as in the case of the relative entropy. We will first 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 4.7 and Corollary 4.8, respectively.

**Corollary 5.12** (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*

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

*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 $b_f = 0$ using Theorem 5.3 and point 1 of Proposition 5.4. Further

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

since $\rho^{1/2}\sigma^{-1}\rho^{1/2} \leq m_\sigma^{-1}\,\mathbb{1}$. Applying now Eq. (3.3) gives the claimed result. □

Employing the above result we obtain a divergence bound for the BS-entropy which constitutes the analogue to the one of the relative entropy Corollary 4.8. Note that even the divergence bounds obtained in both cases are very similar, except for the presence of a factor $m_\sigma^{-1}$ in the second term of the bound.

**Corollary 5.13** (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*

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

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

*Proof.* In the context of Corollary 5.12, we just set $\rho_1 = \rho$ and $\rho_2 = \sigma$, giving us that $\frac{1}{2}\|\rho_1 - \rho_2\|_1 \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon \leq 1$. 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. □

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 4.10 and Theorem 4.13 for the BS-entropy, due to their high technicality and the complexity of the continuity bounds that we would obtain with our method. However, following a procedure similar to the one in the case of the relative entropy would produce analogous continuity bounds in the case of the BS-entropy.

Chapter 6

# Applications and Outlook

In this chapter, we use the continuity bounds and derive results for various contexts in the field of quantum information. Among them are results on quantum Markov chains, minimal distances to separable states and Rains bounds. We then conclude by giving an outlook on future work and open problems that we came across during working on the continuity bounds of the BS-entropy.

## 6.1 Applications

We begin the application section with the results on approximate quantum Markov chains.

### 6.1.1 Approximate Quantum Markov Chains

For this section $\mathcal{H}_{ABC} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ is a tripartite Hilbert space and $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_{ABC})$ a quantum state. We will use the result on the conditional entropy to obtain an upper bound on the conditional mutual information between $A$ and $C$ conditioned on $B$ of $\rho_{ABC}$. This upper bound will be in terms of the one norm distance of the Petz recovery of $\rho_{ABC}$ and the state itself.

The well-known property of strong subadditivity of the von Neumann entropy [37] is equivalent to the non-negativity of the conditional mutual information, which is furthermore known [38, 39] 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 $\mathcal{P}_{B\to AB}(\rho_{BC}) = \rho_{AB}^{1/2}\rho_B^{-1/2}\rho_{BC}\rho_B^{-1/2}\rho_{AB}^{1/2}$, we have

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

If we now decompose the CMI of $\rho_{ABC}$ in terms of a difference of conditional entropies and apply the data processing inequality, we obtain

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

Applying now our continuity bound for the CMI from Corollary 4.4 (which provides, in this case, a tighter result than Corollary 4.6) we find 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 [40, Eq. (26)]. Note that, as a direct consequence of Corollary 4.4, we get the following bound for any state $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_{ABC})$:

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

with

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

As in Remark 4.9, we again have

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

for every $x \in [0,1]$. Combining this with the fact that for $\varepsilon \in [0,1]$, $\varepsilon \leq \sqrt{\varepsilon}$, gives us for an upper bound of the CMI

$$I_\rho(A:C|B) \leq \left(\sqrt{2}\log\min\{d_A, d_C\} + 1\right)\left\|\rho_{ABC} - \rho_{AB}^{1/2}\rho_B^{-1/2}\rho_{BC}\rho_B^{-1/2}\rho_{AB}^{1/2}\right\|_1^{1/2} . \tag{6.1}$$

This bound should be compared to lower bounds for the conditional mutual information. On the one hand, Fawzi and Renner proved in [41] the following lower bound for such a quantity in terms of the fidelity $F(\rho,\sigma) = \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2$:

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

where $\mathcal{R}_{B\to AB}$ is another recovery map, the so-called *rotated Petz recovery*, which was explicitly constructed in [42]. Several results have been provided in this line in the past decade. Here we specifically focus on [43], in which Carlen and Vershynina proved:

$$I_\rho(A:C|B) \geq \left(\frac{\pi}{8}\right)^4 \left\|\rho_B^{-1}\right\|_\infty^{-2}\left\|\rho_{ABC}^{-1}\right\|_\infty^{-2}\left\|\rho_{ABC} - \rho_{AB}^{1/2}\rho_B^{-1/2}\rho_{BC}\rho_B^{-1/2}\rho_{AB}^{1/2}\right\|_1^4 , \tag{6.2}$$

Therefore, by combining Eq. (6.1) with Eq. (6.2) 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:

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

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* [44] (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 [45] to prove that a Gibbs state of a one-dimensional local Hamiltonian is an approximate quantum Markov chain, and subsequently, in [46] 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 [47] to study the decay of the CMI for purely generated finitely correlated states.

### 6.1.2 Difference between the relative entropy and the BS-entropy

A byproduct of our continuity bounds for the relative entropy is a quantification of the difference between the relative entropy of two states and the BS-entropy of the same states. This distance is in terms of the difference between the involved states and their image under the Matsumoto map.

**Corollary 6.1.** *Let $\rho \in \mathcal{S}(\mathcal{H})$, $\sigma \in \mathcal{S}_{\geq 2m}(\mathcal{H})$ and $m$ such that $d_{\mathcal{H}}^{-1} > 2m > 0$. 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*

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

*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*

$$|\widehat{D}(\rho\|\sigma) - D(\rho\|\sigma)| \leq \left(\varepsilon + \frac{\delta}{l_m}\right) \log\left(m^{-1}\right)$$
$$+ (1+\varepsilon) h\left(\frac{\varepsilon}{1+\varepsilon}\right) + 2\frac{l_m + \delta}{l_m} f_{m^{-1}, m^{-1}}\left(\frac{\delta}{l_m + \delta}\right), \tag{6.3}$$

*with* $l_{\widetilde{m}} = 1 - \widetilde{m}$*. In particular, if* $[\rho, \sigma] = 0$*,* $\varepsilon$ *and* $\delta$ *can be taken as* $0$ *such that the RHS of Eq. (6.3) is zero.*

*Moreover, we can further simplify the previous bound to*

$$|\widehat{D}(\rho\|\sigma) - D(\rho\|\sigma)| \leq (\sqrt{2} - \log m)\sqrt{\varepsilon} + \frac{5\log^2 m}{1-m}\sqrt{\delta}\,. \tag{6.4}$$

*Proof.* Our argument is a slight variation of Matsumoto's minimal reverse test [33] (see also [48]). 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 $p_i = \lambda_i \mathrm{tr}[\sigma P_i]$, $q_i = \mathrm{tr}[\sigma P_i]$ that

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

Obviously, if $m$ is the minimal eigenvalue of $\sigma$, then $q_i \geq m$ for all $i \in \{1, \ldots, k\}$. Thus, the assertion follows from Theorem 4.13. 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. (6.4) is a direct consequence of Remark 4.14 and, more specifically, Eq. (4.10).

□

### 6.1.3 Minimal distance to separable states

In this section, we show how to reprove the continuity bounds for the relative entropy of entanglement in [2] 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 [2], 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 [34] and Weierstrass' theorem on extreme values of such functions [49, Theorem 2.43]. Examples of $\mathcal{C}$ include $\mathrm{SEP}_{AB}$, 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 $D_{\mathrm{SEP}_{AB}}$ is known as the relative entropy of entanglement [50, 51]. It constitutes a tight upper bound on the distillable entanglement [52, 51]. This is the quantity we focus on for now.

**Lemma 6.2.** *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. Indeed, for $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ let $\sigma_1$ and $\sigma_2$ be states in $\mathcal{C}$ such that

$$D_{\mathcal{C}}(\rho_1) = D(\rho_1\|\sigma_1), \qquad D_{\mathcal{C}}(\rho_2) = D(\rho_2\|\sigma_2)\,.$$

Let $p \in [0,1]$. Then,

$$\begin{aligned} D_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2) &\leq D(p\rho_1 + (1-p)\rho_2\|p\sigma_1 + (1-p)\sigma_2) \\ &\leq pD(\rho_1\|\sigma_1) + (1-p)D(\rho_1\|\sigma_2) \\ &= pD_{\mathcal{C}}(\rho_1) + (1-p)D_{\mathcal{C}}(\rho_2)\,, \end{aligned}$$

where we have used joint convexity for the relative entropy in the second inequality. □

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

**Lemma 6.3.** *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 $\sigma$ the state that achieves the infimum in $D_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2)$. By Theorem 4.1 and point 1 of Proposition 4.2, we obtain that

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

which is the assertion. □

Finally, we need the following estimate:

**Lemma 6.4.** *Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. It holds that*

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

*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,

$$\begin{aligned}
\sup_{\substack{\rho,\sigma\in\mathcal{S}(\mathcal{H})\\ \frac{1}{2}\|\rho_1-\rho_2\|_1=1}} |D_{\mathrm{SEP}_{AB}}(\rho) - D_{\mathrm{SEP}_{AB}}(\sigma)| &\leq \sup_{|\psi\rangle\langle\psi|\in\mathcal{S}(\mathcal{H})} D_{\mathrm{SEP}_{AB}}(|\psi\rangle\langle\psi|)\\
&\leq \sup_{|\psi\rangle\langle\psi|\in\mathcal{S}(\mathcal{H})} D(|\psi\rangle\langle\psi|\,\|\tau_\psi)\\
&= \sup_{|\psi\rangle\langle\psi|\in\mathcal{S}(\mathcal{H})} -\log(\langle\psi|\,\tau_\psi\,|\psi\rangle)\\
&= \log d_A\,.
\end{aligned}$$

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

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

**Theorem 6.5.** *For $\varepsilon \in [0,1]$ and $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, it holds that for $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ with $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon$*

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

*Proof.* This follows from Theorem 3.6, using Lemma 6.2, Lemma 6.3, point 4 of Proposition 4.2, and Lemma 6.4. ☐

Theorem 6.5 recovers the bound [2, Corollary 8], proven with very similar methods, which improved over the earlier bound in [53]. 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 [54, Section 10]. Using the same arguments as for Lemma 6.2, we can prove convexity.

**Lemma 6.6.** *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 6.7.** *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, $g_d(p) := \frac{d}{p^{1/d}}h(p) - \log\bigl(1 - p^{1/d}\bigr)$ 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 $\sigma$, the best approximation of $\rho$ in $\mathcal{C}$. To this end, we will use a strategy inspired by [53]. Let $\sigma_s$ be the state achieving the infimum in

$$\inf_{\gamma \in \mathcal{C}} \widehat{D}\Bigl(\rho \,\Big\|\, s\gamma + (1-s)\frac{\mathbb{1}}{d}\Bigr)$$

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

$$\widehat{D}_{\mathcal{C}}(\rho) \leq \widehat{D}\Bigl(\rho \,\Big\|\, s\sigma_s + (1-s)\frac{\mathbb{1}}{d}\Bigr)\,.$$

Furthermore, with $\hat{\sigma}$ a state such that $\widehat{D}_{\mathcal{C}}(\rho) = \widehat{D}(\rho\|\hat{\sigma})$,

$$\begin{aligned}\widehat{D}\Bigl(\rho \,\Big\|\, s\sigma_s + (1-s)\frac{\mathbb{1}}{d}\Bigr) &\leq \widehat{D}\Bigl(\rho \,\Big\|\, s\hat{\sigma} + (1-s)\frac{\mathbb{1}}{d}\Bigr)\\ &\leq \widehat{D}_{\mathcal{C}}(\rho) - \log s\,,\end{aligned}$$

as $s\hat{\sigma} + (1-s)\frac{\mathbb{1}}{d} \geq s\hat{\sigma}$ and the logarithm is operator monotone. Note that without loss of generality, we can assume $\hat{\sigma}$ to be invertible, as $\widehat{D}_{\mathcal{C}}(\rho) < \infty$, which implies $\ker\hat{\sigma} \subseteq \ker\rho$. Thus, we can restrict $\hat{\sigma}$ to the support of $\rho$, where $\hat{\sigma}$ is positive definite. Combining this bound with Theorem 5.3, we infer

$$\begin{aligned}\widehat{D}_{\mathcal{C}}(p\rho_1 + (1-p)\rho_2) &\geq \widehat{D}\Bigl(p\rho_1 + (1-p)\rho_2 \Big\| s\sigma_s + (1-s)\frac{\mathbb{1}}{d}\Bigr) + \log s\\ &\geq p\widehat{D}_{\mathcal{C}}(\rho_1) + (1-p)\widehat{D}_{\mathcal{C}}(\rho_2) - \frac{d}{1-s}h(p) + \log s\,.\end{aligned}$$

Here, we have used point 1 of Proposition 5.4. Finally, we have to choose $s$ such that $\frac{d}{1-s}h(p) - \log s$ goes to zero for $p \to 0^+$ 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 C.1 and Lemma C.2. □

*Remark* 6.8. Note that we could have substituted $g_d$ in Lemma 6.7 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

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

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 6.4.

**Lemma 6.9.** *Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. It holds that*

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

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

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

is a separable state. Then,

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

In the first inequality, we have used that $\widehat{D}_{\mathrm{SEP}_{AB}}$ is positive and convex. Note that $\tau_\psi$ is invertible because we can without loss of generality restrict to its support. □

**Theorem 6.10.** *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 $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon$*

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

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

*Proof.* As shown in Lemma C.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 Theorem 3.6 using Lemma 6.6, Lemma 6.7 with Lemma C.1 and Lemma C.2, and Lemma 6.9. □

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. (2.4), we know that

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

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 $\widehat{H}_\rho^{\mathrm{var}}(A|B)$ from the approach in this section.

**Corollary 6.11.** *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 $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon$*

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

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

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

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

since 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. (5.5) and the fact that $\widehat{D}_{\mathcal{C}_0}(\rho) \geq 0$ for all states $\rho$, we obtain

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

The assertion follows from combining the above with Lemma 6.6, Lemma 6.7 with Lemma C.1, Lemma C.2, and Lemma C.3 to apply Theorem 3.6. □

We note that the variational definition of the BS-conditional entropy Eq. (2.4) is less pathological than the one via the partial trace Eq. (2.3). In the sense that the latter exhibits discontinuities on the set of positive semi-definite states (see Proposition 5.6), while the variational BS-conditional entropy does not and is uniformly continuous on the entire set of quantum states.

### 6.1.4 Rains information

Inspired by the Rains bound from entanglement theory [55] the *generalized Rains bound* of a quantum state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ was defined in [56] by

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

where the minimisation is taken over the Rains set

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

In the above definition $\mathbb{D}$ denotes any divergence.

This definition can be easily extended to channels in the following way. For a quantum channel $T_{A' \to B} : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we define

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

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

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

as well as the BS-Rains information by

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

In the rest of this section, we will drop the subindex from the channels whenever it is clear in which systems they act.

In [57], 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 an SDP, as shown in [57]. 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 4.7 and Corollary 5.12, respectively, we can provide continuity results for both the Rains information and the BS-Rains information, respectively, following the lines of Theorem 6.5. 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 shown below in Eq. (6.6). 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

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

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

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

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 $1 \to 1$ norm of the difference between channels. Let us recall that for $T : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ a quantum channel, its $1 \to 1$ norm is given by

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

For $T_{A'\to B}$, the $1 \to 1$ norm coincides with the diamond norm. Now, as a consequence of Lemma 6.4 and Theorem 6.5 from the previous section, we can derive the following continuity bound for the Rains information.

**Theorem 6.12.** *For $\varepsilon \in [0,1]$ and $T^1_{A'\to B}, T^2_{A'\to B} : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ two quantum channels with $\frac{1}{2}\|T^1_{A'\to B} - T^2_{A'\to B}\|_{1\to 1} \leq \varepsilon$, we have:*

$$|R(T^1_{A'\to B}) - R(T^2_{A'\to B})| \leq \varepsilon \log \min\{d_A, d_B\} + (1+\varepsilon) h\Big(\frac{\varepsilon}{1+\varepsilon}\Big) \,. \tag{6.5}$$

*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

$$\begin{aligned}\max_{\substack{\rho_{AB},\,\sigma_{AB}\in\mathcal{S}(\mathcal{H}_{AB})\\ \frac{1}{2}\|\rho_{AB}-\sigma_{AB}\|_1=1}} &|D_{\mathrm{PPT}'(A:B)}(\rho_{AB})-D_{\mathrm{PPT}'(A:B)}(\sigma_{AB})| \\ &\leq \max_{\rho_{AB}\in\mathcal{S}(\mathcal{H}_{AB})} D_{\mathrm{PPT}'(A:B)}(\rho_{AB}) \\ &\leq \max_{\rho_{AB}\in\mathcal{S}(\mathcal{H}_{AB})} D_{\mathrm{SEP}_{AB}}(\rho_{AB}) \\ &\leq \log\min\{d_A,\, d_B\}\,,\end{aligned} \tag{6.6}$$

where in the last inequality we have used Lemma 6.4. Following the lines of Theorem 6.5, 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:

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

Note that since $\mathrm{PPT}'(A:B)$ does not only contain states, but also subnormalized states, Lemma 6.2 and Lemma 6.3 are not directly applicable. One can however verify that the corresponding statements for $\mathrm{PPT}'(A:B)$ still hold using the same arguments. For simplicity, let us denote

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

To estimate an upper bound on the difference that appears in Eq. (6.5), first note that, given $T^1,T^2:\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_{A'})\to\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_B)$ two quantum channels with $\frac{1}{2}\big\|T^1-T^2\big\|_{1\to1}\leq\varepsilon$, and $\rho_A\in\mathcal{S}(\mathcal{H}_A)$ with $\phi_{AA'}$ a purification of it, we have

$$\frac{1}{2}\big\|T^1(\phi_{AA'})-T^2(\phi_{AA'})\big\|_1\leq\frac{1}{2}\big\|T^1-T^2\big\|_{1\to1}\leq\varepsilon\,.$$

Consider now $\rho^1,\rho^2\in\mathcal{S}(\mathcal{H}_A)$ with respective purifications $\phi^1_{AA'},\phi^2_{AA'}$, 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 $i\neq j$,

$$\begin{aligned}|R(T^j)-D_{\mathrm{PPT}'(A:B)}(T^i(\phi^j_{AA'}))| &= |D_{\mathrm{PPT}'(A:B)}(T^j(\phi^j_{AA'}))-D_{\mathrm{PPT}'(A:B)}(T^i(\phi^j_{AA'}))| \\ &\leq b(\varepsilon)\,,\end{aligned}$$

and thus,

$$R(T^i)\geq D_{\mathrm{PPT}'(A:B)}(T^i(\phi^j_{AA'}))\geq R(T^j)-b(\varepsilon)\,.$$

Therefore, we can conclude

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

and consequently

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

□

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:

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

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

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

for $\phi_{AA'}$ a purification of $\rho_A$. We can finally use Lemma 6.9 and Theorem 6.10 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 6.12 and the continuity bound for the BS-entropy of entanglement from Theorem 6.10. Therefore, we omit it, to avoid unnecessary repetitions.

**Theorem 6.13.** *For* $\varepsilon \in [0,1]$ *and* $T^1_{A' \to B}, T^2_{A' \to B} : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *two quantum channels with* $\frac{1}{2}\|T^1 - T^2\|_{1\to 1} \leq \varepsilon$*, 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* $g_d(t) := \frac{d}{t^{1/d}} h(t) - \log\bigl(1 - t^{1/d}\bigr)$.

## 6.2 Outlook

In this thesis, we have introduced a new method to derive results of uniform continuity and explicit continuity bounds for divergences. Our method (cf. Theorem 3.6), is named ALAFF after the functions to which it applies (almost locally affine functions). Based on the ideas by Alicki, Fannes, and Winter which they used to prove continuity bounds for the conditional entropy, it yet works for a much wider class of divergences, namely all those for which we can prove almost concavity. More specifically, our method considers an entropic quantity which is (jointly) convex and almost (jointly) concave, and outputs continuity bounds for such a quantity and any derived entropic quantity.

In particular, in this thesis, we have applied our ALAFF method to the specific cases of the Umegaki and the Belavkin-Staszewski relative entropy. For both of them, we have proven results of almost concavity (for the Umegaki case, our result is shown to be tight), which, combined with their convexity, yielded a plethora of results of continuity bounds for both the Umegaki and BS-entropy, 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 thesis: Is our method applicable to any other family of divergences? We expect this to be the case, since, as shown in Chapter 3, 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 and is expected to exist, for families such as the $\alpha$-sandwiched Rényi divergences, given for two quantum states $\rho$ and $\sigma$ by

$$\widetilde{D}_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\operatorname{tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] & \text{if } \ker\rho\subseteq\ker\sigma \\ +\infty & \text{else} \end{cases},$$

or the $\alpha$-geometric Rényi divergences given by

$$\widehat{D}_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log\operatorname{tr}\left[\sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha\sigma^{1/2}\right] & \text{if } \ker\rho\subseteq\ker\sigma \\ +\infty & \text{else} \end{cases},$$

as they converge to the quantities studied in this thesis. In other words, the $\alpha \to 1$ limit of the $\alpha$-sandwiched Rényi divergences evaluates to the relative, while for the same limit the $\alpha$-geometric Rényi divergences become the BS-entropy.

Also interesting are the findings related to the BS-entropy. First of all, we have seen that the remainder we obtain in the almost concave result (cf. Theorem 5.3) should be improvable, for two reasons. First, it does not coincide with the concave remainder we obtained for the relative entropy (cf. Theorem 4.1) in case the states commute. Second, numerics suggest an almost concave remainder term of the BS-conditional entropy that is independent of the minimal eigenvalue of the involved states (cf. Fig. 5.2), which is again not given for our bound. Although there is room for improvement, there is no doubt that the BS-entropy, and quantities derived from it, are "pathological" in some sense. We have shown that the BS-conditional entropy exhibits discontinuities in the presence of vanishing eigenvalues (cf. Proposition 5.6), as opposed to the conditional entropy, which behaves well in that setting. And further that the BS-mutual information is unbounded in general (cf. Proposition 5.5). Closely related to this discontinuity are the ambiguities when it comes to defining the BS-conditional entropy itself and the dissimilarity between the variational definition and the one via the partial trace. The latter is discontinuous on positive semi-definite states whilst the former is not and both do not agree in general. We know that the infimum in the variational definition is attained (cf. Section 6.1.3), however, not necessarily by the partial trace. Naturally one can ask the question if there exists a CPTP map that saturates the minimization. Directly connected are similar questions about the BS-mutual information and alternative definitions of that quantity we haven't explored in this thesis.

# Bibliography


[1] A. Bluhm, A. Capel, P. Gondolf, and A. Pérez-Hernández, “Continuity of quantum entropic quantities via almost convexity,” 2022. [Online]. Available: https://arxiv.org/abs/2208.00922

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

[3] E. A. Carlen, “Trace inequalities and quantum entropy: An introductory course,” in *Entropy and the Quantum*, ser. Contemporary Mathematics, R. Sims and D. Ueltschi, Eds. American Mathematical Society, 2009, vol. 529, pp. 73–140.

[4] A. Bluhm and Á. Capel, “A strengthened data processing inequality for the belavkin–staszewski relative entropy,” *Reviews in Mathematical Physics*, vol. 32, no. 02, p. 2050005, aug 2019. [Online]. Available: https://doi.org/10.1142%2Fs0129055x20500051

[5] R. Iten, “Relations between different quantum rényi divergences,” 2020. [Online]. Available: https://arxiv.org/abs/2012.08327

[6] D. Sutter, M. Berta, and M. Tomamichel, “Multivariate trace inequalities,” *Commun. Math. Phys.*, vol. 352, pp. 37–58, 2017.

[7] T. Ando and F. Hiai, “Hölder type inequalities for matrices,” *Math. Inequal. Appl.*, vol. 1, no. 1, pp. 1–30, 1998.

[8] A. Rényi *et al.*, “On measures of entropy and information,” in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 547-561. Berkeley, California, USA, 1961.

[9] C. E. Shannon, “A mathematical theory of communication,” *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001.

[10] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[11] M. Tomamichel, *Quantum Information Processing with Finite Resources*. Springer, 2016.

[12] H. Umegaki, "Conditional expectation in an operator algebra IV. Entropy and information," *Kodai Math. Sem. Rep.*, vol. 14, pp. 59–85, 1962.

[13] V. P. Belavkin and P. Staszewski, "$C^*$-algebraic generalization of relative entropy and entropy," *Ann. Inst. Henri Poincaré, section A*, vol. 37, no. 1, pp. 51–58, 1982.

[14] J. Von Neumann, *Mathematische grundlagen der quantenmechanik.* Springer-Verlag, 2013, vol. 38.

[15] 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)*, pp. 118–123, 2021.

[16] S. O. Scalet, A. M. Alhambra, G. Styliaris, and J. I. Cirac, "Computable Rényi mutual information: Area laws and correlations," *Quantum*, vol. 5, p. 541, 2021.

[17] Y. Zhai, B. Yang, and Z. Xi, "Belavkin–Staszewski relative entropy, conditional entropy, and mutual information," *Entropy*, vol. 24, p. 837, 2022.

[18] F. Hiai, M. Mosonyi, D. Petz, and C. Bény, "Quantum $f$-divergences and error correction," *Rev. Mat. Phys.*, vol. 23, no. 07, pp. 691–747, 2011. [Online]. Available: https://doi.org/10.1142%2Fs0129055x11004412

[19] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A: Math. Gen.*, vol. 37, no. 5, pp. L55–L57, 2004.

[20] M. E. Shirokov, "Advanced Alicki-Fannes-Winter method for energy-constrained quantum systems and its use," *Quantum Inf. Process.*, vol. 19, p. 164, 2020.

[21] M. E. Shirokov, "Continuity of characteristics of composite quantum systems: a review," *arXiv preprint: arXiv 2201.11477*, 2022.

[22] 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.*, vol. 111, no. 25, p. 250404, 2013.

[23] G. Lindblad, "Expectations and entropy inequalities for finite quantum systems," *Comm. Math. Phys.*, vol. 39, pp. 111–119, 1974.

[24] M. M. Wilde, *Quantum Information Theory.* Cambridge University Press, 2016.

[25] K. M. R. Audenaert, "Quantum skew divergence," *J. Math. Phys.*, vol. 55, p. 112202, 2014.

[26] 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.*, vol. 52, no. 1, p. 014001, 2019.

[27] K. M. R. Audenaert and N. Datta, "$\alpha$-$z$-Rényi relative entropies," *J. Math. Phys.*, vol. 56, p. 022202, 2015.

[28] A. E. Rastegin, "Upper continuity bounds on the relative $q$-entropy for $q > 1$," *J. Math. Phys.*, vol. 52, p. 062203, 2011. [Online]. Available: https://doi.org/10.1063%2F1.3600535

[29] A. Vershynina, "Upper continuity bound on the quantum quasi-relative entropy," *J. Math. Phys.*, vol. 60, p. 102201, 2019. [Online]. Available: https://doi.org/10.1063%2F1.5114619

[30] K. M. R. Audenaert and J. Eisert, "Continuity bounds on the quantum relative entropy — II," *J. Math. Phys.*, vol. 52, p. 112201, 2011. [Online]. Available: https://doi.org/10.1063%2F1.3657929

[31] K. M. R. Audenaert and J. Eisert, "Continuity bounds on the quantum relative entropy," *J. Math. Phys.*, vol. 46, no. 10, p. 102104, 2005. [Online]. Available: https://doi.org/10.1063%2F1.2044667

[32] O. Bratteli and D. W. Robinson, *Operator algebras and quantum-statistical mechanics II. Equilibrium states. Models in quantum statistical mechanics*, ser. Texts and Monographs in Physics. Springer, 1981.

[33] K. Matsumoto, "Reverse test and characterization of quantum relative entropy," *arXiv preprint arXiv:1010.1030*, 2010.

[34] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, ser. Texts and Monographs in Physics. Springer, 1993.

[35] V. Katariya and M. M. Wilde, "Geometric distinguishability measures limit quantum channel estimation and discrimination," *Quantum Inf. Process.*, vol. 20, p. 78, 2021. [Online]. Available: https://doi.org/10.1007%2Fs11128-021-02992-7

[36] H. Fawzi and O. Fawzi, "Defining quantum divergences via convex optimization," *Quantum*, vol. 5, p. 387, 2021.

[37] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum mechanical entropy," *J. Math. Phys.*, vol. 14, pp. 1938–1941, 1973.

[38] D. Petz, "Monotonicity of quantum relative entropy revisited," *Rev. Math. Phys.*, vol. 15, no. 1, pp. 79–91, 2003.

[39] P. Hayden, R. Jozsa, D. Petz, and A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," *Commun. Math. Phys.*, vol. 246, pp. 359–374, 2004.

[40] D. Sutter and R. Renner, "Necessary criterion for approximate recoverability," *Ann. Henri Poincaré*, vol. 19, no. 10, pp. 3007–3029, 2018.

[41] O. Fawzi and R. Renner, "Quantum conditional mutual information and approximate Markov chains," *Commun. Math. Phys.*, vol. 340, pp. 575–611, 2015.

[42] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, "Universal recovery from a decrease of quantum relative entropy," *Ann. Henri Poincaré*, vol. 19, p. 2955, 2018.

[43] E. A. Carlen and A. Vershynina, "Recovery map stability for the data processing inequality," *J. Phys. A: Math. Theor.*, vol. 53, no. 3, p. 035204, 2020.

[44] D. Sutter, *Approximate Quantum Markov Chains*, ser. SpringerBriefs in Mathematical Physics. Springer, 2018.

[45] K. Kato and F. G. S. L. Brandão, "Quantum approximate Markov chains are thermal," *Commun. Math. Phys.*, vol. 370, pp. 117–149, 2019.

[46] E. P. Hanson, C. Rouzé, and D. Stilck França, "Eventually entanglement breaking Markovian dynamics: Structure and characteristic times," *Ann. Henri Poincaré*, vol. 21, p. 1517–1571, 2020.

[47] P. Svetlichnyy and T. A. B. Kennedy, "Decay of quantum conditional mutual information for purely generated finitely correlated states," *J. Math. Phys.*, vol. 63, p. 072201, 2022.

[48] F. Hiai and M. Mosonyi, "Different quantum $f$-divergencies and the reversibility of quantum operations," *Rev. Math. Phys.*, vol. 29, no. 7, p. 1750023, 2017.

[49] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Springer, 2006.

[50] V. Vedral and M. B. Plenio, "Entanglement measures and purification procedures," *Phys. Rev. A*, vol. 57, pp. 1619–1633, 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.57.1619

[51] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, "Quantifying entanglement," *Phys. Rev. Lett.*, vol. 78, pp. 2275–2279, 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.78.2275

[52] E. M. Rains, "Bound on distillable entanglement," *Phys. Rev. A*, vol. 60, pp. 179–184, 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.60.179

[53] M. J. Donald and M. Horodecki, "Continuity of relative entropy of entanglement," *Phys. Lett. A*, vol. 264, pp. 257–260, 1999.

[54] K. Matsumoto, "A new quantum version of $f$-divergence," in *Reality and Measurement in Algebraic Quantum Theory.* Springer, 2018, pp. 229–273.

[55] E. M. Rains, "A semidefinite program for distillable entanglement," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2921–2933, 2001.

[56] M. Tomamichel, M. M. Wilde, and A. Winter, "Strong converse rates for quantum communication," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 715–727, 2017.

[57] K. Fang and H. Fawzi, "Geometric Rényi divergence and its applications in quantum channel capacities," *Commun. Math. Phys.*, vol. 384, pp. 1615–1677, 2021.

Appendix A

# Supplements Chapter 4

## A.1 Supplements to the proof of Theorem 4.1

We will now show that the result of the inequality in Eq. (4.3) 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

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

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

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

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

$$\Pi_{\rho_1}\rho_1 = \rho_1\Pi_{\rho_1} = 0, \qquad \Pi_{\sigma_1}\sigma_1 = \sigma_1\Pi_{\sigma_1} = 0, \qquad \Pi_{\sigma_1}\rho_1 = \rho_1\Pi_{\sigma_1} = 0\,. \tag{A.1}$$

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 $\mathrm{tr}[\rho_{1,\varepsilon}(\log\sigma - \log\sigma_{1,\delta})]$ we can reduce to the subspace where they are all full rank. We then apply the Peierls-Bogolubov inequality [7] and Eq. (2.2) from Corollary 2.3 to obtain

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

Both of the traces on the LHS and RHS of Eq. (A.2) 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 clearly the limit $\varepsilon \to 0$ requires no more argument as both sides are linear in $\varepsilon$. Hence, we get

$$\operatorname{tr}[\rho_1(\log\sigma - \log\sigma_{1,\delta})] \leq \log \int_{-\infty}^{\infty} dt\, \beta_0(t)\ \operatorname{tr}\left[\rho_1 \sigma_{1,\delta}^{\frac{it-1}{2}} \sigma \sigma_{1,\delta}^{\frac{-it-1}{2}}\right]. \tag{A.3}$$

The limit $\delta \to 0$ 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. (A.1), 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. (A.3)

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

Moreover, for the RHS of Eq. (A.3) 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

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

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

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

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

## A.2 Proof of Proposition 4.2

We first of all note that for all $\rho_1, \rho_2 \in \mathcal{S}(\mathcal{H})$ its true that $\frac{1}{2}\|\rho_1 - \rho_2\|_1 \leq 1$, 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 $\frac{1}{2}\|\rho_1 - \rho_2\|$ in front of the $h$ here already.

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

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

   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 $j, k = 1, 2$, $j \neq k$ we can write

$$\sigma_j^{-1/2} \sigma_k \sigma_j^{-1/2} = \sigma_j^{-1/2} P_{\sigma_j} \sigma_k P_{\sigma_j} \sigma_j^{-1/2} \leq \sigma_j^{-1/2} \widetilde{m}^{-1} \sigma_j \sigma_j^{-1/2} = \widetilde{m}^{-1} P_{\sigma_j} \,, \tag{A.4}$$

   where $P_{\sigma_j}$ is the projection onto the support of $\sigma_j$. What we used in the inequality is that clearly $P_{\sigma_j} \sigma_k P_{\sigma_j} \leq P_{\sigma_j} \leq \widetilde{m}^{-1} \sigma_j$. If we use Eq. (A.4) we find that

$$\begin{aligned} c_j &\leq \int_{-\infty}^{\infty} dt \beta_0(t) \widetilde{m}^{-1} \operatorname{tr}\left[\rho_j \sigma_j^{\frac{it}{2}} P_{\sigma_j} \sigma_j^{\frac{-it}{2}}\right] \\ &= \int_{-\infty}^{\infty} dt \beta_0(t) \widetilde{m}^{-1} \operatorname{tr}\left[\rho_j \sigma_j^{\frac{it}{2}} \sigma_j^{\frac{-it}{2}}\right] = \widetilde{m}^{-1} \,. \end{aligned}$$

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

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

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

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 $p \mapsto \frac{1}{1-p}f_{m_1,m_2}(p) = \frac{p}{1-p}\log(p + m_1(1-p)) + \log(1-p+m_2p)$ we do the same, however, splitting the sum into two parts. First we find that $m_2 \geq 1$ hence $\log(1-p+m_2p) = \log(1+(m_2-1)p)$ is monotone in $p$, i.e. in particular non-decreasing. Second we look at $p \mapsto \frac{p}{1-p}\log(p+m_1(1-p))$. Forming the derivative at $p \in (0,1)$, we get

$$\begin{aligned}
&\frac{1}{(1-p)^2}\Big(\frac{p}{p+(1-p)m_1} + \log(p+m_1(1-p)) - p\Big) \\
&\qquad \geq \frac{1}{(1-p)^2}\Big(\frac{p}{p+(1-p)m_1} + \frac{p+(1-p)m_1-1}{p+m_1(1-p)} - p\Big) \\
&\qquad = \frac{1}{(1-p)^2}\Big(\frac{m_1(1-p)+2p-1}{p+(1-p)m_1} - p\Big) \\
&\qquad = \frac{1}{(1-p)^2}\Big(\frac{m_1(1-p)+2p-1-p(p+(1-p)m_1)}{p+(1-p)m_1}\Big) \\
&\qquad = \frac{1}{(1-p)^2}\Big(\frac{(m_1-1)(p-1)^2}{p+(1-p)m_1}\Big) \\
&\qquad \geq 0\,,
\end{aligned}$$

where 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.

## A.3 Proof of Lemma 4.12

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

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

which immediately gives the kernel inclusion as well as the condition to be lower bounded by $\widetilde{m}\rho$. Therefore, $\Delta^{\pm}(\sigma_1, \sigma_2, \tau) \in \mathcal{S}_0$ which makes $\mathcal{S}_0$ a $s$-perturbed $\Delta$-invariant set. We show the other direction by contrapositive. Let $s < \widetilde{m}$. Since $\widetilde{m} < 1$ we find an $\varepsilon > 0$ and two orthonormal $|0\rangle, |1\rangle \in \operatorname{supp}\rho$, as $\operatorname{rank}\rho \geq 2$, such that $\widetilde{m}\rho < \rho - \frac{\varepsilon}{2}|i\rangle\langle i|$ 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

$$\begin{aligned} |0\rangle\langle 0|^{\perp}\, \Delta^{+}(\sigma_1, \sigma_2, \tau)\, |0\rangle\langle 0|^{\perp} &= |0\rangle\langle 0|^{\perp}\, s\tau\, |0\rangle\langle 0|^{\perp} \geq \widetilde{m}\, |0\rangle\langle 0|^{\perp}\, \rho\, |0\rangle\langle 0|^{\perp} \\ |1\rangle\langle 1|^{\perp}\, \Delta^{-}(\sigma_1, \sigma_2, \tau)\, |1\rangle\langle 1|^{\perp} &= |1\rangle\langle 1|^{\perp}\, s\tau\, |1\rangle\langle 1|^{\perp} \geq \widetilde{m}\, |1\rangle\langle 1|^{\perp}\, \rho\, |1\rangle\langle 1|^{\perp} \end{aligned} \tag{A.5}$$

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 $\Delta^{\pm}(\sigma_1, \sigma_2, \tau) \geq \widetilde{m}\rho$ as $\Delta^{\pm}(\sigma_1, \sigma_2, \tau)$ are in $\mathcal{S}_0$ by assumption. To fulfil Eq. (A.5) we clearly need to choose $s > 0$ and since $s < \widetilde{m}$ we directly obtain the conditions

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

This gives us,

$$\begin{aligned} \operatorname{tr}[\tau] &\geq \operatorname{tr}\Big[|0\rangle\langle 0|^{\perp}\, \tau\, |0\rangle\langle 0|^{\perp} + |0\rangle\langle 0|\, \tau\, |0\rangle\langle 0|\Big] \\ &= \operatorname{tr}\Big[|0\rangle\langle 0|^{\perp}\, \tau\, |0\rangle\langle 0|^{\perp} + |0\rangle\langle 0|\, |1\rangle\langle 1|^{\perp}\, \tau\, |1\rangle\langle 1|^{\perp}\, |0\rangle\langle 0|\Big] \\ &> \operatorname{tr}\Big[|0\rangle\langle 0|^{\perp}\, \rho\, |0\rangle\langle 0|^{\perp} + |0\rangle\langle 0|\, |1\rangle\langle 1|^{\perp}\, \rho\, |1\rangle\langle 1|^{\perp}\, |0\rangle\langle 0|\Big] \\ &= \operatorname{tr}\Big[|0\rangle\langle 0|^{\perp}\, \rho\, |0\rangle\langle 0|^{\perp} + |0\rangle\langle 0|\, \rho\, |0\rangle\langle 0|\Big] \\ &= \operatorname{tr}[P_\rho \rho] = \operatorname{tr}[\rho] = 1, \end{aligned}$$

using 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.

Appendix B

# Supplements Chapter 5

## B.1 Proof of Proposition 5.4

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

$$\hat{c}_j = \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\Big[\rho_j (\rho_j^{1/2} \sigma^{-1} \rho_j^{1/2})^{\frac{it+1}{2}} \rho_j^{-1/2} \sigma \rho_j^{-1/2} (\rho_j^{1/2} \sigma^{-1} \rho_j^{1/2})^{\frac{-it+1}{2}}\Big]$$
$$= \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}[\rho_j] = \int_{-\infty}^{\infty} dt \beta_0(t) = 1$$

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

$$\hat{c}_j \leq \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}\Big[\rho_j (\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2})^{\frac{it+1}{2}} \rho_j^{-1/2} m^{-1} \sigma_j \rho_j^{-1/2} (\rho_j^{1/2} \sigma_j^{-1} \rho_j^{1/2})^{\frac{-it+1}{2}}\Big]$$
$$= m^{-1} \int_{-\infty}^{\infty} dt \beta_0(t) \operatorname{tr}[\rho_j] = m^{-1} .$$

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 [3, Theorem 4.13]. This means, however,

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

which gives us that

$$\left\|(d_A^{-1}\mathbb{1}_A \otimes \rho_{j,B})^{-1}\right\|_\infty \leq \left\|T_A(\rho^{-1})\right\|_\infty \leq \left\|\rho^{-1}\right\|_\infty \leq m^{-1}\,. \tag{B.1}$$

Hence, we have that $\left\|(d_A^{-1} \otimes \rho_{j,B})^{-1}\right\|_\infty^{-1}$ the minimal eigenvalue of $d_A^{-1}\mathbb{1}_A \otimes \rho_{j,B}$ 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. (B.1) 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 A.2.

## B.2 Proof of Proposition 5.5

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

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

where we used the non-negativity of $\widehat{D}(\cdot\|\cdot)$ 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 $-\widehat{D}(\cdot\|\cdot)$ is jointly concave and $\mathrm{tr}_A[\cdot]$ linear which means without loss of generality one can assume $\rho$ to be pure, i.e., a rank one projection. Then

$$\begin{aligned}\widehat{H}_{|\psi\rangle\langle\psi|}(A|B) &= -\widehat{D}(|\psi\rangle\langle\psi|\,\|\,\mathbb{1}_A \otimes P_B)\\ &= -\,\mathrm{tr}\Big[|\psi\rangle\langle\psi|\log|\psi\rangle\langle\psi|^{1/2}\,(\mathbb{1}_A \otimes P_B^{-1})\,|\psi\rangle\langle\psi|^{1/2}\Big]\\ &= -\log\mathrm{tr}\big[|\psi\rangle\langle\psi|\,(\mathbb{1}_A \otimes P_B^{-1})\big] = -\log\mathrm{tr}\big[P_B P_B^{-1}\big]\,,\end{aligned}$$

with $P_B = \mathrm{tr}_A[|\psi\rangle\langle\psi|]$. 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

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

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. $\widehat{D}(\rho||\sigma) \geq 0$ for every pair of states, is a direct consequence of the data processing inequality [18]. Applying $\mathrm{tr}_A[\,\cdot\,]$, 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 $\rho$. 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 not only see that the bound is tight but also that the scaling is of the order of the given bound. For that purpose let $d_A \in \mathbb{N}$ 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

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

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

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

and the Moore-Penrose inverse (in the case of $P_A$ 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

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

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

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

We directly obtain $\left\|\rho_A^{-1}\right\|_\infty = \left\|\rho_B^{-1}\right\|_\infty = \frac{d_A-1}{\varepsilon}$ and by construction $d_A < d_B$, hence the bound in Eq. (5.6) gives

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

We first note that for $\varepsilon = 1-\frac{1}{d_A}$ we get equality in Eq. (B.2). 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 entropy (see Eq. (2.5))

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

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

Appendix C

# Supplements Chapter 6

---

## C.1 Behavior of $g_d$

In this section, we study the function $g_d(p) := \frac{d}{p^{1/d}} h(p) - \log\big(1 - p^{1/d}\big)$ 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 6.1.3.

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

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

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

Continuity, therefore, follows from the definition of the function. □

**Lemma C.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

$$\begin{aligned}
\frac{\partial}{\partial p} g_d(p) &= \frac{1}{p^{1/d}} \left( \frac{p^{2/d-1}}{d(1-p^{1/d})} + (d - 1 + p^{-1}) \log(1-p) - (d-1)\log(p) \right) \\
&=: \frac{1}{p^{1/d}} g_d'(p)\,. \qquad \text{(C.1)}
\end{aligned}$$

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

$$\frac{\partial}{\partial d} g_d'(p) = \frac{p^{2/d-1}\left(d(p^{1/d}-1) + (p^{1/d}-2)\log(p)\right)}{d^3(p^{1/d}-1)^2} + \log(1-p) - \log(p)\,.$$

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 straight forward 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 C.1. ☐

**Lemma C.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 C.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 C.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

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

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

$$\frac{1}{1-p} - \frac{1}{p} \geq 0\,,$$
$$\frac{h(p)}{1-p} + \log(1-p) \geq 0\,,$$
$$p^{-1/d}\log(1-p) - \frac{\log\bigl(1-p^{1/d}\bigr)}{1-p} \geq 0\,.$$

Thus $p \mapsto \frac{g_d(p)}{1-p}$ is non-decreasing on $[1/2, 1)$, which concludes the assertion. □

Appendix D

# Notation and Abbreviations

Complementary to the introduction in Chapter 2 we summarise the abbreviations and notation used in this thesis in the following two tables.

Table D.1: List of abbreviations.

| | |
|---|---|
| BS | Belavkin-Staszewski |
| ALAFF | Almost locally affine |
| CPTP | Completely positive, trace preserving |
| DPI | Data-processing inequality |
| CMI | Conditional mututal information |
| SDP | Semidefinite program |

Table D.2: Notational conventions for mathematical expressions.

| Operators on Hilbert spaces | (see Section 2.1, Section 2.2) |
|---|---|
| $\mathcal{H}$ | Finite dimensional Hilbert space |
| $\langle\cdot\vert\cdot\rangle$ | Inner product |
| $d$ | Dimension of the Hilbert space |
| $\vert\psi\rangle$, $\vert\varphi\rangle$, $\vert i\rangle$ $i \in \mathbb{N}$ | Vectors on a Hilbert space |
| $\mathcal{B}(\mathcal{H})$ | Set of (bounded) linear operators on $\mathcal{H}$ |
| $A, B, N, M, P, \Pi$ | (Bounded) linear operators |
| $T$ | Completely positive, trace-preserving linear map [11] |
| $A^*$ | Adjoint of the operator $A$ |
| $\ker[A]$ | Kernel of the operator $A$ |
| $\mathrm{tr}[A]$ | Trace of the operator $A$ |
| $\vert A\vert$ | $= \sqrt{A^*A}$, absolute value of the operator $A$ |
| $\Vert A\Vert_p$ | Schatten $p$-norm of the operator $A$ |
| $A \geq B$ | Loewner order on the self-adjoint operators |
| $\mathcal{S}(\mathcal{H})$ | Set of quantum states on $\mathcal{H}$ |
| $\rho, \sigma, \tau, \gamma, \omega$ | Quantum states |
| $\mathrm{tr}_C[\rho]$ | Partial trace of the state $\rho$, tracing out the $C$ system |

| | |
|---|---|
| $\rho_C$ | Indicate that the state lives in $\mathcal{S}(\mathcal{H}_C)$, or that it is the image under partial traces, that just leave the $C$ system remaining |
| Entropic quantities | (see Section 2.3) |
| $S(\cdot)$ | Von Neumann entropy |
| $D(\cdot\|\cdot)$ | Umegaki relative entropy |
| $\widehat{D}(\cdot\|\cdot)$ | Belavkin-Staszewski relative entropy |
| $\mathbb{D}(\cdot\|\cdot)$ | Divergence |
| $H(A\|B)$ | Conditional entropy |
| $\widehat{H}(A\|B)$ | BS-conditional entropy |
| $\widehat{H}^{\mathrm{var}}(A\|B)$ | Variational BS-conditional entropy |
| $\mathbb{H}(A\|B)$ | Conditional divergence |
| $\mathbb{H}^{\mathrm{var}}(A\|B)$ | Variational conditional divergence |
| $I(A:B)$ | Mutual information |
| $\widehat{I}(A:B)$ | BS-mutual information |
| $\mathbb{I}(A:B)$ | Mutual information of an arbitrary divergence |
| $I(A:B\|C)$ | Conditional mutual information |
| $\widehat{I}(A:B\|C)$ | BS-conditional mutual information (one-sided) |
| $\mathbb{I}^{\mathrm{os}}(A:B\|C)$ | (One-sided) conditional mutual information of an arbitrary divergence |
| $\mathbb{I}^{\mathrm{ts}}(A:B\|C)$ | (Two-sided) conditional mutual information of an arbitrary divergence |

# Declaration

Ich Paul Elias Gondolf (Matrikel Nr. 5679397) versichere,

1. dass Ich diese Arbeit selbständig verfasst habe,
2. dass Ich keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe,
3. dass Ich wörtliche oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet habe,
4. dass die Arbeit bisher weder gesamt noch in Teilen einer anderen Prüfungsbehörde vorgelegt wurde,
5. dass das in Dateiform eingereichte Exemplar mit dem eingereichten gebundenen Exemplar übereinstimmt.

Tübingen, August 15, 2025,

*Paul Elias Gondolf*