Monotonicity of quantum relative entropy and recoverability

Mario Berta∗ Marius Lemm† Mark M. Wilde‡

August 21, 2015

Abstract

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, arXiv:1410.0664] with the notion of a relative typical subspace from [Bjelakovic and Siegmund-Schultze, arXiv:quant-ph/0307170]. 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.

1 Introduction

The Umegaki relative entropy $D(\rho\|\sigma)$ between a density operator $\rho$ and a positive semi-definite operator $\sigma$ is defined as $\text{Tr}\{\rho [\log \rho - \log \sigma]\}$ whenever $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and as $+\infty$ otherwise. It is a fundamental information measure in quantum information theory [Ume62], from which many other information measures, such as entropy, conditional entropy, and mutual information, can be derived (see, e.g., [BSW15]). When $\sigma$ 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) [HP91, NO00]. 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 [Lin75, Uhl77]:

$$D(\rho\|\sigma) \geq D(N(\rho)\|N(\sigma)),$$

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.
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 satisfied if and only if the following Petz recovery map perfectly recovers $\rho$ from $\mathcal{N}(\rho)$ [Pet86, Pet88] (see also [HJPW04]):

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

(1.2)

with $\mathcal{N}^{\dagger}$ the adjoint of $\mathcal{N}$. (Observe that the Petz recovery map always perfectly recovers $\sigma$ from $\mathcal{N}(\sigma)$.) There are several related inequalities, which are known to be equivalent to (1.1) when $\sigma$ is a density operator (see, e.g., [Rus02]). One equivalent inequality is the monotonicity of relative entropy:

$$
D (\rho_{AB} \| \sigma_{AB} ) \geq D (\rho_{B} \| \sigma_{B} ) ,
$$

(1.3)

where $\rho_{AB}$ and $\sigma_{AB}$ are density operators acting on a tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The operators $\rho_{B}$ and $\sigma_{B}$ 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:

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

(1.4)

where $p_X$ is a probability distribution, $\{ \rho^x \}$ and $\{ \sigma^x \}$ are sets of density operators, $\bar{\rho} \equiv \sum_x p_X (x) \rho^x$, and $\bar{\sigma} \equiv \sum_x p_X (x) \sigma^x$. 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 [LR73a, LR73b]:

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

(1.5)

which can be seen as a special case of (1.1) with $\rho = \omega_{ABC}$, $\sigma = \omega_{AC} \otimes I_B$, and $\mathcal{N} = \text{Tr}_A$, where $\omega_{ABC}$ 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 [LR73b]:

$$
H (A|B)_{\bar{\sigma}} \geq \sum_x p_X (x) H (A|B)_{\rho^x} ,
$$

(1.6)

where $p_X$ is a probability distribution, $\{ \rho^x_{AB} \}$ is a set of density operators, $\bar{\rho}_{AB} \equiv \sum_x p_X (x) \rho^x_{AB}$, 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 [BCY11, WL12, Kim13, LW14a, CL14, ZW14, Zha14, BSW15, SBW14, LW14b, SW14], a breakthrough paper established the following remainder term for strong subadditivity [FR14]:

$$
I (A; B|C)_{\omega} \geq - \log F (\omega_{ABC}, (\mathcal{V}_{AC} \circ \mathcal{R}_{C \rightarrow AC}^P \circ \mathcal{U}_C ) (\omega_{BC} ) ) ,
$$

(1.7)

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 $F(\tau, \varsigma) \equiv \|\sqrt{\tau} \sqrt{\varsigma}\|^2_1$ is the quantum fidelity between positive semi-definite operators $\tau$ and $\varsigma$ [Uhl76], $U_C$ and $V_{AC}$ are unitary channels defined in terms of some unitary operators $U$ and $V_{AC}$ as

$$U_C(\cdot) \equiv U_C(\cdot) U_C^\dagger, \quad (1.8)$$

$$V_{AC}(\cdot) \equiv V_{AC}(\cdot) V_{AC}^\dagger, \quad (1.9)$$

and $R_{C \rightarrow AC}^P$ is the following Petz recovery map:

$$R_{C \rightarrow AC}^P(\cdot) \equiv \omega_{AC}^{1/2} \omega_C^{-1/2}(\cdot) \omega_C^{-1/2} \omega_{AC}^{1/2}. \quad (1.10)$$

In the present paper, our first contribution is to combine the methods of [FR14] and the notion of a relative typical subspace from [BSS03, pages 4-5] in order to establish the following remainder term for the inequality in (1.1):

$$D(\rho; \sigma) - D(N(\rho); N(\sigma)) \geq - \log F(\rho, (V \circ R_{\rho,N}^P \circ U)(N(\rho))), \quad (1.11)$$

where $U$ is a unitary channel acting on the output space of $N$; $R_{\rho,N}^P$ is the Petz recovery map defined in (1.2), and $V$ is a unitary channel acting on the input space of $N$. Thus, the refinement in (1.11) quantifies how well one can recover $\rho$ from $N(\rho)$ by employing the “rotated Petz recovery map” $V \circ R_{\rho,N}^P \circ U$. This result is stated formally as Corollary 3 and can be understood as a generalization of (1.7). We establish a similar refinement of the inequality in (1.3), 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 [BK02, SW02, Tys10, NM10, MN12] or thermodynamics [Ved02, Sag12]. Note that the refinement in (1.7) has already been helpful in improving our understanding of some quantum correlation measures [WL12, LW14b, SW14, WIl14].

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:

$$D(\rho; \sigma) - D(N(\rho); N(\sigma)) \geq - \log F(\rho, R_{\rho,N}^P(N(\rho))), \quad (1.12)$$

$$D(\rho_{AB}; \sigma_{AB}) - D(\rho_B; \sigma_B) \geq - \log F(\rho_{AB}, \sigma_{AB}^{1/2} \sigma_B^{-1/2} \rho_B \sigma_B^{-1/2} \sigma_{AB}^{1/2}), \quad (1.13)$$

$$\sum_x p_X(x) D(\rho^x; \sigma^x) - D(\bar{\rho}; \bar{\sigma}) \geq - 2 \log \sum_x p_X(x) \sqrt{F(\rho^x, (\sigma^x)^{1/2} (\bar{\sigma})^{-1/2} \bar{\rho}(\bar{\sigma})^{-1/2} (\sigma^x)^{1/2})}, \quad (1.14)$$

$$I(A; B| C)_{\omega} \geq - \log F(\omega_{ABC}, \omega_{AC}^{1/2} \omega_C^{-1/2} \omega_{BC} \omega_C^{-1/2} \omega_{AC}^{1/2}), \quad (1.15)$$

$$H(A| B)_{\bar{\rho}} - \sum_x p_X(x) H(A| B)_{\rho^x} \geq - 2 \log \sum_x p_X(x) \sqrt{F(\rho_{AB}^{1/2} \rho_{AB}^{-1/2} \bar{\rho}_{AB} \rho_{AB}^{-1/2} \bar{\rho}_{AB}^{1/2})}. \quad (1.16)$$

In [SBW14, Definition 25], a Rényi information measure was defined to generalize relative entropy differences. The inequalities (1.12)-(1.16) stated above would follow from the monotonicity of this Rényi information measure with respect to the Rényi parameter (see [SBW14, Conjecture 26], [SBW14, Consequences 27 and 28]). A weaker form of (1.12) in terms of trace distance on the right-hand side was first conjectured in [Zha14, 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 [Bur69] \( D^2_B(\omega, \tau) \equiv 2(1 - \sqrt{F(\omega, \tau)}) \) 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 [BSS03, pages 4-5]. Consider spectral decompositions of a density operator \( \rho \) and a positive semi-definite operator \( \sigma \) acting on a finite-dimensional Hilbert space, such that \( \text{supp}(\rho) \subseteq \text{supp}(\sigma) \):

\[
\rho = \sum_x p_X(x) |\psi_x\rangle\langle \psi_x|, \quad (2.1)
\]
\[
\sigma = \sum_y f_Y(y) |\phi_y\rangle\langle \phi_y|. \quad (2.2)
\]

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

\[
T^{\delta,n}_{\rho|\sigma} \equiv \text{span} \left\{ |\phi_{y^n}\rangle : \left| - \frac{1}{n} \log(f_{Y^n}(y^n)) + \text{Tr}\{\rho \log \sigma\} \right| \leq \delta \right\}, \quad (2.3)
\]

where

\[
y^n \equiv y_1 \cdots y_n, \quad (2.4)
\]
\[
f_{Y^n}(y^n) \equiv \prod_{i=1}^n f_Y(y_i), \quad (2.5)
\]
\[
|\phi_{y^n}\rangle \equiv |\phi_{y_1}\rangle \otimes \cdots \otimes |\phi_{y_n}\rangle. \quad (2.6)
\]

We will overload the notation \( T^{\delta,n}_{\rho|\sigma} \) to refer also to the following classical typical set:

\[
T^{\delta,n}_{\rho|\sigma} \equiv \left\{ y^n : \left| - \frac{1}{n} \log(f_{Y^n}(y^n)) + \text{Tr}\{\rho \log \sigma\} \right| \leq \delta \right\}, \quad (2.7)
\]

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 \( T^{\delta,n}_{\rho|\sigma} \) be called \( \Pi^{\delta,n}_{\rho|\sigma} \). Consider that

\[
\text{Tr}\{\rho \log \sigma\} = \text{Tr} \left\{ \rho \log \left( \sum_y f_Y(y) |\phi_y\rangle\langle \phi_y| \right) \right\} \quad (2.8)
\]
\[
= \sum_y \langle \phi_y| \rho |\phi_y\rangle \log f_Y(y). \quad (2.9)
\]
Defining
\[ p_{\tilde{Y}}(y) \equiv \langle \phi_y | \rho | \phi_y \rangle, \]  
we can then write
\[ \text{Tr}\{ \rho \log \sigma \} = \sum_y p_{\tilde{Y}}(y) \log f_Y(y) \]  
\[ = \mathbb{E}_{\tilde{Y}} \left\{ \log f_Y(\tilde{Y}) \right\}. \]  
(2.12)

With this in mind, we can now calculate
\[ \text{Tr} \left\{ \Pi^n_{\rho|\sigma,\delta} \rho^{\otimes n} \right\} = \sum_{y^n \in \mathcal{T}_{\rho|\sigma}} \rho | \sigma \rangle \langle \sigma | \rho \rangle^{\otimes n} \langle \phi_{y^n} | \rho^{\otimes n} | \phi_{y^n} \rangle \]  
\[ = \sum_{y^n \in \mathcal{T}_{\rho|\sigma}} p_{\tilde{Y}}(y^n) \]  
\[ = \Pr \left\{ \tilde{Y} \in \mathcal{T}_{\rho|\sigma} \right\}. \]  
(2.15)

Based on the above reductions, and due to the notion of typicality with respect to the subspace \( \mathcal{T}_{\rho|\sigma} \) defined in (2.3), 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 \( n \), \( \text{Tr} \{ \Pi^n_{\rho|\sigma,\delta} \rho^{\otimes n} \} \geq 1 - \varepsilon \). In fact, the convergence \( \lim_{n \to \infty} \text{Tr} \{ \Pi^n_{\rho|\sigma,\delta} \rho^{\otimes n} \} = 1 \) can be taken exponentially fast in \( n \) for a constant \( \delta \) by employing the Hoeffding inequality \([\text{Hoe63}]\).

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

**Theorem 1** Let \( \rho_{AB} \) be a density operator, \( \sigma_{AB} \) 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:
\[ D(\rho_{AB} \| \sigma_{AB}) - D(\rho_B \| \sigma_B) \geq -\log F(\rho_{AB}, (\mathcal{V}_{AB} \circ \mathcal{R}_{B \rightarrow AB}^P \circ \mathcal{U}_B) (\rho_B)), \]  
for unitary channels \( \mathcal{U}_B \) and \( \mathcal{V}_{AB} \) defined in terms of some unitary operators \( \mathcal{U}_B \) and \( \mathcal{V}_{AB} \) as
\[ \mathcal{U}_B (\cdot) \equiv U_B (\cdot) U_B^\dagger, \]  
\[ \mathcal{V}_{AB} (\cdot) \equiv V_{AB} (\cdot) V_{AB}^\dagger, \]
and with \( \mathcal{R}_{B \rightarrow AB}^P \) the CPTP Petz recovery map:
\[ \mathcal{R}_{B \rightarrow AB}^P (\cdot) \equiv \sigma_{AB}^{1/2} \sigma_B^{-1/2} (\cdot) \sigma_B^{-1/2} \sigma_{AB}^{1/2}. \]  
(3.4)

**Proof.** Our proof of Theorem 1 proceeds very similarly to the proof of [FR14, 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 [FR14]. For convenience of the reader, we recall these statements in Appendix A.

The expression on the left-hand side of (3.1) is equivalent to
\[ -H(A|B)_{\rho} - \text{Tr} \{ \rho_{AB} \log \sigma_{AB} \} + \text{Tr} \{ \rho_B \log \sigma_B \}, \]  
(3.5)
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 \( \Pi^n_{AB} \) and \( \Pi^n_B \), respectively.

We begin by defining
\[ \mathcal{W}_n(X_{A^nB^n}) = \Pi^n_A \Pi^n_{AB} X_{A^nB^n} \Pi^n_{AB} \Pi^n_B. \]  
(3.6)
We employ the shorthand \( \mathcal{W}_n(X_{B^n}) \equiv \mathcal{W}_n(I_A \otimes X_{B^n}) \) throughout. Consider from the gentle measurement lemma [Win99], properties of the trace norm, and relative typicality that
\[ \text{Tr} \{ \mathcal{W}_n(\rho_{AB}^{\otimes n}) \} = \text{Tr} \{ \Pi^n_{AB} \Pi^n_B \rho_{AB}^{\otimes n} \Pi^n_B \Pi^n_{AB} \} \geq \text{Tr} \{ \Pi^n_{AB} \rho_{AB}^{\otimes n} \} - \| \Pi^n_B \rho_{AB}^{\otimes n} \Pi^n_B - \rho_{AB}^{\otimes n} \|_1 \geq 1 - \eta, \]
(3.9)
where \( \eta \) is an arbitrarily small positive number for sufficiently large \( n \). So we apply [FR14, Lemma 2.3] to find that
\[ D(\mathcal{W}_n(\rho_{AB}^{\otimes n}) \parallel \mathcal{W}_n(\rho_B^{\otimes n})) \leq n \left( D(\rho_{AB} \parallel I_A \otimes \rho_B) + \frac{\delta}{2} \right) \]
(3.10)
\[ = n \left( -H(A|B)_{\rho} + \frac{\delta}{2} \right), \]  
(3.11)
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 [FR14, Lemma B.2] then gives that
\[ \frac{1}{\text{Tr} \{ \mathcal{W}_n(\rho_{AB}^{\otimes n}) \}} \sqrt{F}(\mathcal{W}_n(\rho_{AB}^{\otimes n}), \mathcal{W}_n(\rho_B^{\otimes n})) \geq 2^n \left( H(A|B)_{\rho} - \frac{\delta}{2} \right). \]
(3.12)
Use [FR14, Lemma B.6] to remove the projector \( \Pi^n_{AB} \) from the second argument, so that
\[ \frac{1}{\text{Tr} \{ \mathcal{W}_n(\rho_{AB}^{\otimes n}) \}} \sqrt{F}(\mathcal{W}_n(\rho_{AB}^{\otimes n}), \Pi^n_B \rho_B^{\otimes n} \Pi^n_B) \geq 2^n \left( H(A|B)_{\rho} - \frac{\delta}{2} \right), \]
(3.13)
and the trace term can be eliminated at the expense of decreasing the exponent by a constant times \( n \):
\[ \sqrt{F}(\mathcal{W}_n(\rho_{AB}^{\otimes n}), \Pi^n_B \rho_B^{\otimes n} \Pi^n_B) \geq 2^n \left( H(A|B)_{\rho} - \frac{\delta}{2} \right). \]
(3.14)
Let an eigendecomposition of \( \sigma_B^{\otimes n} \) be given as
\[ \sigma^{\otimes n}_B = \sum_{s \in S_n} s \Pi_s, \]  
(3.15)
where \( S_n \) is the set of eigenvalues of \( \sigma_B^{\otimes n} \). By defining
\[ S_{n,\delta} \equiv \left\{ s \in S_n : \left| \frac{1}{n} \log(s) + \text{Tr} \{ \rho_B \log \sigma_B \} \right| \leq \delta \right\}, \]  
(3.16)
we see from (2.3) and the definition of $\Pi_B^n$ that
\[
\Pi_B^n = \sum_{s \in S_{n, \delta}} \Pi_s. \tag{3.17}
\]

Furthermore, it follows from a trivial combinatorial consideration that $|S_{n, \delta}| \leq \text{poly} (n)$. Then consider that $\sum_s \Pi_s = I$ and apply [FR14, Lemma B.7] to get
\[
\sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_B^n \rho_B^{\otimes n} \Pi_B^n) \leq \sum_{s \in S_n} \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_s \rho_B^{\otimes n} \Pi_s) \tag{3.18}
\]
\[
= \sum_{s \in S_{n, \delta}} \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_s \rho_B^{\otimes n} \Pi_s) \tag{3.19}
\]
\[
\leq |S_{n, \delta}| \max_{s \in S_{n, \delta}} \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_s \rho_B^{\otimes n} \Pi_s), \tag{3.20}
\]

where (3.19) 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
\[
\sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_B^n \rho_B^{\otimes n} \Pi_B^n) \leq \text{poly} (n) \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_s \rho_B^{\otimes n} \Pi_s). \tag{3.21}
\]

From the definition of $\Pi_s$ we can write
\[
\Pi_s = \sqrt{s} \left( \sigma_B^{-1/2} \right)^{\otimes n} \Pi_s. \tag{3.22}
\]

From the definition of $S_{n, \delta}$, we have that
\[
\sqrt{s} \leq 2^{\frac{1}{2}n[\text{Tr} (\rho_B \log \sigma_B) + \delta]}, \tag{3.23}
\]

giving that
\[
\sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \Pi_s \rho_B^{\otimes n} \Pi_s)
\]
\[
= \sqrt{s} \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \left( \sigma_B^{-1/2} \right)^{\otimes n} \Pi_s \left( \sigma_B^{-1/2} \right)^{\otimes n}) \tag{3.24}
\]
\[
\leq 2^{\frac{1}{2}n[\text{Tr} (\rho_B \log \sigma_B) + \delta]} \sqrt{F} (W_n (\rho_{AB}^{\otimes n}) , \left( \sigma_B^{-1/2} \right)^{\otimes n} \Pi_s \rho_B^{\otimes n} \Pi_s \left( \sigma_B^{-1/2} \right)^{\otimes n}) \tag{3.25}
\]
\[
= 2^{\frac{1}{2}n[\text{Tr} (\rho_B \log \sigma_B) + \delta]} \sqrt{F} (\Pi_s \left( \sigma_B^{-1/2} \right)^{\otimes n} W_n (\rho_{AB}^{\otimes n}) \left( \sigma_B^{-1/2} \right)^{\otimes n} \Pi_s, \rho_B^{\otimes n}), \tag{3.26}
\]

where the last equality is from [FR14, Lemma B.6]. Now, by [FR14, Lemma 4.2], there exists a unitary $U_B$ such that\(^3\)
\[
\sqrt{F} (\Pi_s \left( \sigma_B^{-1/2} \right)^{\otimes n} W_n (\rho_{AB}^{\otimes n}) \left( \sigma_B^{-1/2} \right)^{\otimes n} \Pi_s, \rho_B^{\otimes n})
\]
\[
\leq \text{poly} (n) \sqrt{F} \left( \left( \sigma_B^{-1/2} \right)^{\otimes n} W_n (\rho_{AB}^{\otimes n}) \left( \sigma_B^{-1/2} \right)^{\otimes n} U_B \rho_B^{\otimes n} (U_B^{\otimes n})^\dagger \right) \tag{3.27}
\]
\[
= \text{poly} (n) \sqrt{F} \left( W_n (\rho_{AB}^{\otimes n}) , \left( \sigma_B^{-1/2} \right)^{\otimes n} U_B \rho_B^{\otimes n} (U_B^{\otimes n})^\dagger \left( \sigma_B^{-1/2} \right)^{\otimes n} \right). \tag{3.28}
\]

\(^3\)Note that the unitary $U_B$ depends on $n$, but we suppress this in the notation for simplicity.
The equality above follows by applying [FR14, Lemma B.6]. Combining everything up until now, we get
\[
2^{\frac{1}{2} n (H(A|B) - \Tr \{ \rho_B \log \sigma_B \} - 2 \delta)} \leq \text{poly} (n) \sqrt{F} \left( \Pi_{AB}^n \Pi_B^P \rho_{AB}^n \Pi^n_{AB}, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) . \tag{3.29}
\]
Let an eigendecomposition of $\sigma_{AB}^n$ be given as
\[
\sigma_{AB}^n = \sum_{p \in \mathcal{P}_n} p \Pi_p , \tag{3.30}
\]
and
\[
\Pi^n_{AB} = \sum_{p \in \mathcal{P}_n} \Pi_p , \tag{3.31}
\]
where these developments follow the same reasoning as (3.15)-(3.17). Now we continue with the fact that $\sum_{p \in \mathcal{P}_n} \Pi_p = I$ and [FR14, Lemma B.7] to get that
\[
\sqrt{F} \left( \Pi_{AB}^n \Pi_B^P \rho_{AB}^n \Pi^n_{AB}, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right)
\]
\[
\leq \sum_{p \in \mathcal{P}_n} \sqrt{F} \left( \Pi_p \Pi_{AB}^n \Pi_B^P \rho_{AB}^n \Pi^n_{AB}, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) \tag{3.32}
\]
\[
= \sum_{p \in \mathcal{P}_n, \delta} \sqrt{F} \left( \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) \tag{3.33}
\]
\[
\leq |P_{n, \delta}| \max_{p \in \mathcal{P}_n, \delta} \sqrt{F} \left( \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) . \tag{3.34}
\]
Then there exists a $p$ such that
\[
\sqrt{F} \left( \Pi_{AB}^n \Pi_B^P \rho_{AB}^n \Pi^n_{AB}, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right)
\]
\[
\leq \text{poly} (n) \sqrt{F} \left( \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) . \tag{3.35}
\]
From the definition of $\Pi_p$ we have that
\[
\Pi_p = \frac{1}{\sqrt{p}} \left( \sigma_{AB}^{1/2} \right)^{\otimes n} \Pi_p , \tag{3.36}
\]
with $\sqrt{p} \geq 2^{\frac{1}{2} n [\Tr \{ \rho_{AB} \log \sigma_{AB} \} - \delta]}$. Then by defining $K \equiv 2^{\frac{1}{2} n [\Tr \{ \rho_{AB} \log \sigma_{AB} \} - \delta]} / \sqrt{p}$, we have that
\[
2^{\frac{1}{2} n [\Tr \{ \rho_{AB} \log \sigma_{AB} \} - \delta]} \sqrt{F} \left( \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right)
\]
\[
= K \sqrt{F} \left( \left( \sigma_{AB}^{1/2} \right)^{\otimes n} \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) \tag{3.37}
\]
\[
\leq \sqrt{F} \left( \left( \sigma_{AB}^{1/2} \right)^{\otimes n} \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) \tag{3.38}
\]
\[
= \sqrt{F} \left( \Pi_p \Pi_B^P \rho_{AB}^n \Pi_B^P \Pi_p, \left( \sigma_{AB}^{1/2} \right)^{\otimes n} \left( \sigma_{B}^{-1/2} \right)^{\otimes n} U_B^\otimes \rho_B^\otimes (U_B^\otimes) \dagger \left( \sigma_{B}^{-1/2} \right)^{\otimes n} \right) . \tag{3.39}
\]
Now by [FR14, Lemma 4.2], there exists a unitary $V_{AB}$ such that \(^4\)

$$\sqrt{F} \left( \Pi_F \Pi_B^\dagger \Pi_{AB}^n \Pi_B \Pi_{AB} \right) \left( \sigma_{AB}^{1/2} \right)^\otimes_n \left( \sigma_B^{-1/2} \right)^\otimes_n U_B^\otimes_n \rho_B^\otimes_n \left( U_B^\otimes_n \right)^\dagger \left( \sigma_B^{-1/2} \right)^\otimes_n \left( \sigma_{AB}^{1/2} \right)^\otimes_n \right) \leq \text{poly} \left( n \right) \sqrt{F} \left( \rho_{AB}^\otimes_n, V_{AB}^\otimes_n \left( \sigma_{AB}^{1/2} \right)^\otimes_n \left( \sigma_B^{-1/2} \right)^\otimes_n U_B^\otimes_n \rho_B^\otimes_n \left( U_B^\otimes_n \right)^\dagger \left( \sigma_B^{-1/2} \right)^\otimes_n \left( \sigma_{AB}^{1/2} \right)^\otimes_n \left( V_{AB}^\otimes_n \right)^\dagger \right). \tag{3.40}$$

Putting everything together, we get that

$$2^{\frac{1}{2} n \left( H\left( A | B \right) \right) - \text{Tr} \left( \rho_B \log \sigma_B \right) + \text{Tr} \left( \rho_{AB} \log \sigma_{AB} \right) - 3\delta} \leq \text{poly} \left( n \right) \sqrt{F} \left( \rho_{AB}^\otimes_n, V_{AB}^\otimes_n \left( \sigma_{AB}^{1/2} \right)^\otimes_n \left( \sigma_B^{-1/2} \right)^\otimes_n U_B^\otimes_n \rho_B^\otimes_n \left( U_B^\otimes_n \right)^\dagger \left( \sigma_B^{-1/2} \right)^\otimes_n \left( \sigma_{AB}^{1/2} \right)^\otimes_n \left( V_{AB}^\otimes_n \right)^\dagger \right) \tag{3.41}$$

$$= \text{poly} \left( n \right) \left[ F \left( \rho_{AB}, V_{AB} \sigma_{AB}^{1/2} \sigma_B^{-1/2} U_B \rho_B U_B^\dagger \sigma_B^{-1/2} \sigma_{AB} \sigma_{AB}^B \right) \right]^n \tag{3.42}$$

$$\leq \text{poly} \left( n \right) \left[ \max_{U_B, V_{AB}} F \left( \rho_{AB}, V_{AB} \sigma_{AB}^{1/2} \sigma_B^{-1/2} U_B \rho_B U_B^\dagger \sigma_B^{-1/2} \sigma_{AB} \sigma_{AB}^B \right) \right]^n. \tag{3.43}$$

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

$$2^{\frac{1}{2} n \left( H\left( A | B \right) \right) - \text{Tr} \left( \rho_B \log \sigma_B \right) + \text{Tr} \left( \rho_{AB} \log \sigma_{AB} \right) - 3\delta} \leq \sqrt{\text{poly} \left( n \right)} \sqrt{F} \left( \rho_{AB}, V_{AB} \sigma_{AB}^{1/2} \sigma_B^{-1/2} U_B \rho_B U_B^\dagger \sigma_B^{-1/2} \sigma_{AB} \sigma_{AB}^B \right). \tag{3.44}$$

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

$$- H\left( A | B \right) - \text{Tr} \left( \rho_B \log \sigma_B \right) + \text{Tr} \left( \rho_{AB} \log \sigma_{AB} \right) = D \left( \rho_{AB} || \sigma_{AB} \right) - D \left( \rho_B || \sigma_B \right), \tag{3.45}$$

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

$$D \left( \rho_{AB} || \sigma_{AB} \right) - D \left( \rho_B || \sigma_B \right) \geq - \log F \left( \rho_{AB}, V_{AB} \sigma_{AB}^{1/2} \sigma_B^{-1/2} U_B \rho_B U_B^\dagger \sigma_B^{-1/2} \sigma_{AB} \sigma_{AB}^B \right). \tag{3.46}$$

\[\text{Remark 2}\] Suppose in Theorem 1 that $\sigma_{AB}$ is a density operator. It remains open to quantify the performance of the rotated Petz recovery map $\mathcal{V}_{AB} \circ \mathcal{R}_{B \to AB}^P \circ \mathcal{U}_B$ on the reduced state $\sigma_B$. 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 [SBW14, Consequence 27] in terms of the following inequality:

$$D \left( \rho_{AB} || \sigma_{AB} \right) - D \left( \rho_B || \sigma_B \right) \geq - \log F \left( \rho_{AB}, \mathcal{R}_{B \to AB}^P \left( \rho_B \right) \right). \tag{3.47}$$

\[\text{Note that the unitary } V_{AB} \text{ depends on } n, \text{ but we suppress this in the notation for simplicity.}\]
If this conjecture is true, then one could perform the Petz recovery map on system $B$ and be guaranteed a perfect recovery of $\sigma_{AB}$ if the state of $B$ is $\sigma_B$, while having a performance limited by (3.47) if the state of $B$ is $\rho_B$. By a modification of the proof of Theorem 1, one can also establish the following lower bound:

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

(3.48)

for some unitaries $U_B$ and $V_{AB}$. The completely positive map $\sigma_{AB}^{1/2} V_{AB} U_B \sigma_B^{-1/2} (\cdot) \sigma_B^{-1/2} U_B^\dagger V_{AB}^\dagger \sigma_{AB}^{1/2}$ recovers $\sigma_{AB}$ perfectly from $\sigma_B$, while having a performance limited by (3.48) when recovering $\rho_{AB}$ from $\rho_B$. It is however unclear whether this map is trace preserving.

4 Remainder term for monotonicity of relative entropy

**Corollary 3** Let $\rho_S$ be a density operator and $\sigma_S$ be a positive semi-definite operator, both acting on a Hilbert space $\mathcal{H}_S$ and such that supp$(\rho_S) \subseteq$ supp$(\sigma_S)$. Let $\mathcal{N}_{S\rightarrow B}$ be a CPTP map taking density operators acting on $\mathcal{H}_S$ to density operators acting on $\mathcal{H}_B$ and such that $\mathcal{N}_{S\rightarrow B}(\sigma_S)$ is a positive definite operator. Then the following inequality refines monotonicity of relative entropy:

$$ D(\rho_{S}||\sigma_{S}) - D(\mathcal{N}_{S\rightarrow B}(\rho_{S})||\mathcal{N}_{S\rightarrow B}(\sigma_{S})) \geq -\log F(\rho_{S}, (\mathcal{V}_{S} \circ \mathcal{R}_{\sigma_{S} \mathcal{N}} \circ \mathcal{U}_{B} \mathcal{N}_{S\rightarrow B}(\rho_{S})) \mathcal{N}_{S\rightarrow B}(\sigma_{S})), $$

(4.1)

for unitary channels $\mathcal{U}_{B}$ and $\mathcal{V}_{S}$ defined in terms of some unitary operators $U_{B}$ and $V_{S}$ as

$$ \mathcal{U}_{B} (\cdot) \equiv U_{B} \cdot U_{B}^\dagger, $$

(4.2)

$$ \mathcal{V}_{S} (\cdot) \equiv V_{S} \cdot V_{S}^\dagger, $$

(4.3)

and with $\mathcal{R}_{\sigma_{S} \mathcal{N}}$ the CPTP Petz recovery map:

$$ \mathcal{R}_{\sigma_{S} \mathcal{N}} (\cdot) \equiv \sigma_{S}^{1/2} N^\dagger \left[ (\mathcal{N}_{S\rightarrow B}(\sigma_{S}))^{-1/2} (\cdot) (\mathcal{N}_{S\rightarrow B}(\sigma_{S}))^{-1/2} \right] \sigma_{S}^{1/2}, $$

(4.4)

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

**Proof of Theorem 3.** 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 [Sti55]. That is, for any channel $\mathcal{N}_{S\rightarrow B}$, there exists a unitary $W_{SE'\rightarrow BE}$ with input systems $SE'$ and output systems $BE$ such that

$$ \mathcal{N}_{S\rightarrow B}(\rho_{S}) = \text{Tr}_{E'} \left\{ W_{SE'\rightarrow BE} (\rho_{S} \otimes |0\rangle \langle 0|_{E'}) W_{SE'\rightarrow BE}^\dagger \right\}. $$

(4.5)

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

$$ \rho_{BE} \equiv W (\rho_{S} \otimes |0\rangle \langle 0|_{E'}) W^\dagger, $$

(4.6)

$$ \sigma_{BE} \equiv W (\sigma_{S} \otimes |0\rangle \langle 0|_{E'}) W^\dagger, $$

(4.7)

so that

$$ \mathcal{N}_{S\rightarrow B}(\rho_{S}) = \rho_{B}, \quad \mathcal{N}_{S\rightarrow B}(\sigma_{S}) = \sigma_{B}. $$

(4.8)
The Kraus operators of $\mathcal{N}_{S\rightarrow B}$ are given as

$$
\mathcal{N}_{S\rightarrow B} (\rho_S) = \sum_i \langle i|_E W (\rho_S \otimes |0\rangle_0 |0\rangle_{E'}) W^\dagger |i\rangle_E
$$

(4.9)

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

(4.10)

so that the adjoint map is given by

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

(4.11)

Furthermore, we have that

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

$$
= D (\rho_S \otimes |0\rangle_0 |0\rangle_{E'} || \sigma_S \otimes |0\rangle_0 |0\rangle_{E'}) - D (\rho_B || \sigma_B)
$$

(4.12)

$$
= D \left( W (\rho_S \otimes |0\rangle_0 |0\rangle_{E'}) W^\dagger || W (\sigma_S \otimes |0\rangle_0 |0\rangle_{E'}) W^\dagger \right) - D (\rho_B || \sigma_B)
$$

(4.13)

$$
= D (\rho_B || \sigma_B) - D (\rho_B || \sigma_B).
$$

(4.14)

Applying Theorem 1, we know that a lower bound on (4.14) is

$$
- \log F \left( \rho_B || \sigma_B \right)
$$

(4.15)

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 $W_{SE'\rightarrow BE} |0\rangle_{E'}$. We justify this as follows. Let $P_n$ denote the support projection of $\rho_{AB}^{\otimes n}$. By (3.40), since the supports of $\Pi_B \Pi_{BS} P_n$, $\rho_{AB}^{\otimes n}$, and $\left[ \sigma^{1/2}_{AB} \sigma_B^{1/2} U_B \rho_B U_B^\dagger \sigma_{BE}^{1/2} \sigma_B^{1/2} \right]^{\otimes n}$ are all contained in the support of $\sigma_{AB}^{\otimes n}$, one can apply [FR14, Lemma 4.2] on the Hilbert space $[\text{supp} (\sigma_{AB})]^{\otimes n} = \text{supp} (\sigma_{AB}^{\otimes n})$ to obtain a unitary $V_{AB}$ on this space, which may be extended to a unitary on the space $H_{AB}^{\otimes n}$ in an arbitrary way. Hence, the maximization in (3.43) can be restricted to unitaries $V_{AB}$ that are isometries on the support of $\sigma_{AB}$. Thus, we indeed have that $V_{BE}$ is an isometry on the support of $\sigma_{AB}$, which can be extended to an isometry on the image of $W_{SE'\rightarrow BE} |0\rangle_{E'}$.

Let us now unravel the term $\sigma_{BE}^{1/2} (\sigma_S)_{1/2} U_B \mathcal{N}_{S\rightarrow B} (\rho_S) U_B^\dagger (\mathcal{N}_{S\rightarrow B} (\sigma_S))^{-1/2}$ in the second argument above. Letting

$$
\omega_B \equiv (\mathcal{N}_{S\rightarrow B} (\sigma_S))^{-1/2} U_B \mathcal{N}_{S\rightarrow B} (\rho_S) U_B^\dagger (\mathcal{N}_{S\rightarrow B} (\sigma_S))^{-1/2},
$$

(4.16)

we then have that

$$
\sigma_{BE}^{1/2} (\sigma_S)_{1/2} U_B \rho_B U_B^\dagger (\sigma_S)^{-1/2} \sigma_{BE}^{1/2}
$$

$$
= \left( W (\sigma_S \otimes |0\rangle_0 |0\rangle_{E'}) W^\dagger \right)^{1/2} \omega_B \left( W (\sigma_S \otimes |0\rangle_0 |0\rangle_{E'}) W^\dagger \right)^{1/2}
$$

(4.17)

$$
= W (\sigma_S \otimes |0\rangle_0 |0\rangle_{E'})^{1/2} W^\dagger \omega_B W (\sigma_S \otimes |0\rangle_0 |0\rangle_{E'})^{1/2} W^\dagger
$$

(4.18)

$$
= W \left( \sigma_S^{1/2} \otimes |0\rangle_0 |0\rangle_{E'} \right) W^\dagger \omega_B W \left( \sigma_S^{1/2} \otimes |0\rangle_0 |0\rangle_{E'} \right) W^\dagger
$$

(4.19)

$$
= W \left( \sigma_S^{1/2} \otimes |0\rangle_0 |0\rangle_{E'} \right) W^\dagger \left[ \omega_B \otimes |1_E\rangle \right] W \left( \sigma_S^{1/2} \otimes |0\rangle_0 |0\rangle_{E'} \right) W^\dagger .
$$

(4.20)
Continuing, the last line above is equal to
\[
W \left( \sigma_S^{1/2} \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger \left[ \omega_B \otimes \sum_i |i\rangle \langle i|_E \right] W \left( \sigma_S^{1/2} \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger \tag{4.21}
\]
\[
= W \left[ \left( \sigma_S^{1/2} \sum_i (0\rangle_{E'} W^\dagger |i\rangle_E \omega_B \langle i|_E W |0\rangle_{E'} \right) \sigma_S^{1/2} \right] \otimes |0\rangle \langle 0|_{E'} \tag{4.22}
\]
\[
= W \left( \left[ \sigma_S^{1/2} \mathcal{N}^\dagger \left( \mathcal{N}_{S\rightarrow B} (\sigma_S) \right)^{-1/2} U_B \mathcal{N}_{S\rightarrow B} (\rho_S) U_B^\dagger \left( \mathcal{N}_{S\rightarrow B} (\sigma_S) \right)^{-1/2} \right] \sigma_S^{1/2} \right) \otimes |0\rangle \langle 0|_{E'} \bigg] W^\dagger . \tag{4.23}
\]

The Petz recovery map is defined as
\[
\mathcal{R}_{\sigma,\mathcal{N}} (\cdot) \equiv \sigma_S^{1/2} \mathcal{N}^\dagger \left[ \left( \mathcal{N}_{S\rightarrow B} (\sigma_S) \right)^{-1/2} (\cdot) \left( \mathcal{N}_{S\rightarrow B} (\sigma_S) \right)^{-1/2} \right] \sigma_S^{1/2} . \tag{4.24}
\]

Then by inspection, (4.23) is equal to
\[
W \left( \left[ \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N}_{S\rightarrow B} (\rho_S) U_B^\dagger \right) \right] \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger . \tag{4.25}
\]
So the fidelity in the remainder term of (4.15) is
\[
F \left( \rho_{BE}, V_{BE} W \left( \left[ \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N} (\rho) U_B^\dagger \right) \right] \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger (V_{BE})^\dagger \right)
\]
\[
= F \left( W (\rho_S \otimes |0\rangle \langle 0|_{E'}) W^\dagger, V_{BE} W \left( \left[ \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N} (\rho_S) U_B^\dagger \right) \right] \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger (V_{BE})^\dagger \right) \tag{4.26}
\]
\[
= F \left( \rho_S, |0\rangle \langle 0|_{E'}, W^\dagger V_{BE} W \left( \left[ \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N} (\rho_S) U_B^\dagger \right) \right] \otimes |0\rangle \langle 0|_{E'} \right) W^\dagger (V_{BE})^\dagger W |0\rangle_{E'} \right) \tag{4.27}
\]
\[
= F \left( \rho_S, V_S \left( \left[ \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N} (\rho_S) U_B^\dagger \right) \right] \right) V_S^\dagger \right) . \tag{4.28}
\]

Given that $V_{BE}$ acts only on the image of the isometry $W_{SE'\rightarrow BE} |0\rangle_{E'}$, the second equality follows because in this case the fidelity is invariant under the partial isometry $|0\rangle_{E'}, W^\dagger$. The last equality follows because we can define a unitary $V_S$ acting on the input space as
\[
V_S \equiv \langle 0|_{E'}, W^\dagger V_{BE} W |0\rangle_{E'} . \tag{4.29}
\]
So the final remainder term for monotonicity of relative entropy is
\[
D (\rho_S || \sigma_S) - D (\mathcal{N} (\rho_S) || \mathcal{N} (\sigma_S)) \geq - \log F \left( \rho_S, V_S \left( \mathcal{R}_{\sigma,\mathcal{N}} \left( U_B \mathcal{N} (\rho_S) U_B^\dagger \right) \right) \right) V_S^\dagger . \tag{4.30}
\]

**Remark 4** Suppose in Theorem 3 that $\sigma_S$ is a density operator. It remains open to quantify the performance of the rotated Petz recovery map $\mathcal{V}_S \circ \mathcal{R}_{\sigma,\mathcal{N}}^P \circ \mathcal{U}_B$ on the state $\mathcal{N}_{S\rightarrow B} (\sigma_S)$. 

12
5 Equivalence of relative entropy inequalities with remainder terms

As discussed in the introduction as well as in Remarks 2 and 4, it would be desirable to have refinements of the inequalities in (1.1) and (1.3)-(1.6) 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 [Bur69], defined as

\[ D_B^2(\rho, \sigma) = 2 \left( 1 - \sqrt{F(\rho, \sigma)} \right), \]

(5.1)

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

**Theorem 5** 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 \(\omega_{ABC}\) be a tripartite density operator such that \(\omega_C\) is positive definite. Then

\[ I(A; B|C)_{\omega} \geq D_B^2(\omega_{ABC}, \mathcal{R}_C^{P} \omega_{AC}(\omega_{BC})) \],

(5.2)

where \(\mathcal{R}_C^{P} \omega_{AC}(\cdot) \equiv \omega_{AC}^{1/2} \omega_{AC}^{-1/2} \omega_{AC}^{-1/2} \omega_{AC}^{1/2}\) denotes the Petz recovery channel.

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

\[ H(A|B)_{\rho} - \sum_x p_X(x) H(A|B)_{\rho^x} \geq \sum_x p_X(x) D_B^2(\rho^x_{AB}, \rho_{AB}^x \rho_B^{-1/2} \rho_{AB}^{-1/2} \rho_B^{-1/2} \rho_{AB}). \]

(5.3)

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

\[ D(\rho_{AB} || \sigma_{AB}) - D(\rho_B || \sigma_B) \geq D_B^2(\rho_{AB}, \mathcal{R}_{\sigma, \text{Tr}_A}^{P}(\rho_B)) \],

(5.4)

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

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 \(\rho_x\) and \(\sigma_x\) density operators such that \(\text{supp}(\rho_x) \subseteq \text{supp}(\sigma_x)\). Let \(\bar{\rho} \equiv \sum_x p_X(x) \rho_x\) and \(\bar{\sigma} \equiv \sum_x p_X(x) \sigma_x\) such that \(\bar{\sigma}\) is positive definite. Then

\[ \sum_x p_X(x) D(\rho_x || \sigma_x) - D(\bar{\rho} || \bar{\sigma}) \geq \sum_x p_X(x) D_B^2(\rho_x, \sigma_x^{1/2} \sigma_x^{-1/2} \bar{\rho} \sigma_x^{-1/2} \sigma_x^{1/2}). \]

(5.5)

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

\[ D(\rho || \sigma) - D(\mathcal{N}(\rho) || \mathcal{N}(\sigma)) \geq D_B^2(\rho, \mathcal{R}_{\sigma, \mathcal{N}}^{P}(\rho)) \],

(5.6)

where \(\mathcal{R}_{\sigma, \mathcal{N}}^{P}(\cdot) \equiv \sigma^{1/2} \mathcal{N}^\dagger \left( [\mathcal{N}(\sigma)]^{-1/2} (\cdot) [\mathcal{N}(\sigma)]^{-1/2} \right) \sigma^{1/2}\) denotes the Petz recovery channel with respect to \(\sigma\) and \(\mathcal{N}\).
Figure 1: It is well known that all of the above fundamental entropy inequalities are equivalent (see, e.g., [Rus02]). Theorem 5 extends this circle of equivalences to apply to refinements of these inequalities in terms of the Petz recovery map.

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 [SBW14, Consequence 28], we obtain $3 \Rightarrow 4 \Rightarrow 5$, leading to $3 \Leftrightarrow 4 \Leftrightarrow 5$. We can get $5 \Rightarrow 1$ by choosing $\rho = \omega_{ABC}$, $\sigma = \omega_{AC} \otimes \omega_B$, and $N = \text{Tr}_A$, so that

$$
\sigma^{1/2}N^\dagger \left( [N(\sigma)]^{-1/2} (\cdot) [N(\sigma)]^{-1/2} \right) \sigma^{1/2} \\
= [\omega_{AC} \otimes \omega_B]^{1/2} \left[ ( [\omega_C \otimes \omega_B]^{-1/2} (\cdot) [\omega_C \otimes \omega_B]^{-1/2} ) \otimes I_A \right] [\omega_{AC} \otimes \omega_B]^{1/2} \\
= \omega_{AC}^{1/2} \omega_C^{-1/2} \cdot \omega_C^{-1/2} \omega_{AC}^{1/2}.
$$

(5.7)  

(5.8)

Then

$$
I(A;B|C)_\omega = D(\omega_{ABC} \| \omega_{AC} \otimes \omega_B) - D(\omega_{BC} \| \omega_{C} \otimes \omega_B) \\
\geq 2 \left( 1 - \sqrt{F} (\omega_{ABC}, R_{\sigma, N}(\omega_{BC})) \right) \\
= 2 \left( 1 - \sqrt{F} (\omega_{ABC}, \omega_{AC}^{1/2} \omega_C^{-1/2} \omega_{BC} \omega_C^{-1/2} \omega_{AC}^{1/2}) \right).
$$

(5.9)  

(5.10)  

(5.11)

The implication $1 \Rightarrow 2$ follows by choosing

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

(5.12)
so that
\[
H(A|B)_{\mathbf{p}} - \sum_x p_x(x)H(A|B)_{\mathbf{p^x}} = I(A; X|B)_{\theta}
\]
\[
\geq 2 \left(1 - \sqrt{F} \left(\theta_{XAB}, \theta_{AB}^{1/2}\theta_B^{-1/2}\theta_{XB}^{-1/2}\theta_{AB}^{1/2}\right)\right)
\]
\[
= 2 \left(1 - \sum_x p_x(x)\sqrt{F} \left(\rho_{AB}^{x}, \rho_{AB}^{1/2}\rho_B^{-1/2}\rho_{XB}^{-1/2}\rho_{AB}^{1/2}\right)\right).
\]

The last remaining implication \(2 \Rightarrow 3\) has the most involved proof, which we establish now by using the idea from [LR73b, Section 3-E]. Throughout our proof, we employ Theorem V.3.3 of [Bha97]. 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
\[
Df(A) : H \to \sum_{\lambda, \eta} f'_{\lambda, \eta} P_A(\lambda) HP_A(\eta),
\]
where \(A = \sum_{\lambda} \lambda P_A(\lambda)\) is the spectral decomposition of \(A\), and \(f'_{\lambda, \eta}\) is the first divided difference function. In particular, if \(x : A(x) \in \mathcal{B}(\mathcal{H})_+\) is a differentiable function on an open interval in \(\mathbb{R}\), with derivative \(A'\), then
\[
\frac{d}{dx} f(A(x)) = \sum_{\lambda, \eta} f'_{\lambda, \eta} P_A(x)(\lambda) A'(x) P_A(x)(\eta),
\]
so that
\[
\frac{d}{dx} \text{Tr} \{f(A(x))\} = \text{Tr} \{f'(A(x)) A'(x)\}.
\]

In particular, if \(A(x) = A + xB\), then
\[
\frac{d}{dx} \text{Tr} \{f(A(x))\} = \text{Tr} \{f'(A(x)) B\}.
\]

We can now proceed. In what follows, we will be taking \(A(x) = \sigma_{AB} + x\rho_{AB}\), where \(\sigma_{AB}\) is a positive definite density operator, \(\rho_{AB}\) 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
\[
H(A|B)_{xG} = xH(A|B)_{G},
\]
where \(x\) is a positive scalar and \(G_{AB}\) is a positive semi-definite operator on systems \(AB\). Let
\[
\xi_{YAB} \equiv \frac{1}{x + 1} \left|0\right>_Y \otimes \sigma_{AB} + \frac{x}{x + 1} \left|1\right>_Y \otimes \rho_{AB},
\]
with \(\sigma_{AB}\) a positive definite density operator and \(\rho_{AB}\) a density operator. Then it follows from homogeneity and concavity with the Petz remainder term (by assumption) that
\[
H(A|B)_{\sigma + x\rho} = (x + 1) H(A|B)_{\xi}
\]
\[
\geq (x + 1) \left[ \frac{1}{x + 1} H(A|B)_{\sigma} + \frac{x}{x + 1} H(A|B)_{\rho} + R(x, \sigma_{AB}, \rho_{AB}) \right]
\]
\[
= H(A|B)_{\sigma} + xH(A|B)_{\rho} + (x + 1) R(x, \sigma_{AB}, \rho_{AB}),
\]
Manipulating the above inequality then gives
\[ 2 \left( 1 - \frac{1}{x+1} \sqrt{F (\sigma_{AB}, \xi_{AB}^{-1/2} \xi_B^{-1/2} \sigma_B^{-1/2} \xi_{AB}^{-1/2})} + \frac{x}{x+1} \sqrt{F (\rho_{AB}, \xi_{AB}^{-1/2} \xi_B^{-1/2} \xi_{AB}^{-1/2})} \right). \]
(5.25)

We evaluate this by using (5.25) to find that
\[ H (A|B)_{\sigma + x\sigma} - H (A|B)_\sigma \geq H (A|B)_{\rho} + \frac{x + 1}{x} R (x, \sigma_{AB}, \rho_{AB}). \]
(5.26)

Taking the limit as \( x \downarrow 0 \) then gives
\[ \lim_{x \downarrow 0} \frac{H (A|B)_{\sigma + x\rho} - H (A|B)_\sigma}{x} = \frac{d}{dx} H (A|B)_{\sigma + x\rho} \bigg|_{x=0} \geq H (A|B)_{\rho} + \lim_{x \downarrow 0} \frac{x + 1}{x} R (x, \sigma_{AB}, \rho_{AB}). \]
(5.27)

We now evaluate the limits separately, beginning with the one on the left hand side. So we consider
\[ \frac{d}{dx} H (A|B)_{\sigma + x\rho} = \frac{d}{dx} [ - \text{Tr} \left\{ (\sigma_{AB} + x\rho_{AB}) \log (\sigma_{AB} + x\rho_{AB}) \right\} + \text{Tr} \left\{ (\sigma_B + x\rho_B) \log (\sigma_B + x\rho_B) \right\}]. \]
(5.28)

We evaluate this by using \( \frac{d}{dy} [ g (y) \log g (y) ] = [ \log g (y) + 1 ] g' (y) \) and (5.19) to find that
\[ \frac{d}{dx} \text{Tr} \left\{ (\sigma_{AB} + x\rho_{AB}) \log (\sigma_{AB} + x\rho_{AB}) \right\} = \text{Tr} \left\{ [ \log (\sigma_{AB} + x\rho_{AB}) + I_{AB} ] \rho_{AB} \right\}, \]
(5.29)

so that
\[ \frac{d}{dx} H (A|B)_{\sigma + x\rho} = - \text{Tr} \{ \rho_{AB} \log (\sigma_{AB} + x\rho_B) \} + \text{Tr} \{ \rho_B \log (\sigma_B + x\rho_B) \}, \]
(5.30)

and thus
\[ \frac{d}{dx} H (A|B)_{\sigma + x\rho} \bigg|_{x=0} = - \text{Tr} \{ \rho_{AB} \log \sigma_{AB} \} + \text{Tr} \{ \rho_B \log \sigma_B \}. \]
(5.31)

Substituting back into the inequality (5.27), we find that
\[ - \text{Tr} \{ \rho_{AB} \log \sigma_{AB} \} + \text{Tr} \{ \rho_B \log \sigma_B \} \geq - \text{Tr} \{ \rho_{AB} \log \rho_{AB} \} + \text{Tr} \{ \rho_B \log \rho_B \} + \lim_{x \downarrow 0} \frac{x + 1}{x} R (x, \sigma_{AB}, \rho_{AB}), \]
(5.32)

which is equivalent to (cf., [LR73b, Eq. (3.2)])
\[ D (\rho_{AB} || \sigma_{AB}) - D (\rho_B || \sigma_B) \geq \lim_{x \downarrow 0} \frac{x + 1}{x} R (x, \sigma_{AB}, \rho_{AB}). \]
(5.33)
Since
\[
\lim_{x \downarrow 0} \sqrt{F} \left( \rho_{AB}, \xi_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \right) = \sqrt{F} \left( \rho_{AB}, \sigma_{AB}^{1/2} \sigma_{AB}^{1/2}, \rho_B \sigma_B^{1/2} \sigma_B^{1/2} \right),
\]
(5.36)
it remains to show that
\[
\lim_{x \downarrow 0} \frac{1 - \sqrt{F} \left( \sigma_{AB}, \xi_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \right)}{x} = \frac{d}{dx} \sqrt{F} \left( \sigma_{AB}, \xi_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \right)_{x=0} = 0. \quad (5.37)
\]
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
\[
\sqrt{F} \left( \sigma_{AB}, \xi_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \right)
= \text{Tr} \left\{ \left( \sigma_{AB}^{1/2} \xi_{AB} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \xi_{AB} \sigma_{AB}^{1/2} \xi_{AB} \xi_{AB}^{1/2} \right)^{1/2} \right\}
= \text{Tr} \left\{ \left( \sigma_{AB}^{1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \left( \sigma_{B} + x \rho_{B} \right)^{1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \sigma_{AB}^{1/2} \right)^{1/2} \right\},
\]
(5.38)
as well as
\[
\frac{d}{dx} \text{Tr} \left\{ \left( G(x) \right)^{1/2} \right\} = \frac{1}{2} \text{Tr} \left\{ G(x)^{-1/2} \frac{d}{dx} G(x) \right\},
\]
(5.40)
which follows from (5.19). Applying the above rule, we get that \( \frac{d}{dx} \) of (5.39) is equal to
\[
\text{Tr} \left\{ \left( \sigma_{AB}^{1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \left( \sigma_{B} + x \rho_{B} \right)^{1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \sigma_{AB}^{1/2} \right)^{-1/2} \sigma_{AB}^{1/2} \frac{d}{dx} \left( \sigma_{AB} + x \rho_{AB} \right)^{1/2} \left( \sigma_{B} + x \rho_{B} \right)^{-1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \sigma_{AB}^{1/2} \right\}.
\]
(5.41)
Now, take the limit as \( x \downarrow 0 \) to find that (5.41) is equal to
\[
\frac{d}{dx} \sqrt{F} \left( \sigma_{AB}, \xi_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \right)_{x=0}
= \text{Tr} \left\{ \left( \sigma_{AB}^{1/2} \xi_{AB}^{1/2}, \rho_B \xi_B^{1/2} \xi_B^{1/2} \xi_{AB} \sigma_{AB}^{1/2} \xi_{AB}^{1/2} \right)^{1/2} \right\} \times \sigma_{AB} \frac{d}{dx} \left( \sigma_{AB} + x \rho_{AB} \right)^{1/2} \left( \sigma_{B} + x \rho_{B} \right)^{-1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \sigma_{AB}^{1/2} \right|_{x=0} \sigma_{AB}^{1/2}
\]
(5.42)
\[
= \text{Tr} \left\{ \frac{1}{2} \sigma_{AB} \frac{d}{dx} \left( \sigma_{AB} + x \rho_{AB} \right)^{1/2} \left( \sigma_{B} + x \rho_{B} \right) \left( \sigma_{B} + x \rho_{B} \right)^{-1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \sigma_{AB}^{1/2} \right|_{x=0} \right\} \left( \sigma_{AB} \right)^{-1/2} \times
\]
(5.43)
\[
= \text{Tr} \left\{ \frac{1}{2} \sigma_{AB} \frac{d}{dx} \left( \sigma_{AB} + x \rho_{AB} \right)^{1/2} \left( \sigma_{B} + x \rho_{B} \right)^{-1/2} \sigma_{B} (\sigma_{B} + x \rho_{B})^{-1/2} (\sigma_{AB} + x \rho_{AB})^{1/2} \right|_{x=0} \right\} \left( \sigma_{AB} \right)^{-1/2}
\]
(5.44)
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:

\[
\lim_{x \to 0} \text{Tr} \left\{ \frac{d}{dx} \left[ (\sigma_{AB} + x\rho_{AB})^{1/2} \right] (\sigma_B + x\rho_B)^{-1/2} \sigma_B (\sigma_B + x\rho_B)^{-1/2} (\sigma_{AB} + x\rho_{AB})^{1/2} \right\} \\
= \text{Tr} \left\{ \frac{d}{dx} \left[ (\sigma_{AB} + x\rho_{AB})^{1/2} \right] \right\vert_{x=0} (\sigma_B + x\rho_B)^{-1/2} \sigma_B (\sigma_B + x\rho_B)^{-1/2} (\sigma_{AB} + x\rho_{AB})^{1/2} \right\} \\
= \text{Tr} \left\{ \frac{d}{dx} \left[ (\sigma_{AB} + x\rho_{AB})^{1/2} \right] \right\vert_{x=0} \sigma_{AB}^{1/2} \right\} \\
= \frac{1}{2} \text{Tr} \{ \rho_{AB} \} \\
= \frac{1}{2},
\]

where the second to last line follows from (5.17). We now consider the second term:

\[
\lim_{x \to 0} \text{Tr} \left\{ (\sigma_{AB} + x\rho_{AB})^{1/2} \frac{d}{dx} \left[ (\sigma_B + x\rho_B)^{-1/2} \right] \sigma_B (\sigma_B + x\rho_B)^{-1/2} (\sigma_{AB} + x\rho_{AB})^{1/2} \right\} \\
= \text{Tr} \left\{ \sigma_{AB}^{1/2} \frac{d}{dx} \left[ (\sigma_B + x\rho_B)^{-1/2} \right] \right\vert_{x=0} \sigma_B (\sigma_B + x\rho_B)^{-1/2} (\sigma_{AB} + x\rho_{AB})^{1/2} \right\} \\
= \text{Tr} \left\{ \sigma_B \frac{d}{dx} \left[ (\sigma_B + x\rho_B)^{-1/2} \right] \right\vert_{x=0} \sigma_B^{3/2} \right\} \\
= -\frac{1}{2} \text{Tr} \left\{ \sigma_B^{-3/2} \sigma_{AB}^{3/2} \right\} \\
= -\frac{1}{2} \text{Tr} \{ \rho_B \} \\
= -\frac{1}{2}.
\]

The third to last line follows from (5.17). 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

\[
\text{Tr} \left\{ \sigma_{AB} + x\rho_{AB} \right\} (\sigma_B + x\rho_B)^{-1/2} \sigma_B (\sigma_B + x\rho_B)^{-1/2} (\sigma_{AB} + x\rho_{AB})^{1/2} \right\vert_{x=0} = 0,
\]

which allows us to conclude that

\[
\frac{d}{dx} \sqrt{F} \left( \sigma_{AB}, \xi_{AB}^{1/2}, \xi_B^{1/2}, \sigma_B^{1/2}, \xi_B^{1/2}, \xi_{AB}^{1/2} \right) \vert_{x=0} = 0,
\]

Hence, we can conclude that the following inequality is a consequence of (5.3):

\[
D (\rho_{AB} || \sigma_{AB}) - D (\rho_B || \sigma_B) \geq 2 \left( 1 - \sqrt{F} \left( \rho_{AB}, \sigma_{AB}^{1/2}, \rho_B^{1/2}, \sigma_B^{1/2} \right) \right).
\]
Note: After the completion of the present paper, the works in [Wil15] and [STH15] appeared, which build upon ideas established in this paper. The main contribution of [Wil15] is to show that the rotated Petz map in Corollary 3 can take a more particular form. Specifically, the unitary channel $U_B$ in Corollary 3 can be taken to commute with $N(σ)$ and the unitary channel $V_S$ can be taken to commute with $σ$. The main contribution of [STH15] is to show that the fidelity remainder term in Corollary 3 can be replaced with the “measured relative entropy” and the rotated Petz map can be replaced with a “twirled Petz map.” Please refer to [Wil15] and [STH15] 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.

A Auxiliary lemmas from [FR14]

In this appendix, for the convenience of the reader, we list verbatim the relevant lemmas that we have used from [FR14].

Lemma 6 (Lemma 2.3 of [FR14]) Let $ρ$ be a density operator, let $σ$ be a non-negative operator on the same space, and let $\{W_n\}_{n∈\mathbb{N}}$ be a sequence of trace non-increasing completely positive maps on the $n$-fold tensor product of this space. If $\text{tr}(W_n(ρ^⊗n))$ decreases less than exponentially in $n$, i.e.,

$$\liminf_{n→∞} e^{ξn} \text{tr}(W_n(ρ^⊗n)) > 0$$  \hspace{1cm} (A.1)

for any $ξ > 0$, then

$$\limsup_{n→∞} \frac{1}{n} D(W_n(ρ^⊗n)∥W_n(σ^⊗n)) ≤ D(ρ∥σ) .$$  \hspace{1cm} (A.2)

Lemma 7 (Lemma 4.2 of [FR14]) Let $ρ_{R^nS^n}$ be a permutation-invariant non-negative operator on $(R ⊗ S)^⊗n$ and let $σ_{RS}$ be a non-negative operator on $R ⊗ S$. Furthermore, let $W_{RS}$ be a permutation-invariant operator on $R^⊗n$ with $∥W_{R^n}∥_∞ ≤ 1$. Then there exists a unitary $U_R$ on $R$ such that

$$\sqrt{F}(ρ_{R^nS^n}, U_R^⊗n σ_{RS}^⊗n(U_R^⊗n)\dagger) ≥ (n + 1)^{-d^2} \sqrt{F}(W_{R^n}ρ_{R^nS^n}W_{R^n}^\dagger, σ_{RS}^⊗n) ,$$  \hspace{1cm} (A.3)

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

Lemma 8 (Lemma B.2 of [FR14]) For any non-negative operators $ρ$ and $σ$

$$D(ρ∥σ) ≥ −2 \log_2 \frac{\sqrt{F}(ρ, σ)}{\text{tr}(ρ)} .$$  \hspace{1cm} (A.4)
Lemma 9 (Lemma B.6 of [FR14]) For any non-negative operators $\rho$ and $\sigma$ and any operator $W$ on the same space we have
\[
\sqrt{F}(\rho, W\sigma W^\dagger) = \sqrt{F}(W^\dagger \rho W, \sigma) .
\] (A.5)

Lemma 10 (Lemma B.7 of [FR14]) Let $\rho$ and $\sigma$ be non-negative operators and let $\{W_d\}_{d \in D}$ be a family of operators such that $\sum_{d \in D} W_d = \text{id}$. Then
\[
\sum_{d \in D} \sqrt{F}(W_d^\dagger \rho W_d, \sigma) \geq \sqrt{F}(\rho, \sigma) .
\] (A.6)

References


