Quantum Information and Computation, Vol. 15, No. 15&16 (2015) 1333–1354

MONOTONICITY OF QUANTUM RELATIVE ENTROPY AND RECOVERABILITY

MARIO BERTA
Institute for Quantum Information and Matter, California Institute of Technology
Pasadena, California 91125, USA

MARIUS LEMM
Mathematics Department, California Institute of Technology
Pasadena, California 91125, USA

MARK M. WILDE
Hearne Institute for Theoretical Physics, Department of Physics and Astronomy
Center for Computation and Technology, Louisiana State University
Baton Rouge, Louisiana 70808, USA

Received June 3, 2015
Revised August 20, 2015

The relative entropy is a principal measure of distinguishability in quantum information theory, with its most important property being that it is non-increasing with respect to noisy quantum operations. Here, we establish a remainder term for this inequality that quantifies how well one can recover from a loss of information by employing a rotated Petz recovery map. The main approach for proving this refinement is to combine the methods of [Fawzi and Renner, 2014] with the notion of a relative typical subspace from [Bjelakovic and Siegmund-Schultze, 2003]. Our paper constitutes partial progress towards a remainder term which features just the Petz recovery map (not a rotated Petz map), a conjecture which would have many consequences in quantum information theory. A well known result states that the monotonicity of relative entropy with respect to quantum operations is equivalent to each of the following inequalities: strong subadditivity of entropy, concavity of conditional entropy, joint convexity of relative entropy, and monotonicity of relative entropy with respect to partial trace. We show that this equivalence holds true for refinements of all these inequalities in terms of the Petz recovery map. So either all of these refinements are true or all are false.

Keywords: quantum relative entropy, Petz recovery map, strong subadditivity, relative typical subspace
Communicated by: R Cleve & A Harrow

1 Introduction

The Umegaki relative entropy $D(\rho\|\sigma)$ between a density operator ρ and a positive semi-definite operator σ is defined as $\mathrm{Tr}\{\rho[\log\rho - \log\sigma]\}$ whenever $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ and as $+\infty$ otherwise.[a] It is a fundamental information measure in quantum information theory [1], from

[a]Recall that a density operator is a positive semi-definite operator with trace equal to one. Throughout this paper, sometimes our statements apply only to positive definite density operators, and we make it clear when this is so.

which many other information measures, such as entropy, conditional entropy, and mutual information, can be derived (see, e.g., [2]). When σ is a density operator, the relative entropy is a measure of statistical distinguishability and receives an operational interpretation in the context of asymmetric quantum hypothesis testing (known as the quantum Stein's lemma) [3, 4]. Being a good measure of distinguishability, the relative entropy does not increase with respect to quantum processing, as is captured in the following inequality, known as monotonicity of relative entropy [5, 6]:

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)), \tag{1}$$

where $\mathcal{N}$ is a linear completely positive trace preserving (CPTP) map (also referred to as a quantum channel). The inequality is known to be saturated if and only if the following Petz recovery map perfectly recovers ρ from $\mathcal{N}(\rho)$ [7, 8] (see also [9]):

$$\mathcal{R}^{P}_{\sigma,\mathcal{N}}(\cdot) \equiv \sigma^{1/2}\mathcal{N}^{\dagger}\left[(\mathcal{N}(\sigma))^{-1/2}(\cdot)(\mathcal{N}(\sigma))^{-1/2}\right]\sigma^{1/2}, \tag{2}$$

with $\mathcal{N}^{\dagger}$ the adjoint of $\mathcal{N}$. (Observe that the Petz recovery map always perfectly recovers σ from $\mathcal{N}(\sigma)$.) There are several related inequalities, which are known to be equivalent to (1) when σ is a density operator (see, e.g., [10]).[b] One equivalent inequality is the monotonicity of relative entropy with respect to partial trace:

$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_B\|\sigma_B), \tag{3}$$

where ρ_{AB} and σ_{AB} are density operators acting on a tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The operators ρ_B and σ_B result from the partial trace: $\rho_B \equiv \text{Tr}_A\{\rho_{AB}\}$ and $\sigma_B \equiv \text{Tr}_A\{\sigma_{AB}\}$. Another equivalent inequality is the joint convexity of relative entropy:

$$\sum_x p_X(x)\, D(\rho^x\|\sigma^x) \geq D(\overline{\rho}\|\overline{\sigma}), \tag{4}$$

where p_X is a probability distribution, $\{\rho^x\}$ and $\{\sigma^x\}$ are sets of density operators, $\overline{\rho} \equiv \sum_x p_X(x)\,\rho^x$, and $\overline{\sigma} \equiv \sum_x p_X(x)\,\sigma^x$. The interpretation of the above inequality is that distinguishability cannot increase under the loss of the classical label x. One other equivalent inequality is the strong subadditivity of quantum entropy [11, 12]:

$$I(A;B|C)_\omega \equiv D(\omega_{ABC}\|\omega_{AC} \otimes I_B) - D(\omega_{BC}\|\omega_C \otimes I_B) \geq 0, \tag{5}$$

which can be seen as a special case of (1) with $\rho = \omega_{ABC}$, $\sigma = \omega_{AC} \otimes I_B$, and $\mathcal{N} = \text{Tr}_A$, where ω_{ABC} is a tripartite density operator acting on the tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. A final equivalent inequality that we mention is the concavity of conditional entropy [12]:

$$H(A|B)_{\overline{\rho}} \geq \sum_x p_X(x)\, H(A|B)_{\rho^x}, \tag{6}$$

[b] The notion that two statements which are known to be true are 'equivalent' of course does not make strict sense logically. So when we say that 'A is equivalent to B' for two statements A and B which are already known to be true (for us A and B will always be some kind of entropy inequalities), we in fact mean the softer (but standard) notion that, if one assumes A, then there exists a relatively direct proof for B and vice versa.

where p_X is a probability distribution, $\{\rho^x_{AB}\}$ is a set of density operators, $\overline{\rho}_{AB} \equiv \sum_x p_X(x)\, \rho^x_{AB}$, and the conditional entropy $H(A|B)_\sigma \equiv -D(\sigma_{AB} \| I_A \otimes \sigma_B)$.

The above inequalities have been critical to the development of quantum information theory. In fact, since so much of quantum information theory relies on these inequalities and given that they are equivalent and apply universally for any states and channels, they are often considered to constitute a fundamental law of quantum information theory. In light of this, we might wonder if there could be refinements of the above inequalities in the form of "remainder terms." While a number of works pursued this direction [13, 14, 15, 16, 17, 18, 19, 2, 20, 21, 22], a breakthrough paper established the following remainder term for strong subadditivity [23]:

$$I(A;B|C)_\omega \geq -\log F\left(\omega_{ABC}, \left(\mathcal{V}_{AC} \circ \mathcal{R}^P_{C\to AC} \circ \mathcal{U}_C\right)(\omega_{BC})\right), \tag{7}$$

where $F(\tau,\varsigma) \equiv \|\sqrt{\tau}\sqrt{\varsigma}\|_1^2$ is the quantum fidelity between positive semi-definite operators τ and ς [24], $\mathcal{U}_C$ and $\mathcal{V}_{AC}$ are unitary channels defined in terms of some unitary operators U_C and V_{AC} as

$$\mathcal{U}_C(\cdot) \equiv U_C(\cdot)\, U_C^\dagger, \tag{8}$$

$$\mathcal{V}_{AC}(\cdot) \equiv V_{AC}(\cdot)\, V_{AC}^\dagger, \tag{9}$$

and $\mathcal{R}^P_{C\to AC}$ is the following Petz recovery map:

$$\mathcal{R}^P_{C\to AC}(\cdot) \equiv \omega_{AC}^{1/2}\omega_C^{-1/2}(\cdot)\,\omega_C^{-1/2}\omega_{AC}^{1/2}. \tag{10}$$

In the present paper, our first contribution is to combine the methods of [23] and the notion of a relative typical subspace from [25, Pages 4-5] in order to establish the following remainder term for the inequality in (1):

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -\log F\left(\rho, \left(\mathcal{V} \circ \mathcal{R}^P_{\sigma,\mathcal{N}} \circ \mathcal{U}\right)(\mathcal{N}(\rho))\right), \tag{11}$$

where $\mathcal{U}$ is a unitary channel acting on the output space of $\mathcal{N}$, $\mathcal{R}^P_{\sigma,\mathcal{N}}$ is the Petz recovery map defined in (2), and $\mathcal{V}$ is a unitary channel acting on the input space of $\mathcal{N}$. Thus, the refinement in (11) quantifies how well one can recover ρ from $\mathcal{N}(\rho)$ by employing the "rotated Petz recovery map" $\mathcal{V} \circ \mathcal{R}^P_{\sigma,\mathcal{N}} \circ \mathcal{U}$. This result is stated formally as Corollary 1 and can be understood as a generalization of (7). We establish a similar refinement of the inequality in (3), stated formally as Theorem 1. Given that the original inequalities without remainder terms have found wide use in quantum information theory, we expect the refinements with remainder terms presented here to find use in some applications of the original inequalities, perhaps in the context of quantum error correction [26, 27, 28, 29, 30] or thermodynamics [31, 32]. Note that the refinement in (7) has already been helpful in improving our understanding of some quantum correlation measures [14, 21, 22, 33].

It would be very useful for applications if the aforementioned refinements of relative entropy inequalities held for the Petz recovery map (and not merely for a rotated Petz recovery map), i.e., if they were of the following form:

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -\log F\left(\rho, \mathcal{R}^{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))\right), \tag{12}$$

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq -\log F\left(\rho_{AB}, \sigma_{AB}^{1/2}\sigma_B^{-1/2}\rho_B\sigma_B^{-1/2}\sigma_{AB}^{1/2}\right), \tag{13}$$

$$\sum_x p_X(x) D(\rho^x\|\sigma^x) - D(\overline{\rho}\|\overline{\sigma}) \geq -2\log\sum_x p_X(x)\sqrt{F(\rho^x, (\sigma^x)^{\frac{1}{2}}(\overline{\sigma})^{-\frac{1}{2}}\,\overline{\rho}\,(\overline{\sigma})^{-\frac{1}{2}}(\sigma^x)^{\frac{1}{2}})}, \tag{14}$$

$$I(A;B|C)_\omega \geq -\log F\left(\omega_{ABC}, \omega_{AC}^{1/2}\omega_C^{-1/2}\omega_{BC}\omega_C^{-1/2}\omega_{AC}^{1/2}\right), \tag{15}$$

$$H(A|B)_{\overline{\rho}} - \sum_x p_X(x) H(A|B)_{\rho^x} \geq -2\log\sum_x p_X(x)\sqrt{F\left(\rho^x_{AB}, \overline{\rho}_{AB}^{1/2}\overline{\rho}_B^{-1/2}\rho^x_B\overline{\rho}_B^{-1/2}\overline{\rho}_{AB}^{1/2}\right)}. \tag{16}$$

In [20, Definition 25], a Rényi information measure was defined to generalize relative entropy differences. The inequalities (12)-(16) stated above would follow from the monotonicity of this Rényi information measure with respect to the Rényi parameter (see [20, Conjecture 26], [20, Consequences 27 and 28]). A weaker form of (12) in terms of trace distance on the right-hand side was first conjectured in [19, Eq. (4.7)].

Our second contribution in this paper is to show that slightly weaker forms of these inequalities, featuring instead the square of the Bures distance [34] $D_B^2(\omega,\tau) \equiv 2(1-\sqrt{F(\omega,\tau)})$ on the right-hand side, are all equivalent (observe that $-\log(F) \geq 2(1-\sqrt{F})$). So either all of these refinements are true or all are false. It remains an important open question to determine which is the case. This second contribution is in principle conjectural, but we believe it is nonetheless important, for two reasons: (1) Obviously, it reduces the work of proving (or even disproving) entropy inequalities with Petz remainder terms to single cases, which can be chosen according to convenience. (2) It furthers the evidence that the Petz remainder term is the natural one.

The next section recalls the notion of a relative typical subspace and the remaining sections give proofs of our claims.

2 Relative typical subspace

We begin by reviewing the notion of a relative typical subspace from [25, Pages 4-5]. Consider spectral decompositions of a density operator ρ and a positive semi-definite operator σ acting on a finite-dimensional Hilbert space, such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$:

$$\rho = \sum_x p_X(x)|\psi_x\rangle\langle\psi_x|, \tag{17}$$

$$\sigma = \sum_y f_Y(y)|\phi_y\rangle\langle\phi_y|. \tag{18}$$

Let us define the relative typical subspace $T^{\delta,n}_{\rho|\sigma}$ for $\delta > 0$ and integer $n \geq 1$ as

$$T^{\delta,n}_{\rho|\sigma} \equiv \text{span}\left\{|\phi_{y^n}\rangle : \left|-\frac{1}{n}\log(f_{Y^n}(y^n)) + \text{Tr}\{\rho\log\sigma\}\right| \leq \delta\right\}, \tag{19}$$

where

$$y^n \equiv y_1 \cdots y_n, \tag{20}$$

$$f_{Y^n}(y^n) \equiv \prod_{i=1}^{n} f_Y(y_i), \tag{21}$$

$$|\phi_{y^n}\rangle \equiv |\phi_{y_1}\rangle \otimes \cdots \otimes |\phi_{y_n}\rangle. \tag{22}$$

We will overload the notation $T_{\rho|\sigma}^{\delta,n}$ to refer also to the following classical typical set:

$$T_{\rho|\sigma}^{\delta,n} \equiv \left\{ y^n : \left| -\frac{1}{n} \log(f_{Y^n}(y^n)) + \mathrm{Tr}\{\rho \log \sigma\} \right| \leq \delta \right\}, \tag{23}$$

with it being clear from the context whether the relative typical subspace or set is being employed.

Let the projection operator corresponding to the relative typical subspace $T_{\rho|\sigma}^{\delta,n}$ be called $\Pi_{\rho|\sigma,\delta}^{n}$. Consider that

$$\mathrm{Tr}\{\rho \log \sigma\} = \mathrm{Tr}\left\{\rho \log \left(\sum_y f_Y(y) |\phi_y\rangle\langle\phi_y| \right)\right\} \tag{24}$$

$$= \sum_y \langle\phi_y| \rho |\phi_y\rangle \log f_Y(y). \tag{25}$$

Defining

$$p_{\widetilde{Y}}(y) \equiv \langle\phi_y| \rho |\phi_y\rangle, \tag{26}$$

we can then write

$$\mathrm{Tr}\{\rho \log \sigma\} = \sum_y p_{\widetilde{Y}}(y) \log f_Y(y) \tag{27}$$

$$= \mathbb{E}_{\widetilde{Y}} \left\{ \log f_Y(\widetilde{Y}) \right\}. \tag{28}$$

With this in mind, we can now calculate

$$\mathrm{Tr}\left\{\Pi_{\rho|\sigma,\delta}^{n} \rho^{\otimes n}\right\} = \sum_{y^n \in T_{\rho|\sigma}^{\delta,n}} \langle\phi_{y^n}| \rho^{\otimes n} |\phi_{y^n}\rangle \tag{29}$$

$$= \sum_{y^n \in T_{\rho|\sigma}^{\delta,n}} p_{\widetilde{Y}^n}(y^n) \tag{30}$$

$$= \Pr_{\widetilde{Y}^n} \left\{ \widetilde{Y}^n \in T_{\rho|\sigma}^{\delta,n} \right\}. \tag{31}$$

Based on the above reductions, and due to the notion of typicality with respect to the subspace $T_{\rho|\sigma}^{\delta,n}$ defined in (19), it follows from the law of large numbers that, for a given small real number $\varepsilon \in (0,1)$, and a sufficiently large value of n, $\mathrm{Tr}\{\Pi_{\rho|\sigma,\delta}^{n} \rho^{\otimes n}\} \geq 1 - \varepsilon$. In fact, the convergence $\lim_{n\to\infty} \mathrm{Tr}\{\Pi_{\rho|\sigma,\delta}^{n} \rho^{\otimes n}\} = 1$ can be taken exponentially fast in n for a constant δ by employing the Hoeffding inequality [35].

3 Remainder term for monotonicity of relative entropy with respect to partial trace

Theorem 1 *Let ρ_{AB} be a density operator, σ_{AB} be a positive semi-definite operator, both acting on a finite-dimensional tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, such that* $\text{supp}(\rho_{AB}) \subseteq \text{supp}(\sigma_{AB})$*, $\sigma_B \equiv \text{Tr}_A\{\sigma_{AB}\}$ is positive definite, and $\rho_B \equiv \text{Tr}_A\{\rho_{AB}\}$. Then the following inequality refines monotonicity of relative entropy with respect to partial trace:*

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq -\log F\left(\rho_{AB}, \left(\mathcal{V}_{AB} \circ \mathcal{R}^P_{B\to AB} \circ \mathcal{U}_B\right)(\rho_B)\right), \quad (32)$$

for unitary channels $\mathcal{U}_B$ and $\mathcal{V}_{AB}$ defined in terms of some unitary operators U_B and V_{AB} as

$$\mathcal{U}_B(\cdot) \equiv U_B(\cdot)U_B^\dagger, \quad (33)$$

$$\mathcal{V}_{AB}(\cdot) \equiv V_{AB}(\cdot)V_{AB}^\dagger, \quad (34)$$

and with $\mathcal{R}^P_{B\to AB}$ the CPTP Petz recovery map:

$$\mathcal{R}^P_{B\to AB}(\cdot) \equiv \sigma_{AB}^{1/2}\sigma_B^{-1/2}(\cdot)\sigma_B^{-1/2}\sigma_{AB}^{1/2}. \quad (35)$$

Proof. Our proof of Theorem 1 proceeds very similarly to the proof of [23, Theorem 5.1], with only a few modifications. We give a full proof for completeness. Our proof makes use of Lemmas 2.3, 4.2, B.2, B.6, and B.7 from [23]. For convenience of the reader, we recall these statements in Appendix 1.

The expression on the left-hand side of (32) is equivalent to

$$-H(A|B)_\rho - \text{Tr}\{\rho_{AB}\log\sigma_{AB}\} + \text{Tr}\{\rho_B\log\sigma_B\}, \quad (36)$$

where $H(A|B)_\rho \equiv H(AB)_\rho - H(B)_\rho$ is the conditional entropy and the entropy is defined as $H(\omega) \equiv -\text{Tr}\{\omega\log\omega\}$. So we need the relative typical projectors $\Pi^n_{\rho_{AB}|\sigma_{AB},\delta}$ and $\Pi^n_{\rho_B|\sigma_B,\delta}$ defined in Section 2. Abbreviate these as Π^n_{AB} and Π^n_B, respectively.

We begin by defining

$$\mathcal{W}_n(X_{A^nB^n}) \equiv \Pi^n_{AB}\Pi^n_B X_{A^nB^n}\Pi^n_B\Pi^n_{AB}. \quad (37)$$

We employ the shorthand $\mathcal{W}_n(X_{B^n}) \equiv \mathcal{W}_n\left(I_A^{\otimes n} \otimes X_{B^n}\right)$ throughout. Consider from the gentle measurement lemma [36], properties of the trace norm, and relative typicality that

$$\begin{aligned}\text{Tr}\left\{\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\right\} &= \text{Tr}\left\{\Pi^n_{AB}\Pi^n_B\rho_{AB}^{\otimes n}\Pi^n_B\right\} && (38)\\ &\geq \text{Tr}\left\{\Pi^n_{AB}\rho_{AB}^{\otimes n}\right\} - \left\|\Pi^n_B\rho_{AB}^{\otimes n}\Pi^n_B - \rho_{AB}^{\otimes n}\right\|_1 && (39)\\ &\geq 1-\eta, && (40)\end{aligned}$$

where η is an arbitrarily small positive number for sufficiently large n. So we apply [23, Lemma 2.3] to find that

$$\begin{aligned}D\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\|\mathcal{W}_n\left(\rho_B^{\otimes n}\right)\right) &\leq n\left(D(\rho_{AB}\|I_A\otimes\rho_B) + \frac{\delta}{2}\right) && (41)\\ &= n\left(-H(A|B)_\rho + \frac{\delta}{2}\right), && (42)\end{aligned}$$

where the above inequality holds for sufficiently large n. A well-known relation between the root fidelity $\sqrt{F}(\omega,\tau) \equiv \|\sqrt{\omega}\sqrt{\tau}\|_1$ and relative entropy [23, Lemma B.2] then gives that

$$\frac{1}{\mathrm{Tr}\left\{\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\right\}}\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \mathcal{W}_n\left(\rho_B^{\otimes n}\right)\right) \geq 2^{\frac{1}{2}n\left(H(A|B)_\rho - \frac{\delta}{2}\right)}. \tag{43}$$

Use [23, Lemma B.6] to remove the projector Π_{AB}^n from the second argument, so that

$$\frac{1}{\mathrm{Tr}\left\{\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\right\}}\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_B^n\rho_B^{\otimes n}\Pi_B^n\right) \geq 2^{\frac{1}{2}n\left(H(A|B)_\rho - \frac{\delta}{2}\right)}, \tag{44}$$

and the trace term can be eliminated at the expense of decreasing the exponent by a constant times n:

$$\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_B^n\rho_B^{\otimes n}\Pi_B^n\right) \geq 2^{\frac{1}{2}n\left(H(A|B)_\rho - \delta\right)}. \tag{45}$$

Let an eigendecomposition of $\sigma_B^{\otimes n}$ be given as

$$\sigma_B^{\otimes n} = \sum_{s\in S_n} s\Pi_s, \tag{46}$$

where S_n is the set of eigenvalues of $\sigma_B^{\otimes n}$. By defining

$$S_{n,\delta} \equiv \left\{s \in S_n : \left|-\frac{1}{n}\log(s) + \mathrm{Tr}\{\rho_B \log\sigma_B\}\right| \leq \delta\right\}, \tag{47}$$

we see from (19) and the definition of Π_B^n that

$$\Pi_B^n = \sum_{s\in S_{n,\delta}} \Pi_s. \tag{48}$$

Furthermore, it follows from a trivial combinatorial consideration that $|S_{n,\delta}| \leq \mathrm{poly}(n)$. Then consider that $\sum_s \Pi_s = I$ and apply [23, Lemma B.7] to get

$$\begin{aligned}
\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_B^n\rho_B^{\otimes n}\Pi_B^n\right) &\leq \sum_{s\in S_n} \sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_s\Pi_B^n\rho_B^{\otimes n}\Pi_B^n\Pi_s\right) &(49)\\
&= \sum_{s\in S_{n,\delta}} \sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_s\rho_B^{\otimes n}\Pi_s\right) &(50)\\
&\leq |S_{n,\delta}| \max_{s\in S_{n,\delta}} \sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_s\rho_B^{\otimes n}\Pi_s\right), &(51)
\end{aligned}$$

where (50) follows because $\Pi_s\Pi_B^n = \Pi_s$ if $s \in S_{n,\delta}$ and it is equal to zero otherwise. So we find that there exists an s such that

$$\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_B^n\rho_B^{\otimes n}\Pi_B^n\right) \leq \mathrm{poly}(n)\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right), \Pi_s\rho_B^{\otimes n}\Pi_s\right). \tag{52}$$

From the definition of Π_s we can write

$$\Pi_s = \sqrt{s}\left(\sigma_B^{-1/2}\right)^{\otimes n}\Pi_s. \tag{53}$$

From the definition of $S_{n,\delta}$, we have that

$$\sqrt{s} \leq 2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_B\log\sigma_B\}+\delta]}, \tag{54}$$

giving that

$$
\begin{aligned}
&\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right),\Pi_s\rho_B^{\otimes n}\Pi_s\right)\\
&=\sqrt{s}\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right),\left(\sigma_B^{-1/2}\right)^{\otimes n}\Pi_s\left(\rho_B^{\otimes n}\right)\Pi_s\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) && (55)\\
&\leq 2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_B\log\sigma_B\}+\delta]}\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right),\left(\sigma_B^{-1/2}\right)^{\otimes n}\Pi_s\rho_B^{\otimes n}\Pi_s\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) && (56)\\
&=2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_B\log\sigma_B\}+\delta]}\sqrt{F}\left(\Pi_s\left(\sigma_B^{-1/2}\right)^{\otimes n}\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\left(\sigma_B^{-1/2}\right)^{\otimes n}\Pi_s,\rho_B^{\otimes n}\right), && (57)
\end{aligned}
$$

where the last equality is from [23, Lemma B.6]. Now, by [23, Lemma 4.2], there exists a unitary U_B such that[c]

$$
\begin{aligned}
&\sqrt{F}\left(\Pi_s\left(\sigma_B^{-1/2}\right)^{\otimes n}\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\left(\sigma_B^{-1/2}\right)^{\otimes n}\Pi_s,\rho_B^{\otimes n}\right)\\
&\leq \mathrm{poly}\left(n\right)\sqrt{F}\left(\left(\sigma_B^{-1/2}\right)^{\otimes n}\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right)\left(\sigma_B^{-1/2}\right)^{\otimes n},U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\right) && (58)\\
&=\mathrm{poly}\left(n\right)\sqrt{F}\left(\mathcal{W}_n\left(\rho_{AB}^{\otimes n}\right),\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right). && (59)
\end{aligned}
$$

The equality above follows by applying [23, Lemma B.6]. Combining everything up until now, we get

$$
\begin{aligned}
&2^{\frac{1}{2}n\left(H(A|B)_\rho-\mathrm{Tr}\{\rho_B\log\sigma_B\}-2\delta\right)}\\
&\quad\leq \mathrm{poly}\left(n\right)\sqrt{F}\left(\Pi_{AB}^n\Pi_B^n\rho_{AB}^{\otimes n}\Pi_B^n\Pi_{AB}^n,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right). && (60)
\end{aligned}
$$

Let an eigendecomposition of $\sigma_{AB}^{\otimes n}$ be given as

$$\sigma_{AB}^{\otimes n}=\sum_{p\in P_n}p\Pi_p, \tag{61}$$

and

$$\Pi_{AB}^n=\sum_{p\in P_{n,\delta}}\Pi_p, \tag{62}$$

where these developments follow the same reasoning as (46)-(48). Now we continue with the fact that $\sum_{p\in P_n}\Pi_p=I$ and [23, Lemma B.7] to get that

$$
\begin{aligned}
&\sqrt{F}\left(\Pi_{AB}^n\Pi_B^n\rho_{AB}^{\otimes n}\Pi_B^n\Pi_{AB}^n,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right)\\
&\leq\sum_{p\in P_n}\sqrt{F}\left(\Pi_p\Pi_{AB}^n\Pi_B^n\rho_{AB}^{\otimes n}\Pi_B^n\Pi_{AB}^n\Pi_p,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) && (63)\\
&=\sum_{p\in P_{n,\delta}}\sqrt{F}\left(\Pi_p\Pi_B^n\rho_{AB}^{\otimes n}\Pi_B^n\Pi_p,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) && (64)\\
&\leq|P_{n,\delta}|\max_{p\in P_{n,\delta}}\sqrt{F}\left(\Pi_p\Pi_B^n\rho_{AB}^{\otimes n}\Pi_B^n\Pi_p,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right). && (65)
\end{aligned}
$$

[c]Note that the unitary U_B depends on n, but we suppress this in the notation for simplicity.

Then there exists a p such that

$$\sqrt{F}\left(\Pi^n_{AB}\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi^n_{AB},\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right)$$
$$\leq \mathrm{poly}\left(n\right)\sqrt{F}\left(\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right). \quad (66)$$

From the definition of Π_p we have that

$$\Pi_p=\frac{1}{\sqrt{p}}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\Pi_p, \quad (67)$$

with $\sqrt{p}\geq 2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_{AB}\log\sigma_{AB}\}-\delta]}$. Then by defining $K\equiv 2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_{AB}\log\sigma_{AB}\}-\delta]}/\sqrt{p}$, we have that

$$2^{\frac{1}{2}n[\mathrm{Tr}\{\rho_{AB}\log\sigma_{AB}\}-\delta]}\sqrt{F}\left(\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p,\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right)$$
$$=K\ \sqrt{F}\left(\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p\left(\sigma_{AB}^{1/2}\right)^{\otimes n},\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) \quad (68)$$
$$\leq\sqrt{F}\left(\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p\left(\sigma_{AB}^{1/2}\right)^{\otimes n},\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\right) \quad (69)$$
$$=\sqrt{F}\left(\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p,\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\right). \quad (70)$$

Now by [23, Lemma 4.2], there exists a unitary V_{AB} such that[d]

$$\sqrt{F}\left(\Pi_p\Pi^n_B\rho^{\otimes n}_{AB}\Pi^n_B\Pi_p,\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\right)\leq$$
$$\mathrm{poly}\left(n\right)\sqrt{F}\left(\rho^{\otimes n}_{AB},V_{AB}^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(V_{AB}^{\otimes n}\right)^\dagger\right). \quad (71)$$

Putting everything together, we get that

$$2^{\frac{1}{2}n\left(H(A|B)_\rho-\mathrm{Tr}\{\rho_B\log\sigma_B\}+\mathrm{Tr}\{\rho_{AB}\log\sigma_{AB}\}-3\delta\right)}$$
$$\leq\mathrm{poly}\left(n\right)\sqrt{F}\left(\rho^{\otimes n}_{AB},V_{AB}^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(\sigma_B^{-1/2}\right)^{\otimes n}U_B^{\otimes n}\rho_B^{\otimes n}\left(U_B^{\otimes n}\right)^\dagger\left(\sigma_B^{-1/2}\right)^{\otimes n}\left(\sigma_{AB}^{1/2}\right)^{\otimes n}\left(V_{AB}^{\otimes n}\right)^\dagger\right) \quad (72)$$
$$=\mathrm{poly}\left(n\right)\left[F\left(\rho_{AB},V_{AB}\sigma_{AB}^{1/2}\sigma_B^{-1/2}U_B\rho_BU_B^\dagger\sigma_B^{-1/2}\sigma_{AB}^{1/2}V_{AB}^\dagger\right)\right]^n \quad (73)$$
$$\leq\mathrm{poly}\left(n\right)\left[\max_{U_B,V_{AB}}F\left(\rho_{AB},V_{AB}\sigma_{AB}^{1/2}\sigma_B^{-1/2}U_B\rho_BU_B^\dagger\sigma_B^{-1/2}\sigma_{AB}^{1/2}V_{AB}^\dagger\right)\right]^n. \quad (74)$$

[d]Note that the unitary V_{AB} depends on n, but we suppress this in the notation for simplicity.

The equality follows because the fidelity is multiplicative with respect to tensor products. In the last line above, we take a maximization over all unitaries in order to remove the dependence of the unitaries on n. Taking the n^{th} root of the last line above, we find that there exists a V_{AB} and U_B such that

$$2^{\frac{1}{2}\left(H(A|B)_{\rho}-\text{Tr}\{\rho_B \log \sigma_B\}+\text{Tr}\{\rho_{AB}\log\sigma_{AB}\}-3\delta\right)} \\ \leq \sqrt[n]{\text{poly}(n)}\sqrt{F}\left(\rho_{AB}, V_{AB}\sigma_{AB}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^{\dagger}\sigma_B^{-1/2}\sigma_{AB}^{1/2}V_{AB}^{\dagger}\right). \quad (75)$$

By taking the limit as n becomes large, using the fact that

$$-\left[H(A|B)_{\rho} - \text{Tr}\{\rho_B \log \sigma_B\} + \text{Tr}\{\rho_{AB}\log\sigma_{AB}\}\right] = D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B), \quad (76)$$

and noting that $\delta > 0$ was arbitrary, this finally yields the desired inequality

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq -\log F\left(\rho_{AB}, V_{AB}\sigma_{AB}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^{\dagger}\sigma_B^{-1/2}\sigma_{AB}^{1/2}V_{AB}^{\dagger}\right). \quad (77)$$

□

Remark 2 *Suppose in Theorem 1 that σ_{AB} is a density operator. It remains open to quantify the performance of the rotated Petz recovery map $\mathcal{V}_{AB}\circ\mathcal{R}^{P}_{B\to AB}\circ\mathcal{U}_B$ on the reduced state σ_B. In particular, if the unitary channels $\mathcal{U}_B$ and $\mathcal{V}_{AB}$ were not necessary (with each instead being equal to the identity channel), then it would be possible to do so. This form of the recovery map was previously conjectured in [20, Consequence 27] in terms of the following inequality:*

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq -\log F\left(\rho_{AB}, \mathcal{R}^{P}_{B\to AB}(\rho_B)\right). \quad (78)$$

If this conjecture is true, then one could perform the Petz recovery map on system B and be guaranteed a perfect recovery of σ_{AB} if the state of B is σ_B, while having a performance limited by (78) *if the state of B is ρ_B. By a modification of the proof of Theorem 1, one can also establish the following lower bound:*

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq -\log F\left(\rho_{AB}, \sigma_{AB}^{1/2}\bar{V}_{AB}\bar{U}_B\sigma_B^{-1/2}\rho_B\sigma_B^{-1/2}\bar{U}_B^{\dagger}\bar{V}_{AB}^{\dagger}\sigma_{AB}^{1/2}\right), \quad (79)$$

for some unitaries $\bar{U}_B$ and $\bar{V}_{AB}$. The completely positive map $\sigma_{AB}^{1/2}\bar{V}_{AB}\bar{U}_B\sigma_B^{-1/2}(\cdot)\sigma_B^{-1/2}\bar{U}_B^{\dagger}\bar{V}_{AB}^{\dagger}\sigma_{AB}^{1/2}$ recovers σ_{AB} perfectly from σ_B, while having a performance limited by (79) *when recovering ρ_{AB} from ρ_B. It is however unclear whether this map is trace preserving.*

4 Remainder term for monotonicity of relative entropy

Corollary 1 *Let ρ_S be a density operator and σ_S be a positive semi-definite operator, both acting on a Hilbert space $\mathcal{H}_S$ and such that* $\text{supp}(\rho_S) \subseteq \text{supp}(\sigma_S)$. *Let $\mathcal{N}_{S\to B}$ be a CPTP map taking density operators acting on $\mathcal{H}_S$ to density operators acting on $\mathcal{H}_B$ and such that $\mathcal{N}_{S\to B}(\sigma_S)$ is a positive definite operator. Then the following inequality refines monotonicity of relative entropy:*

$$D(\rho_S\|\sigma_S) - D(\mathcal{N}_{S\to B}(\rho_S)\|\mathcal{N}_{S\to B}(\sigma_S)) \geq -\log F\left(\rho_S, \left(\mathcal{V}_S\circ\mathcal{R}^{P}_{\sigma,\mathcal{N}}\circ\mathcal{U}_B\right)(\mathcal{N}_{S\to B}(\rho_S))\right), \quad (80)$$

for unitary channels $\mathcal{U}_B$ *and* $\mathcal{V}_S$ *defined in terms of some unitary operators* U_B *and* V_S *as*

$$\mathcal{U}_B(\cdot) \equiv U_B(\cdot)\, U_B^{\dagger}, \tag{81}$$

$$\mathcal{V}_S(\cdot) \equiv V_S(\cdot)\, V_S^{\dagger}, \tag{82}$$

and with $\mathcal{R}^P_{\sigma,\mathcal{N}}$ *the CPTP Petz recovery map:*

$$\mathcal{R}^P_{\sigma,\mathcal{N}}(\cdot) \equiv \sigma_S^{1/2}\mathcal{N}^{\dagger}\left[\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}(\cdot)\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}\right]\sigma_S^{1/2}, \tag{83}$$

where $\mathcal{N}^{\dagger}$ *is the adjoint of* $\mathcal{N}_{S\to B}$*.*

Proof. We begin by recalling that any quantum channel can be realized by tensoring in an ancilla system prepared in a fiducial state, acting with a unitary on the input and ancilla, and then performing a partial trace [38]. That is, for any channel $\mathcal{N}_{S\to B}$, there exists a unitary $W_{SE'\to BE}$ with input systems SE' and output systems BE such that

$$\mathcal{N}_{S\to B}(\rho_S) = \mathrm{Tr}_E\left\{W_{SE'\to BE}\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}_{SE'\to BE}\right\}. \tag{84}$$

For simplicity, we abbreviate the unitary $W_{SE'\to BE}$ as W in what follows. Let ρ_{BE} and σ_{BE} be defined as

$$\rho_{BE} \equiv W\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}, \tag{85}$$

$$\sigma_{BE} \equiv W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}, \tag{86}$$

so that

$$\mathcal{N}_{S\to B}(\rho_S) = \rho_B, \qquad \mathcal{N}_{S\to B}(\sigma_S) = \sigma_B. \tag{87}$$

The Kraus operators of $\mathcal{N}_{S\to B}$ are given as

$$\mathcal{N}_{S\to B}(\rho_S) = \sum_i \langle i|_E\, W\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}\,|i\rangle_E \tag{88}$$

$$= \sum_i \langle i|_E\, W\,|0\rangle_{E'}\,\rho_S\,\langle 0|_{E'}\,W^{\dagger}\,|i\rangle_E\,, \tag{89}$$

so that the adjoint map is given by

$$\mathcal{N}^{\dagger}(\omega_B) = \sum_i \langle 0|_{E'}\,W^{\dagger}\,|i\rangle_E\,\omega_B\,\langle i|_E\,W\,|0\rangle_{E'}\,. \tag{90}$$

Furthermore, we have that

$$D\left(\rho_S\|\sigma_S\right) - D\left(\mathcal{N}_{S\to B}(\rho_S)\|\mathcal{N}_{S\to B}(\sigma_S)\right)$$

$$= D\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\|\sigma_S\otimes|0\rangle\langle 0|_{E'}\right) - D\left(\rho_B\|\sigma_B\right) \tag{91}$$

$$= D\left(W\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}\|W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)W^{\dagger}\right) - D\left(\rho_B\|\sigma_B\right) \tag{92}$$

$$= D\left(\rho_{BE}\|\sigma_{BE}\right) - D\left(\rho_B\|\sigma_B\right). \tag{93}$$

Applying Theorem 1, we know that a lower bound on (93) is

$$-\log F\left(\rho_{BE}, V_{BE}\sigma_{BE}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^{\dagger}\sigma_B^{-1/2}\sigma_{BE}^{1/2}V_{BE}^{\dagger}\right), \tag{94}$$

for some unitaries V_{BE} and U_B. Without loss of generality, V_{BE} can be assumed to be an isometry on the image of $W_{SE'\to BE}|0\rangle_{E'}$. We justify this as follows. Let P_n denote the support projection of $\rho_{AB}^{\otimes n}$. By (71), since the supports of $\Pi_b \Pi_{B^n} P_n$, $\rho_{AB}^{\otimes n}$, and $\left[\sigma_{AB}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^\dagger \sigma_B^{-1/2}\sigma_{AB}^{1/2}\right]^{\otimes n}$ are all contained in the support of $\sigma_{AB}^{\otimes n}$, one can apply [23, Lemma 4.2] on the Hilbert space $[\operatorname{supp}(\sigma_{AB})]^{\otimes n} = \operatorname{supp}\left(\sigma_{AB}^{\otimes n}\right)$ to obtain a unitary V_{AB} on this space, which may be extended to a unitary on the space $\mathcal{H}_{AB}^{\otimes n}$ in an arbitrary way. Hence, the maximization in (74) can be restricted to unitaries V_{AB} that are isometries on the support of σ_{AB}. Thus, we indeed have that V_{BE} is an isometry on the support of σ_{BE}, which can be extended to an isometry on the image of $W_{SE'\to BE}|0\rangle_{E'}$.

Let us now unravel the term $\sigma_{BE}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^\dagger\sigma_B^{-1/2}\sigma_{BE}^{1/2}$ in the second argument above. Letting

$$\omega_B \equiv \left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2} U_B \mathcal{N}_{S\to B}(\rho_S)\, U_B^\dagger \left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}, \tag{95}$$

we then have that

$$\begin{aligned}
&\sigma_{BE}^{1/2}\sigma_B^{-1/2}U_B\rho_B U_B^\dagger\sigma_B^{-1/2}\sigma_{BE}^{1/2} \\
&= \left(W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\right)^{1/2}\omega_B\left(W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\right)^{1/2} &(96)\\
&= W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)^{1/2}W^\dagger\omega_B W\left(\sigma_S\otimes|0\rangle\langle 0|_{E'}\right)^{1/2}W^\dagger &(97)\\
&= W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\omega_B W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger &(98)\\
&= W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\left[\omega_B\otimes I_E\right]W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger. &(99)
\end{aligned}$$

Continuing, the last line above is equal to

$$\begin{aligned}
&W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\left[\omega_B\otimes\sum_i|i\rangle\langle i|_E\right]W\left(\sigma_S^{1/2}\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger &(100)\\
&= W\left[\left(\sigma_S^{1/2}\left[\sum_i\langle 0|_{E'}W^\dagger|i\rangle_E\;\omega_B\;\langle i|_E W|0\rangle_{E'}\right]\sigma_S^{1/2}\right)\otimes|0\rangle\langle 0|_{E'}\right]W^\dagger &(101)\\
&= W\left(\left[\sigma_S^{1/2}\mathcal{N}^\dagger\left[\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}U_B\mathcal{N}_{S\to B}(\rho_S)\,U_B^\dagger\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}\right]\sigma_S^{1/2}\right]\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger. &(102)
\end{aligned}$$

The Petz recovery map is defined as

$$\mathcal{R}_{\sigma,\mathcal{N}}(\cdot) \equiv \sigma_S^{1/2}\mathcal{N}^\dagger\left[\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}(\cdot)\left(\mathcal{N}_{S\to B}(\sigma_S)\right)^{-1/2}\right]\sigma_S^{1/2}. \tag{103}$$

Then by inspection, (102) is equal to

$$W\left(\left[\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}_{S\to B}(\rho_S)\,U_B^\dagger\right)\right]\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger. \tag{104}$$

So the fidelity in the remainder term of (94) is

$$F\left(\rho_{BE}, V_{BE}W\left(\left[\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}\left(\rho\right)U_B^\dagger\right)\right]\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\left(V_{BE}\right)^\dagger\right)$$

$$= F\left(W\left(\rho_S\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger, V_{BE}W\left(\left[\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}\left(\rho_S\right)U_B^\dagger\right)\right]\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\left(V_{BE}\right)^\dagger\right) \tag{105}$$

$$= F\left(\rho_S, \langle 0|_{E'}W^\dagger V_{BE}W\left(\left[\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}\left(\rho_S\right)U_B^\dagger\right)\right]\otimes|0\rangle\langle 0|_{E'}\right)W^\dagger\left(V_{BE}\right)^\dagger W|0\rangle_{E'}\right) \tag{106}$$

$$= F\left(\rho_S, V_S\left(\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}\left(\rho_S\right)U_B^\dagger\right)\right)V_S^\dagger\right). \tag{107}$$

Given that V_{BE} acts only on the image of the isometry $W_{SE'\to BE}|0\rangle_{E'}$, the second equality follows because in this case the fidelity is invariant under the partial isometry $\langle 0|_{E'}W^\dagger$. The last equality follows because we can define a unitary V_S acting on the input space as

$$V_S \equiv \langle 0|_{E'}W^\dagger V_{BE}W|0\rangle_{E'}. \tag{108}$$

So the final remainder term for monotonicity of relative entropy is

$$D\left(\rho_S\|\sigma_S\right) - D\left(\mathcal{N}\left(\rho_S\right)\|\mathcal{N}\left(\sigma_S\right)\right) \geq -\log F\left(\rho_S, V_S\left(\mathcal{R}_{\sigma,\mathcal{N}}\left(U_B\mathcal{N}\left(\rho_S\right)U_B^\dagger\right)\right)V_S^\dagger\right). \tag{109}$$

□

Remark 3 *Suppose in Corollary 1 that σ_S is a density operator. It remains open to quantify the performance of the rotated Petz recovery map $\mathcal{V}_S\circ\mathcal{R}^P_{\sigma,\mathcal{N}}\circ\mathcal{U}_B$ on the state $\mathcal{N}_{S\to B}\left(\sigma_S\right)$.*

5 Equivalence of relative entropy inequalities with remainder terms

As discussed in the introduction as well as in Remarks 2 and 3, it would be desirable to have refinements of the inequalities in (1) and (3)-(6) in terms of the Petz recovery map (and not merely in terms of a rotated Petz recovery map). Here, we establish the following equivalence result, depicted in Figure 1. The remainder terms are given in terms of the square of the Bures distance between two density operators [34], defined as

$$D_B^2\left(\rho,\sigma\right) \equiv 2\left(1-\sqrt{F\left(\rho,\sigma\right)}\right), \tag{110}$$

where $F\left(\rho,\sigma\right)$ is the quantum fidelity.

Theorem 4 *The following inequalities with remainder terms are equivalent (however it is an open question to determine whether any single one of them is true):*

1. ***Strong subadditivity of entropy**. Let ω_{ABC} be a tripartite density operator such that ω_C is positive definite. Then*

$$I(A;B|C)_\omega \geq D_B^2\left(\omega_{ABC}, \mathcal{R}^P_{C\to AC}(\omega_{BC})\right), \tag{111}$$

where $\mathcal{R}^P_{C\to AC}(\cdot) \equiv \omega_{AC}^{1/2}\omega_C^{-1/2}(\cdot)\omega_C^{-1/2}\omega_{AC}^{1/2}$ denotes the Petz recovery channel.

2. ***Concavity of conditional entropy**. Let $p_X\left(x\right)$ be a probability distribution characterizing the ensemble $\{p_X\left(x\right), \rho^x_{AB}\}$ with bipartite density operators ρ^x_{AB}. Let $\overline{\rho}_{AB} \equiv \sum_x p_X(x)\rho^x_{AB}$ such that $\overline{\rho}_B$ is positive definite. Then*

$$H(A|B)_{\overline{\rho}} - \sum_x p_X(x)H(A|B)_{\rho^x} \geq \sum_x p_X(x)D_B^2(\rho^x_{AB}, \overline{\rho}_{AB}^{1/2}\overline{\rho}_B^{-1/2}\rho^x_B\overline{\rho}_B^{-1/2}\overline{\rho}_{AB}^{1/2}). \tag{112}$$

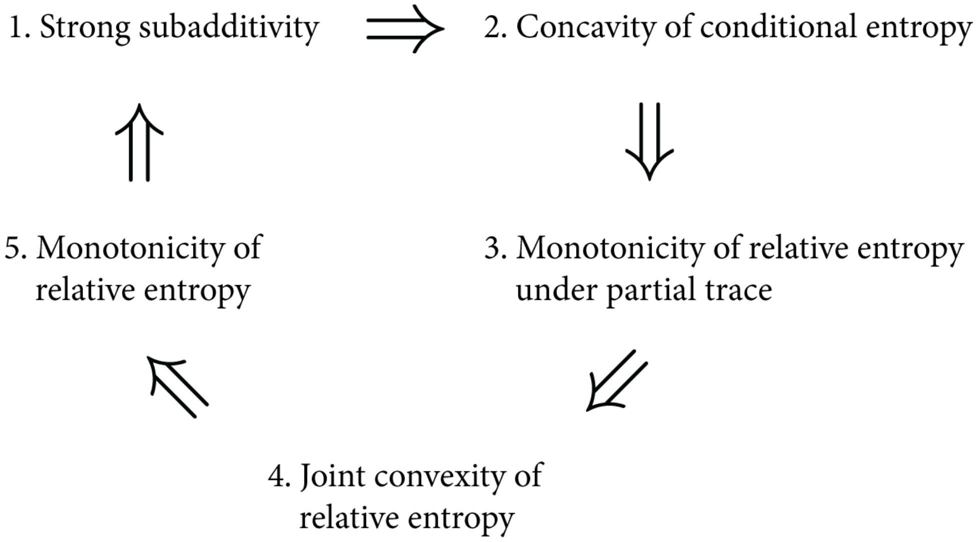

Fig. 1. It is well known that all of the above fundamental entropy inequalities are equivalent (see, e.g., [10]). Theorem 4 extends this circle of equivalences to apply to refinements of these inequalities in terms of the Petz recovery map.

3. ***Monotonicity of relative entropy with respect to partial trace***. *Let ρ_{AB} and σ_{AB} be bipartite density operators such that* $\operatorname{supp}(\rho_{AB}) \subseteq \operatorname{supp}(\sigma_{AB})$ *and σ_B is positive definite. Then*

$$D(\rho_{AB}\|\sigma_{AB}) - D(\rho_B\|\sigma_B) \geq D_B^2\left(\rho_{AB}, \mathcal{R}^P_{\sigma,\mathrm{Tr}_A}(\rho_B)\right), \tag{113}$$

where $\mathcal{R}^P_{\sigma,\mathrm{Tr}_A}(\cdot) \equiv \sigma_{AB}^{1/2}\sigma_B^{-1/2}(\cdot)\sigma_B^{-1/2}\sigma_{AB}^{1/2}$ denotes the Petz recovery channel with respect to σ_{AB} and Tr_A.

4. ***Joint convexity of relative entropy***. *Let $p_X(x)$ be a probability distribution characterizing the ensembles $\{p_X(x), \rho_x\}$, and $\{p_X(x), \sigma_x\}$ with ρ_x and σ_x density operators such that* $\operatorname{supp}(\rho_x) \subseteq \operatorname{supp}(\sigma_x)$. *Let $\overline{\rho} \equiv \sum_x p_X(x)\rho_x$ and $\overline{\sigma} \equiv \sum_x p_X(x)\sigma_x$ such that $\overline{\sigma}$ is positive definite. Then*

$$\sum_x p_X(x) D(\rho_x\|\sigma_x) - D(\overline{\rho}\|\overline{\sigma}) \geq \sum_x p_X(x) D_B^2\left(\rho_x, \sigma_x^{1/2}(\overline{\sigma})^{-1/2}\overline{\rho}(\overline{\sigma})^{-1/2}\sigma_x^{1/2}\right). \tag{114}$$

5. ***Monotonicity of relative entropy***. *Let ρ and σ be density operators such that* $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$, *and $\mathcal{N}$ a CPTP map such that $\mathcal{N}(\sigma)$ is positive definite. Then*

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq D_B^2\left(\rho, \mathcal{R}^P_{\sigma,\mathcal{N}}(\rho)\right), \tag{115}$$

where $\mathcal{R}^P_{\sigma,\mathcal{N}}(\cdot) \equiv \sigma^{1/2}\mathcal{N}^\dagger\left([\mathcal{N}(\sigma)]^{-1/2}(\cdot)[\mathcal{N}(\sigma)]^{-1/2}\right)\sigma^{1/2}$ denotes the Petz recovery channel with respect to σ and $\mathcal{N}$.

Proof. For the proof, we abbreviate the square root of the fidelity F as the root fidelity $\sqrt{F}$. We can easily see that $5 \Rightarrow 3$, and from a variation of the development in [20, Consequence

28], we obtain $3 \Rightarrow 4 \Rightarrow 5$, leading to $3 \Leftrightarrow 4 \Leftrightarrow 5$.[e] We can get $5 \Rightarrow 1$ by choosing $\rho = \omega_{ABC}$, $\sigma = \omega_{AC} \otimes \omega_B$, and $\mathcal{N} = \mathrm{Tr}_A$, so that

$$\begin{aligned}
&\sigma^{1/2}\mathcal{N}^\dagger\left([\mathcal{N}(\sigma)]^{-1/2}(\cdot)[\mathcal{N}(\sigma)]^{-1/2}\right)\sigma^{1/2} \\
&= [\omega_{AC}\otimes\omega_B]^{1/2}\left[\left([\omega_C\otimes\omega_B]^{-1/2}(\cdot)[\omega_C\otimes\omega_B]^{-1/2}\right)\otimes I_A\right][\omega_{AC}\otimes\omega_B]^{1/2} \qquad (116)\\
&= \omega_{AC}^{1/2}\omega_C^{-1/2}(\cdot)\omega_C^{-1/2}\omega_{AC}^{1/2}. \qquad (117)
\end{aligned}$$

Then

$$\begin{aligned}
I(A;B|C)_\omega &= D(\omega_{ABC}\|\omega_{AC}\otimes\omega_B) - D(\omega_{BC}\|\omega_C\otimes\omega_B) \qquad (118)\\
&\geq 2\left(1-\sqrt{F}\left(\omega_{ABC},\mathcal{R}^P_{\sigma,\mathcal{N}}(\omega_{BC})\right)\right) \qquad (119)\\
&= 2\left(1-\sqrt{F}\left(\omega_{ABC},\omega_{AC}^{1/2}\omega_C^{-1/2}\omega_{BC}\omega_C^{-1/2}\omega_{AC}^{1/2}\right)\right). \qquad (120)
\end{aligned}$$

The implication $1 \Rightarrow 2$ follows by choosing

$$\theta_{XAB} \equiv \sum_x p_X(x)\,|x\rangle\langle x|_X \otimes \rho^x_{AB}, \qquad (121)$$

so that

$$\begin{aligned}
H(A|B)_{\overline{\rho}} - \sum_x p_X(x) H(A|B)_{\rho^x} &= I(A;X|B)_\theta \qquad (122)\\
&\geq 2\left(1-\sqrt{F}\left(\theta_{XAB},\theta_{AB}^{1/2}\theta_B^{-1/2}\theta_{XB}\theta_B^{-1/2}\theta_{AB}^{1/2}\right)\right) \qquad (123)\\
&= 2\left(1-\sum_x p_X(x)\sqrt{F}\left(\rho^x_{AB},\overline{\rho}_{AB}^{1/2}\overline{\rho}_B^{-1/2}\rho^x_B\overline{\rho}_B^{-1/2}\overline{\rho}_{AB}^{1/2}\right)\right). \qquad (124)
\end{aligned}$$

The last remaining implication $2 \Rightarrow 3$ has the most involved proof, which we establish now by using the idea from [12, Section 3-E]. Throughout our proof, we employ Theorem V.3.3 of [39]. This theorem states that if f is a differentiable function on an open neighborhood of the spectrum of some self-adjoint operator A, then its derivative Df at A is given by

$$Df(A) : H \to \sum_{\lambda,\eta} f^{[1]}(\lambda,\eta)\,P_A(\lambda)\,H P_A(\eta), \qquad (125)$$

where $A = \sum_\lambda \lambda P_A(\lambda)$ is the spectral decomposition of A, and $f^{[1]}$ is the first divided difference function. In particular, if $x \longmapsto A(x) \in \mathcal{B}(\mathcal{H})_+$ is a differentiable function on an open interval in $\mathbb{R}$, with derivative A', then

$$\frac{d}{dx} f(A(x)) = \sum_{\lambda,\eta} f^{[1]}(\lambda,\eta)\,P_{A(x)}(\lambda)\,A'(x)\,P_{A(x)}(\eta), \qquad (126)$$

so that

$$\frac{d}{dx}\mathrm{Tr}\{f(A(x))\} = \mathrm{Tr}\{f'(A(x))\,A'(x)\}. \qquad (127)$$

[e] Note that [20, Consequence 28] establishes the circle $3 \Leftrightarrow 4 \Leftrightarrow 5$ with a remainder term of $-\log F$.

In particular, if $A(x) = A + xB$, then

$$\frac{d}{dx}\operatorname{Tr}\{f(A(x))\} = \operatorname{Tr}\{f'(A(x))B\}. \tag{128}$$

We can now proceed. In what follows, we will be taking $A(x) = \sigma_{AB} + x\rho_{AB}$, where σ_{AB} is a positive definite density operator, ρ_{AB} is a density operator, and $x \geq 0$. We also make use of the standard fact that the function $f : X \to X^{-1}$ is everywhere differentiable on the set of invertible density operators, and at an invertible X, its derivative is $f'(X) : Y \to -X^{-1}YX^{-1}$.

Consider that the conditional entropy is homogeneous, in the sense that

$$H(A|B)_{xG} = xH(A|B)_G, \tag{129}$$

where x is a positive scalar and G_{AB} is a positive semi-definite operator on systems AB. Let

$$\xi_{YAB} \equiv \frac{1}{x+1}|0\rangle\langle 0|_Y \otimes \sigma_{AB} + \frac{x}{x+1}|1\rangle\langle 1|_Y \otimes \rho_{AB}, \tag{130}$$

with σ_{AB} a positive definite density operator and ρ_{AB} a density operator. Then it follows from homogeneity and concavity with the Petz remainder term (by assumption) that

$$\begin{aligned} H(A|B)_{\sigma+x\rho} &= (x+1)H(A|B)_\xi && (131)\\ &\geq (x+1)\left[\frac{1}{x+1}H(A|B)_\sigma + \frac{x}{x+1}H(A|B)_\rho + R(x,\sigma_{AB},\rho_{AB})\right] && (132)\\ &= H(A|B)_\sigma + xH(A|B)_\rho + (x+1)R(x,\sigma_{AB},\rho_{AB}), && (133)\end{aligned}$$

where

$$R(x,\sigma_{AB},\rho_{AB}) \equiv$$
$$2\left(1 - \left[\frac{1}{x+1}\sqrt{F}\left(\sigma_{AB}, \xi_{AB}^{1/2}\xi_B^{-1/2}\sigma_B\xi_B^{-1/2}\xi_{AB}^{1/2}\right) + \frac{x}{x+1}\sqrt{F}\left(\rho_{AB}, \xi_{AB}^{1/2}\xi_B^{-1/2}\rho_B\xi_B^{-1/2}\xi_{AB}^{1/2}\right)\right]\right). \tag{134}$$

Manipulating the above inequality then gives

$$\frac{H(A|B)_{\sigma+x\rho} - H(A|B)_\sigma}{x} \geq H(A|B)_\rho + \frac{x+1}{x}R(x,\sigma_{AB},\rho_{AB}). \tag{135}$$

Taking the limit as $x \searrow 0$ then gives

$$\lim_{x\searrow 0}\frac{H(A|B)_{\sigma+x\rho} - H(A|B)_\sigma}{x} = \left.\frac{d}{dx}H(A|B)_{\sigma+x\rho}\right|_{x=0} \geq H(A|B)_\rho + \lim_{x\searrow 0}\frac{x+1}{x}R(x,\sigma_{AB},\rho_{AB}). \tag{136}$$

We now evaluate the limits separately, beginning with the one on the left hand side. So we consider

$$\begin{aligned}&\frac{d}{dx}H(A|B)_{\sigma+x\rho}\\ &= \frac{d}{dx}\left[-\operatorname{Tr}\{(\sigma_{AB}+x\rho_{AB})\log(\sigma_{AB}+x\rho_{AB})\} + \operatorname{Tr}\{(\sigma_B+x\rho_B)\log(\sigma_B+x\rho_B)\}\right].\end{aligned} \tag{137}$$

We evaluate this by using $\frac{d}{dy}\left[g\left(y\right)\log g\left(y\right)\right] = \left[\log g\left(y\right)+1\right]g'\left(y\right)$ and (128) to find that

$$\frac{d}{dx}\operatorname{Tr}\left\{\left(\sigma_{AB}+x\rho_{AB}\right)\log\left(\sigma_{AB}+x\rho_{AB}\right)\right\} = \operatorname{Tr}\left\{\left[\log\left(\sigma_{AB}+x\rho_{AB}\right)+I_{AB}\right]\rho_{AB}\right\}, \tag{138}$$

so that

$$\frac{d}{dx}H\left(A|B\right)_{\sigma+x\rho} = -\operatorname{Tr}\left\{\rho_{AB}\log\left(\sigma_{AB}+x\rho_{AB}\right)\right\}+\operatorname{Tr}\left\{\rho_{B}\log\left(\sigma_{B}+x\rho_{B}\right)\right\}, \tag{139}$$

and thus

$$\left.\frac{d}{dx}H\left(A|B\right)_{\sigma+x\rho}\right|_{x=0} = -\operatorname{Tr}\left\{\rho_{AB}\log\sigma_{AB}\right\}+\operatorname{Tr}\left\{\rho_{B}\log\sigma_{B}\right\}. \tag{140}$$

Substituting back into the inequality (136), we find that

$$\begin{aligned}&-\operatorname{Tr}\left\{\rho_{AB}\log\sigma_{AB}\right\}+\operatorname{Tr}\left\{\rho_{B}\log\sigma_{B}\right\}\geq\\&\qquad -\operatorname{Tr}\left\{\rho_{AB}\log\rho_{AB}\right\}+\operatorname{Tr}\left\{\rho_{B}\log\rho_{B}\right\}+\lim_{x\searrow0}\frac{x+1}{x}R\left(x,\sigma_{AB},\rho_{AB}\right),\end{aligned} \tag{141}$$

which is equivalent to (cf., [12, Eq. (3.2)])

$$D\left(\rho_{AB}\|\sigma_{AB}\right)-D\left(\rho_{B}\|\sigma_{B}\right)\geq\lim_{x\searrow0}\frac{x+1}{x}R\left(x,\sigma_{AB},\rho_{AB}\right). \tag{142}$$

So we need to evaluate this last limit to get the remainder term. Consider that

$$\lim_{x\searrow0}\frac{x+1}{x}R\left(x,\sigma_{AB},\rho_{AB}\right) \tag{143}$$

$$=\lim_{x\searrow0}2\left(1+\frac{1-\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\sigma_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)}{x}-\sqrt{F}\left(\rho_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\rho_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)\right). \tag{144}$$

Since

$$\lim_{x\searrow0}\sqrt{F}\left(\rho_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\rho_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)=\sqrt{F}\left(\rho_{AB},\sigma_{AB}^{1/2}\sigma_{B}^{-1/2}\rho_{B}\sigma_{B}^{-1/2}\sigma_{AB}^{1/2}\right), \tag{145}$$

it remains to show that

$$\begin{aligned}&\lim_{x\searrow0}\frac{1-\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\sigma_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)}{x}\\&\qquad=\left.\frac{d}{dx}\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\sigma_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)\right|_{x=0}=0.\end{aligned} \tag{146}$$

Essentially, this derivative vanishes because the fidelity is one at $x=0$ and therefore maximal. In what follows, we explicitly show that the derivative above is equal to zero. Consider that

$$\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_{B}^{-1/2}\sigma_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\right)$$

$$=\operatorname{Tr}\left\{\left(\sigma_{AB}^{1/2}\xi_{AB}^{1/2}\xi_{B}^{-1/2}\sigma_{B}\xi_{B}^{-1/2}\xi_{AB}^{1/2}\sigma_{AB}^{1/2}\right)^{1/2}\right\} \tag{147}$$

$$=\operatorname{Tr}\left\{\left(\sigma_{AB}^{1/2}\left(\sigma_{AB}+x\rho_{AB}\right)^{1/2}\left(\sigma_{B}+x\rho_{B}\right)^{-1/2}\sigma_{B}\left(\sigma_{B}+x\rho_{B}\right)^{-1/2}\left(\sigma_{AB}+x\rho_{AB}\right)^{1/2}\sigma_{AB}^{1/2}\right)^{1/2}\right\}, \tag{148}$$

as well as

$$\frac{d}{dx}\operatorname{Tr}\left\{(G(x))^{1/2}\right\}=\frac{1}{2}\operatorname{Tr}\left\{G(x)^{-1/2}\frac{d}{dx}G(x)\right\}, \tag{149}$$

which follows from (128). Applying the above rule, we get that $\frac{d}{dx}$ of (148) is equal to

$$\operatorname{Tr}\left\{\begin{array}{c}\left(\sigma_{AB}^{1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}(\sigma_B+x\rho_B)^{-1/2}\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\sigma_{AB}^{1/2}\right)^{-1/2}\times\\ \sigma_{AB}^{1/2}\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}(\sigma_B+x\rho_B)^{-1/2}\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\right]\sigma_{AB}^{1/2}\end{array}\right\}. \tag{150}$$

Now, take the limit as $x\searrow 0$ to find that (150) is equal to

$$\begin{aligned}&\frac{d}{dx}\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_B^{-1/2}\sigma_B\xi_B^{-1/2}\xi_{AB}^{1/2}\right)\Big|_{x=0}\\ &=\operatorname{Tr}\left\{\begin{array}{c}\left(\sigma_{AB}^{1/2}\sigma_{AB}^{1/2}\sigma_B^{-1/2}\sigma_B\sigma_B^{-1/2}\sigma_{AB}^{1/2}\sigma_{AB}^{1/2}\right)^{-1/2}\times\\ \sigma_{AB}^{\frac{1}{2}}\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{\frac{1}{2}}(\sigma_B+x\rho_B)^{-\frac{1}{2}}\sigma_B(\sigma_B+x\rho_B)^{-\frac{1}{2}}(\sigma_{AB}+x\rho_{AB})^{\frac{1}{2}}\right]\Big|_{x=0}\sigma_{AB}^{\frac{1}{2}}\end{array}\right\} &(151)\\ &=\operatorname{Tr}\left\{\begin{array}{c}(\sigma_{AB})^{-1}\times\\ \sigma_{AB}^{\frac{1}{2}}\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{\frac{1}{2}}(\sigma_B+x\rho_B)^{-\frac{1}{2}}\sigma_B(\sigma_B+x\rho_B)^{-\frac{1}{2}}(\sigma_{AB}+x\rho_{AB})^{\frac{1}{2}}\right]\Big|_{x=0}\sigma_{AB}^{\frac{1}{2}}\end{array}\right\} &(152)\\ &=\operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}(\sigma_B+x\rho_B)^{-1/2}\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\right]\Big|_{x=0}\right\} &(153)\end{aligned}$$

So we focus on this last expression and note from the derivative product rule that there are four terms to consider. We consider one at a time, beginning with the first term:

$$\begin{aligned}&\lim_{x\searrow 0}\operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}\right](\sigma_B+x\rho_B)^{-1/2}\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\right\}\\ &=\operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}\right]\Big|_{x=0}\sigma_B^{-1/2}\sigma_B\sigma_B^{-1/2}\sigma_{AB}^{1/2}\right\} &(154)\\ &=\operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}\right]\Big|_{x=0}\sigma_{AB}^{1/2}\right\} &(155)\\ &=\frac{1}{2}\operatorname{Tr}\{\rho_{AB}\} &(156)\\ &=\frac{1}{2}, &(157)\end{aligned}$$

where the second to last line follows from (126). We now consider the second term:

$$\lim_{x\searrow 0} \operatorname{Tr}\left\{(\sigma_{AB}+x\rho_{AB})^{1/2}\frac{d}{dx}\left[(\sigma_B+x\rho_B)^{-1/2}\right]\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\right\}$$

$$= \operatorname{Tr}\left\{\sigma_{AB}^{1/2}\frac{d}{dx}\left[(\sigma_B+x\rho_B)^{-1/2}\right]\Big|_{x=0}\sigma_B\sigma_B^{-1/2}\sigma_{AB}^{1/2}\right\} \tag{158}$$

$$= \operatorname{Tr}\left\{\sigma_{AB}\frac{d}{dx}\left[(\sigma_B+x\rho_B)^{-1/2}\right]\Big|_{x=0}\sigma_B\sigma_B^{-1/2}\right\} \tag{159}$$

$$= \operatorname{Tr}\left\{\sigma_{B}\frac{d}{dx}\left[(\sigma_B+x\rho_B)^{-1/2}\right]\Big|_{x=0}\sigma_B\sigma_B^{-1/2}\right\} \tag{160}$$

$$= \operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_B+x\rho_B)^{-1/2}\right]\Big|_{x=0}\sigma_B^{3/2}\right\} \tag{161}$$

$$= -\frac{1}{2}\operatorname{Tr}\left\{\sigma_B^{-3/2}\sigma_B^{3/2}\rho_B\right\} \tag{162}$$

$$= -\frac{1}{2}\operatorname{Tr}\left\{\rho_B\right\} \tag{163}$$

$$= -\frac{1}{2}. \tag{164}$$

The third to last line follows from (126). Combining these results and using that the last two terms resulting from the product rule are Hermitian conjugates of the first two, we find that

$$\operatorname{Tr}\left\{\frac{d}{dx}\left[(\sigma_{AB}+x\rho_{AB})^{1/2}(\sigma_B+x\rho_B)^{-1/2}\sigma_B(\sigma_B+x\rho_B)^{-1/2}(\sigma_{AB}+x\rho_{AB})^{1/2}\right]\Big|_{x=0}\right\}=0, \tag{165}$$

which allows us to conclude that

$$\frac{d}{dx}\sqrt{F}\left(\sigma_{AB},\xi_{AB}^{1/2}\xi_B^{-1/2}\sigma_B\xi_B^{-1/2}\xi_{AB}^{1/2}\right)\Big|_{x=0}=0, \tag{166}$$

Hence, we can conclude that the following inequality is a consequence of (112):

$$D(\rho_{AB}\|\sigma_{AB})-D(\rho_B\|\sigma_B)\geq 2\left(1-\sqrt{F}\left(\rho_{AB},\sigma_{AB}^{1/2}\sigma_B^{-1/2}\rho_B\sigma_B^{-1/2}\sigma_{AB}^{1/2}\right)\right). \tag{167}$$

□

Note: After the completion of the present paper, the works in [37] and [40] appeared, which build upon ideas established in this paper. The main contribution of [37] is to show that the rotated Petz map in Corollary 1 can take a more particular form. Specifically, the unitary channel $\mathcal{U}_B$ in Corollary 1 can be taken to commute with $\mathcal{N}(\sigma)$ and the unitary channel $\mathcal{V}_S$ can be taken to commute with σ. The main contribution of [40] is to show that the fidelity remainder term in Corollary 1 can be replaced with the "measured relative entropy" and the rotated Petz map can be replaced with a "twirled Petz map." Please refer to [37] and [40] for more details.

Acknowledgements

We are especially grateful to Rupert Frank for many discussions on the topic of this paper. We thank the anonymous referees for many suggestions that helped to improve the paper.

We acknowledge additional discussions with Siddhartha Das, Nilanjana Datta, Omar Fawzi, Renato Renner, Volkher Scholz, Kaushik P. Seshadreesan, Marco Tomamichel, and Michael Walter. MMW acknowledges support from startup funds from the Department of Physics and Astronomy at LSU, the NSF under Award No. CCF-1350397, and the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019.

References

1. Hisaharu Umegaki, Conditional expectations in an operator algobra IV (entropy and information). *Kodai Mathematical Seminar Reports,* **14**(2), 59-85 (1962).
2. Mario Berta, Kaushik Seshadreesan, and Mark M. Wilde, Rényi generalizations of the conditional quantum mutual information, *Journal of Mathematical Physics*, **56**(2), 022205 (2015).
3. Hiai, Fumio and Petz, Denes, The proper formula for relative entropy and its asymptotics in quantum probability, *Communications in Mathematical Physics*, **143**(1), 99-114, (1991).
4. Nagaoka, Hirsohi and Ogawa, Tomohiro, Strong converse and Stein's lemma in quantum hypothesis testing, *IEEE Transactions on Information Theory*, **46**(7), 2428-2433, (2000).
5. Goran Lindblad, Completely positive maps and entropy inqualities, *Communictions in Mathematical Physics*, **40**(2), 147-151 (1975).
6. Armin Uhlmann, Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, *Communications in Mathematical Physics*, **54**(1) 21-32 (1977).
7. Denes Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Communications in Mathematical Physics*, **105**(1), 123-131 (1986).
8. Denes Petz, Sufficiency of channels over von Neumann algebras, *Quarterly Journal of Mathematics*, **39**(1), 97-108 (1988).
9. Patrick Hayden, Richard Jozsa, Denes Petz, and Andreas Winter, Structure of states which satisfy strong subadditivity of quantum entropy with equality, *Communications in Mathematical Physics*, **246**(2), 359-374 (2004).
10. Mary Beth Ruskai, Inequalities for quantum entropy: A review with conditions for equality, *Journal of Mathematical Physics*, **43**(9), 4358-4375 (2002).
11. Elliott H. Lieb and Mary Beth Ruskai, A Fundamental Property of Quantum-Mechanical Entropy, *Physical Review Letters*, **30**(10), 434-436 (1973).
12. Elliott H. Lieb and Mary Beth Ruskai, Proof of the Strong Subadditivity of Quantum-Mechanical Entropy, *Journal of Mathematical Physics*, **14**(12), 1938-1941 (1973).
13. Fernando G. S. L. Brandão, Matthias Christandl, and Jon Yard, Faithful Squashed Entanglement, *Communications in Mathematical Physics*, **306**, 805–830 (2011).
14. Andreas Winter and Ke Li, A stronger subadditivity relation?, http://www.maths.bris.ac.uk/∼csajw/stronger_subadditivity.pdf, (2012).
15. Isaac H. Kim, Application of conditional independence to gapped quantum many-body systems, http://www.physics.usyd.edu.au/quantum/Coogee2013, (2013).
16. Ke Li and Andreas Winter, Relative Entropy and Squashed Entanglement, *Communications in Mathematical Physics*, **326**(1), 63-80 (2014).
17. Eric A. Carlen and Elliott H. Lieb, Remainder Terms for Some Quantum Entropy Inequalities, *Journal of Mathematical Physics*, **55**(4), 042201 (2014).
18. Lin Zhang and Junde Wu, A lower bound of quantum conditional mutual information, *Journal of Physics A: Mathematical and Theoretical*, **47**(41), 415303 (2014).
19. Lin Zhang, A stronger monotonicity inequality of quantum relative entropy: A unifying approach via Rényi relative entropy, arXiv:1403.5343v1 (2014).
20. Kaushik P. Seshadreesan, Mario Berta, and Mark M. Wilde, Rényi squashed entanglement, discord, and relative entropy differences, arXiv:1410.1443 (2014).
21. Ke Li and Andreas Winter, Squashed entanglement, k-extendibility, quantum Markov chains, and recovery maps, arXiv:1410.4184 (2014).
22. Kaushik P. Seshadreesan and Mark M. Wilde, Fidelity of recovery, geometric squashed entangle-

ment, and measurement recoverability, arXiv:1410.1441. (2014).
23. Omar Fawzi and Renato Renner, Quantum conditional mutual information and approximate Markov chains, arXiv:1410.0664 (2014).
24. Armin Uhlmann, The "Transition Probability" in the State Space of a *-Algebra, *Reports on Mathematical Physics*, **9**(2), 273-279 (1976).
25. Igor Bjelakovic and Rainer Siegmund-Schultze, Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb, arXiv:quant-ph/0307170 (2003).
26. Howard Barnum and Emanuel Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, *Journal of Mathematical Physics*, **43**(5), 2097 (2002).
27. Benjamin Schumacher and Michael D. Westmoreland, Approximate quantum error correction, *Quantum Information Processing*, **1**(1/2), 5-12 (2002).
28. Jon Tyson, Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates, *Journal of Mathematical Physics*, **51**(9), 092204 (2010).
29. Hui Khoon Ng and Prabha Mandayam, Simple approach to approximate quantum error correction based on the transpose channel, *Physical Review A*, **81** (6), 062342 (2010).
30. Prabha Mandayam and Hui Khoon Ng, Towards a unified framework for approximate quantum error correction, *Physical Review A*, **86**(1), 012335 (2012).
31. Vlatko Vedral, The role of relative entropy in quantum information theory, *Reviews of Modern Physics*, **74** (1), 197-234 (2002).
32. Takahiro Sagawa Second Law-Like Inequalities with Quantum Relative Entropy: An Introduction, *Lectures on Quantum Computing, Thermodynamics and Statistical Physics*, (World Scientific, 2012).
33. Mark M. Wilde, Multipartite quantum correlations and local recoverability, *Proceedings of the Royal Society A*, **471**:20140941, arXiv:1412.0333., (2014).
34. Donald Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras, *Transactions of the American Mathematical Society*, **135**, 199-212 (1969).
35. Wassily Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, **58**(301), 13-30 (1963).
36. Andreas Winter, Coding Theorem and Strong Converse for Quantum Channels, *IEEE Transactions on Information Theory*, **45**(7), 2481-2485 (1999).
37. Mark M. Wilde, Recoverability in quantum information theory, arXiv:1505.04661 (2015).
38. William F. Stinespring, Positive functions on C*-algebras, *Proceedings of the American Mathematical Society*, **6**(2) 211216, (1955).
39. Rajendra Bhatia, *Matrix Analysis* (Springer, 1997).
40. David Sutter, Marco Tomamichel, and Aram W. Harrow, Strengthened Monotonicity of Relative Entropy via Pinched Petz Recovery Map, arXiv:1507.00303 (2015).

Appendix A Auxiliary lemmas from [23]

In this appendix, for the convenience of the reader, we list verbatim the relevant lemmas that we have used from [23].

Lemma A.1 (Lemma 2.3 of [23]) *Let ρ be a density operator, let σ be a non-negative operator on the same space, and let $\{\mathcal{W}_n\}_{n\in\mathbb{N}}$ be a sequence of trace non-increasing completely positive maps on the n-fold tensor product of this space. If $(\mathcal{W}_n(\rho^{\otimes n}))$ decreases less than exponentially in n, i.e.,*

$$\liminf_{n\to\infty} e^{\xi n}\big(\mathcal{W}_n(\rho^{\otimes n})\big) > 0 \tag{A.1}$$

for any $\xi > 0$, then

$$\limsup_{n\to\infty} \frac{1}{n} D\big(\mathcal{W}_n(\rho^{\otimes n}) \| \mathcal{W}_n(\sigma^{\otimes n})\big) \leq D(\rho\|\sigma) \ . \tag{A.2}$$

Lemma A.2 (Lemma 4.2 of [23]) *Let $\rho_{R^n S^n}$ be a permutation-invariant non-negative operator on $(R\otimes S)^{\otimes n}$ and let σ_{RS} be a non-negative operator on $R\otimes S$. Furthermore, let W_{R^n} be a permutation-invariant operator on $R^{\otimes n}$ with $\|W_{R^n}\|_\infty \leq 1$. Then there exists a unitary U_R on R such that*

$$\sqrt{F}\big(\rho_{R^n S^n}, U_R^{\otimes n} \sigma_{RS}^{\otimes n} (U_R^{\otimes n})^\dagger\big) \geq (n+1)^{-d^2} \sqrt{F}\big(W_{R^n} \rho_{R^n S^n} W_{R^n}^\dagger, \sigma_{RS}^{\otimes n}\big) \ , \tag{A.3}$$

where $d = \dim(R)\dim(S)^2$.

Lemma A.3 (Lemma B.2 of [23]) *For any non-negative operators ρ and σ*

$$D(\rho\|\sigma) \geq -2\log_2 \frac{\sqrt{F}(\rho,\sigma)}{\mathrm{tr}(\rho)} \ . \tag{A.4}$$

Lemma A.4 (Lemma B.6 of [23]) *For any non-negative operators ρ and σ and any operator W on the same space we have*

$$\sqrt{F}(\rho, W\sigma W^\dagger) = \sqrt{F}(W^\dagger \rho W, \sigma) \ . \tag{A.5}$$

Lemma A.5 (Lemma B.7 of [23]) *Let ρ and σ be non-negative operators and let $\{W_d\}_{d\in D}$ be a family of operators such that $\sum_{d\in D} W_d = \mathrm{id}$. Then*

$$\sum_{d\in D} \sqrt{F}(W_d^\dagger \rho W_d, \sigma) \geq \sqrt{F}(\rho,\sigma) \ . \tag{A.6}$$