Annales Henri Poincaré

CrossMark

Necessary Criterion for Approximate Recoverability

David Sutter and Renato Renner

Abstract. A tripartite state ρ_{ABC} forms a Markov chain if there exists a recovery map $\mathcal{R}_{B \to BC}$ acting only on the B-part that perfectly reconstructs ρ_{ABC} from ρ_{AB}. To achieve an approximate reconstruction, it suffices that the conditional mutual information $I(A : C|B)_\rho$ is small, as shown recently. Here we ask what conditions are necessary for approximate state reconstruction. This is answered by a lower bound on the relative entropy between ρ_{ABC} and the recovered state $\mathcal{R}_{B \to BC}(\rho_{AB})$. The bound consists of the conditional mutual information and an entropic correction term that quantifies the disturbance of the B-part by the recovery map.

1. Introduction

A recovery map is a trace-preserving completely positive map that reconstructs parts of a composite system. More precisely, for a tripartite state ρ_{ABC} on $A \otimes B \otimes C$, we can consider a recovery map $\mathcal{R}_{B \to BC}$ from B to $B \otimes C$ that reconstructs the C-part from the B-part only. If such a reconstruction is perfectly possible, i.e., if

$$\rho_{ABC} = \mathcal{R}_{B \to BC}(\rho_{AB}) \tag{1}$$

we call ρ_{ABC} a *(quantum) Markov chain* in order $A \leftrightarrow B \leftrightarrow C$.[1]

The structure of Markov chains is well understood. A state ρ_{ABC} is a Markov chain if and only if there exists a decomposition of the B system as

---

[1]We usually omit the identity map and the identity operator in our notation when its use is clear from the context. For example, we write $\mathcal{R}_{B \to BC}(\rho_{AB})$ instead of $(\mathcal{I}_A \otimes \mathcal{R}_{B \to BC})(\rho_{AB})$ and $\rho_B \rho_{AB} \rho_B$ instead of $(\mathrm{id}_A \otimes \rho_B) \rho_{AB} (\mathrm{id}_A \otimes \rho_B)$. We will drop the order of the Markov chain if it is $A \leftrightarrow B \leftrightarrow C$.

$B = \oplus_j (b_j^L \otimes b_j^R)$ such that

$$\rho_{ABC} = \bigoplus_j P(j)\, \rho_{Ab_j^L} \otimes \rho_{b_j^R C}, \tag{2}$$

with states $\rho_{Ab_j^L}$ on $A \otimes b_j^L$, $\rho_{b_j^R C}$ on $b_j^R \otimes C$, and a probability distribution P [19]. A measure that is useful to describe Markov chains is the *conditional mutual information* that is given by

$$I(A:C|B) = \operatorname{tr} \rho_{ABC} \big(\log \rho_{ABC} + \log \rho_B - \log \rho_{AB} - \log \rho_{BC}\big), \tag{3}$$

whenever the trace is defined, i.e., whenever the operator $\rho_{ABC}(\log \rho_{ABC} + \log \rho_B - \log \rho_{AB} - \log \rho_{BC})$ is trace class. One often restricts to the case where the *conditional von Neumann entropy* $H(A|B) = -D(\rho_{AB}\|\mathrm{id}_A \otimes \rho_B)$ is finite, where $D(\rho\|\sigma) := \operatorname{tr}\rho(\log \rho - \log \sigma)$ denotes the relative entropy between ρ and σ. Indeed, in this case, the data processing inequality [24,40] implies that $H(A|BC) = -D(\rho_{ABC}\|\mathrm{id}_A \otimes \rho_{BC})$ is also finite, and hence, the operators $\rho_{ABC}(\log \rho_{AB} - \log \rho_B)$ and $\rho_{ABC}(\log \rho_{ABC} - \log \rho_{BC})$ are both trace class, implying that their difference is trace class, too. We further note that for finite-dimensional Hilbert spaces the conditional mutual information may be written as $I(A:C|B)_\rho := H(AB)_\rho + H(BC)_\rho - H(B)_\rho - H(ABC)_\rho$ where $H(A)_\rho := -\operatorname{tr} \rho_A \log \rho_A$ is the von Neumann entropy of the marginal state on A.

It has been shown that a state ρ_{ABC} is a Markov chain if and only if its conditional mutual information $I(A:C|B)_\rho$ vanishes [30,31]. Furthermore, the *Petz recovery map* (also known as *transpose map*)

$$\mathcal{T}_{B \to BC} : X_B \mapsto \rho_{BC}^{\frac{1}{2}} \big(\rho_B^{-\frac{1}{2}} X_B \rho_B^{-\frac{1}{2}} \otimes \mathrm{id}_C\big) \rho_{BC}^{\frac{1}{2}} \tag{4}$$

recovers such states perfectly, i.e., (1) holds with $\mathcal{R}_{B \to BC} = \mathcal{T}_{B \to BC}$.

Tripartite states ρ_{ABC} that have a small conditional mutual information are called *approximate Markov chains*. The justification for this terminology is a recent result [16] proving that for any state ρ_{ABC} there exists a recovery map $\mathcal{R}_{B \to BC}$ such that

$$I(A:C|B)_\rho \geq -\log F\big(\rho_{ABC}, \mathcal{R}_{B \to BC}(\rho_{AB})\big), \tag{5}$$

where $F(\tau, \omega) := \|\sqrt{\tau}\sqrt{\omega}\|_1^2$ denotes the *fidelity* between τ and ω.[2] Inequality (5) shows that the Markov property (1) approximately holds whenever the conditional mutual information is small. However, there exist tripartite states with a small conditional mutual information whose distance to any Markov chain is nevertheless large [11,20]. As a consequence, approximate quantum Markov chains are not necessarily close to quantum Markov chains. We refer to "Appendix A" for a more detailed explanation of this phenomenon.

Inequality (5) has been refined in a series of works [5,8,21,35–37,41]. More precisely, the initial bound from [16] has been strengthened by replacing the right-hand side of (5) by the measured relative entropy between the original and the recovered state (see (9) below for a definition). This result came with a

---

[2]Recall that for any two states τ and ω we have $F(\tau, \omega) \in [0, 1]$ and that $F(\tau, \omega) = 1$ if and only if $\tau = \omega$ (see, e.g., [28]).

novel proof based on the notion of quantum state redistribution [8]. The proof
has later been simplified by utilizing tools from semidefinite programming [5].
In [36], it was shown that there exists a universal recovery map, i.e., one that
does not depend on the A system, that satisfies (5). Another major step was
the discovery that (5), as well as generalizations thereof, can be obtained by
complex interpolation theory [41], providing further insight into the structure
of the recovery map. In [37], an intuitive proof of (5) based on the *spectral
pinching method* was presented. In [21], it was shown that there exists an
explicit recovery map (of the form (7)) that satisfies (5). The most recent
result [35, Theorem 4.1] shows that for any state ρ_{ABC} we have

$$I(A:C|B)_\rho \geq D_{\mathbb{M}}\big(\rho_{ABC}\|\mathcal{P}_{B\to BC}(\rho_{AB})\big), \tag{6}$$

for the explicit recovery map

$$\mathcal{P}_{B\to BC}(\cdot) := \int_{-\infty}^{\infty} \beta_0(\mathrm{d}t)\,\mathcal{P}_{B\to BC}^{[t]}(\cdot)$$

$$\text{with}\quad \mathcal{P}_{B\to BC}^{[t]}(\cdot) = \rho_{BC}^{\frac{1+it}{2}}\,(\rho_B^{-\frac{1+it}{2}} X_B \rho_B^{-\frac{1-it}{2}} \otimes \mathrm{id}_C)\rho_{BC}^{\frac{1-it}{2}} \tag{7}$$

and the probability measure

$$\beta_0(\mathrm{d}t) := \frac{\pi}{2}\left(\cosh(\pi t) + 1\right)^{-1}\mathrm{d}t \tag{8}$$

on $\mathbb{R}$. $D_{\mathbb{M}}$ denotes the *measured relative entropy*, which is defined as

$$D_{\mathbb{M}}(\rho\|\sigma) := \sup_{M\in\mathcal{M}} D(M(\rho)\|M(\sigma)), \tag{9}$$

where $\mathcal{M}$ is the set of all quantum-classical channels $M(\omega) = \sum_x (\mathrm{tr} M_x \omega)|x\rangle\langle x|$
with $\{M_x\}$ a positive operator valued measure (POVM) and $\{|x\rangle\}$ an orthonor-
mal basis. A simple property of the measured relative entropy ensures that
$D_{\mathbb{M}}(\tau\|\omega) \geq -\log F(\tau,\omega)$ for all states τ,ω [8], which shows that (6) implies (5).
We further note that the recovery map $\mathcal{P}_{B\to BC}$ given in (7) is *universal* in
the sense that it only depends on ρ_{BC} and it satisfies $\mathcal{P}_{B\to BC}(\rho_B) = \rho_{BC}$.
The interested reader can find additional information about the concepts and
achievements around (5) in [34].

   Inequality (5) shows that there always exists a recovery map whose recov-
ery quality (measured in terms of the logarithm of the fidelity) is of the order
of the conditional mutual information. This shows that a small conditional
mutual information is a sufficient condition for a state to be approximately
recoverable. In other words, (5) gives an entropic characterization for the set
of tripartite states that can be approximately recovered.

   In this work, we are interested in an opposite statement. This corresponds
to an inequality that bounds the distance between ρ_{ABC} and any reconstructed
state $\mathcal{R}_{B\to BC}(\rho_{AB})$ from below with an entropic functional of ρ_{ABC} and the
recovery map $\mathcal{R}_{B\to BC}$ that involves the conditional mutual information. Such
an inequality is the converse to (5) and gives a necessary condition for approx-
imate recoverability.

1.1. Main Result

For any trace-preserving completely positive map $\mathcal{E}$ on a system S we denote by $\mathrm{Inv}(\mathcal{E})$ the set of density operators τ on S which are left invariant under the action of $\mathcal{E}$, i.e.,

$$\mathrm{Inv}(\mathcal{E}) := \{\tau : \mathcal{E}(\tau) = \tau\}. \tag{10}$$

We may now quantify the deviation of any state ρ from the set $\mathrm{Inv}(\mathcal{E})$ by

$$\Lambda_{\max}(\rho\|\mathcal{E}) := \inf_{\tau \in \mathrm{Inv}(\mathcal{E})} D_{\max}(\rho\|\tau) , \tag{11}$$

where $D_{\max}(\omega\|\sigma) := \inf\{\lambda \in \mathbb{R} : \omega \leq 2^\lambda \sigma\}$ denotes the *max-relative entropy*. The $\Lambda_{\max}$-quantity has the property that it is zero if and only if $\mathcal{E}$ leaves ρ invariant[3], i.e.,

$$\Lambda_{\max}(\rho\|\mathcal{E}) = 0 \quad \Longleftrightarrow \quad \mathcal{E}(\rho) = \rho. \tag{12}$$

Main Result We prove that for any state ρ_{ABC} on $A \otimes B \otimes C$ and any recovery map $\mathcal{R}_{B \to BC}$ from the B system to the $B \otimes C$ system we have

$$D\big(\rho_{ABC}\|\mathcal{R}_{B \to BC}(\rho_{AB})\big) + \Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B}) \geq I(A : C|B)_\rho , \tag{13}$$

where $D(\tau\|\sigma) := \mathrm{tr}\,\tau \log \tau - \mathrm{tr}\,\tau \log \sigma$ if $\mathrm{supp}(\tau) \subseteq \mathrm{supp}(\sigma)$ and $+\infty$ otherwise denotes the *relative entropy*, and $\mathcal{R}_{B \to B} := \mathrm{tr}_C \circ \mathcal{R}_{B \to BC}$ is the action of the recovery map $\mathcal{R}_{B \to BC}$ on B. We refer to Theorem 3.1 for a more precise statement.

Cases where the $\Lambda_{\max}$-Term Vanishes To interpret the term $\Lambda_{\max}$ in (13), note that the recovery map $\mathcal{R}_{B \to BC}$ generally not only reads the content of system B in order to generate C, but also disturbs it. $\Lambda_{\max}$ quantifies the amount of this disturbance of B, taking system A as a reference. In particular, $\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B}) = 0$ if $\mathcal{R}_{B \to BC}$ is "read only" on B, i.e., if $\rho_{AB} = \mathcal{R}_{B \to B}(\rho_{AB})$. Inequality (13) then simplifies to

$$D\big(\rho_{ABC}\|\mathcal{R}_{B \to BC}(\rho_{AB})\big) \geq I(A : C|B)_\rho . \tag{14}$$

We further note that in case $\mathcal{R}_{B \to BC}$ is a recovery map that is "read only" on B its output state $\sigma_{ABC} := \mathcal{R}_{B \to BC}(\rho_{AB})$ is a Markov chain since

$$H(A|B)_\rho \leq H(A|BC)_\sigma \leq H(A|B)_\sigma = H(A|B)_\rho , \tag{15}$$

where the two inequality steps follow from the data processing inequality [22,23] applied for $\mathcal{R}_{B \to BC}$ and tr_C, respectively and hence $I(A : C|B)_\sigma = H(A|B)_\sigma - H(A|BC)_\sigma = 0$.

---

[3]Note that the max-relative entropy has a definiteness property which ensures that for a sequence $(\omega_k)_{k \in \mathbb{N}}$ of states such that $\lim_{k \to \infty} D_{\max}(\tau\|\omega_k) = 0$ we have $\lim_{k \to \infty} \omega_k = \tau$. This follows from the fact that $-\log F(\tau, \omega) \leq D_{\max}(\tau\|\omega)$ [2,3] and the definiteness property of the fidelity [1,39], i.e., $\lim_{k \to \infty} F(\tau, \omega_k) = 1$ implies $\lim_{k \to \infty} \omega_k = \tau$.

1.2. Tightness of the Main Result

We next discuss several aspects concerning the tightness of (13). This will also give a better understanding about the role of the $\Lambda_{\max}$-term. We first note that by combining (6) with (13) we obtain

$$D_{\mathbb{M}}\big(\rho_{ABC}\|\mathcal{P}_{B\to BC}(\rho_{AB})\big) \leq I(A:C|B)_{\rho} \tag{16}$$

$$\leq \inf_{\mathcal{R}_{B\to BC}} \big\{ D\big(\rho_{ABC}\|\mathcal{R}_{B\to BC}(\rho_{AB})\big) + \Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B\to B}) \big\}, \tag{17}$$

where the recovery map $\mathcal{P}_{B\to BC}$ on the left-hand side is given by (7) and the infimum is over all recovery maps $\mathcal{R}_{B\to BC}$ that map B to $B \otimes C$. The main difference between the lower and upper bound for the conditional mutual information given by (16) and (17), respectively, is the $\Lambda_{\max}$-term.

Classical Case Inequalities (16) and (17) hold with equality in case ρ_{ABC} is a classical state, i.e.,

$$\rho_{ABC} = \sum_{a,b,c} P_{ABC}(a,b,c)|a\rangle\langle a|_A \otimes |b\rangle\langle b|_B \otimes |c\rangle\langle c|_C , \tag{18}$$

for some probability distribution P_{ABC}. To see this, we first note that if ρ_{ABC} is classical (in which case ρ_{ABC} and all its marginals commute pairwise) a straightforward calculation gives

$$I(A:C|B)_{\rho} = D\big(\rho_{ABC}\|\mathcal{T}_{B\to BC}(\rho_{AB})\big), \tag{19}$$

for the Petz recovery map $\mathcal{T}_{B\to BC}$ defined in (4). Furthermore, if ρ_{ABC} is classical $\mathcal{T}_{B\to BC}(\rho_{AB}) = \rho_{BC}\rho_B^{-1}\rho_{AB}$. We further see that $\mathrm{tr}_C \mathcal{T}_{B\to BC}(\rho_{AB}) = \mathcal{T}_{B\to B}(\rho_{AB}) = \rho_{AB}$ and hence

$$\Lambda_{\max}(\rho_{AB}\|\mathcal{T}_{B\to B}) = 0. \tag{20}$$

This shows that in the classical case (17) is an equality and that the Petz recovery map $\mathcal{T}_{B\to BC}$ minimizes the right-hand side of (17).

   We further note that in the classical case the measured relative entropy coincides with the relative entropy and the rotated Petz recovery map $\mathcal{P}_{B\to BC}$ that satisfies (16) simplifies to the Petz recovery map $\mathcal{T}_{B\to BC}$. This together with (19) then shows that (16) holds with equality in the classical case.

Necessity of the $\Lambda_{\max}$-Term A natural question regarding (13) is whether the $\Lambda_{\max}$-term is necessary. Here we show that this is indeed the case by constructing an example proving that a large conditional mutual information does not imply that all recovery maps are bad and hence the $\Lambda_{\max}$-term is indispensable.

   More precisely, in Sect. 4.1 we construct a generic example showing that for any constant $\kappa < \infty$ there exists a classical state ρ_{ABC} (i.e., a state of the form (18)) such that

$$\kappa\, D_{\max}\big(\rho_{ABC}\|\mathcal{R}_{B\to BC}(\rho_{AB})\big) < I(A:C|B)_{\rho}, \tag{21}$$

for some recovery map $\mathcal{R}_{B\to BC}$ that satisfies $\mathcal{R}_{B\to BC}(\rho_B) = \rho_{BC}$. A similar construction (also given in Sect. 4.1) shows that there exists another classical state ρ_{ABC} such that

$$\kappa\, D_{\max}\big(\mathcal{R}_{B\to BC}(\rho_{AB})\|\rho_{ABC}\big) < I(A:C|B)_{\rho}, \tag{22}$$

for some recovery map $\mathcal{R}_{B \to BC}$ that satisfies $\mathcal{R}_{B \to BC}(\rho_B) = \rho_{BC}$.

These constructions therefore show that an additional term like $\Lambda_{\max}(\rho_{AB} \| \mathcal{R}_{B \to B})$, which measures the deviation from a "read only" map on B, is necessary to obtain a lower bound on the relative entropy between a state and its reconstructed version. The example has an even stronger implication. It shows that the $\Lambda_{\max}$-term is necessary even if one tries to bound the max-relative entropy between a state and its reconstructed version, i.e., $D_{\max}(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB}))$, which cannot be smaller than $D(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB}))$, from below.[4] The two strict inequalities (21) and (22) show that the $\Lambda_{\max}$-term is also necessary if one would allow for swapping the two arguments of the relative (or even max-relative) entropy. Furthermore, restricting the set of recovery maps such that they satisfy $\mathcal{R}_{B \to BC}(\rho_B) = \rho_{BC}$ still requires the $\Lambda_{\max}$-term.

Since for classical states (19) holds, these examples also show that for the task of minimizing the relative entropy between ρ_{ABC} and its reconstructed state $\mathcal{R}_{B \to BC}(\rho_{AB})$ the Petz recovery map can be far from being optimal—even in the classical case. The examples further show that considering recovery maps that leave the B system invariant (i.e., they only "read" the B-part) is a considerable restriction. We refer to Sect. 4.1 for more information about these examples.

Optimality of the $\Lambda_{\max}$-Term Even in the case where ρ_{ABC} is not classical, (13) is still close to optimal. We present two arguments why this is the case. First, we show that the $\Lambda_{\max}$-term cannot be replaced by a relative entropy measure that is smaller than the max-relative entropy. More precisely, (13) is violated if the max-relative entropy in the definition of $\Lambda_{\max}(\rho_{AB} \| \mathcal{R}_{B \to B})$ is replaced with any α-Rényi relative entropy for any $\alpha \in [\frac{1}{2}, \infty)$. We refer to Sect. 4.2 for more information.

The $\Lambda_{\max}$-term in (13) quantifies the (max-relative entropy) distance between ρ_{AB} and its closest state that is invariant under $\mathcal{R}_{B \to B}$. A natural question is if (13) remains valid if the $\Lambda_{\max}$-term is replaced by the max-relative entropy distance between ρ_{AB} and $\mathcal{R}_{B \to B}(\rho_{AB})$, i.e., $D_{\max}(\rho_{AB} \| \mathcal{R}_{B \to B}(\rho_{AB}))$. This however is ruled out. To see this we recall that by the example mentioned above in (21) there exists a tripartite state ρ_{ABC} and a recovery map $\mathcal{R}_{B \to BC}$ such that

$$2 D_{\max}\big(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB})\big) < I(A : C|B)_\rho. \tag{23}$$

The data processing inequality for the max-relative entropy [14,38] and the fact that the max-relative entropy cannot be smaller than the relative entropy then imply

$$D\big(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB})\big) < I(A : C|B)_\rho - D_{\max}\big(\rho_{AB} \| \mathcal{R}_{B \to B}(\rho_{AB})\big), \tag{24}$$

which shows that (13) is no longer valid for the modified $\Lambda_{\max}$-term described above.

---

[4]The max-relative entropy and its properties are discussed in more detail in Sect. 2. It is the largest sensible relative entropy measure.

1.3. Related Results

Using the continuity of the conditional entropy, it is possible to derive an upper bound for the conditional mutual information of a state ρ_{ABC} in terms of its distance to any reconstructed state $\sigma_{ABC} := \mathcal{R}_{B \to BC}(\rho_{AB})$, where $\mathcal{R}_{B \to BC}$ denotes an arbitrary recovery map [4,16]. This leads to a lower bound on the relative entropy between ρ_{ABC} and $\mathcal{R}_{B \to BC}(\rho_{AB})$ that however depends on the dimension of the A system. To see this, let $[0,1] \ni x \mapsto h(x) := -x \log x - (1 - x) \log(1-x)$ denote the *binary entropy function* and let $\Delta(\tau, \omega) := \frac{1}{2} \|\tau - \omega\|_1$ be the *trace distance* between τ and ω. The data processing inequality [22,23] implies that

$$I(A:C|B)_\rho = H(A|B)_\rho - H(A|BC)_\rho \le H(A|BC)_\sigma - H(A|BC)_\rho. \qquad (25)$$

By the improved Alicki-Fannes inequality [43, Lemma 2] we find

$$I(A:C|B)_\rho \le 2\Delta(\rho, \sigma) \log(\dim A) + \big(1 + \Delta(\rho, \sigma)\big) h\left(\frac{\Delta(\rho, \sigma)}{1 + \Delta(\rho, \sigma)}\right) \qquad (26)$$

$$\le 2\sqrt{\Delta(\rho, \sigma)}\big(\log(\dim A) + 1\big), \qquad (27)$$

where we used that $(1+x)h(\frac{x}{1+x}) \le 2\sqrt{x}$ for all $x \in [0,1]$ and $\Delta(\rho, \sigma) \in [0,1]$. Together with Pinsker's inequality [13,32] this gives

$$D\big(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB})\big) \ge \frac{2}{\ln 2} \Delta\big(\rho_{ABC}, \mathcal{R}_{B \to BC}(\rho_{AB})\big)^2$$

$$\ge \frac{I(A:C|B)_\rho^4}{8 \ln 2 \big(\log(\dim A) + 1\big)^4}. \qquad (28)$$

The fact that this bound explicitly depends on the dimension of the system A is unsatisfactory. Furthermore, the example discussed above in (21) shows that such a dependence on the dimension is unavoidable.

A different approach to derive an upper bound for the conditional mutual information of a state ρ_{ABC} in terms of its distance to a reconstructed state $\mathcal{R}_{B \to BC}(\rho_{AB})$ was taken in [15, Theorem 11 and Remark 12] (see also [35, Proposition F.1]). It was shown that for any state ρ_{ABC}

$$\int_{-\infty}^{\infty} \beta_0(\mathrm{d}t)\, \bar{D}_2\big(\rho_{ABC} \| \mathcal{P}_{B \to BC}^{[t]}(\rho_{AB})\big) \ge I(A:C|B)_\rho, \qquad (29)$$

where β_0 and $\mathcal{P}_{B \to BC}^{[t]}$ are given in (8) and (7), respectively and $\bar{D}_2(\tau\|\omega) := \log \operatorname{tr} \tau^2 \omega^{-1}$ denotes Petz' Rényi relative entropy of order 2 [29]. The examples discussed above imply that the left-hand side of (29) can be much larger than the relative entropy between ρ_{ABC} and $\mathcal{R}_{B \to BC}(\rho_{AB})$ for the optimal recovery map $\mathcal{R}_{B \to BC}$. In other words, rotated Petz recovery maps are generally far from optimal recovery maps.

2. One-Shot Relative Entropies

The goal of this section is to derive a triangle-like inequality for the relative entropy (see Lemma 2.1) which will be used in the proof of our main result, i.e.,

Theorem 3.1. To understand Lemma 2.1, we need to review a few properties of one-shot relative entropy measures.

2.1. Preliminaries

Let $S(A)$ and $P(A)$ denote the set of density and nonnegative operators on A, respectively. For any linear operator L on A, the *trace norm* is given by $\|L\|_1 := \mathrm{tr}|L|$ with $|L| := \sqrt{L^\dagger L}$. For $\rho, \sigma \in P(A)$ we write $\rho \ll \sigma$ if the support of ρ is contained in the support of σ. Within this document our Hilbert spaces are assumed to be separable. We define the *min-relative entropy* [33] as

$$D_{\min}(\rho\|\sigma) := -\log \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1^2 = -\log F(\rho, \sigma) \tag{30}$$

and the *max-relative entropy* [14,33] as

$$D_{\max}(\rho\|\sigma) := \inf\{\lambda \in \mathbb{R} : \rho \leq 2^\lambda \sigma\}. \tag{31}$$

As the names suggest, the min-relative entropy cannot be larger than the max-relative entropy, or more precisely we have

$$D_{\min}(\rho\|\sigma) \leq D(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma), \tag{32}$$

with strict inequalities in the generic case [27,38]. The max-relative entropy turns out to be the largest relative entropy measure that satisfies the data processing inequality and is additive under tensor products [38, Section 4.2.4]. We also note that it follows immediately from the definition that the max-relative entropy cannot increase if the same positive map is applied to both arguments (see also [26, Theorem 2] for a more general statement).

The min- and max-relative entropies can be seen as the extreme points of a family of relative entropies called *minimal quantum Rényi relative entropy* (also known as *sandwiched Rényi relative entropy*) [27,42]. For $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty)$ and $\rho, \sigma \in P(A)$, this family is defined as

$$D_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1} \log \frac{1}{\mathrm{tr}\rho} \mathrm{tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha & \text{if } \rho \ll \sigma \vee \alpha < 1 \\ \infty & \text{otherwise}. \end{cases} \tag{33}$$

It can be shown [27] that

$$D_{\frac{1}{2}}(\rho\|\sigma) = D_{\min}(\rho\|\sigma), \quad \lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma),$$
$$\text{and} \quad \lim_{\alpha \to \infty} D_\alpha(\rho\|\sigma) = D_{\max}(\rho\|\sigma). \tag{34}$$

Furthermore the minimal quantum Rényi relative entropy is monotone in $\alpha \in [\frac{1}{2}, \infty)$ [27, Theorem 7], i.e.,

$$D_\alpha(\rho\|\sigma) \leq D_{\alpha'}(\rho\|\sigma) \quad \text{for} \quad \alpha \leq \alpha'. \tag{35}$$

2.2. Triangle-Like Inequality for Relative Entropy

It is well known that the relative entropy does not satisfy the triangle inequality. For the three (classical) qubit states $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{4}\mathrm{id}_2$, $\sigma = \frac{1}{2}|1\rangle\langle 1| + \frac{1}{4}\mathrm{id}_2$, and $\omega = \frac{1}{2}\mathrm{id}_2$, we have $D(\rho\|\sigma) > D(\rho\|\omega) + D(\omega\|\sigma)$. The following lemma proves a triangle-like inequality for the minimal quantum Rényi relative entropy.

Lemma 2.1. *Let A be a separable Hilbert space, let $\rho \in \mathrm{S}(A)$, $\sigma, \omega \in \mathrm{P}(A)$ and let $\alpha \in [\frac{1}{2}, 1]$. Then*

$$D_\alpha(\rho\|\sigma) \leq D_\alpha(\rho\|\omega) + D_{\max}(\omega\|\sigma). \tag{36}$$

Proof. For $\alpha \in [\frac{1}{2}, 1)$, the function $t \mapsto t^{\frac{1-\alpha}{\alpha}}$ is operator monotone on $[0, \infty)$ [6, Theorem V.1.9]. Furthermore, the function $X \mapsto \mathrm{tr}X^\alpha$ is monotone on the set of Hermitian operators on a separable Hilbert space, since the function $X \mapsto X^\alpha$ is operator monotone [6]. By definition of the max-relative entropy, we find

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \mathrm{tr}\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha \leq D_\alpha(\rho\|\omega) + D_{\max}(\omega\|\sigma). \tag{37}$$

for $\alpha < 1$. The case $\alpha = 1$ then follows by continuity. $\qquad\square$

Remark 2.2. We note that if A is a finite-dimensional Hilbert space then (36) is valid for all $\alpha \in [\frac{1}{2}, \infty)$. This follows from the fact that $t \mapsto t^{\frac{1-\alpha}{\alpha}}$ is operator anti-monotone [38] for $\alpha > 1$ and that the function $X \mapsto \mathrm{tr}X^\alpha$ is monotone on the set of Hermitian operators [9, Theorem 2.10].

Very recently, a similar triangle-like inequality for Rényi relative entropies that additionally involves trace-preserving completely positive maps has been established in [10]. The following remarks show that Lemma 2.1 is optimal and that there is not much flexibility to prove triangle-like inequalities for the relative entropy different than (36).

Remark 2.3. Lemma 2.1 is optimal in the sense that (36) is no longer valid if $D_{\max}$ is replaced with D_α for any $\alpha \in [\frac{1}{2}, 1]$. To see this, let $p \in (0, 1)$ and consider three classical distributions on $\{0, 1\} \times \{0, 1\}$ defined by

$$P_{XY}(x, y) := \begin{cases} p & \text{if } x = y = 0 \\ \frac{1-p}{3} & \text{otherwise} \end{cases},$$

$$Q_{XY}(x, y) := \begin{cases} p & \text{if } x = y = 1 \\ \frac{1-p}{3} & \text{otherwise} \end{cases},$$

$$S_{XY}(x, y) = \frac{1}{4}. \tag{38}$$

A simple calculation shows that

$$D(P\|Q) = \frac{4p - 1}{3} \log \frac{3p}{1 - p} \tag{39}$$

$$D(P\|S) = p \log 4p + (1 - p) \log \frac{4(1 - p)}{3} \tag{40}$$

$$D_\alpha(S\|Q) = \frac{1}{\alpha - 1} \log \left(\frac{3^\alpha}{4^\alpha(1 - p)^{\alpha - 1}} + \frac{1}{4^\alpha p^{\alpha - 1}} \right). \tag{41}$$

Choosing $p = 1 - 2^{-\alpha}$ reveals that

$$D(P\|Q) > D(P\|S) + D_\alpha(S\|Q) \quad \text{for all} \quad \alpha \in [\tfrac{1}{2}, \infty). \tag{42}$$

In the limit $\alpha \to \infty$, the strict inequality (42) becomes an equality.

Remark 2.4. The statement of Lemma 2.1 is no longer true if the max-relative entropy and the relative entropy on the right-hand side of (36) are exchanged. To see this consider the three classical binary probability distributions

$$P(x) := \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{otherwise} \end{cases},$$

$$Q(x) := \begin{cases} 1 - \varepsilon & \text{if } x = 0 \\ \varepsilon & \text{otherwise} \end{cases},$$

$$S(x) := \begin{cases} 1 - \frac{p}{2} & \text{if } x = 0 \\ \frac{p}{2} & \text{otherwise} \end{cases}, \tag{43}$$

with $p, \varepsilon \in (0, 1)$. This gives

$$D(P\|Q) = (1 - p) \log \frac{1 - p}{1 - \varepsilon} + p \log \frac{p}{\varepsilon} \tag{44}$$

$$D(S\|Q) = \frac{2 - p}{2} \log \frac{2 - p}{2 - 2\varepsilon} + \frac{p}{2} \log \frac{p}{2\varepsilon} \tag{45}$$

$$D_{\max}(P\|S) = \max \left\{ \log \frac{2(1 - p)}{2 - p}, \log 2 \right\} = 1. \tag{46}$$

For $p = \frac{7}{8}$ and $\varepsilon = \frac{1}{8}$ we find that

$$D(P\|Q) > D_{\max}(P\|S) + D(S\|Q). \tag{47}$$

This shows that it is crucial which term in Lemma 2.1 carries a max-relative entropy.

Remark 2.5. The relative entropy satisfies a triangle-like inequality different from Lemma 2.1. For the *log-Euclidean α-Rényi divergence* $D_\alpha^\flat(\omega\|\sigma) := \frac{1}{1-\alpha} \log \operatorname{tr} e^{\alpha \log \rho + (1-\alpha) \log \sigma}$ it is known [25] that

$$D(\rho\|\sigma) \leq \frac{\alpha}{\alpha - 1} D(\rho\|\omega) + D_\alpha^\flat(\omega\|\sigma) \quad \text{for} \quad \alpha \in (1, \infty). \tag{48}$$

We also note that $D_\infty^\flat(\omega\|\sigma) \leq D_{\max}(\omega\|\sigma)$ which shows that in the limit $\alpha \to \infty$ we obtain Lemma 2.1 for the case $\alpha = 1$.

3. Main Result and Proof

Theorem 3.1. *Let A, B, and C be separable Hilbert spaces, let $\rho_{ABC} \in S(A \otimes B \otimes C)$, and let $\mathcal{R}_{B \to BC}$ be a trace-preserving completely positive map from B to $B \otimes C$. Then*

$$D\big(\rho_{ABC}\|\mathcal{R}_{B \to BC}(\rho_{AB})\big) + \Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B}) \geq I(A : C|B)_\rho. \tag{49}$$

The quantity $\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B})$ is defined in (11) and $\mathcal{R}_{B \to B} := \operatorname{tr}_C \circ \mathcal{R}_{B \to BC}$. To prove the assertion of Theorem 3.1, we make use of a known lemma stating that the conditional mutual information of a tripartite density operator is bounded from above by the smallest relative entropy distance to Markov chains. Let $MC(A \otimes B \otimes C)$ denote the set of Markov chains on $A \otimes B \otimes C$, i.e., tripartite density operators $\rho_{ABC} \in S(A \otimes B \otimes C)$ that satisfy (1).

Lemma 3.2 ([20, Theorem 4]). *Let $\rho_{ABC} \in \mathrm{S}(A \otimes B \otimes C)$. Then*

$$I(A:C|B)_\rho \leq \inf_{\mu \in \mathrm{MC}} D(\rho_{ABC}\|\mu_{ABC}).\tag{50}$$

Proof. The proof we provide here follows the lines of a proof by Jenčová (see the short note after the acknowledgements in [20]), but extends it to general separable spaces.

Let $\mu_{ABC} \in \mathrm{MC}$ and assume without loss of generality that the relative entropy $D(\rho_{ABC}\|\mu_{ABC})$ is finite. (If there is no such state then the infimum in (50) equals infinity and the statement is trivial.) Due to the data processing inequality [24, 40] we have

$$0 \leq D(\rho_B\|\mu_B) \leq D(\rho_{AB}\|\mu_{AB}) \leq D(\rho_{ABC}\|\mu_{ABC})\tag{51}$$

and

$$0 \leq D(\rho_B\|\mu_B) \leq D(\rho_{BC}\|\mu_{BC}) \leq D(\rho_{ABC}\|\mu_{ABC}).\tag{52}$$

In particular, the relative entropies $D(\rho_{AB}\|\mu_{AB})$, $D(\rho_{BC}\|\mu_{BC})$, and $D(\rho_B\|\mu_B)$ are finite. We thus have

$$D(\rho_{ABC}\|\mu_{ABC}) + D(\rho_B\|\mu_B) - D(\rho_{AB}\|\mu_{AB}) - D(\rho_{BC}\|\mu_{BC})$$
$$= \mathrm{tr}\Big(\rho_{ABC}\big(\log\rho_{ABC} - \log\mu_{ABC} + \log\rho_B - \log\mu_B - \log\rho_{AB} + \log\mu_{AB}$$
$$- \log\rho_{BC} + \log\mu_{BC}\big)\Big).$$

Using the Markov chain property (2) for μ_{ABC}, i.e.,

$$\mu_{ABC} = \bigoplus_j P(j)\mu_{Ab_j^L} \otimes \mu_{b_j^R C} \quad \text{for} \quad B = \bigoplus_j b_j^L \otimes b_j^R,\tag{53}$$

it is straightforward to verify that

$$\log\mu_{ABC} + \log\mu_B - \log\mu_{AB} - \log\mu_{BC} = 0.\tag{54}$$

The above can thus be simplified to

$$D(\rho_{ABC}\|\mu_{ABC}) + D(\rho_B\|\mu_B) - D(\rho_{AB}\|\mu_{AB}) - D(\rho_{BC}\|\mu_{BC})$$
$$= \mathrm{tr}\Big(\rho_{ABC}\big(\log\rho_{ABC} + \log\rho_B - \log\rho_{AB} - \log\rho_{BC}\big)\Big) = I(A:C|B).$$

It follows from (51) and (52) that

$$D(\rho_{ABC}\|\mu_{ABC}) \geq I(A:C|B),\tag{55}$$

which concludes the proof. $\qquad\square$

In order to prove Theorem 3.1, we need one more lemma that relates the distance to Markov chains and the $\Lambda_{\max}$-quantity defined in (11).

Lemma 3.3. *Let $\rho_{AB} \in \mathrm{P}(A \otimes B)$ and $\mathcal{R}_{B\to BC}$ be a trace-preserving completely positive map. Then*

$$\inf_{\mu \in \mathrm{MC}} D_{\max}\big(\mathcal{R}_{B\to BC}(\rho_{AB})\|\mu_{ABC}\big) \leq \Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B\to B}).\tag{56}$$

Proof. For the proof, we first assume that the system A has a finite dimension, so that conditional entropies of the form $H(A|B)$ are finite. The data processing inequality for the max-relative entropy [14,17,38] implies that

$$
\begin{aligned}
\inf_{\mu_{ABC}} & \{D_{\max}\big(\mathcal{R}_{B\to BC}(\rho_{AB})\|\mu_{ABC}\big) : \mu_{ABC} \in \mathrm{MC}\} \\
& \leq \inf_{\tau_{AB}} \{D_{\max}\big(\mathcal{R}_{B\to BC}(\rho_{AB})\|\mathcal{R}_{B\to BC}(\tau_{AB})\big) \\
& \quad : \mathcal{R}_{B\to BC}(\tau_{AB}) \in \mathrm{MC}, \tau_{AB} \in \mathrm{S}(A \otimes B)\} \tag{57} \\
& \leq \inf_{\tau_{AB}} \{D_{\max}(\rho_{AB}\|\tau_{AB}) : \mathcal{R}_{B\to BC}(\tau_{AB}) \in \mathrm{MC}, \tau_{AB} \in \mathrm{S}(A \otimes B)\}. \tag{58}
\end{aligned}
$$

Furthermore, because the data processing inequality for the conditional entropy [22,23] implies that $H(A|BC)_{\mathcal{R}_{B\to BC}(\tau_{AB})} \geq H(A|B)_{\tau_{AB}}$ for any $\tau_{AB} \in \mathrm{S}(A \otimes B)$, we also have

$$
\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B\to B}) \implies H(A|BC)_\mu \geq H(A|B)_\mu \quad \text{for } \mu_{ABC} = \mathcal{R}_{B\to BC}(\tau_{AB}). \tag{59}
$$

Note that the inequality on the right-hand side of the implication must, again by the data processing inequality, be an equality, which means that $I(A : C|B)_\mu = 0$ and, hence, that $\mu \in \mathrm{MC}$. This proves the general implication

$$
\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B\to B}) \implies \mathcal{R}_{B\to BC}(\tau_{AB}) \in \mathrm{MC}. \tag{60}
$$

We now use it to obtain

$$
\begin{aligned}
\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B\to B}) & \\
& = \inf_{\tau_{AB}} \{D_{\max}(\rho_{AB}\|\tau_{AB}) : \tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B\to B})\} \tag{61} \\
& \geq \inf_{\tau_{AB}} \{D_{\max}(\rho_{AB}\|\tau_{AB}) : \mathcal{R}_{B\to BC}(\tau_{AB}) \in \mathrm{MC}, \tau_{AB} \in \mathrm{S}(A \otimes B)\}. \tag{62}
\end{aligned}
$$

Combining this with (58) completes the proof for the case where the system A is finite-dimensional.

To extend the claim to general separable Hilbert spaces, consider a sequence of finite-rank projectors $(\Pi_A^k)_{k\in\mathbb{N}}$ on A with $\Pi_A^k \leq \Pi_A^{k+1}$ for any $k \in \mathbb{N}$ that, for $k \to \infty$, converges to the identity in the weak, and hence also in the strong, operator topology [18]. It follows from the monotonicity of the max-relative entropy under positive maps and (56) for finite-dimensional A that

$$
\begin{aligned}
\inf_{\mu\in\mathrm{MC}} & D_{\max}\big(\Pi_A^k \mathcal{R}_{B\to BC}(\rho_{AB})\Pi_A^k \| \Pi_A^k \mu_{ABC} \Pi_A^k\big) \\
& \leq \inf_{\mu\in\mathrm{MC}} D_{\max}\big(\Pi_A^k \mathcal{R}_{B\to BC}(\rho_{AB})\Pi_A^k \| \mu_{ABC}\big) \tag{63} \\
& \leq \Lambda_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \mathcal{R}_{B\to B}). \tag{64}
\end{aligned}
$$

The right-hand side can be bounded for any $k \in \mathbb{N}$ by

$$\Lambda_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \mathcal{R}_{B \to B})$$

$$= \inf_{\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B \to B})} D_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \tau_{AB}) \tag{65}$$

$$\leq \inf_{\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B \to B})} D_{\max}\left(\Pi_A^k \rho_{AB} \Pi_A^k \| \frac{\Pi_A^k \tau_{AB} \Pi_A^k}{\mathrm{tr}\, \Pi_A^k \tau_{AB} \Pi_A^k}\right) \tag{66}$$

$$= \inf_{\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B \to B})} \left\{ D_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \Pi_A^k \tau_{AB} \Pi_A^k) + \log \mathrm{tr}\, \Pi_A^k \tau_{AB} \Pi_A^k \right\} \tag{67}$$

$$\leq \inf_{\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B \to B})} D_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \Pi_A^k \tau_{AB} \Pi_A^k), \tag{68}$$

where the first inequality uses that $\Pi_A^k \tau_{AB} \Pi_A^k / \mathrm{tr}\, \Pi_A^k \tau_{AB} \Pi_A^k \in \mathrm{Inv}(\mathcal{R}_{B \to B})$. The final step follows because τ_{AB} is a density operator and hence $\mathrm{tr}\, \Pi_A^k \tau_{AB} \leq 1$ for any projector Π_A^k on A. Using once again the monotonicity of the max-relative entropy under positive maps we find with the above

$$\Lambda_{\max}(\Pi_A^k \rho_{AB} \Pi_A^k \| \mathcal{R}_{B \to B}) \leq \inf_{\tau_{AB} \in \mathrm{Inv}(\mathcal{R}_{B \to B})} D_{\max}(\rho_{AB} \| \tau_{AB}) \tag{69}$$

$$= \Lambda_{\max}(\rho_{AB} \| \mathcal{R}_{B \to B}). \tag{70}$$

To conclude the proof, it thus suffices to establish that

$$\inf_{\mu \in \mathrm{MC}} D_{\max}\big(\mathcal{R}_{B \to BC}(\rho_{AB}) \| \mu_{ABC}\big)$$

$$\leq \lambda := \limsup_{k \to \infty} \inf_{\mu \in \mathrm{MC}} D_{\max}\big(\Pi_A^k \mathcal{R}_{B \to BC}(\rho_{AB}) \Pi_A^k \| \Pi_A^k \mu_{ABC} \Pi_A^k\big). \tag{71}$$

Because the max-relative entropy cannot increase if the same positive map is applied to both arguments, the max-relative entropy is non-decreasing for increasing k, and the $\limsup$ may therefore be replaced by a lim. Hence, there exists a sequence $(\mu^k)_{k \in \mathbb{N}}$ of density operators in MC such that

$$\lambda = \lim_{k \to \infty} D_{\max}\big(\Pi_A^k \mathcal{R}_{B \to BC}(\rho_{AB}) \Pi_A^k \| \Pi_A^k \mu_{ABC}^k \Pi_A^k\big), \tag{72}$$

and we can assume without loss of generality that $\Pi_A^k \mu_{ABC}^k \Pi_A^k = \mu_{ABC}^k$. From here we proceed analogously to the proof of Lemma 11 in [18]. In particular, we use that the space, $T(\mathcal{H})$, of trace-class operators on $\mathcal{H} = A \otimes B \otimes C$ (equipped with the trace norm) is isometrically isomorphic to the dual of the space $K(\mathcal{H})$ of compact operators on $\mathcal{H}$ (equipped with the operator norm), with the isomorphism $\tau \mapsto \psi_\tau$ given by $\psi_\tau(\kappa) = \mathrm{tr}\, \kappa \tau$, and that, by the Banach-Alaoglu theorem, the closed unit ball on $T(\mathcal{H})$ is therefore compact with respect to the weak* topology. This implies that there exists a subsequence $(\mu^k)_{k \in \Gamma \subset \mathbb{N}}$ that converges in the weak* topology to an element $\mu \in T(\mathcal{H})$, i.e.,

$$\lim_{k \to \infty} \mathrm{tr}\, \kappa \mu^k = \mathrm{tr}\, \kappa \mu \qquad (k \in \Gamma) \tag{73}$$

for al $\kappa \in K(\mathcal{H})$. Because, for any $k \in \mathbb{N}$, μ^k is a density operator, μ is also a density operator. The convergence (73) also implies

$$\lim_{k \to \infty} \operatorname{tr} \kappa \Pi_A^k (2^\lambda \mu_{ABC}^k - \mathcal{R}_{B \to BC}(\rho_{AB})) \Pi_A^k$$
$$= \operatorname{tr} \kappa (2^\lambda \mu_{ABC} - \mathcal{R}_{B \to BC}(\rho_{AB})) \quad (k \in \Gamma) \tag{74}$$

for any $\kappa \in K(\mathcal{H})$. By the definition of the max-relative entropy, the sequence on the left-hand side must converge to a nonnegative real for any $\kappa \geq 0$. This implies (71). $\square$

Proof of Theorem 3.1. Let μ_{ABC} be a Markov chain. Combining Lemma 3.2 with Lemma 2.1 applied for $\alpha = 1$, $\rho = \rho_{ABC}$, $\sigma = \mu_{ABC}$ and $\omega = \mathcal{R}_{B \to BC}(\rho_{AB})$ gives

$$I(A : C|B)_\rho \leq D\big(\rho_{ABC} \| \mathcal{R}_{B \to BC}(\rho_{AB})\big) + \inf_{\mu \in \mathrm{MC}} D_{\max}\big(\mathcal{R}_{B \to BC}(\rho_{AB}) \| \mu_{ABC}\big). \tag{75}$$

Lemma 3.3 then proves the assertion of Theorem 3.1.[5] $\square$

4. On the Tightness of the Main Result

In this section, we construct examples that show two things. First, there exist classical tripartite states with a large conditional mutual information that, however, can be recovered well. This shows the necessity of the $\Lambda_{\max}$-term in the main bound (49)—even if the relative entropy was replaced by the largest possible relative entropy measure, i.e., the max-relative entropy. Furthermore, the violation of such a bound without the $\Lambda_{\max}$-term can be made arbitrarily large. Second, our example shows that (49) is no longer valid if the max-relative entropy in the definition of $\Lambda_{\max}(\rho_{AB} \| \mathcal{R}_{B \to B})$ is replaced with any α-Rényi relative entropy for any $\alpha \in [\frac{1}{2}, \infty)$.

Both examples will be classical, i.e., we consider tripartite states of the form (18). Such states are special as the corresponding density operators of the states and all its marginals are simultaneously diagonalizable. As a result, we can use the classical notion of a distribution to describe such states.

4.1. A Large Conditional Mutual Information Does Not Imply Bad Recovery

Let $\mathcal{X} = \{1, 2, \ldots, 2^n\}$ for $n \in \mathbb{N}$, $p, q \in [0, 1]$ such that $p + q \leq 1$, and consider two independent random variables E_Z and E_Y on $\{0, 1\}$ and $\{0, 1, 2\}$, respectively, such that $\mathbb{P}(E_Z = 0) = p+q$, $\mathbb{P}(E_Y = 0) = p$, and $\mathbb{P}(E_Y = 1) = q$. Let $X \sim \mathcal{U}(\mathcal{X})$, where $\mathcal{U}(\mathcal{X})$ denotes the uniform distribution on $\mathcal{X}$ and define two random variables by

$$Z := \begin{cases} X & \text{if } E_Z = 0 \\ U_Z & \text{otherwise} \end{cases} \quad \text{and} \quad Y := \begin{cases} X & \text{if } E_Y = 0 \\ Z & \text{if } E_Y = 1 \\ U_Y & \text{otherwise}, \end{cases} \tag{76}$$

---

[5]We note that (75) is stronger than (49) and therefore may be of independent interest.

where $U_Y \sim \mathcal{U}(\mathcal{X})$ and $U_Z \sim \mathcal{U}(\mathcal{X})$ are independent. This defines a tripartite distribution P_{XYZ}. A simple calculation reveals that

$$
\begin{aligned}
H(X|YE_Y E_Z) & \\
&= pH(X|XE_Z) + qH(X|ZE_Z) + (1-p-q)H(X|U_Y E_Z) \quad (77)\\
&= q\big((p+q)H(X|X) + (1-p-q)H(X|U_Z)\big) + (1-p-q)H(X) \quad (78)\\
&= n(1-p-q)(1+q)\,. \quad (79)
\end{aligned}
$$

Similarly we find

$$
H(X|YZE_Y E_Z) = q(1-p-q)H(X|U_Z) + (1-p-q)(1-p-q)H(X|U_Y) \tag{80}
$$

$$
= n(1-p-q)(1-p)\,. \tag{81}
$$

We thus obtain

$$
\begin{aligned}
I(X:Z|Y)_P &= H(X|Y) - H(X|YZ) \tag{82}\\
&\geq H(X|YE_Y E_Z) - H(X|YZE_Y E_Z) - I(X:E_Y E_Z|YZ) \tag{83}\\
&\geq n(1-p-q)(p+q) - \log 6\,. \tag{84}
\end{aligned}
$$

We next define a recovery map $\mathcal{R}_{Y \to Y'Z'}$ that creates a tuple of random variables (Y', Z') out of Y. Let the recovery map be such that

$$
\begin{aligned}
(Y', Z') := (p^2 + q + pq)(Y,Y) + \frac{1}{2}\left(1 - p^2 - q - pq\right)(Y,U)\\
+ \frac{1}{2}\left(1 - p^2 - q - pq\right)(U', Y)\,, \tag{85}
\end{aligned}
$$

where U, U' are independent uniformly distributed on $\mathcal{X}$. Let

$$
Q_{XY'Z'} := \mathcal{R}_{Y \to Y'Z'}(P_{XY}) \tag{86}
$$

denote the distribution that is generated when applying the recovery map (described above) to P_{XY}. In the following, we will assume that n is sufficiently large. It can be verified easily that $Q_{Y'Z'} = P_{YZ}$. Since P_{XYZ} and $Q_{XY'Z'}$ are classical distributions we have $D_{\max}(P_{XYZ}\|Q_{XY'Z'}) = \max_{x,y,z} \log \frac{P_{XYZ}(x,y,z)}{Q_{XY'Z'}(x,y,z)}$. We note that $\mathbb{P}(X = Y) = p + pq + q^2$ according to the distribution P_{XY} and hence

$$
\begin{aligned}
&D_{\max}(P_{XYZ}\|Q_{XY'Z'})\\
&= \max\Bigg\{ \log \frac{(p+q)^2}{\mathbb{P}(X=Y)(p^2+q+pq)}, \log \frac{(1-p-q)q}{\mathbb{P}(X \neq Y)(p^2+q+pq)},\\
&\quad \log \frac{(p+q)(1-p-q)}{\mathbb{P}(X=Y)\frac{1}{2}(1-p^2-q-pq)},\\
&\quad \log \frac{(1-p-q)p}{\mathbb{P}(X=Y)\frac{1}{2}(1-p^2-q-pq)}, \log \frac{(1-p-q)^2}{\mathbb{P}(X \neq Y)(1-p^2-q-pq)} \Bigg\} \tag{87}
\end{aligned}
$$

and

$$D_{\max}(Q_{XY'Z'}\|P_{XYZ})$$
$$= \max\left\{\log\frac{\mathbb{P}(X=Y)(p^2+q+pq)}{(p+q)^2}, \log\frac{\mathbb{P}(X\neq Y)(p^2+q+pq)}{(1-p-q)q},\right.$$
$$\log\frac{\mathbb{P}(X=Y)\frac{1}{2}(1-p^2-q-pq)}{(p+q)(1-p-q)},$$
$$\left.\log\frac{\mathbb{P}(X=Y)\frac{1}{2}(1-p^2-q-pq)}{(1-p-q)p}, \log\frac{\mathbb{P}(X\neq Y)(1-p^2-q-pq)}{(1-p-q)^2}\right\}. \quad (88)$$

We are now ready to state the conclusion of this example. For $\kappa < \infty$, $p = \frac{1}{2}$, $q = 0$, and n sufficiently large we find by combining (84) with (87)

$$\kappa\,D_{\max}\big(P_{XYZ}\|\mathcal{R}_{Y\to YZ}(P_{XY})\big) = \kappa < \frac{n}{4} - \log 6 \le I(X:Z|Y)_P. \quad (89)$$

For $\kappa < \infty$, $p = q = \frac{1}{4}$, and n sufficiently large (84) and (88) imply

$$\kappa\,D_{\max}\big(\mathcal{R}_{Y\to YZ}(P_{XY})\|P_{XYZ}\big) = \kappa\log\frac{15}{8} < \frac{n}{4} - \log 6 \le I(X:Z|Y)_P. \quad (90)$$

This shows that there exist classical tripartite distributions P_{XYZ} with a large conditional mutual information $I(X:Y|Z)_P$ and a recovery map $\mathcal{R}_{Y\to YZ}$ such that $\mathcal{R}_{Y\to YZ}(P_{XY})$ is close to P_{XYZ} and $\mathcal{R}_{Y\to YZ}(P_Y) = P_{YZ}$. The closeness is measured with respect to the max-relative entropy.

4.2. Tightness of the $\Lambda_{\max}$-Term

In this section, we construct a classical example showing that our main result, i.e., (49) is essentially tight in the sense that it is no longer valid if the max-relative entropy in the definition of $\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B\to B})$, given in (11), is replaced with an α-Rényi relative entropy for any $\alpha < \infty$. More precisely, for $\alpha \in [1,\infty]$ we define

$$\Lambda_\alpha(\rho\|\mathcal{E}) := \inf_{\tau\in\mathrm{Inv}(\mathcal{E})} D_\alpha(\rho\|\tau). \quad (91)$$

For $\alpha = \infty$ we have $\Lambda_\infty(\rho\|\mathcal{E}) = \Lambda_{\max}(\rho\|\mathcal{E})$. In this section, we show that for all $\alpha < \infty$ there exits a (classical) tripartite state ρ_{ABC} and a recovery map $\mathcal{R}_{B\to BC}$ that satisfies $\mathcal{R}_{B\to BC}(\rho_B) = \rho_{BC}$ such that

$$D\big(\rho_{ABC}\|\mathcal{R}_{B\to BC}(\rho_{AB})\big) < I(A:C|B)_\rho - \Lambda_\alpha(\rho_{AB}\|\mathcal{R}_{B\to B}). \quad (92)$$

To see this, consider the following classical example (where we switch to the classical notation). Let $\mathcal{S} = \{0,\ldots,2^n-1\}$ and consider a tripartite distribution Q_{XYZ} defined via the random variables $X \sim \mathcal{U}(\mathcal{S})$ and $X = Y = Z$. Let Q'_{XYZ} be the distribution defined via the random variables $X \sim \mathcal{U}(\mathcal{S})$, $Y \sim \mathcal{U}(\mathcal{S})$ where X and Y are independent and $Z = (X+Y) \mod 2^n$. For $p \in [0,1]$, we define a binary random variable E such that $\mathbb{P}(E=0) = p$. Consider the distribution

$$P_{XYZ} = \begin{cases} Q_{XYZ} & \text{if } E=0 \\ Q'_{XYZ} & \text{if } E=1. \end{cases} \quad (93)$$

We next define two recovery maps $\tilde{\mathcal{R}}_{Y \to Y'Z'}$ and $\bar{\mathcal{R}}_{Y \to Y'Z'}$ that create the tuples (Y', Z') out of Y such that

$$(Y', Z') = (Y, Y) \qquad \text{and} \qquad (Y', Z') = \big(U, (Y - U) \bmod 2^n\big), \qquad (94)$$

where $U \sim \mathcal{U}(\mathcal{S})$, respectively. We then define another recovery map as

$$\mathcal{R}_{Y \to Y'Z'} := p\tilde{\mathcal{R}}_{Y \to Y'Z'} + (1 - p)\bar{\mathcal{R}}_{Y \to Y'Z'}. \qquad (95)$$

We note that the recovery map satisfies $\mathcal{R}_{Y \to Y'Z'}(P_Y) = P_{YZ}$. A simple calculation shows that

$$H(X|YE)_P = pH(X|Y)_Q + (1 - p)H(X|Y)_{Q'} = (1 - p)n \qquad (96)$$

and

$$H(X|YZE)_P = pH(X|YZ)_Q + (1 - p)H(X|YZ)_{Q'} = 0. \qquad (97)$$

We thus find

$$I(X : Z|Y)_P = H(X|Y) - H(X|YZ) \qquad (98)$$
$$\geq H(X|YE) - H(X|YZE) - I(X : E|YZ) \qquad (99)$$
$$\geq (1 - p)n - h(p). \qquad (100)$$

The distribution $\mathcal{R}_{Y \to Y'Z'}(P_{XY})$ generated by applying the recovery map to P_{XY} can be decomposed as

$$\mathcal{R}_{Y \to Y'Z'}(P_{XY})$$
$$= p\left(p\tilde{S}_{XYZ} + (1 - p)\bar{S}_{XYZ}\right) + (1 - p)\left(p\tilde{S}'_{XYZ} + (1 - p)\bar{S}'_{XYZ}\right) \quad (101)$$

where $\tilde{S}_{XYZ} = \tilde{\mathcal{R}}_{Y \to Y'Z'}(Q_{XY})$, $\bar{S}_{XYZ} = \bar{\mathcal{R}}_{Y \to Y'Z'}(Q_{XY})$, $\tilde{S}'_{XYZ} = \tilde{\mathcal{R}}_{Y \to Y'Z'}(Q'_{XY})$, and $\bar{S}'_{XYZ} = \bar{\mathcal{R}}_{Y \to Y'Z'}(Q'_{XY})$. The joint convexity of the relative entropy [12, Theorem 2.7.2] then implies

$$D\big(P_{XYZ}\|\mathcal{R}_{Y \to Y'Z'}(P_{XY})\big)$$
$$\leq pD(Q_{XYZ}\|p\tilde{S}_{XYZ} + (1 - p)\bar{S}_{XYZ}) + (1 - p)D(Q'_{XYZ}\|p\tilde{S}'_{XYZ} + (1 - p)\bar{S}'_{XYZ})$$
$$\qquad (102)$$

A simple calculation shows that

$$D\big(Q_{XYZ}\|p\tilde{S}_{XYZ} + (1 - p)\bar{S}_{XYZ}\big)$$
$$= \sum_{x=y=z} Q_{XYZ}(x, y, z) \log \frac{Q_{XYZ}(x, y, z)}{p\tilde{S}_{XYZ}(x, y, z) + (1 - p)\bar{S}_{XYZ}(x, y, z)}$$
$$\leq \frac{2^{-n}}{p2^{-n}} = \log \frac{1}{p} \qquad (103)$$

and

$$D(Q'_{XYZ} \| p\tilde{S}'_{XYZ} + (1-p)\bar{S}'_{XYZ})$$

$$= \sum_{x,y,z=x+y \bmod 2^n} Q'_{XYZ}(x,y,z) \log \frac{Q'_{XYZ}(x,y,z)}{p\tilde{S}'_{XYZ}(x,y,z) + (1-p)\bar{S}'_{XYZ}(x,y,z)}$$

$$(104)$$

$$\leq \frac{2^{-2n}}{p 2^{-2n}} = \log \frac{1}{p} . \tag{105}$$

We thus have

$$D\big(P_{XYZ} \| \mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) \leq \log \frac{1}{p} . \tag{106}$$

We note that the recovery map $\mathcal{R}_{Y \to Y'} = \mathrm{tr}_{Z'} \circ \mathcal{R}_{Y \to Y'Z'}$ leaves the uniform distribution Q'_{XY} invariant, i.e., $\mathcal{R}_{Y \to Y'}(Q'_{XY}) = Q'_{XY}$. As a result we find

$$\Lambda_\alpha(P_{XY} \| \mathcal{R}_{Y \to Y'}) \leq D_\alpha(P_{XY} \| Q'_{XY})$$

$$= \frac{1}{\alpha - 1} \log \big(2^{-n}(1-p)^\alpha(2^n - 1) + 2^{-n}(1 - p + p2^n)^\alpha\big), \tag{107}$$

where the final step follows by definition of the α-Rényi relative entropy and a straightforward calculation.

Recall that we need to prove (92), which in the classical notation reads as

$$D\big(P_{XYZ} \| \mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) + \Lambda_\alpha(P_{XY} \| \mathcal{R}_{Y \to Y'}) < I(X : Z|Y)_P , \tag{108}$$

for all $\alpha < \infty$. As mentioned in (35), the α-Rényi relative entropy is monotone in α which shows that it suffices to prove (108) for all $\alpha \in (\alpha_0, \infty)$, where $\alpha_0 \geq 0$ can be arbitrarily large.

Combining (106) and (107) shows that for any $\alpha \in (\alpha_0, \infty)$ where α_0 is sufficiently large, $p = \alpha^{-2}$, and $n = \alpha$

$$D\big(P_{XYZ} \| \mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) + \Lambda_\alpha(P_{XY} \| \mathcal{R}_{Y \to Y'})$$

$$\leq 2 \log \alpha + \frac{1}{\alpha - 1} \log \big(1 + 2^{-\alpha}(1 + \alpha^{-2}2^\alpha)^\alpha\big) , \tag{109}$$

where we used that $(1 - \alpha^{-2})^\alpha(2^\alpha - 1) \leq 2^\alpha$ for $\alpha \geq 1$. Using the simple inequality $\log(1 + x) \leq \log x + \frac{2}{x}$ for $x \geq 1$ gives

$$D\big(P_{XYZ} \| \mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) + \Lambda_\alpha(P_{XY} \| \mathcal{R}_{Y \to Y'})$$

$$\leq 2 \log \alpha - \frac{\alpha}{\alpha - 1} + \frac{\alpha}{\alpha - 1} \log \left(1 + \frac{2^\alpha}{\alpha^2}\right) + \frac{2}{\alpha - 1} 2^\alpha \left(1 + \frac{2^\alpha}{\alpha^2}\right)^{-\alpha} \tag{110}$$

$$\leq 2 \log \alpha - \frac{\alpha}{\alpha - 1} + \frac{\alpha}{\alpha - 1} \log \left(1 + \frac{2^\alpha}{\alpha^2}\right) + 2^{-\alpha} , \tag{111}$$

where the final step is valid since α is assumed to be sufficiently large. Using once more $\log(1 + x) \leq \log x + \frac{2}{x}$ for $x \geq 1$ gives

$$D\big(P_{XYZ}\|\mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) + \Lambda_\alpha(P_{XY}\|\mathcal{R}_{Y \to Y'})$$

$$\leq 2\log\alpha + \frac{\alpha}{\alpha - 1}\left(\alpha - 2\log\alpha - 1 + \frac{2\alpha^2}{2^\alpha}\right) + 2^{-\alpha} \tag{112}$$

$$= \alpha - \frac{2}{\alpha - 1}\log\alpha + 2^{-\alpha}\text{poly}(\alpha)\,, \tag{113}$$

where $\text{poly}(\alpha)$ denotes an arbitrary polynomial in α. As a result, we obtain for a sufficiently large α

$$D\big(P_{XYZ}\|\mathcal{R}_{Y \to Y'Z'}(P_{XY})\big) + \Lambda_\alpha(P_{XY}\|\mathcal{R}_{Y \to Y'}) < \alpha - \frac{2}{\alpha} \tag{114}$$

$$\leq \alpha - \alpha^{-1} - h(\alpha^{-2}) \tag{115}$$

$$\leq I(X : Z|Y)_P\,. \tag{116}$$

The two steps (114) and (115) are both valid because α is sufficiently large. The final step uses (100).

This example shows that (49) is no longer valid if the $\Lambda_{\max}$-term is replaced with a Λ_α-term for any $\alpha \in [\frac{1}{2}, \infty)$.[6] Note also that this example implies Remark 2.3 on the tightness of the triangle-like inequality for the relative entropy.

5. Open Questions

In this article, we introduced a new entropic quantity $\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B})$ that measures how much the map $\mathcal{R}_{B \to B}$ disturbs the B system, taking system A as a reference. It would be interesting to better understand this quantity and its properties. For example in case ρ_{ABC} is a state whose marginals are all flat[7], is it possible to bound $\Lambda_{\max}(\rho_{AB}\|\mathcal{R}_{B \to B})$ in terms of $D(\rho_{ABC}\|\mathcal{R}_{B \to BC}(\rho_{AB}))$ from above? This would considerably simplify our main result (49) for this special case, which is of interest, e.g., in applications to condensed matter physics.

---

[6]The example does not work in the limit $\alpha \to \infty$.

[7]A state ω is called flat if all its eigenvalues are either zero or equal to the same constant.

Appendix A. Approximate Markov Chains Can be far from Markov Chains

As mentioned in the introduction, it is known [11,20] that there exist tripartite states with a small conditional mutual information whose distance to any Markov chain is nevertheless large. For example, consider a state $\rho_{S_1,\ldots S_d} = |\psi\rangle\langle\psi|_{S_1,\ldots S_d}$ on $S_1 \otimes \cdots \otimes S_d$ with $\dim S_k = d > 1$ for all $k = 1, \ldots, d$, where

$$|\psi\rangle_{S_1,\ldots S_d} := \sqrt{\frac{1}{d!}} \sum_{\pi \in \mathcal{S}_d} \text{sign}(\pi)|\pi(1)\rangle \otimes \cdots \otimes |\pi(d)\rangle \qquad (117)$$

is the *Slater determinant*, $\mathcal{S}_d$ denotes the group of permutations of d objects, and $\text{sign}(\pi) := (-1)^L$, where L is the number of transpositions in a decomposition of the permutation π. The chain rule and the trivial upper bound for the mutual information show that we have

$$I(S_1 : S_2 \ldots S_d)_\rho = \sum_{k=2}^{d} I(S_1 : S_k | S_2 \ldots S_{k-1})_\rho \le 2 \log d. \qquad (118)$$

Because the mutual information is nonnegative, there exists $k \in \{2, \ldots, d\}$ such that

$$I(S_1 : S_k | S_2 \ldots S_{k-1})_\rho \le \frac{2}{d-1} \log d, \qquad (119)$$

which can be arbitrarily small (as d gets large). By definition, the reduced state $\rho_{S_1 S_k}$ is the *antisymmetric state* on $S_1 \otimes S_k$ that is far from separable [7, p. 53]. More precisely, for any separable state $\sigma_{S_1 S_k}$ on $S_1 \otimes S_k$ we have $\Delta(\rho_{S_1 S_k}, \sigma_{S_1 S_k}) \ge \frac{1}{2}$, where $\Delta(\tau, \omega) := \frac{1}{2}\|\tau - \omega\|_1$ denotes the *trace distance* between τ and ω. For any state $\mu_{S_1 \ldots S_k}$ on $S_1 \otimes \cdots \otimes S_k$ that forms a Markov chain in order $S_1 \leftrightarrow S_2 \ldots S_{k-1} \leftrightarrow S_k$, it follows by (2) that its reduced state $\mu_{S_1 S_k}$ on $S_1 \otimes S_k$ is separable. Using the monotonicity of the trace distance under trace-preserving completely positive maps [28, Theorem 9.2] we thus find

$$\Delta(\rho_{S_1 \ldots S_k}, \mu_{S_1 \ldots S_k}) \ge \Delta(\rho_{S_1 S_k}, \mu_{S_1 S_k}) \ge \frac{1}{2}, \qquad (120)$$

showing that the state $\rho_{S_1 \ldots S_k}$ despite having a conditional mutual information that is arbitrarily small (see (119)) is far from any Markov chain.

As discussed in the introduction, states with a small conditional mutual information are called approximate Markov chains (which is justified by (5)). The example in this appendix shows that approximate quantum Markov chains are not necessarily close to quantum Markov chains.

References

[1] Alberti, P.M.: A note on the transition probability over C^*-algebras. Lett. Math. Phys. **7**(1), 25–32 (1983). https://doi.org/10.1007/BF00398708

[2] Araki, H., Masuda, T.: Positive cones and L_p-spaces for von Neumann algebras. Publ. Res. Inst. Math. Sci. **18**(2), 759–831 (1982). https://doi.org/10.2977/prims/1195183577

[3] Berta, M., Scholz, V.B., Tomamichel, M.: Rényi divergences as weighted noncommutative vector valued L_p-spaces (2016). arXiv:1608.05317

[4] Berta, M., Seshadreesan, K.P., Wilde, M.M.: Rényi generalizations of the conditional quantum mutual information. J. Math. Phys. **56**(2), 022205 (2015). https://doi.org/10.1063/1.4908102

[5] Berta, M., Tomamichel, M.: The fidelity of recovery is multiplicative. IEEE Trans. Inf. Theory **62**(4), 1758–1763 (2016). https://doi.org/10.1109/TIT.2016.2527683

[6] Bhatia, R.: Matrix Analysis. Springer, New York (1997). https://doi.org/10.1007/978-1-4612-0653-8

[7] Brandão, F.G.S.L., Christandl, M., Harrow, A.W., Walter, M.: The mathematics of entanglement (2016). arXiv:1604.01790

[8] Brandão, F.G.S.L., Harrow, A.W., Oppenheim, J., Strelchuk, S.: Quantum conditional mutual information, reconstructed states, and state redistribution. Phys. Rev. Lett. **115**(5), 050501 (2015). https://doi.org/10.1103/PhysRevLett.115.050501

[9] Carlen, E.: Trace inequalities and quantum entropy: an introductory course. Contemp. Math. (2009). https://doi.org/10.1090/conm/529

[10] Christandl, M., Müller-Hermes, A.: Relative entropy bounds on quantum, private and repeater capacities. Commun. Math. Phys. **353**(2), 821–852 (2017). https://doi.org/10.1007/s00220-017-2885-y

[11] Christandl, M., Schuch, N., Winter, A.: Entanglement of the antisymmetric state. Commun. Math. Phys. **311**(2), 397–422 (2012). https://doi.org/10.1007/s00220-012-1446-7

[12] Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Hoboken (2006)

[13] Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. Stud. Sci. Math. Hung. **2**, 299–318 (1967)

[14] Datta, N.: Min- and max-relative entropies and a new entanglement monotone. IEEE Trans. Inf. Theory **55**(6), 2816–2826 (2009). https://doi.org/10.1109/TIT.2009.2018325

[15] Dupuis, F., Wilde, M.M.: Swiveled Rényi entropies. Quantum Inf. Process. **15**(3), 1309–1345 (2016). https://doi.org/10.1007/s11128-015-1211-x

[16] Fawzi, O., Renner, R.: Quantum conditional mutual information and approximate Markov chains. Commun. Math. Phys. **340**(2), 575–611 (2015). https://doi.org/10.1007/s00220-015-2466-x

[17] Frank, R.L., Lieb, E.H.: Monotonicity of a relative Rényi entropy. J. Math. Phys. **54**(12), 122201 (2013). https://doi.org/10.1063/1.4838835

[18] Furrer, F., Åberg, J., Renner, R.: Min- and max-entropy in infinite dimensions. Commun. Math. Phys. **306**(1), 165–186 (2011). https://doi.org/10.1007/s00220-011-1282-1

[19] Hayden, P., Jozsa, R., Petz, D., Winter, A.: Structure of states which satisfy strong subadditivity of quantum entropy with equality. Commun. Math. Phys. **246**(2), 359–374 (2004). https://doi.org/10.1007/s00220-004-1049-z

[20] Ibinson, B., Linden, N., Winter, A.: Robustness of quantum Markov chains. Commun. Math. Phys. **277**(2), 289–304 (2008). https://doi.org/10.1007/s00220-007-0362-8

[21] Junge, M., Renner, R., Sutter, D., Wilde, M.M., Winter, A.: Universal recovery from a decrease of quantum relative entropy. Accepted for publication in Annales Henri Poincaré (2018). arXiv:1509.07127

[22] Lieb, E.H., Ruskai, M.B.: A fundamental property of quantum-mechanical entropy. Phys. Rev. Lett. **30**, 434–436 (1973). https://doi.org/10.1103/PhysRevLett.30.434

[23] Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum-mechanical entropy. J. Math. Phys. **14**(12), 1938–1941 (1973). https://doi.org/10.1063/1.1666274

[24] Lindblad, G.: Completely positive maps and entropy inequalities. Commun. Math. Phys. **40**(2), 147–151 (1975). https://doi.org/10.1007/BF01609396

[25] Mosonyi, M., Ogawa, T.: Strong converse exponent for classical-quantum channel coding (2014). arXiv:1409.3562

[26] Müller-Hermes, A., Reeb, D.: Monotonicity of the quantum relative entropy under positive maps. Ann. Henri Poincaré **18**(5), 1777–1788 (2017). https://doi.org/10.1007/s00023-017-0550-9

[27] Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S., Tomamichel, M.: On quantum Rényi entropies: a new generalization and some properties. J. Math. Phys. **54**(12), 12220 (2013). https://doi.org/10.1063/1.4838856

[28] Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)

[29] Petz, D.: Quasi-entropies for finite quantum systems. Rep. Math. Phys. **23**(1), 57–65 (1986). https://doi.org/10.1016/0034-4877(86)90067-4

[30] Petz, D.: Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. Commun. Math. Phys. **105**(1), 123–131 (1986). https://doi.org/10.1007/BF01212345

[31] Petz, D.: Monotonicity of quantum relative entropy revisited. Rev. Math. Phys. **15**(01), 79–91 (2003). https://doi.org/10.1142/S0129055X03001576

[32] Pinsker, M.S.: Information and Information Stability of Random Variables and Processes. Izvestiya Akademii Nauk, Moskva (1960)

[33] Renner, R.: Security of quantum key distribution. Ph.D. thesis, ETH Zurich (2005). arXiv:quant-ph/0512258

[34] Sutter, D.: Approximate quantum Markov chains. Ph.D. thesis, ETH Zurich (2018). arXiv:1802.05477

[35] Sutter, D., Berta, M., Tomamichel, M.: Multivariate trace inequalities. Commun. Math. Phys. **352**(1), 37–58 (2017). https://doi.org/10.1007/s00220-016-2778-5

[36] Sutter, D., Fawzi, O., Renner, R.: Universal recovery map for approximate Markov chains. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. (2016). https://doi.org/10.1098/rspa.2015.0623

[37] Sutter, D., Tomamichel, M., Harrow, A.W.: Strengthened monotonicity of relative entropy via pinched Petz recovery map. IEEE Trans. Inf. Theory **62**(5), 2907–2913 (2016). https://doi.org/10.1109/TIT.2016.2545680

[38] Tomamichel, M.: Quantum Information Processing with Finite Resources, Volume 5 of Springer Briefs in Mathematical Physics. Springer (2015). https://doi.org/10.1007/978-3-319-21891-5

[39] Uhlmann, A.: The "transition probability" in the state space of a*-algebra. Rep. Math. Phys. **9**(2), 273–279 (1976). https://doi.org/10.1016/0034-4877(76)90060-4

[40] Uhlmann, A.: Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. Commun. Math. Phys. **54**(1), 21–32 (1977). https://doi.org/10.1007/BF01609834

[41] Wilde, M.M.: Recoverability in quantum information theory. Proc. R. Soc. Lond. A Math. Phys. Eng. Sci. **471**(2182), 20150338 (2015). https://doi.org/10.1098/rspa.2015.0338

[42] Wilde, M.M., Winter, A., Yang, D.: Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. Commun. Math. Phys. **331**(2), 593–622 (2014). https://doi.org/10.1007/s00220-014-2122-x

[43] Winter, A.: Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. Commun. Math. Phys. **347**(1), 291–313 (2016). https://doi.org/10.1007/s00220-016-2609-8

David Sutter and Renato Renner
Institute for Theoretical Physics
ETH Zurich
Zurich
Switzerland
e-mail: sutterd@itp.phys.ethz.ch

Renato Renner
e-mail: renner@phys.ethz.ch