INEQUALITIES FOR QUANTUM DIVERGENCIES AND THE AUDENAERT–DATTA CONJECTURE

ERIC A. CARLEN, RUPERT L. FRANK, AND ELLIOTT H. LIEB

ABSTRACT. Given two density matrices ρ and σ, there are a number of different expressions that reduce to the α-Rényi relative entropy of ρ with respect to σ in the classical case; i.e., when ρ and σ commute. Only those expressions for which the Data Processing Inequality (DPI) is valid are of potential interest as quantum divergences in quantum information theory. Audenaert and Datta have made a conjecture on the validity of the DPI for an interesting family of quantum generalizations of the α-Rényi relative entropies, the $\alpha - z$–Rényi relative entropies. They and others have contributed to the partial solution of this conjecture. We review the problem, its context, and the methods that have been used to obtain the results that are known at present, presenting a unified treatment of developments that have unfolded in a number of different papers.

1. The Audenaert–Datta conjecture and known results

The sender of a message over a classical noisy communication channel encodes it into a sequence of characters from an alphabet – possibly just $\{0, 1\}$ – that are physically represented by levels (amplitude or frequency) in a transmitted signal. Because of noise, the received levels will be random variables with continuous and overlapping distributions. The alphabet and encoding are known to the recipient, who is, however, faced with the problem of deciding whether each incoming random signal level represents, say, 0 or 1.

Shannon's theory tells the sender and the recipient just how much redundancy they must employ for the recipient to correctly extract the message from the noisy signal that is received with an arbitrarily small probability of error. For example, suppose the communication channel they use is such that sending 0 through the channel results in a random variable with a density ρ centered on 0, while sending 1 through the channel results in a random variable with density σ centered on 1. Suppose the noise is such that the different random variables produced at each step of the transmission are independent. Let X_n denote the n-th signal received. If the densities ρ and σ overlap (and they will if the noise is Gaussian) the receiver cannot tell for sure what was sent on a single observation. But if the sender repeats the transmission of the

U.S. National Science Foundation grants DMS-1363432 (R.L.F.), DMS-1764254 (E.A.C.), PHY-1265118 (E.H.L.) are acknowledged.

same signal m times, and the receiver knows it is the same signal being sent m times, and makes optimal use of the observed signals, then the probability of reading the wrong bit will go to zero exponentially fast as m increases.

Thus, the following problem is fundamental to classical communication theory: Given two probability distributions ρ and σ on $\mathbb{R}$, and a sequence of random variables X_n drawn from one of these two distributions, decide on the basis of the observations $\{X_1, \dots, X_N\}$ whether ρ or σ is the distribution from which the random sequence is being drawn. That is, one has to come up with a set $A_N \subset \mathbb{R}^N$ so that if $(X_1(\omega), \dots, X_N(\omega)) \in A_N$, then one accepts that ρ is the governing distribution, while otherwise, one accepts that σ is the governing distribution.

There are two kinds of errors that one can make: accepting σ when ρ is the governing distribution, and accepting ρ when σ is the governing distribution. Of course, if one takes A_N to be all of $\mathbb{R}^N$, then one never makes the first kind of error, but one will make the second kind of error whenever σ is the governing distribution. Therefore, fix some small $\varepsilon > 0$, and require of A_N that $\int_{A_N} \rho^{\otimes N} > 1 - \varepsilon$. Then among all such choices for A_N, choose one that (nearly) minimizes $\log \int_{A_N} \sigma^{\otimes N}$. That is, define

$$\beta_{\varepsilon,N}(\rho, \sigma) = \inf \left\{ \log \int_{A_N} \sigma^{\otimes N} \ : \ A_N \subset \mathbb{R}^N \text{ is such that } \int_{A_N} \rho^{\otimes N} \geq 1 - \varepsilon \right\}. \quad (1)$$

Then one has [25, 26]

$$\limsup_{N\to\infty} \frac{1}{N} \beta_{N,\varepsilon}((\rho, \sigma) \leq -D(\rho||\sigma) \quad \text{and} \quad \liminf_{N\to\infty} \frac{1}{N} \beta_{N,\varepsilon}((\rho, \sigma) \geq -\frac{1}{1-\varepsilon} D(\rho||\sigma)\,, \quad (2)$$

where $D(\rho||\sigma)$ is the *relative entropy* or *Kullbach–Liebler divergence* of ρ with respect to σ:

$$D(\rho||\sigma) = \int_{\mathbb{R}} \rho(x)(\log \rho(x) - \log \sigma(x)) \mathrm{d}x\,. \quad (3)$$

If one chooses N so that $e^{-ND(\rho||\sigma)} < \varepsilon$, then one can expect to have made both types of errors small, of order ε.

It is an easy consequence of Jensen's inequality that $D(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$. However, it is not in general true that $D(\rho||\sigma) = D(\sigma||\rho)$, and so the relative entropy is not a metric on the space of probability distributions. The asymmetry directly reflects the asymmetry in the question that is answered in terms of the relative entropy, namely: *If one chooses the acceptance rule A_N so that the probability of correctly accepting ρ is at least $1 - \varepsilon$, how small, as a function of N, can one make the probability of incorrectly accepting ρ when σ is sent?* In this simple setting of independent random variables, the fact that the relative entropy arises as the answer to this question is a consequence of Cramér's theorem on large deviations, and it gives the relative entropy an *operational meaning*.

A *(classical) divergence* is a function on pairs of probability densities taking values in $[0, \infty]$ that is somehow connected with how "distinguishable" the two densities are, and as above, such functions need not be symmetric. A mathematical definition

of classical divergences was given by Rényi [38] who introduced the Rényi relative entropies as a family of such divergences. For $\alpha \in (0,1)$, the α-Rényi entropy is defined as

$$D_\alpha(\rho||\sigma) = \frac{1}{\alpha - 1} \log \left(\int_{\mathbb{R}} \rho^\alpha(x) \sigma^{1-\alpha}(x) \mathrm{d}x \right) . \tag{4}$$

Later, Csizlár [13] gave an operational meaning to the Rényi relative entropies, but going into this would be too large a digression. Suffice it to say that a great many bounds on error probabilities have been given in terms of Rényi entropies. Our focus here is on the quantum aspects of this problem, and especially, quantum analogs of a certain monotonicity property that classical divergences must have. The basic monotonicity property can be easily explained at an intuitive level for both the classical and quantum cases.

Any divergence is supposed to give, or at least bound, the "best asymptotic rate of distinguishability" between ρ and σ in some operational context. Let $P(x,y)$ be a non-negative kernel with $\int P(x,y)\mathrm{d}x = 1$ for all y. Define $P\rho(x) := \int_{\mathbb{R}} P(x,y)\rho(y)\mathrm{d}y$ and $P\sigma(x) := \int_{\mathbb{R}} P(x,y)\sigma(y)\mathrm{d}y$. Any classical divergence $D(\rho||\sigma)$ must satisfy

$$D(P\rho||P\sigma) \leq D(\rho||\sigma) \tag{5}$$

since otherwise, one could apply the operation P to ρ and σ, and use whatever procedure one is using to distinguish ρ and from σ, and get a better result. The inequality (5) which reflects the fact that applying a further random corruption of the signal can only make it harder to discern what is being sent, is known as the *(classical) Data Processing Inequality*. For the classical divergences discussed above, this is true on account of Jensen's inequality and the joint concavity of the integrands in ρ and σ. There are other important properties that divergences should have; e.g., certain additivity properties over products, but we focus instead on (5) and its quantum analog.

In the quantum setting, probability densities are replaced by density matrices; that is, by non-negative trace class matrices ρ with unit trace, and integrals are replaced by traces. One natural quantum analog of the relative entropy $D(\rho||\sigma)$, known as the *Umegaki relative entropy* [42], is defined to be

$$D(\rho||\sigma) = \mathrm{Tr}[\rho(\log\rho - \log\sigma)] , \tag{6}$$

which is closely related to the *von Neumann entropy* of a density matrix ρ:

$$S(\rho) = -\mathrm{Tr}[\rho \log \rho] . \tag{7}$$

The fact that $D(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$ is still true in this setting, but it is no longer a direct application of Jensen's inequality. Indeed, not every classical entropy inequality is valid in the quantum setting. For example, a marginal of a classical joint probability distribution never has an entropy exceeding that of the joint distribution itself. In the quantum setting, however, this is false.

Even when a classical entropy inequality has a valid quantum analog, its proof in the quantum setting may be much more difficult. Probably the first conjecture about a

classical entropy inequality that did not obviously hold true in quantum mechanics was made by Lanford and Robinson, namely, Strong Subadditivity of the von Neumann entropy (SSA). This was proved in [28] based on a convexity theorem for certain trace functionals in [27]. It is a remarkable fact that while some of the inequalities in classical information theory carry through in quantum information theory and some do not, it is often the more complicated ones that carry through, SSA for instance, while some of the simpler ones fail. SSA is essentially equivalent to the joint convexity of the Umegaki relative entropy, and the quantum analog of (5), the Data Processing Inequality. Wehrl's paper [43] provides a good review of the what was known on entropy and relative entropy inequalities, classical and quantum, through 1978, and it is an especially good source on differences between the classical and quantum settings.

The *quantum Data Processing Inequality for a divergence D* says that for completely positive trace preserving maps $\mathcal{E}$ and all density matrices ρ and σ, one has

$$D(\mathcal{E}\rho||\mathcal{E}\sigma) \leq D(\rho||\sigma) \ . \tag{8}$$

Completely positive trace preserving (CPTP) maps are also known as quantum operations, and are the general class of state transformations possible in an open quantum system [24]. Applying any such operation can only make the states harder to distinguish, and hence if D is to have an operational meaning, it must satisfy the Data Processing Inequality.

While the von Neumann entropy (7) is the natural analog of the classical entropy $S(\rho) = -\int \rho \log \rho$, the situation is more complicated when one considers relative entropy and other divergences, since these involve two density matrices ρ and σ that need not commute. Thus, there are many ways one might try to write down a quantum analog of $\rho \log(\rho/\sigma)$, for example.

Relative entropy arises in the answers to all sorts of questions in classical probability theory, and it is not evident that it is Umegaki's quantum generalization that answers the corresponding question in the quantum setting. Among the many other expressions that reduce to it when ρ and σ commute are

$$\mathrm{Tr}[\rho \log(\sigma^{-1/2}\rho\sigma^{-1/2})] \qquad \text{and} \qquad \mathrm{Tr}[\rho \log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \ . \tag{9}$$

The Data Processing Inequality may be used to winnow the large field of putative divergences leaving a reduced class for which one might hope to find operational meanings. For example, the trace function on the right in (9) does satisfy (8), as does the Umegaki relative entropy, but the trace function on the left does not [12]. It turns out that the Umegaki relative entropy is the "right choice" as far as the direct quantum analog of the classical decision problem that was discussed above is concerned, as shown by Hiai and Petz [21]: Consider two density matrices ρ and σ on a finite dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{P}(\mathcal{H}^{\otimes N})$ be the set of orthogonal projections onto subspaces of $\mathcal{H}^{\otimes N}$. Fix $\varepsilon > 0$, and define

$$\beta_{\varepsilon,N}(\rho,\sigma) = \inf\left\{ \log \mathrm{Tr}[\sigma^{\otimes N} A_N] \ : \ A_N \in \mathcal{P}(\mathcal{H}^{\otimes N}) \ , \ \mathrm{Tr}[\rho^{\otimes N} A_N] > 1-\varepsilon \right\}. \tag{10}$$

The result of Hiai and Petz [21] is that then (2) is valid also in this quantum setting where now $D(\rho||\sigma)$ denotes the Umegaki relative entropy. (They actually prove this in a somewhat more general setting.) Another of their results is that

$$\mathrm{Tr}[\rho(\log\rho - \log\sigma)] \leq \mathrm{Tr}[\rho\log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \ , \tag{11}$$

and the inequality is strict when ρ and σ do not commute. Hence while the quantity on the right in (11) may seem to be a natural extension of the notion of relative entropy to the quantum setting, and while it does satisfy (8), it is not the one that is relevant to the decision problem that we have been discussing – which is not to say that is is not the relevant quantity for some other problem.

Thus, when generalizing classical entropy inequalities to the quantum setting, there is the difficulty that non-commutativity prevents one from directly mimicking the classical proofs, but also the non-commutativity is the source of a multiplicity of distinct quantum analogs of classical entropies or divergences: Which analog is meaningful in which settings?

This situation naturally arises when one considers quantum analogs of the Rényi relative entropies. The obvious generalization of (4), namely,

$$D_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\left(\mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}]\right) \ , \tag{12}$$

turns out to have the same operational meaning that was given for (4) by Ciszlár; this was proved by Mosonyi and Hiai [34]. However, (12) is not the only quantum analog of (4) to have an operational meaning.

Another generalization of the relative Rényi entropy was introduced recently by Müller-Lennert, Dupuis, Szehr, Fehr, Tomamichel [36] and Wilde, Winter, Yang [44]. They introduces the *sandwiched Rényi entropies*

$$\widetilde{D}_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\left(\mathrm{Tr}[(\sigma^{(1-\alpha)/2\alpha}\rho\sigma^{(1-\alpha)/2\alpha})^\alpha]\right) \ , \tag{13}$$

For certain values of the the parameter, an operational meaning has been given in [35].

Audenaert and Datta realized that all of these different quantum Rényi entropies – and more – could be brought together in a two parameter family, the $\alpha - z$ Rényi entropies defined by

$$D_{\alpha,z}(\rho||\sigma) := \frac{1}{\alpha-1}\log\left(\mathrm{Tr}[(\sigma^{(1-\alpha)/2z}\rho^{\alpha/z}\sigma^{(1-\alpha)/2z})^z]\right) \ . \tag{14}$$

Evidently, $D_\alpha(\rho||\sigma) = D_{\alpha,1}(\rho||\sigma)$ and $\widetilde{D}_\alpha(\rho||\sigma) = D_{\alpha,\alpha}(\rho||\sigma)$. The α-z Rényi relative entropies have appeared earlier in a paper by Jaksic, Ogata, Pautrat and Pillet [22].

Audenaert and Datta raised the question:

> For which values of α and z does $D_{\alpha,z}$ satisfy the quantum Data Processing Inequality (8)?

By an argument of Lindblad and Uhlmann, based on Stinespring's theorem, this question is equivalent to whether the trace functionals appearing in the $D_{\alpha,z}$ have certain convexity or concavity properties. This will be explained below in detail, but in a nutshell, for our finite dimensional setting, the Stinespring Representation Theorem [40] allows one to express CPTP operations in terms of isometric injections, unitary conjugations and partial traces, in such a way that once one knows the monotonicity (8) under partial traces – one example of a CPTP map – one knows it in general. Since partial traces can be written as averages over unitary conjugations [41], it suffices to prove $D(\mathcal{E}\rho||\mathcal{E}\sigma) \leq D(\rho||\sigma)$ when $\mathcal{E}$ takes the form

$$\mathcal{E}(\rho) = \frac{1}{M}\sum_{j=1}^{M} U_j \rho U_j^*$$

where $U_1, \dots, U_M$ are unitary. Now consider, for example, the trace functional in (12), $\mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$. Evidently, for each j,

$$\mathrm{Tr}[((U_j\rho U_j^*)^\alpha (U_j \sigma U_j^*)^{1-\alpha}] = \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}] \ ,$$

and therefore if $(\rho, \sigma) \mapsto \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ is jointly concave, the standard quantum Rényi relative entropy (12) with $\alpha \in (0,1)$ satisfies $D(\mathcal{E}\rho||\mathcal{E}\sigma) \leq D(\rho||\sigma)$.

However, the joint concavity of $(\rho, \sigma) \mapsto \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ is a special case of the Lieb Concavity Theorem [27], which says that for $0 \leq p, q$ with $p + q \leq 1$, $(A, B) \mapsto \mathrm{Tr}[A^p B^q]$ is jointly concave. One may also consider Rényi relative entropies for $\alpha > 1$. In this case, the factor of $\frac{1}{1-\alpha}$ in (12) is negative, and so the DPI will follow if $(\rho, \sigma) \mapsto \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ is *convex*, and for $\alpha \in (1, 2)$ this is true by the Ando Convexity Theorem [1], which says that $(A, B) \mapsto \mathrm{Tr}[A^p B^q]$ is jointly convex for $-1 \leq q \leq 0$ and $0 \leq q + p \leq 1$. (Since we shall be concerned with the relation between convexity and concavity inequalities in what follows, it is worth remarking, that the proof of Ando's Convexity Theorem relies on the concavity result of [27].) Since the Umegaki relative entropy is the $\alpha \to 1$ limit of the Rényi relative entropy, these convexity results also yield the Data Processing Inequality for it as well, as first proved by Lindblad [31, 32].

To prove the data processing inequality for the $\alpha - z$ Rényi entropies one requires concavity and convexity of more complicated trace functionals. (In fact, as explained below, concavity/convexity is necessary as well as sufficient for the data processing inequality.) The convexity theorem that implies the data processing inequality for certain of the sandwiched Rényi entropies was proved in [16] (see also [6]). In their paper [5] Audenaert and Datta deduced the data processing inequality in certain cases from concavity theorems of Hiai and conjectured a precise parameter regime for the validity of the data processing inequality.

Let $\mathcal{M}_N$ denote the set of complex $N \times N$ matrices and $\mathcal{P}_N$ denote the subset of positive definite matrices. Their conjecture refers to the trace function at the heart of the definition of the $\alpha - z$ Rényi entropies, and may be stated as follows:

Conjecture 1. *If $1 \leq p \leq 2$ and $-1 \leq q < 0$, then for any $K \in \mathcal{M}_N$*

$$\mathcal{P}_N \times \mathcal{P}_N \ni (A, B) \mapsto \mathrm{Tr}\left(B^{q/2} K^* A^p K B^{q/2}\right)^{1/(p+q)}$$

is jointly convex.

Currently, this conjecture is known for $q = p - 1$ [1], for $p = 1$ [16] and for $p = 2$ [9]. It is open in the remaining cases.

Conjecture 1 is stated in a different, but equivalent form in [5], namely, as convexity of $\mathcal{P}_N \ni A \mapsto \mathrm{Tr}\left(A^{q/2} K^* A^p K A^{q/2}\right)^{1/(p+q)}$. The equivalence of this version with the one stated above will be explained at the end of this introduction.

The concavity of $(A, B) \mapsto \mathrm{Tr}[A^p B^q]$ with $0 \leq p, q$ and $p + q \leq 1$ lies at the basis of all of the examples of quantum data processing inequalities that we have discussed. A number of different proofs of this fundamental result were given by other authors, and the two that are most relevant for our purposes appeared within a few years of [27]. One of these was due to Araki [2, 3]. His proof introduced a method making use of the *relative modular operator* and other tools coming from modern theory of von Neumann algebras, and has the merit of providing generalizations of the inequality to this setting. The paper [2] is especially clear on the matricial case and is accessible without prior knowledge of the the modern theory of von Neumann algebras. An even more pedestrian account of this approach, which is now well known in the quantum information theory community, can be found in [37].

Another methodology was introduced by Epstein [15], who employs the theory of Herglotz functions, also known as Pick or Nevanlinna functions. These are functions φ that are analytic in the open upper half plane, $\mathbb{C}_+$, and that have a positive imaginary part. Such functions have a canonical integral representation, as recalled below, and from this a number of their properties follow. These functions have played a central role in the theory of operator inequalities since the 1934 theorem of Löwner [33] stating that a function $f : \mathbb{R} \to \mathbb{R}$ is such that for all self adjoint matrices A, B of the same size, $f(A) \geq f(B)$ whenever $A \geq B$ (in the usual ordering of self-adjoint matrices) if and only if f has an analytic continuation to a Herglotz function. The deep part of Löwner's Theorem is the "only if" part: It is an immediate consequence of the integral representation for Herglotz functions that if f is Herglotz, not only is $A \mapsto f(A)$ operator monotone, it is operator concave. A full account of this theory, with three distinct proofs of Löwner's Theorem, can be found in the book of Donoghue [14]. A forthcoming book of Simon [39] will present 11 distinct proofs, three of them new.

One of the merits of Epstein's method, explained in Section 3, is that he was able to use it to prove a conjecture that had been made in [27], namely that for fixed self adjoint B, the map $A \mapsto \mathrm{Tr}[(BA^pB)^{1/p}]$ is concave for $p \in (0, 1)$. At the end of the first paragraph of his paper, Epstein wrote: "The applicability of the method obviously extends beyond the examples treated here." Hiai [17, 18] has carried out a thorough development of the method, bringing in a number of significant new ideas. However,

the method still has limitations. It was conjectured in [10] and proved in [11] that $A \mapsto \mathrm{Tr}[(BA^pB)^{1/p}]$ is convex for $1 < p < 2$. Since $z \mapsto z^p$ is a Herglotz function only for $0 \le p \le 1$, Epstein's argument cannot be adapted to this case. The method of proof, which turns out to be useful also for some cases of the Audenaert-Datta Conjecture, relied on variational arguments, specifically Legendre transforms, and, most crucially, the fact that if $f(x,y)$ is jointly convex on $\mathbb{R}^m \times \mathbb{R}^m$, then $g(y) := \inf_{x\in\mathbb{R}^m}\{f(x,y)\}$ is convex in y. This variational method is explained in Section 4.

One of the intriguing aspects of the story of progress on the Audenaert-Datta Conjectures is the interplay between the "analyticity method" and the "variational method". It appears that the analyticity method is especially useful for proving concavity and the variational method is more useful for proving convexity, but this is not meant to be an absolute distinction. However, it is our belief that understanding these two methodologies is a worthwhile endeavor also for work outside the Audenaert–Datta program. There are quite a number of papers [17, 11, 18, 16, 5, 9, 19] over which trace inequalities related to the Audenaert and Datta conjecture are spread out. These papers, both before and after the Audenaert–Datta paper [5], are often not self-contained, and one of our goals is to try to tell the story seamlessly and thereby make the results from the literature more easily available also to newcomers to the field. There are no new results in this paper, but we hope to present a newly coherent account of recent advances.

In the following we will study a more general problem than the one occurring in Conjecture 1, with three parameters instead of two. For $A, B \in \mathcal{P}_N$, $K \in \mathcal{M}_N$ and parameters $p, q, s \in \mathbb{R}$ we define

$$\Psi_{p,q,s}(A,B) := \mathrm{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^s ,$$

and the problem is to determine the values of p, q, s such that $\Psi_{p,q,s}$ is jointly convex or concave. Because of symmetries, it suffices to consider p, q, s such that

$$p \ge q \qquad \text{and} \qquad s > 0\,. \tag{15}$$

To see this, note that by an approximation argument, we can assume that K is invertible, and then $\left(B^{q/2}K^*A^pKB^{q/2}\right)^s = \left(B^{-q/2}K^{-1}A^{-p}K^{-*}B^{-q/2}\right)^{-s}$, so that with K^{-*} replacing K on the right,

$$\Psi_{p,q,s}(A,B) = \Psi_{-p,-q,-s}(A,B)\ .$$

Next since $B^{q/2}K^*A^pKB^{q/2}$ and $A^{p/2}KB^qK^*A^{p/2}$ have the same non-zero eigenvalues with the same mutiplicities, with K^* replacing K on the right,

$$\Psi_{p,q,s}(A,B) = \Psi_{q,p,s}(B,A). \tag{16}$$

The following theorem, which is the main subject of this paper, summarizes our current knowledge about concavity and convexity properties of $\Psi_{p,q,s}$.

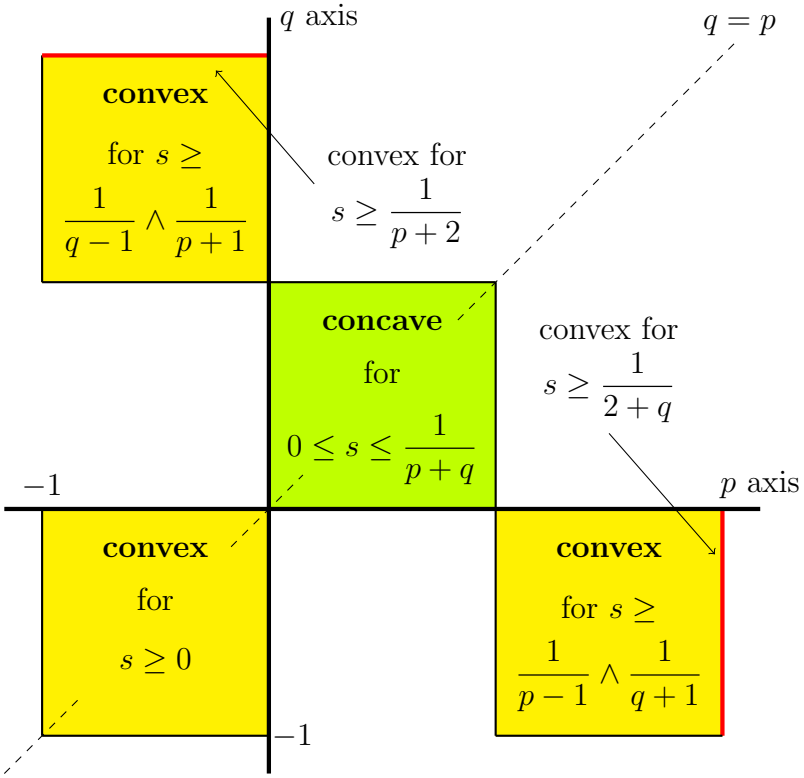
Theorem 2. *Let $K \in \mathcal{M}_N$ be arbitrary.*

(1) *If $0 \le q \le p \le 1$ and $0 < s \le 1/(p+q)$, then $\Psi_{p,q,s}$ is jointly concave.*

(2) *If $-1 \le q \le p \le 0$ and $s > 0$, then $\Psi_{p,q,s}$ is jointly convex.*

(3) *If $1 \le p < 2$, $-1 \le q \le 0$ and $s \ge \min\{1/(p-1), 1/(q+1)\}$, then $\Psi_{p,q,s}$ is jointly convex. If $p = 2$, $-1 \le q \le 0$ and $s \ge 1/(2+q)$, then $\Psi_{p,q,s}$ is jointly convex.*

In part (3), for $p = 1$ the condition $s \ge \min\{1/(p-1), 1/(q+1)\}$ is to be understood as $s \ge 1/(q+1)$. The information contained in the thoerem, and extended by symmetry in p and q, is summarized in the following figure:



Convexity and concavity for $\Psi_{p,q,s}$

History. We discuss the three cases in the theorem separately, plus the case $s = 1$ which historically came first and played an important role in the development of the field of matrix analysis.

(0) Case $s = 1$: This is due to Lieb [27] for $0 \le q \le p \le 1$ with $p + q \le 1$, as well as for $-1 \le q \le p \le 0$, and due to Ando [1] for $1 \le p \le 2$, $-1 \le q \le 0$ with $p + q \ge 1$.

(1) Case $0 \leq q \leq p \leq 1$: Partial results by Hiai for $1 \leq s \leq 1/(p+q)$ [17, Theorem 2.3] and for $1/2 \leq s \leq 1/(p+q)$ [18, Theorem 2.1 (1)] and by the authors for $0 < s \leq 1/(1+q)$ [9, Theorem 4.4] (see also [16, Proposition 3] for $p = 1$, $s = 1/(1+q)$). Later Hiai found a proof [19, Theorem 2.1] that covers the full range. We reproduce this proof in Section 3.

(2) Case $-1 \leq q \leq p \leq 0$: This is due to Hiai. After a partial result for $1/2 \leq s \leq -1/(p+q)$ [18, Theorem 2.1 (2)], the full result appears in [19, Theorem 2.1]. We reproduce this proof in Section 4. The key to Hiai's proof is to consider the equivalent result for *negative* s: Of the two equivalent results that are related by changing the signs of p, q and s, *only one is amenable to treatment by the variational method*.

(3) Case $1 \leq p \leq 2$, $-1 \leq q \leq 0$: Result for $p = 1$, $s = 1/(1+q)$ due to the last two authors [16, Proposition 3]. Remaining results due to the authors [9]. We reproduce the proofs in Section 4.

We now complement Theorem 2 with necessary conditions.

Proposition 3. *Let $s > 0$ and $p \geq q$ with $(p,q) \neq (0,0)$.*

(1) *If $\mathcal{P}_2 \times \mathcal{P}_2 \ni (A,B) \mapsto \Psi_{p,q,s}(A,B)$ is jointly concave for $K = 1$, then $0 \leq q \leq p \leq 1$ and $0 < s \leq 1/(p+q)$.*

(2) *If $\mathcal{P}_4 \times \mathcal{P}_4 \ni (A,B) \mapsto \Psi_{p,q,s}(A,B)$ is jointly convex for $K = 1$, then either $-1 \leq q \leq p \leq 0$ and $s > 0$ or $1 \leq p \leq 2$, $-1 \leq q \leq 0$, $(p,q) \neq (1,-1)$ and $s \geq 1/(p+q)$.*

This proposition is due to Hiai [18, Propositions 5.1(2) and 5.4(2)]. It is natural to conjecture that these necessary conditions are also sufficient.

Conjecture 4. *If $1 \leq p \leq 2$, $-1 \leq q < 0$ and $s \geq 1/(p+q)$, then for any $K \in \mathcal{M}_N$*

$$\mathcal{P}_N \times \mathcal{P}_N \ni (A,B) \mapsto \mathrm{Tr}\left(B^{q/2} K^* A^p K B^{q/2}\right)^s$$

is jointly convex.

Note that for $s = 1/(p+q)$ this is Conjecture 1 of Audenaert and Datta. The remaining case of the conjecture is

$$1 < p < 2\,,\ -1 \leq q < 0\,,\ 1/(p+q) \leq s < \min\{1/(p-1), 1/(q+1)\}\,.$$

(In fact, the case $s = 1$ can be excluded as a conjecture, due to a theorem of Ando.)

We now turn to a different, but related problem. For $A \in \mathcal{P}_N$, $K \in \mathcal{M}_N$ and parameters $p, s \in \mathbb{R}$ we define

$$\Upsilon_{p,s}(A) := \mathrm{Tr}(K^* A^p K)^s\,,$$

that is, $\Upsilon_{p,s}(A) = \Psi_{p,q,s}(A,1)$ for any q. As before, we can and will restrict ourselves to the case $s > 0$.

We will be interested in convexity and concavity properties of $\Upsilon_{p,s}$. While those are a consequence of similar properties of $\Psi_{p,q,s}$, we will conversely prove them first and use them in our discussion of $\Psi_{p,q,s}$.

Proposition 5. *Let K be arbitrary.*

(1) *If $0 \leq p \leq 1$ and $0 < s \leq 1/p$, then $\Upsilon_{p,s}$ is concave.*

(2) *If $-1 \leq p \leq 0$ and $s > 0$, then $\Upsilon_{p,s}$ is convex.*

(3) *If $1 \leq p \leq 2$ and $s \geq 1/p$, then $\Upsilon_{p,s}$ is convex.*

History.

(1) Case $0 \leq p \leq 1$: Initial result for $s = 1/p$ due to Epstein [15]. The first and third authors proved the result under the assumption $s \geq 1$ [11, Theorem 1.1]. The full result is due to Hiai [18, Theorem 4.1 (1)].

(2) Case $-1 \leq p < 0$: Result due to Hiai [18, Theorem 4.1 (2)].

(3) Case $1 \leq p \leq 2$: Result due to the first and third author [11, Theorem 1.1].

We next show that the conditions in Proposition 5 are also necessary.

Proposition 6. *Let $s > 0$ and $p \neq 0$.*

(1) *If $\mathcal{P}_2 \ni A \mapsto \Upsilon_{p,s}(A)$ is concave for any invertible K, then $0 < p \leq 1$.*

(2) *If $\mathcal{P}_4 \ni A \mapsto \Upsilon_{p,s}(A)$ is convex for any invertible K, then either $-1 \leq p < 0$ and $s > 0$ or $1 \leq p \leq 2$ and $s \geq 1/p$.*

This proposition is due to Hiai [18, Propositions 5.1(1) and 5.4(1)]. Earlier, the first and third authors [11] had used a similar argument to show that $\Upsilon_{p,s}$ is not convex or concave for $p > 2$. Related arguments also appear in [7].

We conclude this introduction by explaining why Conjecture 1 is equivalent to the form in [5], which corresponds to taking $A = B$ in our form. Given $A, B \in \mathcal{P}_N$ and $K \in \mathcal{M}_N$ let

$$C_{A,B} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \in \mathcal{P}_{2N} \qquad \text{and} \qquad L_K = \begin{pmatrix} 0 & K \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{2N}\,.$$

Then

$$\operatorname{Tr}\left(C_{A,B}^{q/2} L_K^* C_{A,B}^p L_K C_{A,B}^{q/2}\right)^s = \operatorname{Tr}\left(B^{q/2} K^* A^p K B^{q/2}\right)^s\,.$$

Thus, since $(A, B) \mapsto C_{A,B}$ is linear, convexity of $C \mapsto \operatorname{Tr}(C^{q/2} L^* C L^p C^{q/2})^s$ on $\mathcal{P}_{2N}$, implies joint convexity of $(A, B) \mapsto \operatorname{Tr}\left(B^{q/2} K^* A^p K B^{q/2}\right)^s$ on $\mathcal{P}_N \times \mathcal{P}_N$.

2. Application: Monotonicity of the $\alpha - z$ relative entropies

In this section we present an application of Theorem 2 to a problem motivated by quantum information theory. For $\rho, \sigma \in \mathcal{P}_N$ and $\alpha, z > 0$ with $\alpha \neq 1$, we consider the so-called $\alpha - z$-relative Rényi entropies

$$D_{\alpha,z}(\rho\|\sigma) = \frac{1}{\alpha - 1} \ln \frac{\operatorname{Tr}\left(\sigma^{(1-\alpha)/(2z)} \rho^{\alpha/z} \sigma^{(1-\alpha)/(2z)}\right)^z}{\operatorname{Tr} \rho} \,.$$

(One also obtains interesting quantities by taking the limit $\alpha \to 1$, possibly simultaneously with $z \to 0$ [5], but for the sake of brevity we exclude this case.) These functionals appeared in [22, Sec. 3.3] and were further studied in [5], where the question was raised whether the $\alpha - z$-relative Rényi entropies are monotone under completely positive, trace preserving maps (CPTP), that is, whether for any such map $\mathcal{E}$ and any $\rho, \sigma \in \mathcal{P}_N$ one has

$$D_{\alpha,z}(\rho\|\sigma) \geq D_{\alpha,z}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \,.$$

Proposition 7. *Let $\alpha, z > 0$ with $\alpha \neq 1$ and define*

$$p = \alpha/z \qquad \text{and} \qquad q = (1-\alpha)/z \,.$$

Then $D_{\alpha,z}$ is monotone under CPTP maps on $\mathcal{P}_N$ for all N if and only if $\Psi_{p,q,1/(p+q)}$ with $K = 1$ is jointly convex (if $\alpha > 1$) or jointly concave (if $\alpha < 1$) on $\mathcal{P}_N \times \mathcal{P}_N$ for all N.

Proof. For simplicity of notation, we write $\Psi := \Psi_{p,q,1/(p+q)}$ with $K = 1$ in the following. Clearly, $D_{\alpha,z}$ is monotone under CPTP maps if and only if Ψ is monotone decreasing (if $\alpha > 1$) or increasing (if $\alpha < 1$) under CPTP maps. Therefore the proposition follows from what we prove in the following two steps.

Step 1. We show that for $\alpha > 1$, if Ψ is monotone decreasing under CPTP maps on $\mathcal{P}_{2N}$, then Ψ is jointly convex on $\mathcal{P}_N \times \mathcal{P}_N$. A similar assertion holds for $\alpha < 1$ if decreasing and convex are replaced by increasing and concave.

We give the proof for $\alpha > 1$. For $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathcal{P}_N$ and $\theta \in [0,1]$ we consider the operators

$$\rho = (1-\theta)\rho_0 \otimes |\uparrow\rangle\langle\uparrow| + \theta\rho_1 \otimes |\downarrow\rangle\langle\downarrow| \qquad \text{and} \qquad \sigma = (1-\theta)\sigma_0 \otimes |\uparrow\rangle\langle\uparrow| + \theta\sigma_1 \otimes |\downarrow\rangle\langle\downarrow|$$

on $\mathbb{C}^N \otimes \mathbb{C}^2$, where $|\uparrow\rangle, |\downarrow\rangle$ denote a basis of $\mathbb{C}^2$. Applying the assumed monotonicity with the channel that takes the partial trace over $\mathbb{C}^2$, we conclude that

$$\Psi((1-\theta)\rho_0 + \theta\rho_1, (1-\theta)\sigma_0 + \theta\sigma_1) \leq (1-\theta)\Psi(\rho_0, \sigma_0) + \theta\Psi(\rho_1, \sigma_1) \,,$$

which means that Ψ is jointly convex.

Step 2. We show that for $\alpha > 1$, if Ψ is jointly convex on $\mathcal{P}_{N^2} \times \mathcal{P}_{N^2}$, then Ψ is monotone decreasing under CPTP maps on $\mathcal{P}_N$. A similar assertion holds for $\alpha < 1$ if convex and decreasing are replaced by concave and increasing.

Again, we assume $\alpha > 1$. Following a method of Lindblad and Uhlmann, we use Stinespring's theorem [40] to obtain an integer $N' \leq N^2$, a non-negative matrix

$\tau \in \mathcal{M}_{N'}$ with $\operatorname{Tr}\tau = 1$ (which can be chosen to be rank one) and a unitary U on $\mathbb{C}^N \otimes \mathbb{C}^{N'}$ such that

$$\mathcal{E}(\gamma) = \operatorname{Tr}_2 U\,(\gamma \otimes \tau)\,U^*\,.$$

Thus, if du denotes normalized Haar measure on all unitaries on $\mathbb{C}^{N'}$, then

$$\mathcal{E}(\gamma) \otimes (N')^{-1}1_{\mathbb{C}^{N'}} = \int (1 \otimes u)U\,(\gamma \otimes \tau)\,U^*(1 \otimes u^*)\,du\,. \tag{17}$$

By the tensor property of Ψ,

$$\Psi(\mathcal{E}(\rho), \mathcal{E}(\sigma)) = \Psi(\mathcal{E}(\rho) \otimes (N')^{-1}1_{\mathbb{C}^{N'}}\,, \mathcal{E}(\sigma) \otimes (N')^{-1}1_{\mathbb{C}^{N'}})\,. \tag{18}$$

By (17) and the assumed convexity the expression in (18) is bounded from above by

$$\int \Psi((1 \otimes u)U\,(\rho \otimes \tau)\,U^*(1 \otimes u^*), (1 \otimes u)U\,(\sigma \otimes \tau)\,U^*(1 \otimes u^*))\,du\,.$$

By unitary invariance and the tensor property of Ψ, the integrand here equals

$$\Psi((1\otimes u)U\,(\rho \otimes \tau)\,U^*(1\otimes u^*), (1\otimes u)U\,(\sigma \otimes \tau)\,U^*(1\otimes u^*)) = \Psi(\rho\otimes\tau, \sigma\otimes\tau) = \Psi(\rho, \sigma)\,,$$

and therefore, recalling that Haar measure is normalized, we conclude that

$$\Psi(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \Psi(\rho, \sigma)\,,$$

which means that Ψ is monotone decreasing under CPTP maps. □

Combining this proposition with Theorem 2 and Proposition 3 we obtain the following monotonicity result.

Corollary 8. *$D_{\alpha,z}$ is monotone under completely positive, trace preserving maps if one of the following holds*

$$\begin{aligned} &0 < \alpha < 1 \text{ and } z \geq \max\{\alpha, 1-\alpha\}\,,\\ &1 < \alpha \leq 2 \text{ and } z \in \{\alpha/2, 1, \alpha\}\,,\\ &2 \leq \alpha \leq \infty \text{ and } z = \alpha\,. \end{aligned}$$

Conversely, if $D_{\alpha,z}$ is monotone under completely positive, trace preserving maps, then one of the following holds

$$\begin{aligned} &0 < \alpha < 1 \text{ and } z \geq \max\{\alpha, 1-\alpha\}\,,\\ &1 < \alpha \leq 2 \text{ and } \alpha/2 \leq z \leq \alpha\,,\\ &2 \leq \alpha < \infty \text{ and } \alpha - 1 \leq z \leq \alpha\,. \end{aligned}$$

If Conjecture 1 is true, then monotonicity holds also in the cases $1 < \alpha \leq 2$, $\alpha/2 < z < \alpha$, as well as $2 \leq \alpha < \infty$, $\alpha - 1 \leq z < \alpha$, that is, in the full range allowed by the second part of the corollary.

Remark 9. Hiai and Mosonyi have made further progress [20] on the conjecture under the additional assumption that the CPTP map in question is unital and preserves either ρ or σ.

3. The complex analysis method

In this section we show part (1) of Theorem 2, namely that

$$(A,B) \mapsto \operatorname{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^s$$

is jointly concave for $0 \leq q \leq p \leq 1$ and $0 < s \leq 1/(p+q)$. Note that this also proves part (1) of Theorem 5.

This was shown under the extra assumption $s \geq 1/2$ in [18] using the complex analysis method and under the extra assumption $s \leq 1/(1+q)$ in [9, Theorem 4.4] using the variational method. Here we follow Hiai's proof [19, Theorem 2.1], using the complex analysis method, which works for the full range of exponents.

Proposition 10. *Let $0 \leq p, q \leq 1$. Then*

$$(A,B) \mapsto \operatorname{Tr}\left(1 + \left(B^{q/2}K^*A^pKB^{q/2}\right)^{-1/(p+q)}\right)^{-1}$$

is jointly concave.

Before proving this proposition we use it to deduce the concavity part of Theorem 2.

Proof of Theorem 2. (1). Multiplying A by a power of t we deduce from Proposition 10 that

$$(A,B) \mapsto \operatorname{Tr}\left(1 + t\left(B^{q/2}K^*A^pKB^{q/2}\right)^{-1/(p+q)}\right)^{-1}$$

is jointly concave for any $t > 0$. Multiplying by t and letting $t \to \infty$ we deduce that

$$(A,B) \mapsto \operatorname{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^{1/(p+q)}$$

is jointly concave. This is the assertion for $s = 1/(p+q)$.

Now let $\sigma := s(p+q) < 1$. Then

$$x^\sigma = \frac{\sin(\pi\sigma)}{\pi} \int_0^\infty \frac{1}{1+t/x}\, t^{-1+\sigma}\, dt \qquad \text{for all } x \geq 0$$

and therefore

$$\operatorname{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^s = \frac{\sin(\pi\sigma)}{\pi} \int_0^\infty \operatorname{Tr}\left(1 + t\left(B^{q/2}K^*A^pKB^{q/2}\right)^{-1/(p+q)}\right)^{-1} t^{-1+\sigma}\, dt\,.$$

As observed at the beginning of the proof, the integrand is jointly concave and therefore so is $\operatorname{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^s$. This concludes the proof of part (1) of Proposition 2. □

We now turn to Hiai's proof of Proposition 10. It is based on Epstein's method, which relies on the following lemma from complex analysis.

Lemma 11. *Let φ be analytic in $\mathbb{C}_+$ with $\operatorname{Im}\varphi \geq 0$ and assume that it extends continuously to a real function on (R,∞) for some $R > 0$. Then*

$$\frac{d^2}{d\xi^2}\left(\xi\varphi(1/\xi)\right) \leq 0 \qquad \textit{for all } \xi \in (0, 1/R)\,.$$

Proof of Lemma 11. By the Nevanlinna theorem [14] there are numbers $a \geq 0$ and $b \in \mathbb{R}$ and a non-negative measure μ with $\int_{\mathbb{R}}(t^2+1)^{-1}\,d\mu(t) < \infty$ such that

$$\varphi(z) = az + b + \int_{\mathbb{R}} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) d\mu(t) \qquad \text{for all } z \in \mathbb{C}_+ \,.$$

Since

$$d\mu(t) = \pi^{-1}\mathrm{w} - \lim_{\varepsilon \to 0+} \operatorname{Im} \varphi(t+i\varepsilon) dt$$

and since φ is assumed real on (R,∞), we infer that $\operatorname{supp}\mu \subset (-\infty, R]$. Using this fact one can show that the above representation formula for $\varphi(z)$ is also valid for $z \in (R,\infty)$ and, therefore,

$$\xi\varphi(1/\xi) = a + b\xi + \int_{\mathbb{R}} \left(\frac{\xi^2}{\xi t - 1} - \frac{\xi t}{t^2+1} \right) d\mu(t) \qquad \text{for all } \xi \in (0, 1/R) \,.$$

By dominated convergence it follows that $\xi \mapsto \xi\varphi(1/\xi)$ is smooth on $(0,R)$ and that

$$\frac{d^2}{d\xi^2}\left(\xi\varphi(1/\xi)\right) = 2\int_{\mathbb{R}} \frac{d\mu(t)}{(\xi t - 1)^3} \qquad \text{for all } \xi \in (0, 1/R) \,.$$

Since $(\xi t - 1)^{-3} \leq 0$ for $\xi < 1/R$ and $t \in \operatorname{supp}\mu \subset (-\infty, R]$, we obtain the claimed concavity. □

Proof of Proposition 10. By an approximation argument we may assume that K is invertible. Let C and D be positive definite and G and H Hermitian. We will show that

$$\frac{d^2}{d\xi^2} \operatorname{Tr} \left(1 + \left((D+\xi H)^{q/2} K^* (C+\xi G)^p K (D+\xi H)^{q/2} \right)^{-1/(p+q)} \right)^{-1} \leq 0$$

for all sufficiently small $\xi > 0$, which will prove the proposition. To achieve this, we write

$$\operatorname{Tr} \left(1 + \left((D+\xi H)^{q/2} K^* (C+\xi G)^p K (D+\xi H)^{q/2} \right)^{-1/(p+q)} \right)^{-1} = \xi\varphi(1/\xi)$$

with

$$\varphi(x) := \operatorname{Tr} \left(x^{-1} + \left((xD+H)^{q/2} K^* (xC+G)^p K (xD+H)^{q/2} \right)^{-1/(p+q)} \right)^{-1} \tag{19}$$

and appeal to Lemma 11. Thus we need to show that φ is well-defined on (R,∞) for some $R > 0$ and has an analytic extension to $\mathbb{C}_+$ with non-negative imaginary part.

Let us introduce some notation. For $-\pi \leq \alpha < \beta \leq \pi$ we set

$$S_{\alpha,\beta} := \left\{ re^{i\theta} \in \mathbb{C} : \ r > 0 \ , \ \alpha < \theta < \beta \right\}$$

and (dropping the dimension N from the notation)

$$\mathcal{S}_{\alpha,\beta} := \left\{ M \in \mathcal{M}_N : \ \operatorname{Im}\left(e^{-i\alpha}M\right) > 0 \,, \ \operatorname{Im}\left(e^{-i\beta}M\right) < 0 \right\} .$$

For $z \in \mathbb{C}_+$ we define

$$F(z) := (zD+H)^{q/2} K^* (zC+G)^p K (zD+H)^{q/2} \,.$$

This is well-defined since $\mathrm{Im}(zC+G) = (\mathrm{Im}\, z)C > 0$ and similarly $\mathrm{Im}(zD+H) > 0$, and for such matrices the p-th and $q/2$-th root are well-defined by analytic functional calculus. Moreover, F is analytic in $\mathbb{C}_+$.

Since $\mathrm{Im}(zC+G) > 0$ and $\mathrm{Im}(zD+H) > 0$, one has (see, for instance, [15, Lemma 1])

$$(zC+G)^p \in \mathcal{S}_{0,p\pi} \qquad \text{and} \qquad (zD+H)^{q/2} \in \mathcal{S}_{0,q\pi/2}\,,$$

and, since K is invertible, also $K^*(zC+G)^pK \in \mathcal{S}_{0,p\pi}$. Therefore (see, for instance, [15, Lemma 2] or, for a simpler proof, [4, Lemma 10])

$$\operatorname{spec} F(z) \in S_{0,(p+q)\pi}\,.$$

By the analytic functional calculus we can define $F(z)^{-1/(p+q)}$ for $z \in \mathbb{C}_+$ and by the spectral mapping theorem we obtain

$$\operatorname{spec} F(z)^{-1/(p+q)} \in \mathbb{C}_-\,.$$

Therefore $z^{-1} + F(z)^{-1/(p+q)}$ is invertible for $z \in \mathbb{C}_+$ and

$$\operatorname{spec}\left(z^{-1} + F(z)^{-1/(p+q)}\right)^{-1} \in \mathbb{C}_+\,.$$

This proves that

$$\varphi(z) := \operatorname{Tr}\left(z^{-1} + F(z)^{-1/(p+q)}\right)^{-1}$$

is analytic in $\mathbb{C}_+$ and that $\operatorname{Im}\varphi(z) \geq 0$ for $z \in \mathbb{C}_+$.

Let $x > \max\{\lambda_{\max}(G)/\lambda_{\min}(C), \lambda_{\max}(H)/\lambda_{\min}(D)\} =: R$, where $\lambda_{\max\,/\,\min}(M)$ denote the largest and smallest eigenvalue of a Hermitian matrix M. Then F extends continuously from $\mathbb{C}_+$ to (R,∞) and $F(x)$ is a positive definite matrix for $x > R$. Therefore φ extends continuously from $\mathbb{C}_+$ to (R,∞) and the continuation is given by the right side of (19). Note that $\varphi(x)$ is real (in fact, positive) for $x > R$. □

4. The variational method

In this section we show parts (2) and (3) of Theorem 2, namely that

$$(A,B) \mapsto \operatorname{Tr}\left(B^{q/2}K^*A^pKB^{q/2}\right)^s$$

is jointly convex for $-1 \leq q \leq p \leq 0$ and $s > 0$ and for $1 \leq p \leq 2$, $-1 \leq q \leq 0$ and $s \geq \min\{1/(p-1), 1/(q+1)\}$. We also prove parts (2) and (3) of Proposition 5.

We begin with the proof of Proposition 5, since this is simpler and since this will also be needed in the proof of Theorem 2. We use a variational method which originates in the work [11]. It is based on two ingredients. The first one is a variational characterization of the trace of a power.

Lemma 12. *If $X \in \mathcal{P}_N$, then*

$$\operatorname{Tr} X^s = \sup_{Y \geq 0}\left(s \operatorname{Tr} XY - (s-1)\operatorname{Tr} Y^{s/(s-1)}\right) \qquad \text{if } s > 1 \text{ or } s < 0\,, \tag{20}$$

$$\operatorname{Tr} X^s = \inf_{Y > 0}\left(s \operatorname{Tr} XY + (1-s)\operatorname{Tr} Y^{-s/(1-s)}\right) \qquad \text{if } 0 < s < 1\,. \tag{21}$$

Proof. We provide two different proofs.

First proof. We first assume $s > 1$. Then by Hölder's and Young's inequalities for any $Y \geq 0$,

$$\operatorname{Tr} XY \leq (\operatorname{Tr} X^s)^{1/s} \left(\operatorname{Tr} Y^{s/(s-1)}\right)^{(s-1)/s} \leq \frac{1}{s} \operatorname{Tr} X^s + \frac{s-1}{s} \operatorname{Tr} Y^{s/(s-1)} .$$

This proves $\geq$ in (20), and for $\leq$ it suffices to choose $Y = X^s$.

Next, let $0 < s < 1$. Then by Hölder's and Young's inequalities for any $Y > 0$,

$$\begin{aligned} \operatorname{Tr} X^s &= \operatorname{Tr}(Y^{s/2} X^s Y^{s/2}) Y^{-s} \leq \left(\operatorname{Tr}(Y^{s/2} X^s Y^{s/2})^{1/s}\right)^s \left(\operatorname{Tr} Y^{-s/(1-s)}\right)^{1-s} \\ &\leq s \operatorname{Tr}(Y^{s/2} X^s Y^{s/2})^{1/s} + (1-s) \operatorname{Tr} Y^{-s/(1-s)} . \end{aligned}$$

By the Lieb–Thirring inequality [30] (see also [8, Theorem 7.4]),

$$\operatorname{Tr}(Y^{s/2} X^s Y^{s/2})^{1/s} \leq \operatorname{Tr} XY ,$$

so we obtain $\geq$ in (21), and for $\leq$ it suffices to choose $Y = X^s$.

Finally, for $s < 0$ we apply (21) with s replaced by $s/(s-1) \in (0,1)$ and with the roles of X and Y interchanged. We obtain

$$\operatorname{Tr} Y^{s/(s-1)} \leq \frac{s}{s-1} \operatorname{Tr} XY + \frac{1}{1-s} \operatorname{Tr} X^s .$$

This proves $\geq$ in (20), and for $\leq$ we choose again $Y = X^s$.

Second proof. We provide the details only in the case $0 < s < 1$ (since in that case before we had to use the Lieb–Thirring inequality). It is easy to see that the infimum on the right side of (21) is attained by some Y_0 and, differentiating with respect to Y, this minimizer satisfies

$$sX - sY_0^{-1/(1-s)} = 0 ,$$

that is, $Y_0 = X^{-1+s}$. Inserting this we obtain

$$\inf_{Y>0} \left(s \operatorname{Tr} XY + (1-s) \operatorname{Tr} Y^{-s/(1-s)}\right) = s \operatorname{Tr} XY_0 + (1-s) \operatorname{Tr} Y_0^{-s/(1-s)} = \operatorname{Tr} X^s ,$$

as claimed. □

The second ingredient in this section is a result about suprema and infima of convex functions.

Lemma 13. *Let X be a convex subset of a vector space, Y a set and $f : X \times Y \to \mathbb{R}$ a function such that $f(\cdot, y)$ is convex for any $y \in Y$.*

1. *Then $x \mapsto \sup_{y \in Y} f(x,y)$ is convex.*
2. *If Y is a convex subset of a vector space and f is jointly convex on $X \times Y$, then $x \mapsto \inf_{y \in Y} f(x,y)$ is convex.*

Proof. Let $x_0, x_1 \in X$ and $0 < \theta < 1$.

(1) Abbreviating $g(x) := \sup_{y \in Y} f(x,y)$, we have for any $y \in Y$,

$$f((1-\theta)x_0 + \theta x_1, y) \leq (1-\theta) f(x_0, y) + \theta f(x_1, y) \leq (1-\theta) g(x_0) + \theta g(x_1) .$$

Taking the supremum over y we obtain $g((1-\theta)x_0 + \theta x_1) \leq (1-\theta) g(x_0) + \theta g(x_1)$.

(2) Let $h(x) := \inf_{y\in Y} f(x,y)$. Let $\varepsilon > 0$ and choose $y_0, y_1 \in Y$ such that

$$f(x_0,y_0) \le h(x_0) + \varepsilon \qquad \text{and} \qquad f(x_1,y_1) \le h(x_1) + \varepsilon\,.$$

Then, by joint convexity,

$$\begin{aligned} h((1-\theta)x_0 + \theta x_1) &\le f((1-\theta)x_0 + \theta x_1, (1-\theta)y_0 + \theta y_1) \\ &\le (1-\theta) f(x_0,y_0) + \theta f(x_1,y_1) \\ &\le (1-\theta) h(x_0) + \theta h(x_1) + \varepsilon\,. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we obtain $h((1-\theta)x_0 + \theta x_1) \le (1-\theta)h(x_0) + \theta h(x_1)$. □

We now use these tools to deduce the convexity assertions in Proposition 5.

Proof of Proposition 5. (2). Let $-1 \le p < 0$. We begin with the more difficult case $s < 1$ (which is the only case that will be used in the proof of Theorem 2). Then, by (21),

$$\begin{aligned} \Upsilon_{p,s}(A) &= \inf_{Y>0} \left(s \operatorname{Tr} K^* A^p K Y + (1-s) \operatorname{Tr} Y^{-s/(1-s)} \right) \\ &= \inf_{C>0} \left(s \operatorname{Tr} K^* A^p K C^{1-p} + (1-s) \operatorname{Tr} C^{-s(1-p)/(1-s)} \right). \end{aligned}$$

By Ando's convexity theorem [1] (see also [8, Theorem 6.2]), $(A,C) \mapsto \operatorname{Tr} K^* A^p K C^{1-p}$ is jointly convex. Moreover, since $-s(1-p)/(1-s) < 0$, $C \mapsto \operatorname{Tr} C^{-\frac{s(1-p)}{1-s}}$ is convex. Thus, by part (2) of Lemma 13, $\Upsilon_{p,s}$ is convex.

We now consider the case $s > 1$ and therefore

$$\Upsilon_{p,s}(A) = \sup_{Y\ge 0} \left(s \operatorname{Tr} K^* A^p K Y - (s-1) \operatorname{Tr} Y^{s/(s-1)} \right).$$

Since $A \mapsto A^p$ is operator convex, $A \mapsto \operatorname{Tr} K^* A^p K Y$ is convex. Thus, by part (1) of Lemma 13, $\Upsilon_{p,s}$ is convex.

The case $s = 1$ is even simpler and follows directly from the operator convexity of $A \mapsto A^p$. □

Proof of Proposition 5. (3). The argument is similar to that in part (1) and we refer to [11] for details . This result will not be needed in the proof of Theorem 2. □

Before we turn to the proof of Theorem 2 we recall a joint operator convexity statement from [9].

Lemma 14. *For any $-1 \le q \le 0$ and any $K \in \mathcal{M}_N$ the map*

$$\mathcal{P}_N \times \mathcal{P}_N \ni (A,B) \mapsto A K B^q K^* A$$

is jointly convex.

In fact, in [9, Theorem 3.2] we also proved that the restriction $-1 \le q \le 0$ is necessary and that convexity does not hold if A is raised to a non-zero power.

In [9] we observed that this lemma follows from [27, Corollary 2.1]. (The presence of the operators K and K^* does not present a problem. They can be dealt with by

doubling of dimension as, for instance, in [9, Lemma 1.1]. In this way they can be made into unitary operators and then they can be absorbed into B.) Also, in [9, Remark 3.5] we explained a simple alternative proof which reduces the case $-1 < q < 0$ to the well-known case $q = -1$ [23, 29].

We now prove the convexity assertions in Theorem 2.

Proof of Theorem 2. (2). We assume that $-1 \leq q \leq p \leq 0$ and that $s > 0$. By an approximation argument we may assume that K is invertible and we denote $L := K^{-*}$. Then, by (20),

$$\begin{aligned}\Psi_{p,q,s}(A,B) &= \operatorname{Tr}\left(B^{-q/2}L^*A^{-p}LB^{-q/2}\right)^{-s} \\ &= \sup_{Y\geq 0}\left(-s\operatorname{Tr}B^{-q/2}L^*A^{-p}LB^{-q/2}Y + (s+1)\operatorname{Tr}Y^{\frac{s}{s+1}}\right) \\ &= \sup_{Z\geq 0}\left(-s\operatorname{Tr}L^*A^{-p}LZ + (s+1)\operatorname{Tr}(B^{q/2}ZB^{q/2})^{\frac{s}{s+1}}\right).\end{aligned}$$

Since $A \mapsto A^{-p}$ is operator concave, $A \mapsto -s\operatorname{Tr}L^*A^{-p}LZ$ is convex for any Z. Moreover, by part (1) of Proposition 5, $B \mapsto \operatorname{Tr}(Z^{1/2}B^qZ^{1/2})^{\frac{s}{s+1}} = \operatorname{Tr}(B^{q/2}ZB^{q/2})^{\frac{s}{s+1}}$ is convex for any Z. (We apply the lemma with A replaced by B, K by $Z^{1/2}$, p by q and s by $s/(s+1)$.) Thus, by part (1) of Lemma 13, $\Psi_{p,q,s}$ is convex. □

Proof of Theorem 2. (3). We break the proof into three steps.

Step 1. We assume $1 \leq p \leq 2$, $-1 \leq q \leq 0$ and $s \geq 1/(1+q)$. Then, by (20),

$$\begin{aligned}\Psi_{p,q,s}(A,B) &= \sup_{Y\geq 0}\left(s\operatorname{Tr}B^{q/2}K^*A^pKB^{q/2}Y - (s-1)\operatorname{Tr}Y^{s/(s-1)}\right) \\ &= \sup_{Z\geq 0}\left(s\operatorname{Tr}K^*A^pKZ - (s-1)\operatorname{Tr}\left(B^{-q/2}ZB^{-q/2}\right)^{s/(s-1)}\right).\end{aligned}$$

Since $A \mapsto A^p$ is operator convex, $A \mapsto s\operatorname{Tr}K^*A^pKZ$ is convex for any $Z \geq 0$. Moreover, by part (1) of Proposition 5 and since $\frac{s}{s-1} \leq -\frac{1}{q}$, $B \mapsto \operatorname{Tr}\left(Z^{1/2}B^{-q}Z^{1/2}\right)^{\frac{s}{s-1}} = \operatorname{Tr}\left(B^{-q/2}ZB^{-q/2}\right)^{\frac{s}{s-1}}$ is concave. Thus, by part (1) of Lemma 13, $\Psi_{p,q,s}$ is convex.

Step 2. We assume $1 < p \leq 2$, $-1 \leq q \leq 0$ and $s \geq 1/(p-1)$. First, assume that $s = 1$, so that necessarily $p = 2$. The joint convexity follows from (16) and the fact that according to Lemma 14 $(A,B) \mapsto AKB^qK^*A$ is operator convex.

Now let $s > 1$ (and still $s \geq 1/(p-1)$). Then, by (16) and (20),

$$\begin{aligned}\Psi_{p,q,s}(A,B) &= \sup_{Y>0}\left(s\operatorname{Tr}A^{p/2}KB^qK^*A^{p/2}Y - (s-1)\operatorname{Tr}Y^{s/(s-1)}\right) \\ &= \sup_{Z>0}\left(s\operatorname{Tr}AKB^qK^*AZ - (s-1)\operatorname{Tr}\left(A^{1-p/2}ZA^{1-p/2}\right)^{s/(s-1)}\right).\end{aligned}$$

Again by Lemma 14, $(A,B) \mapsto \operatorname{Tr}AKB^qK^*AZ$ is jointly convex. Moreover, by part (1) of Proposition 5 and since $s/(s-1) \leq 1/(2-p)$, $A \mapsto \operatorname{Tr}\left(Z^{1/2}A^{2-p}Z^{1/2}\right)^{s/(s-1)} = \operatorname{Tr}\left(A^{1-p/2}ZA^{1-p/2}\right)^{s/(s-1)}$ is concave. Thus, by part (1) of Lemma 13, $\Psi_{p,q,s}$ is convex.

Step 3. We assume that $p = 2$ and $1/(2+q) \leq s < 1$. Then, by (16) and (21),

$$\begin{aligned}\Psi_{2,q,s}(A,B) &= \inf_{Y>0} \left(s \operatorname{Tr} AKB^q K^* AY + (1-s) \operatorname{Tr} Y^{-s/(1-s)} \right) \\ &= \inf_{Z>0} \left(s \operatorname{Tr} AKB^q K^* AZ^{1-1/s} + (1-s) \operatorname{Tr} Z \right).\end{aligned}$$

By [27, Corollary 2.1] and since $q + 1 - 1/s \geq -1$, $(A,B,Z) \mapsto \operatorname{Tr} AKB^q K^* AZ^{1-1/s}$ is jointly convex. (Note that the operator K, which is not present in [27, Corollary 2.1], can be dealt with as explained after Lemma 14.) Thus, by part (2) of Lemma 13, $\Psi_{2,q,s}$ is convex. □

Remark 15. Both in Steps 1 and 2 of the previous proof we applied part (1) of Proposition 5. We have proved the latter, as a special case of part (1) of Theorem 2, using the complex analysis method. It is interesting to note, however, that in Steps 1 and 2 of the previous proof we applied part (1) of Proposition 5 only with a power $s \geq 1$. For such powers part (1) of Proposition 5 can be proved also using the variational method; see [11, Theorem 1.1].

Remark 16. In the special case $p = 2$, $-1 \leq q \leq 0$, $s \geq 1$ there is a proof which is only based on Lemma 14 and which does not use Proposition 5; see [9, Remark 4.3].

Remark 17. Let us give an alternative proof for $p = 2$ and $1/(2+q) \leq s < 1$, which does not use the deep [27, Corollary 2.1], but only the special case $q = -1$ of Lemma 14, plus part (3) of Theorem 2 for $p = 1$. (Recall that Lemma 14 can be deduced from [27, Corollary 2.1], but that its special case $q = -1$ is rather simple and well-known.) We do *not* use (16), but use directly (21) to write

$$\begin{aligned}\Psi_{2,q,s}(A,B) &= \inf_{Y>0} \left(s \operatorname{Tr} B^{q/2} K^* A^2 K B^{q/2} Y + (1-s) \operatorname{Tr} Y^{-s/(1-s)} \right) \\ &= \inf_{Z>0} \left(s \operatorname{Tr} K^* A^2 K Z^{-1} + (1-s) \operatorname{Tr} (B^{-q/2} Z^{-1} B^{-q/2})^{-s/(1-s)} \right).\end{aligned}$$

By Lemma 14, $(A,Z) \mapsto \operatorname{Tr} K^* A^2 K Z^{-1}$ is jointly convex. Moreover, by Step 1 in the proof of part (3) of Theorem 2, $(B,Z) \mapsto \operatorname{Tr}(Z^{1/2} B^q Z^{1/2})^{\frac{s}{1-s}} = \operatorname{Tr}(B^{-q/2} Z^{-1} B^{-q/2})^{-\frac{s}{1-s}}$ is jointly convex. Thus, by part (2) of Lemma 13, $\Psi_{2,q,s}$ is convex.

5. Necessary conditions

In this section we reproduce the arguments from [7, 11, 19] to prove Propositions 6 and 3 containing necessary conditions for concavity and convexity.

Proof of Proposition 6. (1) Clearly, concavity of $\Upsilon_{p,s}$ on $\mathcal{P}_1$ implies that $(0,\infty) \ni a \mapsto a^{ps}$ is concave and therefore $0 < ps \leq 1$. Moreover, taking

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \qquad \text{and} \qquad K = \begin{pmatrix} 1 & 0 \\ 1 & \varepsilon \end{pmatrix}$$

with numbers a, b and ε and letting $\varepsilon \to 0$, we deduce from concavity on $\mathcal{P}_2$ that $(0,\infty) \times (0,\infty) \ni (a,b) \mapsto (a^p + b^p)^s$ is jointly concave. By differentiating twice with respect to a and evaluating at $a \ll b$ we find that $p \leq 1$.

(2) Clearly, convexity of $\Upsilon_{p,s}$ on $\mathcal{P}_1$ implies that $(0,\infty) \ni a \mapsto a^{ps}$ is convex and therefore $0 < ps \leq 1$ or $ps < 0$. If $s = 1$, then the convexity of $\Upsilon_{p,s}$ on $\mathcal{P}_2$ for any (invertible) K implies that $\mathcal{P}_2 \ni A \mapsto A^p$ is operator convex. As is well known, this implies that $-1 \leq p < 0$ or $1 \leq p \leq 2$. Now let $s \neq 1$. Taking

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \qquad \text{and} \qquad K = \begin{pmatrix} 1 & 0 \\ 1 & \varepsilon \end{pmatrix}$$

where now a, b and 1 are 2×2 matrices and letting $\varepsilon \to 0$, we deduce from convexity on $\mathcal{P}_4$ that $\mathcal{P}_2 \times \mathcal{P}_2 \ni (A, B) \mapsto \mathrm{Tr}(A^p + B^p)^s$ is jointly convex. In particular, for any $t > 0$, $\mathcal{P}_2 \ni A \mapsto \mathrm{Tr}(tA^p + B^p)^s$ is convex. Using

$$\mathrm{Tr}(tA^p + B^p)^s = \mathrm{Tr}\, B^{ps} + st\, \mathrm{Tr}\, B^{p(s-1)}A^p + o(t) \qquad \text{as } t \to 0$$

we deduce that $\mathcal{P}_2 \ni A \mapsto \mathrm{Tr}\, B^{p(s-1)}A^p$ is convex, which (since $s \neq 1$) means that $\mathcal{P}_2 \ni A \mapsto A^p$ is operator convex. As before, this implies $-1 \leq p < 0$ or $1 \leq p \leq 2$. □

Proof of Proposition 3. (1) Clearly, concavity of $\Psi_{p,q,s}$ on $\mathcal{P}_1 \times \mathcal{P}_1$ implies that $(0,\infty) \ni a \mapsto a^{(p+q)s}$ is concave and therefore $0 \leq (p+q)s \leq 1$. Writing an invertible K as $K = U|K|$ with U unitary, chosing $B^{q/2} = |K|$ and absorbing U into A we deduce from the concavity of $\mathcal{P}_2 \ni A \mapsto \Psi_{p,q,s}(A, B)$ that $\mathcal{P}_2 \ni A \mapsto \Upsilon_{p,s}(A)$ is concave. According to part (1) of Proposition 6 this implies $0 \leq p \leq 1$. Exchanging the roles of p and q we find that $0 \leq q \leq 1$.

(2) Clearly, convexity of $\Psi_{p,q,s}$ on $\mathcal{P}_1 \times \mathcal{P}_1$ implies that $(0,\infty) \times (0,\infty) \ni (a, b) \mapsto a^{ps}b^{qs}$ is jointly concave. By an elementary analysis of the Hessian we conclude that, if $p \geq 0$, then $q \leq 0$ and $(p+q)s \geq 1$. Similarly as in the first part of the proof, part (2) of Proposition 6 implies either $-1 \leq p \leq 0$ or $1 \leq p \leq 2$ and, exchanging the roles of p and q, either $-1 \leq q \leq 0$ or $1 \leq q \leq 2$. This corresponds to four disjoint squares in the (p, q) plane. The square $-1 \leq p \leq 0$, $1 \leq q \leq 2$ is excluded by our assumption $p \geq q$ and the square $1 \leq p, q \leq 2$ is excluded by the above elementary analysis. This concludes the proof. □

References

[1] T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products*. Linear Algebra Appl. **26** (1979), 203–241.

[2] H. Araki, *Inequalities in Von Neumann Algebras* Les rencontres physiciens-mathématiciens de Strasbourg, RCP25 22 (1975), 1–25.

[3] H. Araki, *Relative Entropy of States of von Neumann Algebras*. Publ. RIMS, Kyoto Univ. **11** (1976), 809–833.

[4] K. M. R. Audenaert, N. Datta, *$\alpha - z$-Rényi relative entropies*. Preprint, arXiv:1310.7178v3

[5] K. M. R. Audenaert, N. Datta, *$\alpha - z$-Rényi relative entropies*. J. Math. Phys. **56** (2015), no. 2, 022202, 16 pp.

[6] S. Beigi. *Sandwiched Rényi divergence satisfies data processing inequality*. J. Math. Phys. **54** (2013), no. 12, 122202.

[7] T. N. Bekjan, *On joint convexity of trace functions*. Linear Algebra Appl. **390** (2004), 321–327.

[8] E. A. Carlen, *Trace inequalities and quantum entropy: an introductory course*. In: Entropy and the Quantum, Contemp. Math., vol. 529, Amer. Math. Soc., Providence, RI, 2010, pp. 73–140.

[9] E. A. Carlen, R. L. Frank, E. H. Lieb, *Some operator and trace function convexity theorems*. Linear Algebra Appl. **490** (2016), 174–185.

[10] E. A. Carlen, E. H. Lieb, *A Minkowski Type Trace Inequality and Strong Subadditivity of Quantum Entropy*, Advances in the Mathematical Sciences, AMS Translations, **189** Series 2, (1999) 59–68.

[11] E. A. Carlen, E. H. Lieb, *A Minkowski type trace inequality and strong subadditivity of quantum entropy. II. Convexity and concavity*. Lett. Math. Phys. **83** (2008), no. 2, 107–126.

[12] E. A. Carlen, E. H. Lieb, *Some trace inequalities for exponential and logarithmic functions*, Bull. Math. Sci., to appear.

[13] I. Csiszár: *Generalized cutoff rates and Rényi's information measures*, IEEE Trans. Inf. Theory **41** (1995), 26–34,

[14] W. F. Donoghue, *Monotone Matrix Functions and Analytic Continuation*, Die Grundlehren der Math. Wiss., vol. 207, Springer, Berlin, 1974

[15] H. Epstein, *Remarks on two theorems of E. Lieb*. Comm. Math. Phys. **31** (1973), 317–325.

[16] R. L. Frank, E. H. Lieb, *Monotonicity of a relative Rényi entropy*. J. Math. Phys. **54** (2013), no. 12, 122201, 5 pp.

[17] F. Hiai, *Concavity of certain matrix trace functions*. Taiwanese J. Math. **5** (2001), no. 3, 535–554.

[18] F. Hiai, *Concavity of certain matrix trace and norm functions*. Linear Algebra Appl. **439** (2013), no. 5, 1568–1589.

[19] F. Hiai, *Concavity of certain matrix trace and norm functions. II*. Linear Algebra Appl. **496** (2016), 193–220.

[20] F. Hiai, M. Mosonyi, *Quantum F-divergences and error correction* Reviews in Mathematical Physics, **29**, (2017), 691-747

[21] F. Hiai and D. Petz, *The proper formula for relative entropy and its asymptotics in quantum probability*, Comm. Math. Phys. **413** (2006), 99–114.

[22] V. Jaksic, Y. Ogata, Y. Pautrat and C.-A. Pillet, *Entropic fluctuations in quantum statistical mechanics. An Introduction*. In: Quantum Theory from Small to Large Scales: Lecture Notes of the Les Houches Summer School: Volume **95**, August 2010, Oxford University Press, 2012.

[23] J. Kiefer, *Optimum experimental designs*. J. Roy. Statist. Soc. Ser. B **21** (1959) 272–310.

[24] K. Kraus, *General state changes in quantum theory*. Ann. Phys. **64** (1971), 311–335.

[25] S. Kullback and R.A. Leibler, *On information and sufficiency*, Annals of Math. Stat. **22** (1951), 79–86.

[26] S. Kullback *Lower bound for discrimination information in terms of variation*, IEEE Trans. on Inf. Theory, **13**, 1967, 126–127. Correction, **16**, (1970) p. 652.

[27] E. H. Lieb, *Convex trace functions and the Wigner–Yanase–Dyson conjecture*. Advances in Math. **11** (1973), 267–288.

[28] E. H. Lieb, M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*. With an appendix by B. Simon. J. Mathematical Phys. **14** (1973), 1938–1941.

[29] E. H. Lieb, M. B. Ruskai, *Some operator inequalities of the Schwarz type*. Adv. Math. **12** (1974), 269–273.

[30] E. H. Lieb, W. Thirring, *Inequalities for the moments of the eigenvalues of the Schröderinger Hamiltonian and their relation to Sobolev inequalities*. Studies in Math. Phys., Essays in Honor of Valentine Bargmann, Princeton, 269–303 (1976)

[31] G. Lindblad, *Expectations and Entropy Inequalities for Finite Quantum Systems*, Comm. math. Phys. **39** (1974), 111–119.

[32] G. Lindblad, *Completely Positive Maps and Entropy Inequalities,* Comm. Math. Phys. **40** (1975), 147–151.

[33] K. Löwner, *Über monotone Matrixfunktionen*, Math. Zeit. **38** (1934), 177–216.

[34] M. Mosonyi, F. Hiai, *On the quantum Rényi relative entropies and related capacity formulas* IEEE Trans. Inf. Th. **57** (2011), 2474–2487.

[35] M. Mosonyi, T. Ogawa, *Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies*. Comm. Math. Phys. **334** (2015), 161–1648

[36] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, M. Tomamichel, *On quantum Rényi entropies: a new generalization and some properties*. J. Math. Phys. **54** (2013), no. 12, 122203, 20 pp.

[37] M. A. Nielsen, D. Petz, *A Simple Proof of the Strong Subadditivity Inequality*, Quantum Information and Computation **5** (2005), 507–513.

[38] A. Rényi. *On measures of entropy and information*. Proc. 4th Berkeley Symp. on Math. Statist. and Probability, Vol. 1, 547–561, University of California Press (1961). Available online at http://projecteuclid.org/euclid.bsmsp/1200512181.

[39] B. Simon, *Loewner's Theorem on Matrix Functions*, forthcoming.

[40] W. F. Stinespring, *Positive functions on C^*-algebras*. Proc. Amer. Math. Soc. **6** (1955), 211–216.

[41] A. Uhlmann, *Endlich Dimensionale Dichtematrizen, II,* Wiss. Z. Karl-Marx University Leipzig, Math-Naturwiss. **22** (1973), no. 2, 139.

[42] H. Umegaki, *Conditional expectation in an operator algebra, IV (entropy and information)*. Kodai Math. Sem. Rep. **14** (1962), 59–85

[43] A. Wehrl, *General properties of entropy*. Rev. Modern Phys. **50** (1978), no. 2, 221–260.

[44] M. M. Wilde, A. Winter, D. Yang, *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy*. Comm. Math. Phys. **331** (2014), no. 2, 593–622.

(Eric A. Carlen) Department of Mathematics, Hill Center, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA

E-mail address: carlen@math.rutgers.edu

(Rupert L. Frank) Mathematisches Institut, Ludwig-Maximilans Universität München, Theresienstr. 39, 80333 München, Germany, and Mathematics 253-37, Caltech, Pasadena, CA 91125, USA

E-mail address: rlfrank@caltech.edu

(Elliott H. Lieb) Departments of Mathematics and Physics, Princeton University, Washington Road, Princeton, NJ 08544, USA

E-mail address: lieb@princeton.edu