Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb

Igor Bjelaković[1] and Rainer Siegmund-Schultze[2]

[1]Technische Universität München, Theoretische Informationstechnik,
Arcisstraße 21, 80333 München, Germany
igor.bjelakovic@tum.de
[2]Technische Universität Berlin, Institut für Mathematik,
Straße des 17. Juni 136, 10623 Berlin, Germany
siegmund@math.tu-berlin.de

Abstract

We derive the monotonicity of the quantum relative entropy by an elementary operational argument based on Stein's lemma in quantum hypothesis testing. For the latter we present an elementary and short proof that requires the law of large numbers only. Joint convexity of the quantum relative entropy is proven too, resulting in a self-contained elementary version of Tropp's approach to Lieb's concavity theorem, according to which the map $a \mapsto \mathrm{tr}(\exp(h + \log a))$ is concave on positive operators for self-adjoint h.

Contents

1 Introduction

Inequalities for quantum mechanical entropies and related concave trace functions play a fundamental role in quantum information theory. The groundbreaking results on these inequalities by Lieb [14] and by Lieb and Ruskai [15] together with the extension of the fundamental operational ideas and concepts from Shannon's information theory [22] to the quantum realm have made the rapid development of quantum information theory possible. For example, many optimality proofs in quantum information theory rely on one of the fundamental inequalities established in [14] and [15]. The proofs presented in [14] and [15] are masterpieces of matrix analysis.

In this paper we approach some of the major inequalities for quantum entropies from the point of view of Shannon's theory. It turns out that the monotonicity property of the quantum relative entropy [15, 17, 27] is an elementary and intuitive consequence of the quantum version of Stein's lemma [11, 18], which gives an operational interpretation to the quantum relative entropy as a distinguishability measure on the set of quantum states. Stein's lemma, in turn, can be established by an elementary argument on less than two pages: The new proof that we present below requires only the law of large numbers and a simple estimate on "overlaps" of certain projections with respect to some given quantum state. It is this simple proof of Stein's lemma that we consider as the main contribution of the paper. The observation that the monotonicity of the relative entropy can be derived from Stein's lemma has been made in the technical report [3] by the authors almost a decade ago. There we have given an elementary but, unfortunately, somewhat non-transparent proof of Stein's lemma.

Once the monotonicity is established, we have access to other fundamental properties of quantum entropies as described in Ruskai's review [21, Sec. V]. We restrict our attention to the joint convexity of the quantum relative entropy [16, 27] for which we include a short proof. Indeed, the joint convexity and monotonicity of the relative entropy are equivalent [21, Sec. V] and therefore we could easily obtain it directly from Stein's lemma. We prefer, however, to follow the nice derivation from the monotonicity mentioned in [21, Sec. V].

Our motivation for including the joint convexity of relative entropy is the recent work [25] by Tropp where he derives one of Lieb's concavity theorems [14], according to which the map $a \mapsto \mathrm{tr}(\exp(h + \log a))$ is concave on the positive cone for fixed self-adjoint h, from the joint convexity of the relative entropy. We present a short and slightly streamlined version of Tropp's argument in Section 4. Together with our operational derivation of the monotonicity of the quantum relative entropy we obtain a self-contained access to some of the fundamental inequalities for quantum entropies and trace functions from the information-theoretic perspective.

Besides the fact that Lieb's concavity theorem played a crucial role in [15] it became of great importance for establishing sharp tail concentration bounds for sums of random matrices [24]. The latter development, in turn, has its origin in the famous Ahlswede-Winter bound [1] that arose in the context of the theory of identification via quantum channels.

We resisted the temptation of producing an extremely short paper, which could be done given the elementary nature of the arguments that are used. Instead, our leitmotif was to give a self-contained presentation of the results at a slow pace so that anybody knowing the law of large numbers and being familiar with basic linear algebra can easily follow our arguments. Only exception being Remark 10 in Section 3 where we assume the familiarity with the definition and simple properties of completely positive maps which can be easily picked up in Bhatia's beautiful book [2, Ch. 3]. We should note, however, that no result in the paper depends on the inequality presented in Remark 10. It is included for completeness only and can safely be skipped without any consequence for the subsequent parts of the paper.

2 An elementary proof of quantum Stein's lemma

For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$[1], $\varepsilon \in (0,1)$, and $n \in \mathbb{N}$ we define

$$\beta_{\varepsilon,n}(\rho,\sigma) := \min\left\{\mathrm{tr}(\sigma^{\otimes n}a) : a \in [0, \mathbf{1}_{\mathcal{H}^{\otimes n}}], \mathrm{tr}(\rho^{\otimes n}a) \geq 1-\varepsilon\right\}, \tag{1}$$

where $[0, \mathbf{1}_{\mathcal{H}^{\otimes n}}]$ denotes the set of (self-adjoint) operators a on $\mathcal{H}^{\otimes n}$ with $0 \leq a \leq \mathbf{1}_{\mathcal{H}^{\otimes n}}$.

The quantities $\beta_{\varepsilon,n}(\rho,\sigma)$ obtain their natural interpretation in terms of statistical hypothesis testing. We suppose that the system under consideration is prepared either according to the state ρ or to the state σ and we can perform measurements/observations on n independently prepared systems whose state is then $\rho^{\otimes n}$ or $\sigma^{\otimes n}$. According to Quantum Mechanics, a binary observable of the n-partite system is a map $E : \{0,1\} \to [0, \mathbf{1}_{\mathcal{H}^{\otimes n}}]$ such that $E(0)+E(1) = \mathbf{1}_{\mathcal{H}^{\otimes n}}$. Given a state $\tau \in \mathcal{S}(\mathcal{H}^{\otimes n})$ Quantum Mechanics assigns the probabilities

$$\mathrm{tr}(\tau E(i)) \in [0,1] \qquad (i \in \{0,1\}) \tag{2}$$

for obtaining the outcome i when measuring the observable E and when the system is prepared in the state τ. Notice that this is consistent with the requirement $E(0) + E(1) = \mathbf{1}_{\mathcal{H}^{\otimes n}}$ leading to $\mathrm{tr}(\tau E(0)) + \mathrm{tr}(\tau E(1)) = 1$ for all $\tau \in \mathcal{S}(\mathcal{H}^{\otimes n})$.

Given states $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ the probabilities for obtaining the outcome 0 for the observable E are given by $\mathrm{tr}(\rho^{\otimes n}a)$ and $\mathrm{tr}(\sigma^{\otimes n}a)$ with $a := E(0)$ given that the n-partite system is prepared either in state $\rho^{\otimes n}$ or in state $\sigma^{\otimes n}$.

Suppose that we use the observable E as a decision rule, meaning that when obtaining the outcome 0 we decide that the state was ρ and else we decide in favor of σ. Then $\mathrm{tr}(\sigma^{\otimes n}a)$ (again abbreviating $a = E(0)$) represents the error probability of our decision rule and the numbers $\beta_{\varepsilon,n}(\rho,\sigma)$ are the minimum error probabilities for the decision rules deciding in favor of ρ with high probability (assuming that ε is close to 0).

The quantum relative entropy of $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ is given by[2]

$$D(\rho||\sigma) := \begin{cases} \mathrm{tr}(\rho\log\rho - \rho\log\sigma) & \text{if } \ker\sigma \subseteq \ker\rho \\ +\infty & \text{else.} \end{cases} \tag{3}$$

Theorem 1 (Quantum Stein's lemma [11], [18]): Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be states with $\ker\sigma \subseteq \ker\rho$. Then for any $\varepsilon \in (0,1)$ we have

$$\lim_{n\to\infty} \frac{1}{n}\log\beta_{\varepsilon,n}(\rho,\sigma) = -D(\rho||\sigma). \tag{4}$$

Remark 2: Stein's Lemma shows that, roughly, $\beta_{\varepsilon,n}(\rho,\sigma) \approx e^{-nD(\rho||\sigma)}$ giving an operational interpretation to the quantum relative entropy as the largest rate at which the error probability decays to 0 exponentially fast.

The proof of Theorem 1 that is presented below relies on two simple lemmas that are proven first.

Let $\tau_1, \tau_2 \in \mathcal{S}(\mathcal{H})$ and assume that $\ker\tau_2 \subseteq \ker\tau_1$ holds. We set

$$M(\tau_1||\tau_2) := -\mathrm{tr}(\tau_1\log\tau_2). \tag{5}$$

[1] $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}(\mathcal{H}) : \rho \geq 0, \mathrm{tr}(\rho) = 1\}$ denotes the set of density operators or states on $\mathcal{H}$ whereas $\mathcal{L}(\mathcal{H})$ stands for the set of linear maps from $\mathcal{H}$ to itself.

[2] All logarithms are to the base e.

Remark 3: Notice that

$$M(\tau_1||\tau_2) - S(\tau_1) = D(\tau_1||\tau_2), \tag{6}$$

and

$$M(\tau_1||\tau_1) = S(\tau_1) \tag{7}$$

hold, where $S(\tau_1) := -\mathrm{tr}(\tau_1 \log \tau_1)$ is the von Neumann entropy of τ_1.

Let $\mu_1, \ldots, \mu_d$ be the eigenvalues of τ_2 counted with their multiplicities and $e_1, \ldots, e_d$ a complete orthonormal set of corresponding eigenvectors. For $[d] := \{1, \ldots, d\}$ and $x^n := (x_1, \ldots, x_n) \in [d]^n$ we set

$$\mu_{x^n} := \prod_{i=1}^{n} \mu_{x_i} \qquad e_{x^n} := \bigotimes_{i=1}^{n} e_{x_i}. \tag{8}$$

Additionally, for $\delta > 0$ we set

$$\begin{aligned} T_{\delta,n}(\tau_1,\tau_2) &:= \left\{ x^n \in [d]^n : M(\tau_1||\tau_2) - \delta < -\frac{1}{n}\log \mu_{x^n} < M(\tau_1||\tau_2) + \delta \right\} \qquad (9) \\ &= \left\{ x^n \in [d]^n : e^{-n(M(\tau_1||\tau_2)+\delta)} < \mathrm{tr}(\tau_2^{\otimes n}|e_{x^n}\rangle\langle e_{x^n}|) < e^{-n(M(\tau_1||\tau_2)-\delta)} \right\}. \end{aligned}$$

Finally we introduce the following projection

$$p_{\delta,n}(\tau_1,\tau_2) := \sum_{x^n \in T_{\delta,n}(\tau_1,\tau_2)} |e_{x^n}\rangle\langle e_{x^n}|. \tag{10}$$

Lemma 4: For all $\tau_1, \tau_2 \in \mathcal{S}(\mathcal{H})$ with $\ker \tau_2 \subseteq \ker \tau_1$ and all $\delta > 0$ we have:

1. $p_{\delta,n}(\tau_1,\tau_2)\tau_2^{\otimes n} = \tau_2^{\otimes n} p_{\delta,n}(\tau_1,\tau_2)$ for all $n \in \mathbb{N}$.
2. $p_{\delta,n}(\tau_1,\tau_2)\tau_2^{\otimes n} p_{\delta,n}(\tau_1,\tau_2) \le e^{-n(M(\tau_1||\tau_2)-\delta)} p_{\delta,n}(\tau_1,\tau_2)$ for all $n \in \mathbb{N}$
3. $p_{\delta,n}(\tau_1,\tau_2)\tau_2^{\otimes n} p_{\delta,n}(\tau_1,\tau_2) \ge e^{-n(M(\tau_1||\tau_2)+\delta)} p_{\delta,n}(\tau_1,\tau_2)$ for all $n \in \mathbb{N}$
4. $\lim_{n\to\infty} \mathrm{tr}(\tau_1^{\otimes n} p_{\delta,n}(\tau_1,\tau_2)) = 1$.

Proof: The first three claims in the lemma are obvious from the definition of $p_{\delta,n}(\tau_1,\tau_2)$. The last assertion follows from the law of large numbers: First of all, we can w.l.o.g. assume that τ_2 is invertible due to our assumption that $\ker \tau_2 \subseteq \ker \tau_1$. Let $X_1, \ldots, X_n$ be independent, identically distributed (i.i.d.) random variables taking values in $[d]$ with distribution

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \mathrm{tr}(\tau_1^{\otimes n}|e_{x^n}\rangle\langle e_{x^n}|) = \prod_{i=1}^{n} \mathrm{tr}(\tau_1|e_{x_i}\rangle\langle e_{x_i}|). \tag{11}$$

For $i = 1, \ldots, n$ we introduce

$$U_i := -\log \mu_{X_i} \tag{12}$$

i.e. $U_i = f \circ X_i$ with the function $f : [d] \to \mathbb{R}$, $f(x) = -\log \mu_x$.
$U^n := (U_1, \ldots, U_n)$ is an i.i.d. collection of random variables and

$$\mathbb{E}(U_i) = \sum_{x\in[d]} \mathrm{tr}(\tau_1|e_x\rangle\langle e_x|)(-\log \mu_x) = M(\tau_1||\tau_2) \tag{13}$$

for all $i = 1, \ldots, n$. Moreover, it is clear that

$$\begin{aligned}\operatorname{tr}(\tau_1^{\otimes n} p_{\delta,n}(\tau_1,\tau_2)) &= \Pr(U^n \in T_{\delta,n}(\tau_1,\tau_2)) \\ &= \Pr\left(\left|\sum_{i=1}^{n} U_i - nM(\tau_1||\tau_2)\right| < n\delta\right) \end{aligned} \tag{14}$$

Eqn. (13), (14), and the law of large numbers imply that

$$\lim_{n\to\infty} \operatorname{tr}(\tau_1^{\otimes n} p_{\delta,n}(\tau_1,\tau_2)) = 1, \tag{15}$$

as desired.

□

Remark 5: In the proof of Theorem 1 we will have to apply Lemma 4 for the pairs $(\tau_1,\tau_2) = (\rho,\sigma)$ and $(\tau_1,\tau_2) = (\rho,\rho)$ simultaneously. For the latter case we introduce a separate projection:

$$p_{\delta,n}(\rho) := p_{\delta,n}(\rho,\rho). \tag{16}$$

The next lemma is a fusion and a slight generalization of Lemma 6 in [10] and Lemma 8 in [4]. The proof is a standard application of the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product on the space of linear operators over a finite-dimensional Hilbert space and is relegated to the Appendix A.

Lemma 6: Let $\mathcal{K}$ be a Hilbert space over $\mathbb{C}$ with $\dim \mathcal{K} < \infty$. Let $p, q \in \mathcal{L}(\mathcal{K})$ with $0 \leq p, q \leq \mathbf{1}$ and $\tau \in \mathcal{S}(\mathcal{K})$. Then

1. $\operatorname{tr}(\tau pqp) \geq \operatorname{tr}(\tau q) - 2\sqrt{\operatorname{tr}(\tau(\mathbf{1}_{\mathcal{K}} - p))}$.
2. If $u \in \mathcal{L}(\mathcal{K})$ is any projection commuting with τ (i.e. $u\tau = \tau u$) and satisfying $\tau u \leq cu$ for some $c \in \mathbb{R}_+$ then

$$\operatorname{tr}(pqp) \geq \frac{1}{c}\left(\operatorname{tr}(\tau q) - 2\sqrt{\operatorname{tr}(\tau(\mathbf{1}_{\mathcal{K}} - p))} - \operatorname{tr}(\tau(\mathbf{1}_{\mathcal{K}} - u))\right).$$

Proof of Theorem 1: In a first step we show that for all $\varepsilon \in (0,1)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n}(\rho,\sigma) \leq -D(\rho||\sigma).$$

Let $\delta > 0$ be given. Then on account of Lemma 4.4 applied simultaneously to $(\tau_1,\tau_2) = (\rho,\sigma)$ and $(\tau_1,\tau_2) = (\rho,\rho)$, Remark 5, and Lemma 6.1 for any $\varepsilon \in (0,1)$ there is $n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n_0(\varepsilon)$ we have

$$\begin{aligned}\operatorname{tr}(\rho^{\otimes n} p_{\delta,n}(\rho,\sigma) p_{\delta,n}(\rho) p_{\delta,n}(\rho,\sigma)) &\geq \operatorname{tr}(\rho^{\otimes n} p_{\delta,n}(\rho)) \\ &\quad -2\sqrt{\operatorname{tr}(\rho^{\otimes n}(\mathbf{1}_{\mathcal{H}^{\otimes n}} - p_{\delta,n}(\rho,\sigma)))} \\ &> 1 - \varepsilon. \end{aligned} \tag{17}$$

On the other hand, Lemma 4.1 and 4.2 imply $((\tau_1,\tau_2) = (\rho,\sigma))$

$$\sigma^{\otimes n} p_{\delta,n}(\rho,\sigma) \leq e^{-n(M(\rho||\sigma)-\delta)} p_{\delta,n}(\rho,\sigma) \tag{18}$$

and consequently

$$p_{\delta,n}(\rho)\sigma^{\otimes n}p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho) \le e^{-n(M(\rho||\sigma)-\delta)}p_{\delta,n}(\rho)p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho). \tag{19}$$

Taking trace in (19) and observing that $\sigma^{\otimes n}$ and $p_{\delta,n}(\rho,\sigma)$ commute (cf. Lemma 4.1 with $(\tau_1,\tau_2)=(\rho,\sigma)$) we obtain

$$\begin{aligned}\mathrm{tr}(\sigma^{\otimes n}p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho)p_{\delta,n}(\rho,\sigma)) &\le e^{-n(M(\rho||\sigma)-\delta)}\mathrm{tr}(p_{\delta,n}(\rho)p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho))\\ &\le e^{-n(M(\rho||\sigma)-\delta)}\mathrm{tr}(p_{\delta,n}(\rho))\\ &\le e^{-n(M(\rho||\sigma)-\delta)}e^{n(S(\rho)+\delta)}\mathrm{tr}(\rho^{\otimes n}p_{\delta,n}(\rho))\\ &\le e^{-n(D(\rho||\sigma)-2\delta)},\end{aligned} \tag{20}$$

where in the second line we have used

$$\mathrm{tr}(p_{\delta,n}(\rho)p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho)) \le \mathrm{tr}(p_{\delta,n}(\rho)) \tag{21}$$

which follows from $p_{\delta,n}(\rho,\sigma)\le \mathbf{1}_{\mathcal{H}^{\otimes n}}$ and $(p_{\delta,n}(\rho))^2 = p_{\delta,n}(\rho)$. In the third line we have used Lemma 4.3 with $(\tau_1,\tau_2)=(\rho,\rho)$ together with $M(\rho||\rho)=S(\rho)$. The final line holds because $M(\rho||\sigma)-S(\rho)=D(\rho||\sigma)$ and $\mathrm{tr}(\rho^{\otimes n}p_{\delta,n}(\rho))\le 1$. Defining

$$a_n := p_{\delta,n}(\rho,\sigma)p_{\delta,n}(\rho)p_{\delta,n}(\rho,\sigma)\in[0,\mathbf{1}_{\mathcal{H}^{\otimes n}}] \tag{22}$$

we obtain from (17) and (20) that

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta_{\varepsilon,n}(\rho,\sigma)\le \limsup_{n\to\infty}\frac{1}{n}\log\mathrm{tr}(\sigma^{\otimes n}a_n)\le -D(\rho||\sigma)+2\delta. \tag{23}$$

Since $\delta>0$ is arbitrary we can conclude that

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta_{\varepsilon,n}(\rho,\sigma)\le -D(\rho||\sigma). \tag{24}$$

We turn now to the proof of

$$\liminf_{n\to\infty}\frac{1}{n}\log\beta_{\varepsilon,n}(\rho,\sigma)\ge -D(\rho||\sigma).$$

Let $\varepsilon\in(0,1)$ be arbitrary and let

$$q_n := \arg\min\beta_{\varepsilon,n}(\rho,\sigma). \tag{25}$$

From Lemma 4.1 and Lemma 4.3 with $(\tau_1,\tau_2)=(\rho,\sigma)$ we can infer that

$$\begin{aligned}\sigma^{\otimes n} &\ge \sigma^{\otimes n}p_{\delta,n}(\rho,\sigma)\\ &\ge e^{-n(M(\rho||\sigma)+\delta)}p_{\delta,n}(\rho,\sigma)\end{aligned} \tag{26}$$

implying

$$q_n^{1/2}\sigma^{\otimes n}q_n^{1/2}\ge e^{-n(M(\rho||\sigma)+\delta)}q_n^{1/2}p_{\delta,n}(\rho,\sigma)q_n^{1/2}, \tag{27}$$

and taking the trace

$$\begin{aligned}\beta_{\varepsilon,n}(\rho,\sigma) &= \mathrm{tr}(\sigma^{\otimes n}q_n)\quad \text{(cf. (25))}\\ &\ge e^{-n(M(\rho||\sigma)+\delta)}\mathrm{tr}(q_n p_{\delta,n}(\rho,\sigma))\\ &= e^{-n(M(\rho||\sigma)+\delta)}\mathrm{tr}(p_{\delta,n}(\rho,\sigma)q_n p_{\delta,n}(\rho,\sigma)).\end{aligned} \tag{28}$$

Here we have used the cyclicity of the trace in the first and second line and, additionally, in the last line that $(p_{\delta,n}(\rho,\sigma))^2 = p_{\delta,n}(\rho,\sigma)$ holds.
We will lower-bound the last term in (28). Recall that Lemma 4.1 and Lemma 4.2 in the case $(\tau_1,\tau_2) = (\rho,\rho)$ guarantee that

$$p_{\delta,n}(\rho)\rho^{\otimes n} = \rho^{\otimes n} p_{\delta,n}(\rho) \text{ and } \rho^{\otimes n} p_{\delta,n}(\rho) \le e^{-n(S(\rho)-\delta)} p_{\delta,n}. \tag{29}$$

Then Lemma 6.2 and eq. (25) show that there is $n_1(\varepsilon) \in \mathbb{N}$ such that

$$\begin{aligned} \operatorname{tr}(p_{\delta,n}(\rho,\sigma) q_n p_{\delta,n}(\rho,\sigma)) &\ge e^{n(S(\rho)-\delta)} (\operatorname{tr}(\rho^{\otimes n} q_n) - 2\sqrt{\operatorname{tr}(\rho^{\otimes n}(\mathbf{1}_{\mathcal{H}^{\otimes n}} - p_{\delta,n}(\rho,\sigma)))} \\ &\quad -\operatorname{tr}(\rho^{\otimes n}(\mathbf{1}_{\mathcal{H}^{\otimes n}} - p_{\delta,n}(\rho)))) \\ &\ge e^{n(S(\rho)-\delta)} \cdot \frac{1-\varepsilon}{2} \end{aligned} \tag{30}$$

for all $n \ge n_1(\varepsilon)$ by Lemma 4.4 applied to $p_{\delta,n}(\rho)$ and $p_{\delta,n}(\rho,\sigma)$.
The inequalities (28),(30), and $M(\rho||\sigma) - S(\rho) = D(\rho||\sigma)$ show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n}(\rho,\sigma) \ge -D(\rho||\sigma) - 2\delta \tag{31}$$

holds for all $\delta > 0$ and we are done.

□

3 Monotonicity and joint convexity of quantum relative entropy

In this section we will show how Stein's lemma, Theorem 1, can be used to show the monotonicity of the relative entropy under the partial trace as well as the joint convexity.
Let $\mathcal{H}_1$, $\mathcal{H}_2$ two finite-dimensional Hilbert spaces over $\mathbb{C}$. The partial trace is given by the map $\operatorname{tr}_2 : \mathcal{L}(\mathcal{H}_1) \otimes \mathcal{L}(\mathcal{H}_2) \to \mathcal{L}(\mathcal{H}_1)$, $\operatorname{tr}_2 := \operatorname{id}_{\mathcal{L}(\mathcal{H}_1)} \otimes \operatorname{tr}_{\mathcal{L}(\mathcal{H}_2)}$, where $\operatorname{tr}_{\mathcal{L}(\mathcal{H}_2)}$ denotes the trace on $\mathcal{L}(\mathcal{H}_2)$. Notice that for any state $\tau \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and any $a \in \mathcal{L}(\mathcal{H}_1)$

$$\operatorname{tr}_{\mathcal{L}(\mathcal{H}_1)}(\operatorname{tr}_2(\tau)a) = \operatorname{tr}_{\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2)}(\tau(a \otimes \mathbf{1}_{\mathcal{H}_2})) \tag{32}$$

holds. Introducing the map $E : \mathcal{L}(\mathcal{H}_1) \to \mathcal{L}(\mathcal{H}_1) \otimes \mathcal{L}(\mathcal{H}_2)$, $E(a) := a \otimes \mathbf{1}_{\mathcal{H}_2}$ this can be written compactly as

$$\operatorname{tr}_{\mathcal{L}(\mathcal{H}_1)}(\operatorname{tr}_2(\tau)a) = \operatorname{tr}_{\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2)}(\tau E(a)). \tag{33}$$

Notice that the map E has the following properties which are readily checked: For all $a, b \in \mathcal{L}(\mathcal{H}_1)$ we have $E(\mathbf{1}_{\mathcal{H}_1}) = \mathbf{1}_{\mathcal{H}_1\otimes\mathcal{H}_2}$, $E(a^*) = E(a)^*$, and $E(ab) = E(a)E(b)$ and this already implies that E is positive, i.e. preserves the positive semi-definiteness of operators. However, the latter property is also obvious from the definition of E.
For $n \in \mathbb{N}$ and all $\tau \in \mathcal{S}((\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes n})$, $a \in \mathcal{L}(\mathcal{H}_1)^{\otimes n}$ we have

$$\operatorname{tr}_{\mathcal{L}(\mathcal{H}_1)^{\otimes n}}(\operatorname{tr}_2^{\otimes n}(\tau)a) = \operatorname{tr}_{(\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2))^{\otimes n}}(\tau E^{\otimes n}(a)). \tag{34}$$

$E^{\otimes n}$ inherits the following properties: For all $a, b \in \mathcal{L}(\mathcal{H}_1)^{\otimes n}$

$$E^{\otimes n}(\mathbf{1}_{\mathcal{H}_1^{\otimes n}}) = \mathbf{1}_{(\mathcal{H}_1\otimes\mathcal{H}_2)^{\otimes n}}, \quad E^{\otimes n}(a^*) = (E^{\otimes n}(a))^*, \tag{35}$$

and

$$E^{\otimes n}(ab) = E^{\otimes n}(a)E^{\otimes n}(b), \tag{36}$$

implying that $E^{\otimes n}$ is positive too[3]. From this we see that

$$P_n := E^{\otimes n}([0, \mathbf{1}_{\mathcal{H}_1^{\otimes n}}]) \subset [0, \mathbf{1}_{(\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes n}}], \tag{37}$$

where $[0, \mathbf{1}_{\mathcal{H}_1^{\otimes n}}]$ is the set of all (self-adjoint) operators $a \in \mathcal{L}(\mathcal{H}_1)^{\otimes n}$ with $0 \leq a \leq \mathbf{1}_{\mathcal{H}_1^{\otimes n}}$ and $[0, \mathbf{1}_{(\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes n}}]$ is defined correspondingly.

Theorem 7 (Monotonicity under partial trace [15]): Let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ be states. Then

$$D(\rho||\sigma) \geq D(\mathrm{tr}_2(\rho)||\mathrm{tr}_2(\sigma)). \tag{38}$$

Proof: In the case that $\ker \sigma \not\subseteq \ker \rho$ we have $D(\rho||\sigma) = +\infty$ and the inequality (38) is trivially true.

Suppose that $\ker \sigma \subseteq \ker \rho$. We will use the abbreviation tr for $\mathrm{tr}_{\mathcal{L}(\mathcal{H}_1) \otimes \mathcal{L}(\mathcal{H}_2)}$ and $\mathrm{tr}_{\mathcal{L}(\mathcal{H}_1)}$ as well as for tensored versions thereof in what follows. It will always be clear from the context on which space the trace is acting. Moreover we set

$$\rho_1 := \mathrm{tr}_2(\rho), \quad \sigma_1 := \mathrm{tr}_2(\sigma).$$

Then for any $\varepsilon \in (0,1)$ and all $n \in \mathbb{N}$ we have

$$\begin{aligned} \beta_{\varepsilon,n}(\rho_1, \sigma_2) &= \min\{\mathrm{tr}(\sigma_1^{\otimes n} a) : a \in [0, \mathbf{1}_{\mathcal{H}_1^{\otimes n}}], \mathrm{tr}(\rho_1^{\otimes n} a) \geq 1 - \varepsilon\} \\ &= \min\{\mathrm{tr}(\sigma^{\otimes n} E^{\otimes n}(a)) : a \in [0, \mathbf{1}_{\mathcal{H}_1^{\otimes n}}], \mathrm{tr}(\rho^{\otimes n} E^{\otimes n}(a)) \geq 1 - \varepsilon\} \\ &= \min\{\mathrm{tr}(\sigma^{\otimes n} a) : a \in P_n : \mathrm{tr}(\rho^{\otimes n} a) \geq 1 - \varepsilon\} \\ &\geq \min\{\mathrm{tr}(\sigma^{\otimes n} a) : a \in [0, \mathbf{1}_{(\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes n}}], \mathrm{tr}(\rho^{\otimes n} a) \geq 1 - \varepsilon\} \\ &= \beta_{\varepsilon,n}(\rho, \sigma), \end{aligned} \tag{39}$$

where in the second line we have used the relation (34), in the third line we used (37), while in the fourth line we used the elementary fact that the minimum value of a given function decreases if we enlarge the set we are minimizing over. Taking log of both sides of (39), dividing by n, and taking the limit shows, according to Theorem 1, that

$$D(\rho_1||\sigma_2) \leq D(\rho||\sigma)$$

as desired.

□

Remark 8: The inequality (39) has a nice intuitive interpretation: The minimum probability of error can only increase if we have access to a smaller set of measurements upon which we can base our decisions.

[3] The properties in (35) and (36) are obvious on the set $G(n, \mathcal{H}_1) := \{\otimes_{i=1}^n a_i : a_i \in \mathcal{L}(\mathcal{H}_1), i = 1, \ldots, n\}$ and extend by linearity to the whole $\mathcal{L}(\mathcal{H}_1)^{\otimes n}$ since $G(n, \mathcal{H}_1)$ is a generating set for $\mathcal{L}(\mathcal{H}_1)^{\otimes n}$. The positivity follows from $E^{\otimes n}(b^* b) = (E^{\otimes n}(b))^* E^{\otimes n}(b)$ and the fact that every positive semi-definite $a \in \mathcal{L}(\mathcal{H}_1)^{\otimes n}$ can be written as $a = b^* b$ for suitable $b \in \mathcal{L}(\mathcal{H}_1)^{\otimes n}$

Remark 9: Notice that for the proof of Theorem 7 we do not need the full power of Theorem 1. It is sufficient to know that there is an increasing subsequence $(n_k)_{k\in\mathbb{N}}$ of non-negative integers and a decreasing sequence $(\varepsilon_{n_k})_{k\in\mathbb{N}}$ with $\varepsilon_{n_k} \in (0,1)$ and $\lim_{k\to\infty} \varepsilon_{n_k} = 0$ such that

$$\lim_{k\to\infty} \frac{1}{n_k} \log \beta_{\varepsilon_{n_k},n_k}(\rho,\sigma) = -D(\rho,\sigma)$$

holds.

Remark 10: Theorem 7 has a natural generalization to completely positive trace-preserving maps due to Lindblad [17]: For any such map $T : \mathcal{L}(\mathcal{H}_1) \to \mathcal{L}(\mathcal{H}_2)$ and $\rho, \sigma \in \mathcal{S}(\mathcal{H}_1)$ we have

$$D(\rho||\sigma) \geq D(T(\rho)||T(\sigma)). \tag{40}$$

The following generalization of the proof of Theorem 7 to the situation of inequality (40) was suggested to us by Janis Nötzel. We just have to replace the maps tr_2 and E by T and its dual $T_* : \mathcal{L}(\mathcal{H}_2) \to \mathcal{L}(\mathcal{H}_1)$ defined via

$$\mathrm{tr}(aT(a')) = \mathrm{tr}(T_*(a)a') \tag{41}$$

for $a \in \mathcal{L}(\mathcal{H}_2), a' \in \mathcal{L}(\mathcal{H}_1)$. The map T_* is completely positive and unital, the latter meaning that $T_*(\mathbf{1}_{\mathcal{H}_2}) = \mathbf{1}_{\mathcal{H}_1}$, implying that for the set

$$P_n := T_*^{\otimes n}([0, \mathbf{1}_{\mathcal{H}_2^{\otimes n}}]) \tag{42}$$

we have

$$P_n \subseteq [0, \mathbf{1}_{\mathcal{H}_1^{\otimes n}}]. \tag{43}$$

Then a similar reasoning as in the inequality chain (39) leads to

$$\beta_{\varepsilon,n}(T(\rho), T(\sigma)) \geq \beta_{\varepsilon,n}(\rho,\sigma) \tag{44}$$

for all $n \in \mathbb{N}$ and all $\varepsilon \in (0,1)$ and we obtain (40) via Stein's lemma.

Our next step is to show that the monotonicity of the relative entropy under partial trace, Theorem 7, implies the joint convexity of the relative entropy. To this end we extend slightly the definition of the relative entropy to positive semi-definite operators and show that Theorem 7 carries over to this generalized situation. For any pair of positive semi-definite operators $a, b \in \mathcal{L}(\mathcal{H})$ we set

$$D(a||b) := \begin{cases} \mathrm{tr}(a \log a - a \log b) & \text{if } \ker b \subseteq \ker a \\ +\infty & \text{else.} \end{cases} \tag{45}$$

Corollary 11 (Monotonicity under partial trace II): For any pair $a, b \in \mathcal{L}(\mathcal{H}_1) \otimes \mathcal{L}(\mathcal{H}_2)$ of positive semi-definite operators we have

$$D(a||b) \geq D(\mathrm{tr}_2(a)||\mathrm{tr}_2(b)). \tag{46}$$

Proof: First note that in the case $\ker b \not\subseteq \ker a$ there is nothing to prove since $D(a||b) = +\infty$.

Let us suppose that $\ker b \subseteq \ker a$ holds. If $a = 0$ there is again nothing to prove because

$$D(a||b) = 0 = D(\mathrm{tr}_2(a)||\mathrm{tr}_2(b)). \tag{47}$$

So, we can assume that $a \neq 0$ and consequently $b \neq 0$. Some simple algebra shows that for $\alpha, \beta \in \mathbb{R}$, $\alpha, \beta > 0$ it holds that

$$D(\alpha \cdot a||\beta \cdot b) = \alpha D(a||b) + \mathrm{tr}(a)\alpha \log \frac{\alpha}{\beta}. \tag{48}$$

On the other hand, it follows from the definition of the partial trace that

$$\mathrm{tr}_{\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2)}(a) = \mathrm{tr}_{\mathcal{L}(\mathcal{H}_1)}(\mathrm{tr}_2(a)) \tag{49}$$

for all positive semi-definite $a \in \mathcal{L}(\mathcal{H}_1) \otimes \mathcal{L}(\mathcal{H}_2)$.
Now we set $\alpha := \frac{1}{\mathrm{tr}_{\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2)}(a)}, \beta := \frac{1}{\mathrm{tr}_{\mathcal{L}(\mathcal{H}_1)\otimes\mathcal{L}(\mathcal{H}_2)}(b)}$ and obtain from Theorem 7

$$D(\alpha \cdot a||\beta \cdot b) \geq D(\alpha \cdot \mathrm{tr}_2(a)||\beta \cdot \mathrm{tr}_2(b)). \tag{50}$$

Taking into account (48) and (49) leads to

$$D(a||b) \geq D(\mathrm{tr}_2(a)||\mathrm{tr}_2(b)) \tag{51}$$

and we are done.

□

Corollary 12 (Joint convexity [16]): Let $\mathcal{H}$ be a finite-dimensional Hilbert space over $\mathbb{C}$, $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathcal{L}(\mathcal{H})$ positive semi-definite operators, and $\lambda_1, \ldots, \lambda_k \in \mathbb{R}_+$ with $\sum_{i=1}^k \lambda_i = 1$. Then

$$D\left(\sum_{i=1}^k \lambda_i a_i || \sum_{i=1}^k \lambda_i b_i\right) \leq \sum_{i=1}^k \lambda_i D(a_i||b_i). \tag{52}$$

Proof: If for some $i \in \{1, \ldots, k\}$ $\ker b_i \nsubseteq \ker a_i$ then the right hand side of (52) equals $+\infty$ and (52) holds.
We may, therefore, suppose that for all $i \in \{1, \ldots, k\}$ $\ker b_i \subseteq \ker a_i$ and define positive semi-definite operators $a', b' \in \mathcal{L}(\mathcal{H}) \otimes \mathcal{L}(\mathbb{C}^k)$ by

$$a' := \sum_{i=1}^k \lambda_i a_i \otimes |e_i\rangle\langle e_i|, \quad b' := \sum_{i=1}^k \lambda_i b_i \otimes |e_i\rangle\langle e_i|,$$

where $\{e_1, \ldots, e_k\}$ is an orthonormal basis of $\mathbb{C}^k$. Note that $\ker b' \subseteq \ker a'$ and it is not hard to see that

$$D(a'||b') = \sum_{i=1}^k \lambda_i D(a_i||b_i) \tag{53}$$

holds. Let $\mathrm{tr}_2 : \mathcal{L}(\mathcal{H}) \otimes \mathcal{L}(\mathbb{C}^k) \to \mathcal{L}(\mathcal{H})$ denote the partial trace. Then

$$\mathrm{tr}_2(a') = \sum_{i=1}^k \lambda_i a_i, \quad \mathrm{tr}_2(b') = \sum_{i=1}^k \lambda_i b_i, \tag{54}$$

and Corollary 11 shows that

$$D(a'||b') \geq D(\mathrm{tr}_2(a')||\mathrm{tr}_2(b')), \tag{55}$$

which is nothing else than (52) by (53) and (54).

□

4 Lieb's concavity theorem: Tropp's argument

In this section we shall outline Tropp's argument [25] leading to the following theorem of Lieb [14, Theorem 6]:

Theorem 13: Let $\mathcal{H}$ be a finite dimensional Hilbert-space over $\mathbb{C}$ and let $h \in \mathcal{L}(\mathcal{H})$ be self-adjoint. Then the map $a \mapsto \operatorname{tr}\exp(h + \log a)$ is concave on the positive-definite cone of $\mathcal{L}(\mathcal{H})$.

Tropp's proof of Theorem 13 is based on a sequence of lemmas which we will present first.

Lemma 14: 1. (Klein's Inequality [13]) Let $a, b \in \mathcal{L}(\mathcal{H})$ be positive semi-definite operators. Then

$$D(a||b) \geq \operatorname{tr}(a - b). \tag{56}$$

2. (Variational Formula for Trace) For any positive-definite $b \in \mathcal{L}(\mathcal{H})$ we have

$$\operatorname{tr}(b) = \max_{x \in \mathcal{L}(\mathcal{H}), x \geq 0} \operatorname{tr}(x \log b - x \log x + x). \tag{57}$$

Proof: 1. We may suppose that $\ker b \subseteq \ker a$ since otherwise the inequality (56) is clearly true. Moreover we can assume that $a \neq 0$ because (56) is trivially satisfied in the case $a = 0$. Since $a \neq 0$ implies $b \neq 0$ we see that $\operatorname{tr}(a), \operatorname{tr}(b) > 0$ and Corollary 11 applied with $\mathcal{H}_1 := \mathbb{C}$, $\mathcal{H}_2 = \mathcal{H}$ shows that

$$D(a||b) \geq D(\operatorname{tr}(a)||\operatorname{tr}(b)) = \operatorname{tr}(a) \log \frac{\operatorname{tr}(a)}{\operatorname{tr}(b)} \geq \operatorname{tr}(a) - \operatorname{tr}(b), \tag{58}$$

where the last inequality follows from the numerical inequality $\log x \geq -\frac{1}{x} + 1$ valid for all positive numbers x.
2. Note that the inequality $\operatorname{tr}(b) \geq \operatorname{tr}(x \log b - x \log x + x)$ is nothing else than (56) and equality holds for $x = b$.

□

The final lemma we need for the proof of Theorem 13 is Lemma 2.3 from [5]. We omit the elementary proof.

Lemma 15: Let $f : K_1 \times K_2 \to \mathbb{R}$ be a jointly concave function such that for each $y \in K_2$ there is $x' \in K_1$ such that

$$f(x', y) = \sup_{x \in K_1} f(x, y),$$

i.e. sup is attained for each y and is in fact max. Then the function $y \mapsto \max_{x \in K_1} f(x, y)$ is concave.

Proof of Theorem 13: We apply Lemma 14.2 with $b := \exp(h + \log a)$ and end up with

$$\begin{aligned} \operatorname{tr}(\exp(h + \log a)) &= \max_{x \geq 0} \operatorname{tr}(x(h + \log a) - x \log x + x) \\ &= \max_{x \geq 0} (\operatorname{tr}(xh) - D(x||a) + \operatorname{tr}(x)). \end{aligned} \tag{59}$$

The right hand side of (59) is concave in a by Lemma 15 due to the fact that the map $(x, a) \mapsto \operatorname{tr}(xh) - D(x||a) + \operatorname{tr}(x)$ jointly concave for fixed h by Corollary 11.

□

Remark 16: It is an interesting aside to have a look at Lindblad's proof [16] of the joint convexity of the relative entropy (which Tropp [25] cites) and Tropp's argument as a whole.
Lindblad's starting point is a special case of [14, Theorem 1] stating that on pairs of positive-semidefinite operators the map $(a,b) \mapsto \mathrm{tr}(a^{1-p}b^p)$ is jointly concave for any $p \in [0,1]$. He then observes that the derivative of that map with respect to p at $p=0$ is $-D(a||b)$ from which the joint convexity of the relative entropy follows. Tropp shows how to derive the concavity of $a \mapsto \mathrm{tr}\exp(h+\log a)$ (h self-adjoint) on the positive cone from the joint convexity of the relative entropy. Thus, when seen in a sequence, the arguments of Lindblad and Tropp show in few lines that the joint concavity of $(a,b) \mapsto \mathrm{tr}(a^{1-p}b^p)$ ($p \in [0,1]$, a,b positive-semidefinite) implies the concavity of $a \mapsto \mathrm{tr}\exp(h+\log a)$ (h self-adjoint, a positive-definite).
In a similar vein, following the proofs of Corollaries 2.1 or 2.1 in Effros' paper [8] we can see that Theorem 13 can be easily deduced from the operator convexity of $f(x) = x\log x$ or $g(x) = -x^p$, for $p \in [0,1]$.

5 Historical remarks and related work

Stein's lemma in the classical form appears for the first time in Chernoff's (!) work [6]. In the quantum realm, Hiai and Petz [11] have shown that for all $\varepsilon \in (0,1)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n}(\rho,\sigma) \leq -D(\rho||\sigma),$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n}(\rho,\sigma) \geq -\frac{D(\rho||\sigma)}{1-\varepsilon} \tag{60}$$

hold. The proof of the inequality (60) presented in [11] relies on the monotonicity of the quantum relative entropy. Ogawa and Nagaoka [18] obtained the strong converse, i.e. they showed that $1-\varepsilon$ on the right hand side of (60) can be replaced by 1 thus leading to

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n}(\rho,\sigma) = -D(\rho||\sigma)$$

for all $\varepsilon \in (0,1)$. The proof in [18] relies on the monotonicity of quantum quasi-entropies which present a generalization of relative entropy [19], [20].
The monotonicity of the relative entropy under the partial trace (MPT) has been shown by Lieb and Ruskai in [15]. Their proof relies on the concavity of $a \mapsto \mathrm{tr}(\exp(h+\log a))$ (h self-adjoint) on the positive cone of $\mathcal{L}(\mathcal{H})$ which was proven by Lieb in [14]. Uhlmann [26] observed that this implies so called strong subadditivity of von Neumann entropy. The paper [15] by Lieb and Ruskai, in turn, contains an argument showing that the strong subadditivity of von Neumann entropy implies MPT.
Joint convexity of the relative entropy was established by Lindblad [16]. Lindblad's proof uses another theorem of Lieb [14] which states that the map $(a,b) \mapsto \mathrm{tr}(a^{1-p}b^p)$, $p \in [0,1]$, is jointly concave in (a,b) for positive semi-definite $a,b \in \mathcal{L}(\mathcal{H})$. The monotonicity of the quantum relative entropy under the action of completely positive trace-preserving maps was established by Lindblad

in [17]. Uhlmann [27] derives the monotonicity as well as the joint convexity of the relative entropy in the general setting of operator algebras via interpolation theory. An ingenious analytic proof of joint convexity of the relative entropy is discovered by Simon [23, Ch. 8].
In [8] Effros gives very short and elegant proofs of joint convexity of relative entropy and several results of Lieb from [14] based on the notion of operator convex functions and Jensen's inequality for operators proven by Hansen and Pedersen [9].
More historical facts of interest as well as other analytic approaches to the properties of quantum relative entropy and interrelation among the entropy inequalities can be picked up in the nice review [21] by Ruskai.
Finally, note that the ansatz to derive inequalities for matrices or entropy from operational, information theoretic, or probabilistic interpretation of the quantities in question is not new at all. Already Dembo, Cover, and Thomas in [7] derived several matrix inequalities from the properties of multivariate gaussian distributions.
Much closer in spirit to our work is Winter's [28] derivation of the famous Holevo bound [12] from the coding theorem with the strong converse for channels with classical input and quantum mechanical output.

Acknowledgement We thank Holger Boche, Gisbert Janßen, Janis Nötzel, and Moritz Wiese for encouragement to publish this work and many discussions, comments, and suggestions during the process of the preparation of the manuscript. We thank especially Janis Nötzel for sharing with us his proof of the monotonicity under completely positive trace-preserving maps reproduced in Remark 10.
While preparing the first version [3] of the present paper we had several discussions with Tyll Krüger, Ruedi Seiler, Arleta Szkoła, and Andreas Winter and we thank them for that. Mary Beth Ruskai contributed several remarks and corrections concerning the history of and interrelation between the entropy inequalities.
The support by the Deutsche Forschungsgemeinschaft (DFG) via grant BO 1734/20-1 and Bundesministerium für Bildung und Forschung (BMBF) via grant 01BQ1050 is gratefully acknowledged.

A Proof of Lemma 6

The proof relies on the following simple fact: For all $a, b \in \mathcal{L}(\mathcal{K})$ we have

$$|\mathrm{tr}(a^*b)| = |\mathrm{tr}(b^*a)| \leq \sqrt{\mathrm{tr}(a^*a)}\sqrt{\mathrm{tr}(b^*b)}, \tag{61}$$

where the equality stems from the fact that $\mathrm{tr}(x^*) = \overline{\mathrm{tr}(x)}$ and the inequality is nothing else than the Cauchy-Schwarz inequality for the Hilbert Schmidt inner product $(a, b) \mapsto \langle a, b\rangle_{HS} := \mathrm{tr}(a^*b)$ on the space $\mathcal{L}(\mathcal{K})$.
In what follows we will write $\mathbf{1}$ for $\mathbf{1}_{\mathcal{K}}$. For p, q as given in the statement of the lemma we have

$$0 \leq (\mathbf{1} - p)q(\mathbf{1} - p) = -q + q(\mathbf{1} - p) + (\mathbf{1} - p)q + pqp, \tag{62}$$

which is readily verified. Multiplying (62) from left and from right by $\tau^{1/2}$, taking trace of both sides, and rearranging leads to

$$\begin{aligned}\mathrm{tr}(\tau q) &\le \mathrm{tr}(\tau^{1/2}q(\mathbf{1}-p)\tau^{1/2}) + \mathrm{tr}(\tau^{1/2}(\mathbf{1}-p)q\tau^{1/2}) + \mathrm{tr}(\tau pqp)\\ &= |\mathrm{tr}(\tau^{1/2}q(\mathbf{1}-p)\tau^{1/2})| + |\mathrm{tr}(\tau^{1/2}(\mathbf{1}-p)q\tau^{1/2})| + \mathrm{tr}(\tau pqp)\\ &= 2|\mathrm{tr}(\tau^{1/2}(\mathbf{1}-p)q\tau^{1/2})| + \mathrm{tr}(\tau pqp)\\ &\le 2\sqrt{\mathrm{tr}(\tau(\mathbf{1}-p)^2)}\sqrt{\mathrm{tr}(\tau q^2)} + \mathrm{tr}(\tau pqp)\\ &\le 2\sqrt{\mathrm{tr}(\tau(\mathbf{1}-p))} + \mathrm{tr}(\tau pqp), \end{aligned} \tag{63}$$

where in the third line we have used the left relation in (61), in the fourth we used the Cauchy-Schwarz inequality (right half of (61)) with $a^* = \tau^{1/2}(\mathbf{1}-p), b = q\tau^{1/2}$, while in the last line we estimated $(\mathbf{1}-p)^2 \le \mathbf{1}-p$ since $0 \le \mathbf{1}-p \le \mathbf{1}$, and $\mathrm{tr}(\tau q^2) \le \mathrm{tr}(\tau q) \le 1$.
This shows

$$\mathrm{tr}(\tau pqp) \ge \mathrm{tr}(\tau q) - 2\sqrt{\mathrm{tr}(\tau(\mathbf{1}-p))}. \tag{64}$$

For the second part of the lemma we simply observe that

$$\begin{aligned}\mathrm{tr}(pqp) &= \mathrm{tr}(upqp) + \mathrm{tr}((\mathbf{1}-u)pqp)\\ &\ge \mathrm{tr}(upqp) \qquad (\text{since } \mathrm{tr}((\mathbf{1}-u)pqp) \ge 0)\\ &\ge \frac{1}{c}\mathrm{tr}(\tau upqp) \qquad (\text{since } \tau u \le cu)\\ &= \frac{1}{c}(\mathrm{tr}(\tau pqp) - \mathrm{tr}(\tau(\mathbf{1}-u)(pqp)))\\ &\ge \frac{1}{c}(\mathrm{tr}(\tau pqp) - \mathrm{tr}(\tau(\mathbf{1}-u))), \end{aligned} \tag{65}$$

in the last line we have used the fact that $pqp \le \mathbf{1}$. Now, a combination of (65) and (64) finishes the proof of the second part of the lemma.

B Singular case of Stein's lemma

In this appendix we provide the variant of Theorem 1 for the case $\ker\sigma \not\subseteq \ker\rho$.

Lemma 17: Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ with $\ker\sigma \not\subseteq \ker\rho$. Then to each $\varepsilon \in (0,1)$ there is $n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}, n \ge n_0(\varepsilon)$

$$\beta_{\varepsilon,n}(\rho,\sigma) = 0. \tag{66}$$

Proof: Since $\ker\sigma \not\subseteq \ker\rho$ there is $e \in \ker\sigma$, $e \notin \ker\rho$ with $||e|| = 1$. Then, clearly,

$$\mathrm{tr}(\sigma|e\rangle\langle e|) = 0, \qquad \mathrm{tr}(\rho|e\rangle\langle e|) = a \in (0,1]. \tag{67}$$

We set

$$p_0 := |e\rangle\langle e|, \qquad p_1 := \mathbf{1}_{\mathcal{H}} - p_0. \tag{68}$$

For $n \in \mathbb{N}$ we introduce

$$T_n := \{x^n \in \{0,1\}^n : x^n \text{ contains at least one } 0\}, \tag{69}$$

and the projection

$$q_n := \sum_{x^n \in T_n} p_{x^n} \tag{70}$$

where $p_{x^n} := \otimes_{i=1}^n p_{x_i}$. Notice that

$$q_n + p_1^{\otimes n} = \mathbf{1}_{\mathcal{H}^{\otimes n}} = \mathbf{1}_{\mathcal{H}}^{\otimes n} \tag{71}$$

which leads to

$$\mathrm{tr}(\rho^{\otimes n} q_n) = \mathrm{tr}(\rho^{\otimes n}(\mathbf{1}_{\mathcal{H}}^{\otimes n} - p_1^{\otimes n})) = 1 - (1-a)^n \tag{72}$$

by (67). On the other hand it is clear by the definition of q_n and T_n and by (67) that

$$\mathrm{tr}(\sigma^{\otimes n} q_n) = 0. \tag{73}$$

Consequently, by (72) and (73) for any $\varepsilon \in (0,1)$ there is $n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}, n \geq n_0(\varepsilon)$

$$\beta_{\varepsilon,n}(\rho,\sigma) \leq \mathrm{tr}(\sigma^{\otimes n} q_n) = 0 \tag{74}$$

□

References

[1] R. Ahlswede, A. Winter: Strong converse for identification via quantum channels, *IEEE Trans. Inf. Th.* Vol. 48, No.3, 569-579 (2002)

[2] R. Bhatia: Positive Definite Matrices, *Princeton Series in Applied Mathematics*, Princeton University Press, Princeton, New Jersey (2007)

[3] I. Bjelaković, Ra. Siegmund-Schultze: A new proof of the monotonicity of quantum relative entropy for finite dimensional systems, preprint available at: http://arxiv.org/abs/quant-ph/0307170

[4] I. Bjelaković, J.-D. Deuschel. T. Krüger, R. Seiler, Ra. Siegmund-Schultze, A. Szkoła: Typical support and Sanov large deviations of correlated states, *Commun. Math. Phys.* 279, 559-584 (2008)

[5] 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, 107-126 (2008)

[6] H. Chernoff: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Ststist.* Vol. 23, No. 4, 493-507 (1952)

[7] A. Dembo, T.M. Cover, J.A. Thomas: Information theoretic inequalities, *IEEE Trans. Inf. Th.* Vol. 37, No. 6, 1501-1518, (1991)

[8] E.G. Effros: A matrix convexity approach to some celebrated quantum inequalities, *Proc. Nat. Acad. Sci. USA*, Vol. 106, No. 4,1006-1008 (2009)

[9] F. Hansen, G. Pedersen: Jensen's inequality for operators and Löwner's theorem, *Math. Ann.* 258, 229-241 (1982)

[10] M. Hayashi, H. Nagaoka: General formulas for capacity of classical-quantum channels, *IEEE Trans. Inf. Th.* Vol. 49, No. 7, 1753-1768 (2003)

[11] F. Hiai, D. Petz: The proper formula for relative entropy and its asymptotics in quantum probability, *Commun. Math. Phys.* 143, 99-114 (1991)

[12] A.S. Holevo: Bounds for the quantity of information transmitted by a quantum channel, *Probl. Inform. Transm.*, Vol. 9, No. 3, 177–183 (1973)

[13] O. Klein: Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre, *Z. Phys.* Vol. 72, No. 11-12, 767-775 (1931)

[14] E.H. Lieb: Convex trace functions and the Wigner-Yanase-Dyson conjecture, *Adv. Math.* 11, 267-288 (1973)

[15] E.H. Lieb, M.B. Ruskai: Proof of the strong subadditivity of quantum-mechanical entropy, *J. Math. Phys.* Vol. 14, No. 12, 1938-1941 (1973)

[16] G. Lindblad: Expectations and entropy inequalities for finite quantum systems, *Commun. Math. Phys.* 39, 111-119 (1974)

[17] G. Lindblad: Completely positive maps and entropy inequalities, *Commun. Math. Phys.* 40, 147-151 (1975)

[18] T. Ogawa, H. Nagaoka: Strong converse and Stein's lemma in quantum hypothesis testing, *IEEE Trans. Inf. Th.*, Vol. 46, No. 7, 2428-2433 (2000)

[19] D. Petz: Quasi-entropies for states of a von Neumann algebra, *Publ. RIMS, Kyoto Univ.* 21, 787-800 (1985)

[20] D. Petz: Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* 23, 57-65 (1986)

[21] M.B. Ruskai: Inequalitues for quantum entropy: A review with conditions for equality, *J. Math. Phys.* Vol.43, No. 9, 4358-4375 (2002)

[22] C.E. Shannon: A mathematical theory of communication, *Bell Sys. Tech. J.* 27, 379-425 & 623-656 July & October (1948)

[23] B. Simon: Trace Ideals and Their Applications, *London Mathematical Society Lecture Notes Series 35*, Cambridge University Press, Cambridge (1979)

[24] J.A. Tropp: User-friendly tail bounds for sums of random matrices, *Found. Comput. Math* , DOI 10.1007/s10208-011-9099-z, arXiv:1004.4389v7 (2011)

[25] J.A. Tropp: From joint convexity of quantum relative entropy to a concavity theorem of Lieb, *Proc. of the AMS* 140, 1757-1760 (2012)

[26] A. Uhlmann: Endlich-dimensionale Dichtematrizen II, *Wiss. Z. Karl-Marx-Universität Leipzig, Math.-Naturwiss. R.* 22, H. 2, 139-180 (1973)

[27] A. Uhlmann: Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, *Commun. Math. Phys.* 54, 21-32 (1977)

[28] A. Winter: Coding theorem and strong converse for quantum channels, *IEEE Trans. Inf. Th.* Vol. 45, No. 7, 2481-2485 (1999)