



# Multivariate Trace Inequalities

David Sutter<sup>1</sup> , Mario Berta<sup>2</sup>, Marco Tomamichel<sup>3</sup>

<sup>1</sup> Institute for Theoretical Physics, ETH Zurich, Zurich, Switzerland.

E-mail: [sutterd@itp.phys.ethz.ch](mailto:sutterd@itp.phys.ethz.ch)

<sup>2</sup> Institute for Quantum Information and Matter, Caltech, California, USA

<sup>3</sup> School of Physics, The University of Sydney, Sydney, Australia

Received: 30 April 2016 / Accepted: 11 August 2016

Published online: 18 October 2016 – © The Author(s) 2016. This article is published with open access at [Springerlink.com](http://Springerlink.com)

**Abstract:** We prove several trace inequalities that extend the Golden–Thompson and the Araki–Lieb–Thirring inequality to arbitrarily many matrices. In particular, we strengthen Lieb’s triple matrix inequality. As an example application of our four matrix extension of the Golden–Thompson inequality, we prove remainder terms for the monotonicity of the quantum relative entropy and strong sub-additivity of the von Neumann entropy in terms of recoverability. We find the first explicit remainder terms that are tight in the commutative case. Our proofs rely on complex interpolation theory as well as asymptotic spectral pinching, providing a transparent approach to treat generic multivariate trace inequalities.

## 1. Introduction

Trace inequalities are mathematical relations between different multivariate trace functionals. Often these relations are straightforward equalities if the involved matrices commute—and can be difficult to prove for the non-commuting case.

Arguably one of the most powerful trace inequalities is the celebrated *Golden–Thompson (GT) inequality* [23,64]. It states that for any two Hermitian matrices  $H_1$  and  $H_2$  we have

$$\mathrm{tr} \exp(H_1 + H_2) \leq \mathrm{tr} \exp(H_1) \exp(H_2). \quad (1)$$

We note that in case  $H_1$  and  $H_2$  commute (1) holds with equality. Inequality (1) has been generalized in various directions (see, e.g., [3,12,32,38,39,42,57,59]). For example, it has been shown that it remains valid by replacing the trace with any unitarily invariant norm [41,58,65] and an extension to three non-commuting matrices was suggested in [44].

The GT inequality has found applications ranging from statistical physics [64] and random matrix theory [1,67] to quantum information theory [45,46].

Straightforward extensions of this inequality to three matrices are incorrect, as discussed in Appendix A. In this work, for any  $n \in \mathbb{N}$ , Hermitian matrices  $\{H_k\}_{k=1}^n$  and any  $p \geq 1$ , we show that

$$\log \left\| \exp \left( \sum_{k=1}^n H_k \right) \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_0(t) \log \left\| \prod_{k=1}^n \exp((1+it)H_k) \right\|_p, \quad (2)$$

where  $\|\cdot\|_p$  denotes the Schatten  $p$ -norm and  $\beta_0$  is a fixed probability distribution on  $\mathbb{R}$ , independent of the other parameters. An extensive discussion of this result is provided in Sect. 3.2. The precise statement is given in Corollary 3.3. Note that the expression  $\exp((1+it)H_k)$  decomposes as  $\exp(H_k) \exp(itH_k)$  where the latter is a unitary rotation. Since the Schatten  $p$ -norm is unitarily invariant it follows that the integrand in (2) is independent of  $t$  for  $n = 2$ . Inequality (2) thus constitutes an  $n$ -matrix extension of the GT inequality and further simplifies to (1) for  $n = 2$  and  $p = 2$ . For  $n = 3$  and  $p = 2$  our result strengthens *Lieb's triple matrix inequality* [44], as shown in Lemma 3.4.

The GT inequality can be seen as a limiting case of the more general *Araki–Lieb–Thirring (ALT) inequality* [4,43]. The latter states that, for any positive semi-definite matrices  $A_1$  and  $A_2$ , and  $q > 0$ ,

$$\mathrm{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \leq \mathrm{tr} \left( A_1^{\frac{1}{2}} A_2 A_1^{\frac{1}{2}} \right)^q \quad \text{if } r \in (0, 1]. \quad (3)$$

The inequality holds in the opposite direction for  $r \geq 1$  by an appropriate substitution.<sup>1</sup> The GT inequality for Schatten  $p$ -norms is implied by the Lie–Trotter product formula in the limit  $r \rightarrow 0$ . The ALT inequality has also been extended in various directions (see, e.g., [2,39,70]).

In this work, we provide an  $n$ -matrix extension of the ALT inequality. For any  $n \in \mathbb{N}$ , positive semi-definite matrices  $\{A_k\}_{k=1}^n$  and any  $p \geq 1$ , we show that

$$\log \left\| \prod_{k=1}^n A_k^{\frac{1}{r}} \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_r(t) \log \left\| \prod_{k=1}^n A_k^{1+it} \right\|_p \quad \text{if } r \in (0, 1], \quad (4)$$

where  $\beta_r$  are a family of probability distributions on  $\mathbb{R}$ , independent of the other parameters. In this article we use the convention that  $0^z = 0$  for any  $z \in \mathbb{C}$ . We refer to Theorem 3.2 for a precise statement and discussion. Our extension of the GT inequality again follows in the limit  $r \rightarrow 0$  by the Lie–Trotter product formula. We also provide an extension of the ALT and GT inequality for general square matrices (see Theorem 3.5).

We apply our results to quantum information theory and show how it can be used to prove strong sub-additivity. This yields remainder terms on the monotonicity of relative entropy in terms of recoverability, strengthening the Fawzi–Renner bound [21] and subsequent improvements [9,11,35,62,63,71]. We find that for any positive semi-definite operator  $\sigma$ , and any trace-preserving completely positive map  $\mathcal{N}$ , there exists a trace-preserving completely positive recovery map  $\mathcal{R}_{\sigma,\mathcal{N}}$  that satisfies

$$D(\rho \parallel \sigma) - D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) \geq D_{\mathbb{M}}(\rho \parallel \mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N}(\rho)) \quad (5)$$

for any quantum state  $\rho$ . Here the bound is given in terms of the measured relative entropy,  $D_{\mathbb{M}}(\cdot \parallel \cdot)$ , as in [63]. The recovery map is the explicit universal (i.e., independent of  $\rho$ )

<sup>1</sup> This can be seen by considering the substitution  $A_i^{\frac{r}{2}} \leftarrow A_i$  for  $i \in \{1, 2\}$ ,  $\frac{q}{r} \leftarrow q$ , and  $\frac{1}{r} \leftarrow r$ .

rotated Petz recovery map introduced in [35]. We thus provide the first explicit lower bound that is tight in the commutative case. A precise statement and further results are presented in Sect. 4.

We believe that the proof techniques used in this article, based on asymptotic spectral pinching and complex interpolation theory, yield a transparent method to derive multivariate trace inequalities, which should be applicable beyond the extensions of the GT and ALT inequalities studied here.

Section 2 introduces the method of asymptotic spectral pinching and explains how it can be used to prove trace inequalities. Section 3 then explains how trace inequalities can be derived via complex interpolation theory. Readers interested in the proof of (2) and (4) may directly proceed to Sect. 3.

## 2. Trace Inequalities via Asymptotic Spectral Pinching

One contribution of this article is the presentation of a transparent method, based on *asymptotic spectral pinching*, that can be used to prove several trace inequalities. The results in this section hold for  $p$ -norms with  $p > 0$  and are in this sense slightly more general than the results mentioned in the introduction. However, the asymptotic spectral pinching method does not yield an explicit form of the distributions  $\beta_0$  and  $\beta_r$  in (2) and (4), respectively.

**2.1. The asymptotic spectral pinching method.** Let ‘ $\succeq$ ’ denote the Löwner partial order on positive semi-definite matrices. Any positive semi-definite matrix  $A$  has a decomposition  $A = \sum_{\lambda} \lambda P_{\lambda}$  where  $\lambda \in \text{spec}(A) \subset \mathbb{R}$  are unique eigenvalues and  $P_{\lambda}$  are mutually orthogonal projectors. The *spectral pinching map* with respect to  $A$  is

$$\mathcal{P}_A : X \mapsto \sum_{\lambda \in \text{spec}(A)} P_{\lambda} X P_{\lambda}. \quad (6)$$

Such maps are trace-preserving, completely positive, unital, self-adjoint, and can be viewed as dephasing operations that remove off-diagonal blocks of a matrix. Moreover, they satisfy the following properties: (i)  $\mathcal{P}_A[X]$  commutes with  $A$  for any  $X \geq 0$ , (ii)  $\text{tr } \mathcal{P}_A[X]A = \text{tr } XA$  for any  $X \geq 0$ , and (iii) we have the pinching inequality [28],

$$\mathcal{P}_A[X] = \sum_{\lambda \in \text{spec}(A)} P_{\lambda} X P_{\lambda} = \frac{1}{|\text{spec}(A)|} \sum_{y=1}^{|\text{spec}(A)|} U_y X U_y^{\dagger} \geq \frac{1}{|\text{spec}(A)|} X \quad \text{for all } X \geq 0, \quad (7)$$

where  $\text{spec}(A) = \{\lambda_1, \dots, \lambda_{|\text{spec}(A)|}\}$  and  $U_y = \sum_{z=1}^{|\text{spec}(A)|} \exp(\frac{i2\pi yz}{|\text{spec}(A)|}) P_{\lambda_z}$  are unitaries. The inequality step in (7) follows from the facts that  $U_y X U_y^{\dagger} \geq 0$  and  $U_{|\text{spec}(A)|} = \text{id}$ . The following observation is crucial. Let  $A$  be a positive semi-definite  $d \times d$  matrix. The cardinality  $|\text{spec}(A^{\otimes m})|$  grows polynomially in  $m$  due to the fact that the number of distinct eigenvalues of  $A^{\otimes m}$  is bounded by the number of different types of sequences of  $d$  symbols of length  $m$ , a concept widely used in information theory. More precisely [15, Lemma II.1] gives

$$|\text{spec}(A^{\otimes m})| \leq \binom{m+d-1}{d-1} \leq \frac{(m+d-1)^{d-1}}{(d-1)!} = O(\text{poly}(m)), \quad (8)$$

where  $\text{poly}(m)$  denotes an arbitrary polynomial in  $m$ . Another useful property of the pinching operation is that it exhibits the following integral representation.

**Lemma 2.1.** *Let  $A$  be positive definite. There exists a probability measure  $\mu$  on  $\mathbb{R}$  such that*

$$\mathcal{P}_A[X] = \int \mu(dt) A^{it} X A^{-it} \quad \text{for all } X \geq 0. \quad (9)$$

The proof of Lemma 2.1 is given in Appendix B. More information about pinching maps can be found in [13, Section 4.4] or [66, Section 2.6.3].

*2.2. Illustrative example: intuitive proof of Golden–Thompson inequality.* Here we illustrate the technique by proving the original GT inequality (1). Let  $A$  and  $B$  be two positive definite matrices. For any  $m \in \mathbb{N}$  we find the following chain of inequalities

$$\log \text{tr} \exp(\log A + \log B) = \frac{1}{m} \log \text{tr} \exp(\log A^{\otimes m} + \log B^{\otimes m}) \quad (10)$$

$$\leq \frac{1}{m} \log \text{tr} \exp(\log \mathcal{P}_{B^{\otimes m}}[A^{\otimes m}] + \log B^{\otimes m}) + \frac{\log \text{poly}(m)}{m} \quad (11)$$

$$= \frac{1}{m} \log \text{tr} \mathcal{P}_{B^{\otimes m}}[A^{\otimes m}] B^{\otimes m} + \frac{\log \text{poly}(m)}{m} \quad (12)$$

$$= \log \text{tr} AB + \frac{\log \text{poly}(m)}{m}. \quad (13)$$

The first equality (10) follows because the trace is multiplicative under tensor products. The sole inequality in (11) follows by the pinching inequality (7), i.e., Property (iii), together with the fact that the logarithm is operator monotone and  $\text{tr} \exp(\cdot)$  is monotone. Equality (12) uses Property (i) which ensures that  $\mathcal{P}_{B^{\otimes m}}[A^{\otimes m}]$  commutes with  $B^{\otimes m}$  and GT thus holds as an equality for these matrices. Equality (13) employs Property (ii) and again the multiplicativity of the trace under tensor products. Considering the limit  $m \rightarrow \infty$  directly implies the GT inequality (1).

As we will see later, this proof already suggests an extension of the GT inequality to  $n$  matrices by iterative pinching. Let us emphasize the high-level intuition of the proof method presented above. We know that the GT inequality is trivial if the operators commute, and spectral pinching forces our operators to commute. At the same time the pinching should hopefully not destroy the operator which it acts on too much. This is indeed the case (guaranteed by the pinching inequality) if we consider an  $m$ -fold tensor product of our operators and the limit  $m \rightarrow \infty$ .

*2.3. A convexity result for Schatten quasi-norms.* Let us define the Schatten  $p$ -norm of any matrix  $L$  as

$$\|L\|_p := (\text{tr}|L|^p)^{\frac{1}{p}} \quad \text{for } p \geq 1, \quad (14)$$

where  $|L| := \sqrt{L^\dagger L}$ . We extend this definition to all  $p > 0$ , but note that  $\|L\|_p$  is not a norm for  $p \in (0, 1)$  since it does not satisfy the triangle inequality. In the limit  $p \rightarrow \infty$  we recover the operator norm and for  $p = 1$  we obtain the trace norm.

Schatten norms are functions of the singular values and thus unitarily invariant. They satisfy  $\|L\|_p = \|L^\dagger\|_p$  and  $\|L\|_{2p}^2 = \|LL^\dagger\|_p = \|L^\dagger L\|_p$ . They are also multiplicative under tensor products. We note that the Schatten  $p$ -norm with  $p \geq 1$  is the unique norm that is unitarily invariant and multiplicative under tensor products [5, Theorem 4.2].<sup>2</sup>

Due to the triangle inequality  $p$ -norms for  $p \geq 1$  are convex. In particular, for any probability measure  $\mu$  on a measurable space  $(\mathcal{X}, \Sigma)$  and a collection  $\{L_x\}_{x \in \mathcal{X}}$  of matrices, we have

$$\left\| \int \mu(\mathrm{d}x) L_x \right\|_p \leq \int \mu(\mathrm{d}x) \|L_x\|_p. \quad (15)$$

Quasi-norms with  $p \in (0, 1)$  are no longer convex. However, we show that these quasi-norms still satisfy an asymptotic convexity property for tensor products of matrices in the following sense. We believe that this result may be of independent interest.

**Lemma 2.2.** *Let  $p \in (0, 1)$ ,  $\mu$  be a probability measure on  $(\mathcal{X}, \Sigma)$  and consider a collection  $\{A_x\}_{x \in \mathcal{X}}$  of positive semi-definite matrices. Then*

$$\frac{1}{m} \log \left\| \int \mu(\mathrm{d}x) A_x^{\otimes m} \right\|_p \leq \frac{1}{m} \log \int \mu(\mathrm{d}x) \|A_x^{\otimes m}\|_p + \frac{\log \text{poly}(m)}{m}. \quad (16)$$

The proof is given in Appendix C. Combining this with (15) shows that for all  $p > 0$  we have the following quasi-convexity property

$$\frac{1}{m} \log \left\| \int \mu(\mathrm{d}x) A_x^{\otimes m} \right\|_p \leq \log \sup_{x \in \mathcal{X}} \|A_x\|_p + \frac{\log \text{poly}(m)}{m}. \quad (17)$$

**2.4. Main results and proofs via pinching.** In this section we present two results obtained via the spectral pinching method, which are extensions of the ALT and the GT inequality, respectively, for arbitrarily many matrices. We want to emphasize that in addition to the fact that Theorem 2.3 is valid for Schatten quasi-norms, i.e.,  $p \in (0, 1)$ , the proof technique via pinching has the advantage of being transparent and intuitive.

**Theorem 2.3.** *Let  $p > 0$ ,  $r \in (0, 1]$ ,  $n \in \mathbb{N}$  and consider a collection  $\{A_k\}_{k=1}^n$  of positive semi-definite matrices. Then*

$$\left\| \left| \prod_{k=1}^n A_k^r \right|^{\frac{1}{r}} \right\|_p \leq \sup_{t \in \mathbb{R}^n} \left\| \prod_{k=1}^n A_k^{1+it_k} \right\|_p. \quad (18)$$

Before we present the proof, let us give an equivalent statement that follows by a simple substitution  $p \leftarrow 2q$  and  $A_k \leftarrow \sqrt{A_k}$  for  $q > 0$ , namely

$$\begin{aligned} & \text{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} \cdots A_{n-1}^{\frac{r}{2}} A_n^r A_{n-1}^{\frac{r}{2}} \cdots A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \\ & \leq \sup_{t \in \mathbb{R}^{n-2}} \text{tr} \left( A_1^{\frac{1}{2}} A_2^{\frac{1+it_2}{2}} \cdots A_{n-1}^{\frac{1+it_{n-1}}{2}} A_n A_{n-1}^{\frac{1-it_{n-1}}{2}} \cdots A_2^{\frac{1-it_2}{2}} A_1^{\frac{1}{2}} \right)^q. \end{aligned} \quad (19)$$

For  $n = 2$  the right-hand side of (19) is independent of  $t$  and we recover the ALT inequality in (3).

<sup>2</sup> Two properties that are crucial for the asymptotic spectral pinching method.

*Proof of Theorem 2.3.* We prove the result for positive definite matrices and note that the generalization to positive semi-definite matrices follows by continuity under the convention that  $0^z = 0$  for any  $z \in \mathbb{C}$ . For convenience of exposition we provide only the proof of Theorem 2.3 for three matrices (i.e,  $n = 3$ ). The generalization to  $n$  matrices follows by appropriately iterating the technical steps presented below. Using the multiplicativity of the trace under tensor products, we write

$$\log \operatorname{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} A_3^r A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} = \frac{1}{m} \log \operatorname{tr} \left( (A_1^{\frac{r}{2}})^{\otimes m} (A_2^{\frac{r}{2}})^{\otimes m} (A_3^r)^{\otimes m} (A_2^{\frac{r}{2}})^{\otimes m} (A_1^{\frac{r}{2}})^{\otimes m} \right)^{\frac{q}{r}} \quad (20)$$

Then, employing the pinching inequality and the monotonicity of  $\operatorname{tr}(\cdot)^{q/r}$ , we find

$$\begin{aligned} & \log \operatorname{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} A_3^r A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \\ & \leq \frac{1}{m} \log \operatorname{tr} \left( (A_1^{\frac{r}{2}})^{\otimes m} (A_2^{\frac{r}{2}})^{\otimes m} \mathcal{P}_{A_2^{\otimes m}} [(A_3^r)^{\otimes m}] (A_2^{\frac{r}{2}})^{\otimes m} (A_1^{\frac{r}{2}})^{\otimes m} \right)^{\frac{q}{r}} + o(1) \end{aligned} \quad (21)$$

$$\leq \frac{1}{m} \log \operatorname{tr} \left( (A_1^{\frac{r}{2}})^{\otimes m} (A_2^{\frac{r}{2}})^{\otimes m} \mathcal{P}_{A_2^{\otimes m}} [A_3^{\otimes m}]^r (A_2^{\frac{r}{2}})^{\otimes m} (A_1^{\frac{r}{2}})^{\otimes m} \right)^{\frac{q}{r}} + o(1) \quad (22)$$

$$= \frac{1}{m} \log \operatorname{tr} \left( (A_1^{\frac{r}{2}})^{\otimes m} ((A_2^{\frac{1}{2}})^{\otimes m} \mathcal{P}_{A_2^{\otimes m}} [A_3^{\otimes m}] (A_2^{\frac{1}{2}})^{\otimes m})^r (A_1^{\frac{r}{2}})^{\otimes m} \right)^{\frac{q}{r}} + o(1), \quad (23)$$

where  $o(1)$  simply denotes an additive term that vanishes as  $m \rightarrow \infty$ . The second inequality uses the fact that  $t \mapsto t^r$  is operator concave for  $r \in (0, 1]$ . The final step uses property (i) of pinching maps. Repeating these steps shows that

$$\begin{aligned} & \log \operatorname{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} A_3^r A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \\ & \leq \frac{1}{m} \log \operatorname{tr} \left( (A_1^{\frac{1}{2}})^{\otimes m} \mathcal{P}_{A_1^{\otimes m}} [(A_2^{\frac{1}{2}})^{\otimes m} \mathcal{P}_{A_2^{\otimes m}} [A_3^{\otimes m}] (A_2^{\frac{1}{2}})^{\otimes m}] (A_1^{\frac{1}{2}})^{\otimes m} \right)^q + o(1). \end{aligned} \quad (24)$$

The integral representation of pinching maps (see Lemma 2.1) ensures that there exist probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}$  such that

$$\begin{aligned} & \log \operatorname{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} A_3^r A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \\ & \leq \frac{1}{m} \log \operatorname{tr} \left( \int \mu(dt_1) \int \nu(dt_2) (A_1^{\frac{1}{2}+it_1})^{\otimes m} (A_2^{\frac{1}{2}+it_2})^{\otimes m} A_3^{\otimes m} (A_2^{\frac{1}{2}-it_2})^{\otimes m} (A_1^{\frac{1}{2}-it_1})^{\otimes m} \right)^q + o(1) \end{aligned} \quad (25)$$

$$\leq \sup_{t \in \mathbb{R}^2} \log \operatorname{tr} \left( A_1^{\frac{1}{2}+it_1} A_2^{\frac{1}{2}+it_2} A_3 A_2^{\frac{1}{2}-it_2} A_1^{\frac{1}{2}-it_1} \right)^q + o(1) \quad (26)$$

$$= \sup_{t \in \mathbb{R}} \log \operatorname{tr} \left( A_1^{\frac{1}{2}} A_2^{\frac{1}{2}+it} A_3 A_2^{\frac{1}{2}-it} A_1^{\frac{1}{2}} \right)^q + o(1), \quad (27)$$

where the second inequality uses Lemma 2.2. The final step follows from the fact that Schatten (quasi) norms are unitarily invariant. Considering the limit  $m \rightarrow \infty$  implies (19) and thus completes the proof.  $\square$

The multivariate Lie–Trotter product formula (see, e.g., [10, Problem IX.8.5]) states that

$$\lim_{r \searrow 0} \left( \exp(rL_1) \exp(rL_2) \dots \exp(rL_n) \right)^{\frac{1}{r}} = \exp \left( \sum_{k=1}^n L_k \right) \quad (28)$$

for square matrices  $\{L_k\}_{k=1}^n$ . This allows us to derive a multivariate extension of the GT inequality as a limit of the above extended ALT inequality in the limit  $r \rightarrow 0$ . In particular, combining the product formula with Theorem 2.3 implies an extension of the GT inequality to  $n$  matrices.

**Corollary 2.4.** *Let  $p > 0$ ,  $n \in \mathbb{N}$  and consider a collection  $\{H_k\}_{k=1}^n$  of Hermitian matrices. Then*

$$\left\| \exp \left( \sum_{k=1}^n H_k \right) \right\|_p \leq \sup_{t \in \mathbb{R}^n} \left\| \prod_{k=1}^n \exp((1 + it_k)H_k) \right\|_p. \quad (29)$$

For  $n = 2$  the right-hand side term of (29) is independent of  $t$  and we recover the GT inequality (1) for the choice  $p = 2$ .

### 3. Trace Inequalities via Interpolation Theory

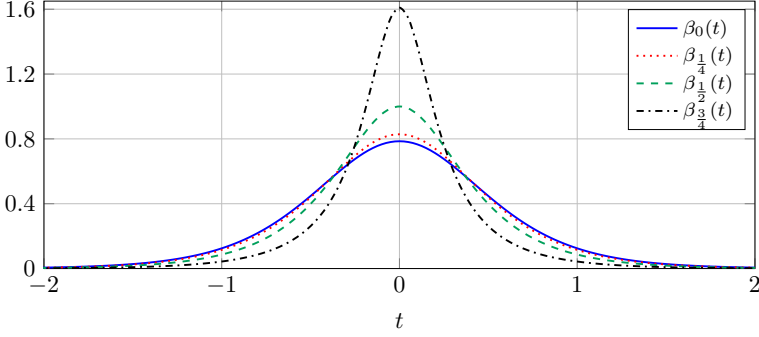
For  $p$ -norms we can prove a more explicit and also more general version of Theorem 2.3 based on an entirely different technique—*complex interpolation theory*.

**3.1. The complex interpolation method.** The main ingredient for most of our proofs in this section is a complex interpolation result for Schatten norms, commonly attributed to Stein [60], and based on Hirschman’s improvement of the Hadamard three-lines theorem [33]. Epstein [20] showed how interpolation theory can be used in matrix analysis.

Complex interpolation theory has recently garnered attention in quantum information theory for proving entropy inequalities. Beigi [6] and Dupuis [17] used variations of the Riesz–Thorin theorem based on Hadamard’s three line theorem to show properties of sandwiched Rényi divergence and conditional Rényi entropy, respectively. Wilde [71] first used these techniques to derive remainder terms for the monotonicity of quantum relative entropy (see Sect. 4 for more details). Extensions and further applications of this approach are discussed by Dupuis and Wilde [18]. Hirschmann’s refinement was first studied in this context by Junge et al. [35], where the following theorem essentially appeared:

**Theorem 3.1** (Stein–Hirschman). *Let  $S := \{z \in \mathbb{C} : 0 \leq \operatorname{Re}(z) \leq 1\}$  and let  $G$  be a map from  $S$  to bounded linear operators on a separable Hilbert space that is holomorphic in the interior of  $S$  and continuous on the boundary. Let  $p_0, p_1 \in [1, \infty]$ ,  $\theta \in (0, 1)$ , define  $p_\theta$  by*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1} \quad \text{and} \quad \beta_\theta(t) := \frac{\sin(\pi\theta)}{2\theta(\cosh(\pi t) + \cos(\pi\theta))}. \quad (30)$$



**Fig. 1.** This plot depicts the probability densities  $\beta_\theta$  defined in (30) for  $\theta \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$

Then, if furthermore  $z \mapsto \|G(z)\|_{p_{\text{Re}(z)}}$  is uniformly bounded on  $S$ ,<sup>3</sup> the following bound holds:

$$\log \|G(\theta)\|_{p_\theta} \leq \int_{-\infty}^{\infty} dt \left( \beta_{1-\theta}(t) \log \|G(it)\|_{p_0}^{1-\theta} + \beta_\theta(t) \log \|G(1+it)\|_{p_1}^\theta \right). \quad (31)$$

For the sake of completeness a proof is given in Appendix D. We note that for any  $\theta \in (0, 1)$  the function  $\beta_\theta$  is non-negative and  $\int_{-\infty}^{\infty} dt \beta_\theta(t) = 1$  so that  $\beta_\theta$  can be interpreted as probability density function on  $\mathbb{R}$ . These distributions are depicted in Fig. 1. Furthermore, the following limits hold:

$$\lim_{\theta \searrow 0} \beta_\theta(t) = \frac{\pi}{2} (\cosh(\pi t) + 1)^{-1} =: \beta_0(t), \quad \text{and} \quad \lim_{\theta \nearrow 1} \beta_\theta(t) = \delta(t) =: \beta_1(t). \quad (32)$$

Here  $\beta_0$  is another probability density function on  $\mathbb{R}$  and  $\delta(t)$  denotes the Dirac  $\delta$ -distribution.

**3.2. Main results and proofs via interpolation theory.** In this section we prove our main results which are extensions of the ALT and the GT inequality to arbitrarily many matrices.

**Theorem 3.2.** *Let  $p \geq 1$ ,  $r \in (0, 1]$ ,  $\beta_r$  as defined in (30),  $n \in \mathbb{N}$ , and consider a collection  $\{A_k\}_{k=1}^n$  of positive semi-definite matrices. Then*

$$\log \left\| \left\| \prod_{k=1}^n A_k^r \right\|_p^{\frac{1}{r}} \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_r(t) \log \left\| \prod_{k=1}^n A_k^{1+it} \right\|_p. \quad (33)$$

*Proof.* The case  $r = 1$  holds trivially with equality, so suppose  $r \in (0, 1)$ . We prove the result for positive definite matrices and note that the generalization to positive semi-definite matrices follows by continuity. We define the function  $G(z) := \prod_{k=1}^n A_k^z =$

<sup>3</sup> In fact, we only need that  $\sup_{z \in S} \exp(-a|\text{Im}(z)|) \log \|G(z)\|_{p_{\text{Re}(z)}} \leq A$  for some constants  $A < \infty$  and  $a < \pi$ .



$\prod_{k=1}^n \exp(z \log A_k)$  which satisfies the regularity assumptions of Theorem 3.1. Furthermore we pick  $\theta = r$ ,  $p_0 = \infty$  and  $p_1 = p$  such that  $p_\theta = \frac{p}{r}$ . We find

$$\begin{aligned} \log \|G(1+it)\|_{p_1}^\theta &= r \log \left\| \prod_{k=1}^n A_k^{1+it} \right\|_p \quad \text{and} \\ \log \|G(it)\|_{p_0}^{1-\theta} &= (1-r) \log \left\| \prod_{k=1}^n A_k^{it} \right\|_\infty = 0, \end{aligned} \quad (34)$$

since the matrices  $A_k^{it}$  are unitary. Moreover, we have

$$\log \|G(\theta)\|_{p_\theta} = \log \left\| \prod_{k=1}^n A_k^r \right\|_{\frac{p}{r}} = r \log \left\| \prod_{k=1}^n A_k^r \right\|_p^{\frac{1}{r}}. \quad (35)$$

Plugging this into Theorem 3.1 yields the desired inequality.  $\square$

Let us now remark on several aspects of this inequality. First, we note that the substitution  $p \leftarrow 2q$  and  $A_k \leftarrow \sqrt{A_k}$  allows to rewrite (33) in a more suggestive form. For  $q \geq \frac{1}{2}$  and  $r \in (0, 1]$ , we have

$$\begin{aligned} &\log \operatorname{tr} \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} \cdots A_{n-1}^{\frac{r}{2}} A_n^r A_{n-1}^{\frac{r}{2}} \cdots A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{q}{r}} \\ &\leq \int_{-\infty}^{\infty} dt \beta_r(t) \log \operatorname{tr} \left( A_1^{\frac{1}{2}} A_2^{\frac{1+it}{2}} \cdots A_{n-1}^{\frac{1+it}{2}} A_n A_{n-1}^{\frac{1-it}{2}} \cdots A_2^{\frac{1-it}{2}} A_1^{\frac{1}{2}} \right)^q. \end{aligned} \quad (36)$$

For  $n = 2$  the term on the right-hand side is independent of  $t$  and we recover the ALT inequality in (3). However, we only recover the result for  $q \geq \frac{1}{2}$  using complex interpolation theory. This can be fixed by proving a multivariate extension of the ALT inequality based on pinching (see Theorem 2.3).

Also note that we can always remove the logarithm in the above inequalities by using its concavity and Jensen's inequality. Moreover, for  $q \in [\frac{1}{2}, 1]$  we may pull the integration inside the quasi-norm (by employing the fact that  $X \mapsto \log \|X\|_p$  is concave for  $p \in [0, 1]$  on the positive definite cone), which yields the following relaxation

$$\begin{aligned} &\left\| \left( A_1^{\frac{r}{2}} A_2^{\frac{r}{2}} \cdots A_{n-1}^{\frac{r}{2}} A_n^r A_{n-1}^{\frac{r}{2}} \cdots A_2^{\frac{r}{2}} A_1^{\frac{r}{2}} \right)^{\frac{1}{r}} \right\|_q \\ &\leq \left\| \int_{-\infty}^{\infty} dt \beta_r(t) A_1^{\frac{1}{2}} A_2^{\frac{1+it}{2}} \cdots A_{n-1}^{\frac{1+it}{2}} A_n A_{n-1}^{\frac{1-it}{2}} \cdots A_2^{\frac{1-it}{2}} A_1^{\frac{1}{2}} \right\|_q. \end{aligned} \quad (37)$$

Next, recall the multivariate Lie–Trotter product formula in (28). Again, this allows us to derive an extension of the GT inequality to arbitrarily many matrices by taking the limit  $r \rightarrow 0$  of (36).

**Corollary 3.3.** *Let  $p \geq 1$ ,  $\beta_0$  as defined in (32),  $n \in \mathbb{N}$  and consider a collection  $\{H_k\}_{k=1}^n$  of Hermitian matrices. Then*

$$\log \left\| \exp \left( \sum_{k=1}^n H_k \right) \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_0(t) \log \left\| \prod_{k=1}^n \exp((1+it)H_k) \right\|_p. \quad (38)$$

Let us take a closer look at the case  $n = 3$  and  $p = 2$ . Substituting  $H_k \leftarrow \frac{1}{2}H_k$  and using the concavity of the logarithm and Jensen's inequality, we relax Corollary 3.3 to

$$\mathrm{tr} \exp(H_1 + H_2 + H_3) \leq \int_{-\infty}^{\infty} dt \beta_0(t) \mathrm{tr} \exp(H_1) \exp\left(\frac{1+it}{2}H_2\right) \exp(H_3) \exp\left(\frac{1-it}{2}H_2\right). \quad (39)$$

This is to be contrasted with Lieb's triple matrix inequality [44], which asserts that

$$\begin{aligned} \mathrm{tr} \exp(H_1 + H_2 + H_3) &\leq \int_0^{\infty} d\lambda \mathrm{tr} \exp(H_1) (\exp(-H_2) + \lambda \mathrm{id})^{-1} \\ &\quad \times \exp(H_3) (\exp(-H_2) + \lambda \mathrm{id})^{-1}. \end{aligned} \quad (40)$$

As the next lemma shows, it turns out that these two expressions are in fact equivalent. We believe that this result might be of independent interest as it allows us to write the Fréchet derivative of the operator logarithm using an integration over rotations.

**Lemma 3.4.** *The following two expressions for the Fréchet derivative of the logarithm are equivalent. For any positive definite operator  $A$  and Hermitian operator  $H$  on a separable Hilbert space, we have*

$$D \log(A)[H] = \int_0^{\infty} d\lambda (A + \lambda \mathrm{id})^{-1} H (A + \lambda \mathrm{id})^{-1} = \int_{-\infty}^{\infty} dt \beta_0(t) A^{-\frac{1}{2}-\frac{it}{2}} H A^{-\frac{1}{2}+\frac{it}{2}}. \quad (41)$$

The proof is given in Appendix E. The above lemma also gives a further means to understand the probability distribution  $\beta_0$  which we obtained from Hirschman's interpolation theorem.<sup>4</sup> Whereas Lieb's triple matrix inequality in (40) has not been extended to more than three matrices, the alternative representation obtained in (39) through Corollary 3.3 naturally extends to arbitrarily many matrices. Finally, it should be noted that Lieb's triple matrix inequality has been shown to be equivalent to many other interesting statements (such as Lieb's concavity theorem [44]), and hence it is valuable to have an entirely different proof of these results.

Corollary 3.3 is valid for Hermitian matrices, but we can extend its scope to general square matrices using the same techniques.

**Theorem 3.5.** *Let  $p \geq 1$ ,  $\beta_0$  as defined in (32),  $n \in \mathbb{N}$  and consider a collection  $\{L_k\}_{k=1}^n$  of square matrices. Define  $\Re(L_k) := \frac{1}{2}(L_k + L_k^\dagger)$ . Then*

$$\log \left\| \exp \left( \sum_{k=1}^n L_k \right) \right\|_p \leq \int_{-\infty}^{\infty} dt \beta_0(t) \log \left\| \prod_{k=1}^n \exp((1+it)\Re(L_k)) \right\|_p. \quad (42)$$

*Proof.* We write  $L_k = \Re(L_k) + i\Im(L_k)$  where  $\Im(L_k) = \frac{1}{2i}(L_k - L_k^\dagger)$ , and note that both  $\Re(L_k)$  and  $\Im(L_k)$  are Hermitian. Now define

$$G(z) := \prod_{k=1}^n \exp(z\Re(L_k) + i\theta\Im(L_k)), \quad (43)$$

<sup>4</sup> In Appendix A we also give numerical evidence that the exact form of the distribution  $\beta_0$  is crucial. Our results indicate that if the distribution is more narrow or more flat then Corollary 3.3 is no longer valid.

which satisfies the regularity assumption of Theorem 3.1. We note that  $G(it)$  is unitary, and thus  $\log \|G(it)\|_\infty$  vanishes. We again pick  $\theta = r \in (0, 1)$ ,  $p_0 = \infty$  and  $p_1 = p$  such that  $p_\theta = \frac{p}{r}$ , and find

$$r \log \left\| \exp \left( r \sum_{k=1}^n L_k \right) \right\|_{\frac{1}{r}}^{\frac{1}{r}} = \log \|G(\theta)\|_{p_\theta} \quad (44)$$

$$\leq \int_{-\infty}^{\infty} dt \beta_r(t) \log \|G(1+it)\|_p^r \quad (45)$$

$$= r \int_{-\infty}^{\infty} dt \beta_r(t) \log \left\| \prod_{k=1}^n \exp((1+it)\Re(L_k) + r\Im(L_k)) \right\|_p. \quad (46)$$

Dividing by  $r$  and taking the limit  $r \rightarrow 0$  then yields the desired result via the Lie–Trotter product formula.  $\square$

We note that for the case of normal matrices  $N$ , the matrices  $\Re(N)$  and  $\Im(N)$  commute, which allows us to slightly simplify the above formula by employing the fact that  $\exp(\Re(N)) = |\exp(N)|$ . For two normal matrices the result then reads

$$\|\exp(N_1 + N_2)\|_p \leq \|\exp(N_1)\|_p \|\exp(N_2)\|_p, \quad (47)$$

generalizing an inequality by Li and Zhao [42]. Finally, we note that (46) can be viewed as an ALT inequality for general square matrices.

#### 4. An Application: Entropy Inequalities

In this section we show that the multivariate extension of the GT inequality derived in Corollary 3.3 can be used to derive remainder terms in terms of recoverability for certain entropy inequalities.

For positive semi-definite matrices  $\rho, \sigma$  with  $\text{tr } \rho = 1$ , Umegaki’s *quantum relative entropy* [69] is defined as  $D(\rho\|\sigma) := \text{tr } \rho(\log \rho - \log \sigma)$  if  $\rho \ll \sigma$  and as  $+\infty$  if  $\rho \not\ll \sigma$ . Here,  $\rho \ll \sigma$  denotes that the support of  $\rho$  is contained in the support of  $\sigma$ . We recall the following variational formula for the relative entropy [53] (see also [7]):

$$D(\rho\|\sigma) = \sup_{\omega > 0} \text{tr } \rho \log \omega + 1 - \text{tr } \exp(\log \sigma + \log \omega). \quad (48)$$

The *measured relative entropy* is given as  $D_{\mathbb{M}}(\rho\|\sigma) := \sup_{(\mathcal{X}, M)} D(P_{\rho, M}\|P_{\sigma, M})$  [7, 16, 31, 52], where the optimization is over positive operator valued measures (POVMs)  $M$  on the power-set of a finite set  $\mathcal{X}$ , the probability mass functions are given by  $P_{\rho, M}(x) = \text{tr } \rho M(x)$ , and  $D(P\|Q)$  is the *Kullback–Leibler divergence* [40]. We recall the following variational formula [7, 54]:

$$D_{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega > 0} \text{tr } \rho \log \omega + 1 - \text{tr } \sigma \omega. \quad (49)$$

A fundamental entropy inequality [46, 47] states that the quantum relative entropy is monotone under trace-preserving and completely positive maps  $\mathcal{N}$ , i.e.,<sup>5</sup>

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq 0. \quad (50)$$

<sup>5</sup> We note that this monotonicity statement remains valid for more general maps  $\mathcal{N}$  [29, 30, 49, 50, 68].

This is closely related to the celebrated strong sub-additivity of quantum entropy [45, 46] stating that

$$I(A : C|B)_\rho := H(\rho_{AB}) + H(\rho_{BC}) - H(\rho_B) - H(\rho_{ABC}) \geq 0 \quad (51)$$

for any positive semi-definite matrix  $\rho_{ABC}$  on a composite Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  with  $\text{tr } \rho_{ABC} = 1$ . Here  $\rho_{AB}$ ,  $\rho_{BC}$ , and  $\rho_B$  are marginals of  $\rho_{ABC}$  obtained via the partial trace, and  $H(\rho) = -\text{tr } \rho \log \rho$  denotes the *von Neumann entropy*.

Motivated by recoverability questions in quantum information theory, (50) and (51) have been refined in a series of recent works [9, 11, 21, 35, 62, 63, 71], making use of complex interpolation theory as well as asymptotic spectral pinching. With the four matrix extension of the GT inequality given by Corollary 3.3, we find the following statement which answers an open question stated in [35].

**Theorem 4.1** (Strengthened monotonicity for partial trace). *Let  $\rho_{AB}$  and  $\sigma_{AB}$  be positive semi-definite matrices on  $\mathcal{H}_A \otimes \mathcal{H}_B$  such that  $\rho_{AB} \ll \sigma_{AB}$  and  $\text{tr } \rho_{AB} = 1$ . Then*

$$D(\rho_{AB} \| \sigma_{AB}) - D(\rho_A \| \sigma_A) \geq D_{\mathbb{M}}(\rho_{AB} \| \mathcal{R}_{\sigma_{AB}, \text{tr}_B}(\rho_A)), \quad (52)$$

with the rotated Petz recovery map given by

$$\begin{aligned} \mathcal{R}_{\sigma_{AB}, \text{tr}_B}(\cdot) &:= \int_{-\infty}^{\infty} dt \beta_0(t) \mathcal{R}_{\sigma_{AB}, \text{tr}_B}^{[t]}(\cdot) \quad \text{and} \\ \mathcal{R}_{\sigma_{AB}, \text{tr}_B}^{[t]}(\cdot) &:= \sigma_{AB}^{\frac{1+i t}{2}} \left( \sigma_A^{-\frac{1+i t}{2}}(\cdot) \sigma_A^{-\frac{1-i t}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{\frac{1-i t}{2}}. \end{aligned} \quad (53)$$

*Proof.* Let us recall Corollary 3.3 applied for  $n = 4$  and  $p = 2$ . Using the concavity of the logarithm and Jensen's inequality, it yields

$$\begin{aligned} &\text{tr} \exp(H_1 + H_2 + H_3 + H_4) \\ &\leq \int_{-\infty}^{\infty} dt \beta_0(t) \text{tr} \exp(H_1) \exp\left(\frac{1+i t}{2} H_2\right) \exp\left(\frac{1+i t}{2} H_3\right) \exp(H_4) \\ &\quad \times \exp\left(\frac{1-i t}{2} H_3\right) \exp\left(\frac{1-i t}{2} H_2\right) \end{aligned} \quad (54)$$

for Hermitian matrices  $\{H_i\}_{i=1}^4$ . Moreover, by definition of the relative entropy for positive definite operators  $\rho_{AB}$  and  $\sigma_{AB}$ , we have

$$D(\rho_{AB} \| \sigma_{AB}) - D(\rho_A \| \sigma_A) = D(\rho_{AB} \| \exp(\log \sigma_{AB} + \log \rho_A \otimes \text{id}_B - \log \sigma_A \otimes \text{id}_B)). \quad (55)$$

For positive semi-definite operators  $\rho_{AB}$  and  $\sigma_{AB}$ , the Hermitian operators  $\log \sigma_{AB}$ ,  $\log \rho_A$  and  $\log \sigma_A$  are well-defined under the convention  $\log 0 = 0$ . Under this convention, the above equality (55) also holds for positive semi-definite operators as long as  $\rho_{AB} \ll \sigma_{AB}$ , which is required by the theorem. By the variational formula for the

relative entropy (48) we thus find

$$D(\rho_{AB} \parallel \sigma_{AB}) - D(\rho_A \parallel \sigma_A) = \sup_{\omega_{AB} > 0} \text{tr } \rho_{AB} \log \omega_{AB} + 1 - \text{tr} \exp(\log \sigma_{AB} + \log \rho_A \otimes \text{id}_B - \log \sigma_A \otimes \text{id}_B + \log \omega_{AB}) \quad (56)$$

$$\geq \sup_{\omega_{AB} > 0} \text{tr } \rho_{AB} \log \omega_{AB} + 1 - \int_{-\infty}^{\infty} dt \beta_0(t) \text{tr } \sigma_{AB}^{\frac{1+it}{2}} \times \left( \sigma_A^{-\frac{1+it}{2}} \rho_A \sigma_A^{-\frac{1-it}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{\frac{1-it}{2}} \omega_{AB} \quad (57)$$

$$= D_{\mathbb{M}} \left( \rho_{AB} \parallel \int_{-\infty}^{\infty} dt \beta_0(t) \sigma_{AB}^{\frac{1+it}{2}} \left( \sigma_A^{-\frac{1+it}{2}} \rho_A \sigma_A^{-\frac{1-it}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{\frac{1-it}{2}} \right), \quad (58)$$

where the single inequality follows by the four matrix extension of the GT inequality in (54). The final step uses the variational formula (49) for the measured relative entropy.  $\square$

We note that the four matrix extension of the GT inequality is the only inequality used in the proof of Theorem 4.1. More properties of the recovery map  $\mathcal{R}_{\sigma_{AB}, \text{tr}_B}$  given by (53) are discussed in [35].

Theorem 4.1 implies two other interesting statements. If we substitute  $\rho_{AB} \leftarrow \rho_{ABC}$ ,  $\rho_A \leftarrow \rho_{AB}$ ,  $\sigma_{AB} \leftarrow \text{id}_A \otimes \rho_{BC}$ , and  $\sigma_A \leftarrow \text{id}_A \otimes \rho_B$  we immediately find a remainder term for the conditional quantum mutual information, namely

$$I(A : C|B)_\rho \geq D_{\mathbb{M}}(\rho_{ABC} \parallel \mathcal{R}_{\rho_{BC}, \text{tr}_C} \otimes \mathcal{I}_A(\rho_{AB})), \quad (59)$$

where  $\mathcal{I}_A$  is the identity map and  $\mathcal{R}_{\rho_{BC}, \text{tr}_C}$  is defined in (53). Moreover, using the Stinespring dilation theorem [61] and the fact that the relative entropy is invariant under isometries, Theorem 4.1 generalizes to the following result.

**Corollary 4.2** (Strengthened monotonicity). *Let  $\rho, \sigma$  be positive semi-definite matrices such that  $\rho \ll \sigma$ ,  $\text{tr } \rho = 1$ , and  $\mathcal{N}$  be a trace-preserving completely positive map acting on these matrices. Then*

$$D(\rho \parallel \sigma) - D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) \geq D_{\mathbb{M}}(\rho \parallel \mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N}(\rho)), \quad (60)$$

with the rotated Petz recovery map given by

$$\begin{aligned} \mathcal{R}_{\sigma, \mathcal{N}}(\cdot) &:= \int_{-\infty}^{\infty} dt \beta_0(t) \mathcal{R}_{\sigma, \mathcal{N}}^{[t]}(\cdot) \quad \text{and} \\ \mathcal{R}_{\sigma, \mathcal{N}}^{[t]}(\cdot) &:= \sigma^{\frac{1+it}{2}} \mathcal{N}^\dagger \left( \mathcal{N}(\sigma)^{-\frac{1+it}{2}} (\cdot) \mathcal{N}(\sigma)^{-\frac{1-it}{2}} \right) \sigma^{\frac{1-it}{2}}. \end{aligned} \quad (61)$$

*Proof.* Let us introduce the Stinespring dilation of  $\mathcal{N}$ , denoted  $U$ , and the states  $\rho_{AB} = U\rho U^\dagger$ ,  $\sigma_{AB} = U\sigma U^\dagger$  such that  $\mathcal{N}(\rho) = \rho_A$  and  $\mathcal{N}(\sigma) = \sigma_B$ . Then, using the fact that the relative entropy is invariant under isometries, we have

$$D(\rho \parallel \sigma) - D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) = D(\rho_{AB} \parallel \sigma_{AB}) - D(\rho_A \parallel \sigma_A) \quad (62)$$

$$\geq D_{\mathbb{M}}(\rho_{AB} \parallel \mathcal{R}_{\sigma_{AB}, \text{tr}_B}(\rho_A)) = D_{\mathbb{M}}(\rho \parallel \mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N}(\rho)), \quad (63)$$

where the inequality is due to Theorem 4.1 and the last equality uses again invariance under isometries and the fact that

$$U^\dagger \mathcal{R}_{\sigma_{AB}, \text{tr}_B}^{[r]}(\cdot) U = U^\dagger U \sigma^{\frac{1+i\epsilon}{2}} U^\dagger \left( \mathcal{N}(\sigma)^{-\frac{1+i\epsilon}{2}}(\cdot) \mathcal{N}(\sigma)^{-\frac{1-i\epsilon}{2}} \otimes \text{id}_B \right) U \sigma^{\frac{1-i\epsilon}{2}} U^\dagger U \quad (64)$$

$$= \sigma^{\frac{1+i\epsilon}{2}} \mathcal{N}^\dagger \left( \mathcal{N}(\sigma)^{-\frac{1+i\epsilon}{2}}(\cdot) \mathcal{N}(\sigma)^{-\frac{1-i\epsilon}{2}} \right) \sigma^{\frac{1-i\epsilon}{2}} = \mathcal{R}_{\sigma, \mathcal{N}}^{[r]}(\cdot). \quad (65)$$

□

We note that Corollary 4.2 is no longer valid if we replace the measured relative entropy in (60) with a relative entropy. This leads us to believe that (60) cannot be further improved.

The right-hand side of (60) can be relaxed using Uhlmann’s *fidelity*,  $F(\rho, \sigma) := \|\sqrt{\rho} \sqrt{\sigma}\|_1^2$ . It is well known that  $D_{\mathbb{M}}(\rho \| \sigma) \geq -\log F(\rho, \sigma)$ .<sup>6</sup> Therefore, Corollary 4.2 implies

$$D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \geq -\log F(\rho, \mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N}(\rho)). \quad (66)$$

Moreover, Corollary 4.2 can be transformed into universal remainder terms (in terms of recoverability with the measured relative entropy) for other entropy inequalities, such as concavity of the conditional entropy and joint convexity of the relative entropy [8]. We refer to [35, Section 5] for a more detailed discussion of these bounds.

We also want to refer the reader to Appendix F where we give a different derivation that yields lower and upper bounds on the difference of relative entropies in Theorem 4.1. This derivation follows the structure of Lieb and Ruskai’s original proof of strong sub-additivity [45, 46], i.e., it uses the Peierls–Bogoliubov inequality followed by an extension of the GT inequality (see also [14]). However, whereas Lieb and Ruskai use the three matrix extension of the GT inequality (Lieb’s triple matrix inequality) we use the four matrix extension of the GT inequality, leading us to a stronger statement.

## 5. Discussion

We discussed two techniques to prove trace inequalities. One is based on asymptotic pinching and the other one uses complex interpolation theory. Both methods lead to transparent and direct proofs of generalized multivariate extensions of the GT and also the more general ALT inequalities. We believe that these methods can be used to prove trace inequalities beyond the extensions of the GT and ALT inequalities studied in this article. For example in [2, 32], complementary GT and ALT inequalities have been shown in terms of matrix means. It is left for future research to investigate if these inequalities can be obtained (and possibly be extended to the multivariate case) via pinching or interpolation theory.

Hansen gave an alternative multivariate extension of the GT inequality [25] that can be considered an interpolation between the original GT inequality (1) and the operator Jensen inequality [26, 27]. It would be interesting to unify his and our result.

Moreover, Lieb showed that his triple matrix inequality (40) is equivalent to many other interesting statements such as several concavity results [44]. As Corollary 3.3

<sup>6</sup> This follows by the monotonicity of quantum Rényi divergence in the order parameter [51] and of the fact that for any two states there exists an optimal measurement that does not increase their fidelity [22, Section 3.3].

generalizes the triple matrix inequality, it is natural to ask if it can be used to prove more general concavity results.

Ahlsweide and Winter noticed that the GT inequality can be used to prove tail bounds for sums of random matrices via the Laplace transform method [1]. As the (original) GT inequality is only valid for two matrices it has to be applied sequentially. Later, Tropp realized that sharper tail bounds can be obtained by using Lieb's concavity theorem instead of the GT inequality [67]. An interesting question is whether the multivariate extension of the GT inequality derived in this article (see Corollary 3.3) can be used to prove tail bounds for random matrices.

*Acknowledgements.* Elliott H. Lieb's talk at the "Beyond I.I.D. in Information Theory" workshop in Banff inspired us to derive an alternative proof for his triple matrix inequality. We thank Christian Majenz for allowing us to include his counterexample for a three matrix GT inequality without rotations. We would also like to thank Jürg Fröhlich, Aram W. Harrow, Spyridon Michalakis, and Renato Renner for useful discussions about trace inequalities. We thank Marius C. Lemm, Thomas Vidick, and Mark M. Wilde for comments on an earlier draft. MB and MT thank the Institute for Theoretical Physics at ETH Zurich for hosting them when this project was initiated. DS acknowledges support by the Swiss National Science Foundation (SNSF) via the National Centre of Competence in Research "QSIT" and by the European Commission via the project "RAQUEL". MB acknowledges funding by the SNSF through a fellowship, funding by the Institute for Quantum Information and Matter (IQIM), an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028), and funding support from the ARO grant for Research on Quantum Algorithms at the IQIM (W911NF-12-1-0521). MT is funded by an ARC Discovery Early Career Researcher Award (DECRA) fellowship and acknowledges support from the ARC Centre of Excellence for Engineered Quantum Systems (EQUS).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A. About the Probability Distribution in Corollary 3.3

As observed in [64], we have  $\text{tr} \exp(H_1 + H_2 + H_3) \not\leq \text{tr} \exp(H_1) \exp(H_2) \exp(H_3)$  in general. Also a more symmetric conjecture does not hold in general [48], i.e.,

$$\text{tr} \exp(H_1 + H_2 + H_3) \not\leq \text{tr} \exp(H_1) \exp\left(\frac{1}{2}H_2\right) \exp(H_3) \exp\left(\frac{1}{2}H_2\right). \quad (67)$$

For the following discussion, let us define

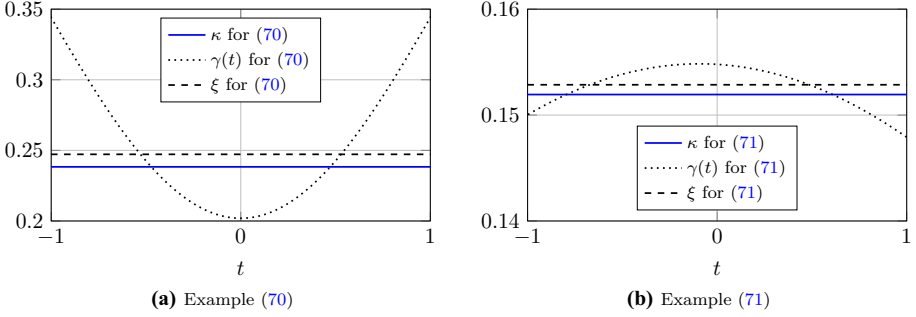
$$\gamma(t) := \text{tr} A_3^{\frac{1}{2}} A_2^{\frac{1}{2} + it} A_1 A_2^{\frac{1}{2} - it} A_3^{\frac{1}{2}} \quad \text{and} \quad \kappa := \text{tr} \exp(\log A_1 + \log A_2 + \log A_3), \quad (68)$$

for three positive definite matrices  $A_1$ ,  $A_2$ , and  $A_3$ . As discussed in (39), Corollary 3.3 implies that

$$\kappa \leq \int_{-\infty}^{\infty} dt \beta_0(t) \gamma(t) =: \xi. \quad (69)$$

It is a natural question to investigate how much freedom we have in choosing a probability distribution (different than  $\beta_0$ ) such that (69) remains valid, where the distribution should be independent of the matrices  $A_1$ ,  $A_2$ , and  $A_3$ . The following two examples indicate that it might be difficult to find a distribution different than  $\beta_0$  that satisfies (69) since it cannot be too narrow (around  $t = 0$ ) but also not too flat, either. Let us consider the positive semi-definite matrices [48]:

$$A_1 = \frac{1}{4} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}, \quad A_2 = \frac{1}{4} \begin{pmatrix} 1 & -2 \\ -2 & 2 \end{pmatrix}, \quad \text{and} \quad A_3 = \frac{1}{4} \begin{pmatrix} 8 & -2 \\ -2 & 1 \end{pmatrix}. \quad (70)$$



**Fig. 2.** This plot compare  $\gamma(t)$  with  $\kappa$  for the matrices defined in (70) and (71). If we want  $\kappa \leq \int \mu(dt)\gamma(t)$  to hold for some probability measure  $\mu$  on  $\mathbb{R}$  that does not depend on  $A_1, A_2$  and  $A_3$ , these two example show that  $\mu$  cannot be too narrow (around  $t = 0$ ) but also not too flat, either

As a second example we consider the positive semi-definite matrices

$$A_1 = \frac{1}{8} \begin{pmatrix} 4 & 2-i \\ 2+i & 3 \end{pmatrix}, \quad A_2 = \frac{1}{60} \begin{pmatrix} 15 & -5-3i \\ -5+3i & 12 \end{pmatrix}, \quad \text{and} \\ A_3 = \frac{1}{20} \begin{pmatrix} 15 & 10-5i \\ 10+5i & 11 \end{pmatrix}. \quad (71)$$

Figure 2 compares  $\kappa$  with  $\gamma(t)$ . We note that the matrices (70) also show that  $\kappa > \gamma(0)$  is possible, i.e., a three matrix extension of the GT inequality without any phases does not hold in general [48].

## B. Proof of Lemma 2.1

We want to write  $\mathcal{P}_A[X] = \sum_{\lambda \in \text{spec}(A)} P_\lambda X P_\lambda$  in the form  $\mathcal{P}_A[X] = \int \mu(dt) A^{it} X A^{-it}$ , where  $\mu$  is a probability measure on  $\mathbb{R}$ . Recalling that  $A = \sum_{\lambda \in \text{spec}(A)} \lambda P_\lambda$  and the fact that the eigenvectors to distinct eigenvalues of positive semi-definite matrices are orthogonal we find

$$A^{it} = \sum_{\lambda \in \text{spec}(A)} \lambda^{it} P_\lambda. \quad (72)$$

and thus

$$A^{it} X A^{-it} = \sum_{\lambda, \lambda' \in \text{spec}(A)} \lambda^{it} (\lambda')^{-it} P_\lambda X P_{\lambda'}. \quad (73)$$

Defining  $\tilde{\lambda} = \log \lambda$  and  $\tilde{\lambda}' = \log \lambda'$  we thus require

$$\begin{aligned} \hat{\mu}(\tilde{\lambda}' - \tilde{\lambda}) &= \int \mu(dt) \exp(-it(\tilde{\lambda}' - \tilde{\lambda})) \\ &= \int \mu(dt) \lambda^{it} (\lambda')^{-it} = \delta\{\lambda = \lambda'\} = \delta\{\tilde{\lambda} = \tilde{\lambda}'\}, \end{aligned} \quad (74)$$

where  $\hat{\mu}$  denotes the Fourier transform of  $\mu$ . We thus require that (i)  $\hat{\mu}(0) = 1$  and (ii)  $\hat{\mu}(\tilde{\lambda} - \tilde{\lambda}') = 0$  for all  $\exp(\tilde{\lambda}), \exp(\tilde{\lambda}') \in \text{spec}(A)$ . Let us define

$$\Delta := \min\{|\tilde{\lambda} - \tilde{\lambda}'| : \exp(\tilde{\lambda}), \exp(\tilde{\lambda}') \in \text{spec}(A), \tilde{\lambda} \neq \tilde{\lambda}'\}. \quad (75)$$

Furthermore for a fixed  $\tau > 0$  we define the triangular function

$$r_\tau(t) = \begin{cases} 1 - \frac{|t|}{\tau}, & \text{if } |t| < \tau \\ 0 & \text{otherwise.} \end{cases} \quad (76)$$



We next pick  $\hat{\mu}(\xi) = r_{\frac{\Delta}{2}}(\xi)$  which clearly satisfies the requirements (i) and (ii) mentioned above. Its inverse Fourier transform can be computed as

$$\mu(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi r_{\frac{\Delta}{2}}(\xi) \exp(it\xi) = \frac{1 - \cos(\Delta t/2)}{\pi \Delta t^2/2}. \quad (77)$$

It is immediate to verify that  $\mu$  as given in (77) is a probability distribution on  $\mathbb{R}$  (i.e.,  $\mu(t) \geq 0$  for all  $t \in \mathbb{R}$  and  $\int_{\mathbb{R}} dt \mu(t) = 1$ ). This is the distribution that satisfies the assertion of Lemma 2.1 and thus completes the proof.

### C. Proof of Lemma 2.2

Let  $\mathcal{H}$  denote the Hilbert space of dimension  $d$  where the matrices  $A_x$  act on. For any  $x$ , consider the spectral decomposition  $A_x = \sum_k \lambda_k |k\rangle\langle k|$  in Dirac bra-ket notation. Introducing an isometric space  $\mathcal{H}'$ , we define the vector  $|v_x\rangle \in \mathcal{H} \otimes \mathcal{H}'$  by  $|v_x\rangle = \sum_k \sqrt{\lambda_k} |k\rangle \otimes |k\rangle$ —i.e., the purification of  $A_x$ . Now note that the projectors  $(|v_x\rangle\langle v_x|)^{\otimes m}$  lie in the symmetric subspace of  $(\mathcal{H} \otimes \mathcal{H}')^{\otimes m}$  whose dimension grows as  $\text{poly}(m)$ .<sup>7</sup> Moreover, we have

$$\int \mu(dx) A_x^{\otimes m} = \int \mu(dx) \text{tr}_{\mathcal{H}'^{\otimes m}} (|v_x\rangle\langle v_x|)^{\otimes m}. \quad (78)$$

Then by Carathéodory's theorem (see, e.g., [19, Theorem 18]) there exists a discrete probability measure  $P$  on  $\mathcal{I} \subset \mathcal{X}$  with  $|\mathcal{I}| = \text{poly}(m)$  such that

$$\int \mu(dx) A_x^{\otimes m} = \sum_{x \in \mathcal{I}} P(x) A_x^{\otimes m} \quad \text{and} \quad \int \mu(dx) \|A_x^{\otimes m}\|_p = \sum_{x \in \mathcal{I}} P(x) \|A_x^{\otimes m}\|_p. \quad (79)$$

If  $p \in (0, 1)$  the Schatten  $p$ -norms only satisfy a weakened version of the triangle inequality (see, e.g., [36, Equation 20]), which states that

$$\left\| \sum_{x=1}^n A_x \right\|_p^p \leq \sum_{x=1}^n \|A_x\|_p^p. \quad (80)$$

Hence, we find the following chain of inequalities

$$\frac{1}{m} \log \left\| \int \mu(dx) A_x^{\otimes m} \right\|_p = \frac{1}{m} \log \left\| \sum_{x \in \mathcal{I}} P(x) A_x^{\otimes m} \right\|_p \quad (81)$$

$$\leq \frac{1}{m} \log \left( \sum_{x \in \mathcal{I}} \|P(x) A_x^{\otimes m}\|_p^p \right)^{\frac{1}{p}} \quad (82)$$

$$= \frac{1}{m} \log \left( |\mathcal{I}|^{\frac{1}{p}} \left( \frac{1}{|\mathcal{I}|} \sum_{x \in \mathcal{I}} \|P(x) A_x^{\otimes m}\|_p^p \right)^{\frac{1}{p}} \right) \quad (83)$$

$$\leq \frac{1}{m} \log \left( |\mathcal{I}|^{\frac{1}{p}-1} \sum_{x \in \mathcal{I}} \|P(x) A_x^{\otimes m}\|_p \right) \quad (84)$$

$$= \frac{1}{m} \log \left( \sum_{x \in \mathcal{I}} P(x) \|A_x^{\otimes m}\|_p \right) + \frac{1}{m} \frac{1-p}{p} \log |\mathcal{I}| \quad (85)$$

$$= \frac{1}{m} \log \left( \int \mu(dx) \|A_x^{\otimes m}\|_p \right) + \frac{\log \text{poly}(m)}{m}, \quad (86)$$

where the second inequality uses that the map  $t \mapsto t^{\frac{1}{p}}$  is convex for  $p \in (0, 1)$ . The final step follows from the fact that  $|\mathcal{I}| = \text{poly}(m)$ .

<sup>7</sup> This follows from the fact that the dimension of the symmetric subspace of  $\mathcal{H}^{\otimes m}$  is equal to the number of types of sequences of  $d$  symbols of length  $m$ , which is polynomial in  $m$  (as shown in (8)).

## D. Proof of Theorem 3.1

We follow the argument given in [35, Appendix A], and take care of the explicit conditions on the Schatten norms of  $G(z)$ . We recall Hirschman's strengthening [33] (see also [24, Lemma 1.3.8]) of Hadamard's three line theorem.

**Lemma D.1** (Hirschman). *Let  $S := \{z \in \mathbb{C} : 0 \leq \operatorname{Re}(z) \leq 1\}$  and let  $g(z)$  be uniformly bounded on  $S$ , holomorphic on the interior of  $S$  and continuous on the boundary. Then for  $\theta \in (0, 1)$ , we have*

$$\log |g(\theta)| \leq \int_{-\infty}^{\infty} dt \beta_{1-\theta}(t) \log |g(it)|^{1-\theta} + \beta_{\theta}(t) \log |g(1+it)|^{\theta}. \quad (87)$$

Moreover, the assumption that the function is uniformly bounded can be relaxed to

$$\sup_{z \in S} \exp(-a|\operatorname{Im}(z)|) \log |g(z)| \leq A \text{ for some constants } A < \infty \text{ and } a < \pi. \quad (88)$$

We are now prepared to prove Theorem 3.1.

*Proof of Theorem 3.1.* For  $x \in [0, 1]$ , define  $q_x$  as the Hölder conjugate of  $p_x$  such that  $p_x^{-1} + q_x^{-1} = 1$ . Hence, using the definition of  $p_x$  in (30), we have

$$\frac{1}{q_x} = \frac{1-x}{q_0} + \frac{x}{q_1}. \quad (89)$$

Now for our fixed  $\theta \in (0, 1)$  the operator  $G(\theta)$  is bounded by assumption and thus allows a polar decomposition,  $G(\theta) = U\Delta$ , where  $\Delta$  is positive semi-definite and  $U$  is a partial isometry [55, Theorem VI.10] satisfying  $\Delta U^{\dagger}U = U^{\dagger}U\Delta = \Delta$ . Then define  $X(z)$  via

$$X(z)^{\dagger} = C^{-p_{\theta}} \left( \frac{1-\bar{z}}{q_0} + \frac{z}{q_1} \right) \Delta^{p_{\theta}} \left( \frac{1-\bar{z}}{q_0} + \frac{z}{q_1} \right) U^{\dagger} \quad \text{with} \quad C := \|\Delta\|_{p_{\theta}} = \|G(\theta)\|_{p_{\theta}} < \infty. \quad (90)$$

We find that  $z \mapsto X(z)$  is anti-holomorphic on  $S$  and

$$\|X(x+iy)\|_{q_x}^{q_x} = \operatorname{tr} \left( C^{-1}\Delta \right)^{p_{\theta}q_x} \left( \frac{1-x}{q_0} + \frac{x}{q_1} \right) = \operatorname{tr} \left( C^{-1}\Delta \right)^{p_{\theta}} = 1. \quad (91)$$

Consequently, the Hilbert-Schmidt inner product  $g(z) := \operatorname{tr} X(z)^{\dagger} G(z)$  is holomorphic and bounded on  $S$  because the Hölder inequality (see, e.g., [34, Theorem 7.8]) yields

$$|g(x+iy)| \leq \|X(x+iy)\|_{q_x} \|G(x+iy)\|_{p_x} \leq \|G(x+iy)\|_{p_x}, \quad (92)$$

Hence, our assumptions on  $G(z)$  imply that  $g(z)$  satisfies the assumptions of Lemma D.1.

It remains to verify the following relations using the Hölder inequality in (92):

$$g(\theta) = \operatorname{tr} X(\theta)G(\theta) = C^{-p_{\theta}} \frac{1}{q_{\theta}} \operatorname{tr} \Delta^{p_{\theta}-1} U^{\dagger} U \Delta = C^{1-p_{\theta}} \operatorname{tr} \Delta^{p_{\theta}} = \|G(\theta)\|_{p_{\theta}}, \quad (93)$$

$$|g(it)| \leq \|G(it)\|_{p_0}, \quad \text{and} \quad |g(1+it)| \leq \|G(1+it)\|_{p_1}. \quad (94)$$

Substituting this into Lemma D.1 yields the desired result.  $\square$

## E. Proof of Lemma 3.4

The first expression for the derivative given in (41) is well known and can be derived using integral representations of the operator logarithm (see, e.g., [13]). Now let  $A = \sum_k \mu_k |k\rangle\langle k|$  for an orthonormal eigenbasis  $\{|k\rangle\}_k$  of  $A$ . The claim is thus equivalent to

$$\int_{-\infty}^{\infty} dt \beta_0(t) \mu_i^{-\frac{1}{2}-\frac{it}{2}} \mu_j^{-\frac{1}{2}+\frac{it}{2}} \langle k|H|\ell\rangle = \int_0^{\infty} d\lambda (\mu_i + \lambda)^{-1} (\mu_j + \lambda)^{-1} \langle k|H|\ell\rangle \quad \forall k, \ell. \quad (95)$$

Thus, it suffices to show that

$$\frac{1}{\sqrt{xy}} \int_{-\infty}^{\infty} dt \beta_0(t) \left(\frac{y}{x}\right)^{\frac{it}{2}} = \int_0^{\infty} d\lambda (x+\lambda)^{-1} (y+\lambda)^{-1} \quad \forall x, y > 0. \quad (96)$$

Since  $\beta_0(t)$  is symmetric in  $t$ , we have

$$\frac{1}{\sqrt{xy}} \int_{-\infty}^{\infty} dt \beta_0(t) \left(\frac{y}{x}\right)^{\frac{it}{2}} = \frac{1}{\sqrt{xy}} \int_{-\infty}^{\infty} dt \beta_0(t) \cos\left(\frac{t}{2} \log\left(\frac{y}{x}\right)\right) = \frac{1}{y-x} \log\left(\frac{y}{x}\right), \quad (97)$$

and the claim follows because we also have

$$\int_0^{\infty} d\lambda (x+\lambda)^{-1} (y+\lambda)^{-1} = \frac{1}{y-x} \log\left(\frac{y}{x}\right). \quad (98)$$

## F. Additional Recoverability Bounds

The purpose of this section is to present two additional entropy inequalities that also follow from the multivariate extension of the GT inequality given by Corollary 3.3. These two bounds have been proven before [18, 35].

**Proposition 1.** *Let  $\rho_{AB}$  and  $\sigma_{AB}$  be positive semi-definite matrices on  $\mathcal{H}_A \otimes \mathcal{H}_B$  such that  $\rho_{AB} \ll \sigma_{AB}$  and  $\text{tr } \rho_{AB} = 1$  and let  $\mathcal{R}_{\sigma_{AB}, \text{tr } B}^{[t]}$  be as defined in (53). Then*

$$- \int_{-\infty}^{\infty} dt \beta_0(t) \log F(\rho_{AB}, \mathcal{R}_{\sigma_{AB}, \text{tr } B}^{[t]}(\rho_A)) \leq D(\rho_{AB} \| \sigma_{AB}) - D(\rho_A \| \sigma_A) \quad (99)$$

$$\leq \int_{-\infty}^{\infty} dt \beta_0(t) D_2(\rho_{AB} \| \mathcal{R}_{\sigma_{AB}, \text{tr } B}^{[t]}(\rho_A)), \quad (100)$$

where  $D_2(\rho \| \sigma) := \log \text{tr } \rho^2 \sigma^{-1}$  is Petz' Rényi relative entropy of order 2.

*Proof.* Klein's inequality [37] states that for any Hermitian matrices  $H_1, H_2$  and for any differentiable convex function  $f: \mathbb{R} \rightarrow \mathbb{R}$ , we have  $\text{tr}(f(H_1) - f(H_2) - (H_1 - H_2)f'(H_2)) \geq 0$ . If we apply Klein's inequality with  $f(\cdot) = \exp(\cdot)$ ,  $H_1 = G_1 + G_2$  and  $H_2 = G_1 + \text{id}_B \text{tr } G_2 \exp(G_1)$  we obtain the Peierls–Bogoliubov inequality (see, e.g., [56]) which tells us that for Hermitian matrices  $G_1$  and  $G_2$ , we have

$$-\text{tr } G_2 \exp(G_1) \geq -\log \text{tr } \exp(G_1 + G_2). \quad (101)$$

We first prove (99). Let  $\rho_{AB}$  and  $\sigma_{AB}$  be positive semi-definite matrices on  $\mathcal{H}_A \otimes \mathcal{H}_B$  such that  $\rho_{AB} \ll \sigma_{AB}$  and  $\text{tr } \rho_{AB} = 1$ . For  $G_1 = \log \rho_{AB}$  and  $G_2 = \frac{1}{2}(\log \rho_A \otimes \text{id}_B + \log \sigma_{AB} - \log \sigma_A \otimes \text{id}_B - \log \rho_{AB})$ , this gives

$$D(\rho_{AB} \| \sigma_{AB}) - D(\rho_A \| \sigma_A) \quad (102)$$

$$\geq -2 \log \text{tr } \exp\left(\frac{1}{2}(\log \rho_A \otimes \text{id}_B + \log \sigma_{AB} - \log \sigma_A \otimes \text{id}_B + \log \rho_{AB})\right) \quad (103)$$

$$\geq -2 \int_{-\infty}^{\infty} dt \beta_0(t) \log \left\| \rho_{AB}^{\frac{1+it}{2}} \sigma_{AB}^{\frac{1+it}{2}} \left( \sigma_A^{-\frac{1+it}{2}} \rho_A^{\frac{1+it}{2}} \otimes \text{id}_B \right) \right\|_1 \quad (104)$$

$$= - \int_{-\infty}^{\infty} dt \beta_0(t) \log F\left(\rho_{AB}, \sigma_{AB}^{\frac{1+it}{2}} \left( \sigma_A^{-\frac{1+it}{2}} \rho_A \sigma_A^{-\frac{1-it}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{\frac{1-it}{2}}\right), \quad (105)$$

where the penultimate step uses the extension of the GT inequality from Corollary 3.3 for  $n = 4$  and  $p = 1$ .

It remains to prove (100). Applying the Peierls–Bogoliubov inequality (101) for  $G_1 = \log \rho_{AB}$  and  $G_2 = \log \sigma_A \otimes \text{id}_B + \log \rho_{AB} - \log \rho_A \otimes \text{id}_B - \log \sigma_{AB}$ , we find

$$D(\rho_{AB} \| \sigma_{AB}) - D(\rho_A \| \sigma_A) \leq \log \text{tr } \exp(2 \log \rho_{AB} + \log \sigma_A - \log \rho_A - \log \sigma_{AB}) \quad (106)$$

$$\leq \int_{-\infty}^{\infty} dt \beta_0(t) \log \text{tr } \rho_{AB}^2 \sigma_{AB}^{-\frac{1+it}{2}} \left( \sigma_A^{\frac{1+it}{2}} \rho_A^{-1} \sigma_A^{\frac{1-it}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{-\frac{1-it}{2}} \quad (107)$$

$$= \int_{-\infty}^{\infty} dt \beta_0(t) D_2\left(\rho_{AB} \left\| \sigma_{AB}^{\frac{1+it}{2}} \left( \sigma_A^{-\frac{1+it}{2}} \rho_A \sigma_A^{-\frac{1-it}{2}} \otimes \text{id}_B \right) \sigma_{AB}^{\frac{1-it}{2}} \right\| \right), \quad (108)$$

where the second inequality follows by Corollary 3.3 applied for  $n = 4$  and  $p = 2$ .  $\square$

Following the same line of arguments as in the proof in the proof of Corollary 4.2, (99) can be extended to the case of arbitrary trace-preserving completely positive maps. This then reproduces a result in [35, Section 3].

## References

1. Ahlswede, R., Winter, A.: Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory* **48**(3), 569–579 (2002). doi:[10.1109/18.985947](https://doi.org/10.1109/18.985947)
2. Ando, T., Hiai, F.: Log majorization and complementary Golden–Thompson type inequalities. *Linear Algebra Appl.* **197**, 113–131 (1994). doi:[10.1016/0024-3795\(94\)90484-7](https://doi.org/10.1016/0024-3795(94)90484-7)
3. Araki, H.: Golden–Thompson and Peierls–Bogolubov inequalities for a general von Neumann algebra. *Commun. Math. Phys.* **34**(3), 167–178 (1973). doi:[10.1007/BF01645678](https://doi.org/10.1007/BF01645678)
4. Araki, H.: On an inequality of Lieb and Thirring. *Lett. Math. Phys.* **19**(2), 167–170 (1990). doi:[10.1007/BF01045887](https://doi.org/10.1007/BF01045887)
5. Aubrun, G., Nechita, I.: The multiplicative property characterizes  $\ell_p$  and  $L_p$  norms. *Conflu. Math.* **03**(04), 637–647 (2011). doi:[10.1142/S1793744211000485](https://doi.org/10.1142/S1793744211000485)
6. Beigi, S.: Sandwiched Rényi divergence satisfies data processing inequality. *J. Math. Phys.* **54**(12) (2013). doi:[10.1063/1.4838855](https://doi.org/10.1063/1.4838855)
7. Berta, M., Fawzi, O., Tomamichel, M.: On variational expressions for quantum relative entropies (2015). [arXiv:1512.02615](https://arxiv.org/abs/1512.02615)
8. Berta, M., Lemm, M., Wilde, M.M.: Monotonicity of quantum relative entropy and recoverability. *Quantum Inf. Comput.* **15**(15), 1333–1354 (2015)
9. Berta, M., Tomamichel, M.: The fidelity of recovery is multiplicative. *IEEE Trans. Inf. Theory* **62**(4), 1758–1763 (2016). doi:[10.1109/TIT.2016.2527683](https://doi.org/10.1109/TIT.2016.2527683)
10. Bhatia, R.: *Matrix Analysis*. Springer, Berlin (1997). doi:[10.1007/978-1-4612-0653-8](https://doi.org/10.1007/978-1-4612-0653-8)
11. Brandão, F.G.S.L., Harrow, A.W., Oppenheim, J., Strelchuk, S.: Quantum conditional mutual information, reconstructed states, and state redistribution. *Phys. Rev. Lett.* **115**(5), 050501 (2015). doi:[10.1103/PhysRevLett.115.050501](https://doi.org/10.1103/PhysRevLett.115.050501)
12. Breitenecker, M., Grümmer, H.R.: Note on trace inequalities. *Commun. Math. Phys.* **26**(4), 276–279 (1972). doi:[10.1007/BF01645522](https://doi.org/10.1007/BF01645522)
13. Carlen, E.: *Trace Inequalities and Quantum Entropy: An Introductory Course*. Contemporary Mathematics (2009). doi:[10.1090/conm/529](https://doi.org/10.1090/conm/529)
14. Carlen, E., Lieb, E.: Remainder terms for some quantum entropy inequalities. *J. Math. Phys.* **55**(4), 042201 (2014). doi:[10.1063/1.4871575](https://doi.org/10.1063/1.4871575)
15. Csiszár, I.: The method of types. *IEEE Trans. Inf. Theory* **44**(6), 2505–2523 (1998). doi:[10.1109/18.720546](https://doi.org/10.1109/18.720546)
16. Donald, M.J.: On the relative entropy. *Commun. Math. Phys.* **105**(1), 13–34 (1986). doi:[10.1007/BF01212339](https://doi.org/10.1007/BF01212339)
17. Dupuis, F.: Chain rules for quantum Rényi entropies. *J. Math. Phys.* **56**(2) (2015). doi:[10.1063/1.4907981](https://doi.org/10.1063/1.4907981)
18. Dupuis, F., Wilde, M.M.: Swiveled Rényi entropies. *Quantum Inf. Process.* **15**(3), 1309–1345 (2016). doi:[10.1007/s11128-015-1211-x](https://doi.org/10.1007/s11128-015-1211-x)
19. Eggleston, H.: *Convexity*. Cambridge University Press, Cambridge (1958). doi:[10.1017/CBO9780511566172](https://doi.org/10.1017/CBO9780511566172)
20. Epstein, H.: Remarks on two theorems of E. Lieb. *Commun. Math. Phys.* **31**(4), 317–325 (1973). doi:[10.1007/BF01646492](https://doi.org/10.1007/BF01646492)
21. Fawzi, O., Renner, R.: Quantum conditional mutual information and approximate Markov chains. *Commun. Math. Phys.* **340**(2), 575–611 (2015). doi:[10.1007/s00220-015-2466-x](https://doi.org/10.1007/s00220-015-2466-x)
22. Fuchs, C.A.: *Distinguishability and accessible information in quantum theory*. PhD Thesis, University of New Mexico (1996). [arXiv:quant-ph/9601020](https://arxiv.org/abs/quant-ph/9601020)
23. Golden, S.: Lower bounds for the Helmholtz function. *Phys. Rev.* **137**, B1127–B1128 (1965). doi:[10.1103/PhysRev.137.B1127](https://doi.org/10.1103/PhysRev.137.B1127)
24. Grafakos, L.: *Classical Fourier Analysis*, 2nd edn. Springer, Berlin (2008). doi:[10.1007/978-0-387-09432-8](https://doi.org/10.1007/978-0-387-09432-8)
25. Hansen, F.: Multivariate extensions of the Golden–Thompson inequality. *Ann. Funct. Anal.* **6**(4), 301–310 (2015). doi:[10.15352/afa/06-4-301](https://doi.org/10.15352/afa/06-4-301)
26. Hansen, F., Kjergård Pedersen, G.: Jensen’s inequality for operators and Löwner’s theorem. *Math. Ann.* **258**(3), 229–241 (1982). doi:[10.1007/BF01450679](https://doi.org/10.1007/BF01450679)
27. Hansen, F., Pedersen, G.K.: Jensen’s operator inequality. *Bull. Lond. Math. Soc.* **35**(4), 553–564 (2003). doi:[10.1112/S0024609303002200](https://doi.org/10.1112/S0024609303002200)
28. Hayashi, M.: Optimal sequence of quantum measurements in the sense of Stein’s lemma in quantum hypothesis testing. *J. Phys. A Math. Gen.* **35**(50), 10759 (2002). doi:[10.1088/0305-4470/35/50/307](https://doi.org/10.1088/0305-4470/35/50/307)

29. Hayashi, M. Quantum Information An Introduction. Springer, Berlin (2006). doi:[10.1007/3-540-30266-2](https://doi.org/10.1007/3-540-30266-2)
30. Hiai, F., Mosonyi, M., Petz, D., Bény, C.: Quantum f-divergences and error correction. *Rev. Math. Phys.* **23**(07), 691–747 (2011). doi:[10.1142/S0129055X11004412](https://doi.org/10.1142/S0129055X11004412)
31. Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.* **143**(1), 99–114 (1991). doi:[10.1007/BF02100287](https://doi.org/10.1007/BF02100287)
32. Hiai, F., Petz, D.: The Golden–Thompson trace inequality is complemented. *Linear Algebra Appl.* **181**, 153–185 (1993). doi:[10.1016/0024-3795\(93\)90029-N](https://doi.org/10.1016/0024-3795(93)90029-N)
33. Hirschman, I.I.: A convexity theorem for certain groups of transformations. *J. Anal. Math.* **2**(2), 209–218 (1952). doi:[10.1007/BF02825637](https://doi.org/10.1007/BF02825637)
34. Joachim, W. Linear Operators in Hilbert Spaces. Springer, Berlin (1980). doi:[10.1007/978-1-4612-6027-1](https://doi.org/10.1007/978-1-4612-6027-1)
35. Junge, M., Renner, R., Sutter, D., Wilde, M.M., Winter, A.: Universal recovery from a decrease of quantum relative entropy (2015). [arXiv:1509.07127](https://arxiv.org/abs/1509.07127)
36. Kittaneh, F.: Norm inequalities for certain operator sums. *J. Funct. Anal.* **143**(2), 337–348 (1997). doi:[10.1006/jfan.1996.2957](https://doi.org/10.1006/jfan.1996.2957)
37. Klein, O.: Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre. *Zeitschr. Phys.* **72**(11), 767–775 (1931). doi:[10.1007/BF01341997](https://doi.org/10.1007/BF01341997)
38. Klimek, S., Lesniewski, A.: A Golden–Thompson inequality in supersymmetric quantum mechanics. *Lett. Math. Phys.* **21**(3), 237–244 (1991). doi:[10.1007/BF00420374](https://doi.org/10.1007/BF00420374)
39. Kosaki, H.: An inequality of Araki–Lieb–Thirring (von Neumann algebra case). *Proc. Am. Math. Soc.* **114**, 477–481 (1992). doi:[10.2307/2159671](https://doi.org/10.2307/2159671)
40. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951). doi:[10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694)
41. Lenard, A.: Generalization of the Golden–Thompson inequality. *Indiana Univ. Math. J.* **21**, 457–467 (1971). doi:[10.1512/iumj.1971.21.21036](https://doi.org/10.1512/iumj.1971.21.21036)
42. Li, H., Zhao, D.: An extension of the Golden–Thompson theorem. *J. Inequal. Appl.* **2014**(1), 1–6 (2014). doi:[10.1186/1029-242X-2014-14](https://doi.org/10.1186/1029-242X-2014-14)
43. Lieb, E., Thirring, W.: Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities, in *Studies in Mathematical Physics*. Princeton University Press, pp. 296–303 (1976). doi:[10.1007/3-540-27056-6\\_16](https://doi.org/10.1007/3-540-27056-6_16)
44. Lieb, E.H.: Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.* **11**(3), 267–288 (1973). doi:[10.1016/0001-8708\(73\)90011-X](https://doi.org/10.1016/0001-8708(73)90011-X)
45. Lieb, E.H., Ruskai, M.B.: A fundamental property of quantum-mechanical entropy. *Phys. Rev. Lett.* **30**, 434–436 (1973). doi:[10.1103/PhysRevLett.30.434](https://doi.org/10.1103/PhysRevLett.30.434)
46. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.* **14**(12), 1938–1941 (1973). doi:[10.1063/1.1666274](https://doi.org/10.1063/1.1666274)
47. Lindblad, G.: Completely positive maps and entropy inequalities. *Commun. Math. Phys.* **40**(2), 147–151 (1975). doi:[10.1007/BF01609396](https://doi.org/10.1007/BF01609396)
48. Majenz, C.: Private communication (2016)
49. Müller-Hermes, A., Reeb, D.: Monotonicity of the quantum relative entropy under positive maps (2015). [arXiv:1512.06117](https://arxiv.org/abs/1512.06117)
50. Müller-Hermes, A., Reeb, D., Wolf, M.M.: Positivity of linear maps under tensor powers. *J. Math. Phys.* **57**(1) (2016). doi:[10.1063/1.4927070](https://doi.org/10.1063/1.4927070)
51. Müller-Lennert, M., Dupuis, F., Szeher, O., Fehr, S., Tomamichel, M.: On quantum Rényi entropies: a new generalization and some properties. *J. Math. Phys.* **54**(12) (2013). doi:[10.1063/1.4838856](https://doi.org/10.1063/1.4838856)
52. Petz, D.: Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Commun. Math. Phys.* **105**(1), 123–131 (1986). doi:[10.1007/BF01212345](https://doi.org/10.1007/BF01212345)
53. Petz, D.: A variational expression for the relative entropy. *Commun. Math. Phys.* **114**(2), 345–349 (1988). doi:[10.1007/BF01225040](https://doi.org/10.1007/BF01225040)
54. Petz, D. Quantum Information Theory and Quantum Statistics. Springer, Berlin (2008). doi:[10.1007/978-3-540-74636-2](https://doi.org/10.1007/978-3-540-74636-2)
55. Reed, M., Simon, B.: Functional Analysis. Elsevier, Academic Press, New York (1980)
56. Ruelle, D. Statistical Mechanics: Rigorous Results. World Scientific Publishing Company, Singapore (1999). doi:[10.1142/4090](https://doi.org/10.1142/4090)
57. Ruskai, M.B.: Inequalities for traces on von Neumann algebras. *Commun. Math. Phys.* **26**(4), 280–289 (1972). doi:[10.1007/BF01645523](https://doi.org/10.1007/BF01645523)
58. Segal, I.: Notes towards the construction of nonlinear relativistic quantum fields III. *Bull. Am. Math. Soc.* **75**, 1390–1395 (1969). doi:[10.1090/S0002-9904-1969-12428-6](https://doi.org/10.1090/S0002-9904-1969-12428-6)
59. Simon, B.: Trace Ideals and Their Applications. Cambridge University Press, Cambridge (1979). doi:[10.1090/surv/120](https://doi.org/10.1090/surv/120)
60. Stein, E.M.: Interpolation of linear operators. *Trans. Am. Math. Soc.* **83**(2), 482–492 (1956). doi:[10.2307/1992885](https://doi.org/10.2307/1992885)

61. Stinespring, W.F.: Positive functions on  $C^*$ -algebras. *Proc. Am. Math. Soc.* **6**(211), 211–216 (1955). doi:[10.1090/S0002-9939-1955-0069403-4](https://doi.org/10.1090/S0002-9939-1955-0069403-4)
62. Sutter, D., Fawzi, O., Renner, R.: Universal recovery map for approximate Markov chains. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **472**(2186) (2016). doi:[10.1098/rspa.2015.0623](https://doi.org/10.1098/rspa.2015.0623)
63. Sutter, D., Tomamichel, M., Harrow, A.W.: Strengthened monotonicity of relative entropy via pinched petz recovery map. *IEEE Trans. Inf. Theory* **62**(5), 2907–2913 (2016). doi:[10.1109/TIT.2016.2545680](https://doi.org/10.1109/TIT.2016.2545680)
64. Thompson, C.J.: Inequality with applications in statistical mechanics. *J. Math. Phys.* **6**(11), 1812–1813 (1965). doi:[10.1063/1.1704727](https://doi.org/10.1063/1.1704727)
65. Thompson, C.J.: Inequalities and partial orders on matrix spaces. *Indiana Univ. Math. J.* **21**, 469–480 (1971). doi:[10.1512/iumj.1971.21.21037](https://doi.org/10.1512/iumj.1971.21.21037)
66. Tomamichel, M. *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics, vol. 5. Springer, Berlin (2015). doi:[10.1007/978-3-319-21891-5](https://doi.org/10.1007/978-3-319-21891-5)
67. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12**(4), 389–434 (2011). doi:[10.1007/s10208-011-9099-z](https://doi.org/10.1007/s10208-011-9099-z)
68. Uhlmann, A.: Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Commun. Math. Phys.* **54**(1), 21–32 (1977). doi:[10.1007/BF01609834](https://doi.org/10.1007/BF01609834)
69. Umegaki, H.: Conditional expectation in an operator algebra. *Kodai Math. Seminar Rep.* **14**, 59–85 (1962). doi:[10.2996/kmj/1138844604](https://doi.org/10.2996/kmj/1138844604)
70. Wang, B.-Y., Zhang, F.: Trace and eigenvalue inequalities for ordinary and Hadamard products of positive semidefinite Hermitian matrices. *SIAM J. Matrix Anal. Appl.* **16**(4), 1173–1183 (1995). doi:[10.1137/S0895479893253616](https://doi.org/10.1137/S0895479893253616)
71. Wilde, M.M.: Recoverability in quantum information theory. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **471**(2182), 20150338 (2015). doi:[10.1098/rspa.2015.0338](https://doi.org/10.1098/rspa.2015.0338)

Communicated by R. Seiringer