Around the quantum conditional mutual information

Omar Fawzi

September 18, 2015

Abstract

These are self notes for a series of two lectures given at a workshop in Toulouse on September 9th. The goal is to present a strengthening of strong subadditivity of the von Neumann entropy. Warning: This document is likely to contain many inaccuracies, please refer to the papers for a more careful treatment.

1 Fundamental properties of the von Neuman entropy

Remark: all Hilbert spaces are finite dimensional in this talk.

Definition 1.1. Let ρ_{ABC} be a density operator acting on $A \otimes B \otimes C$. To refer to the marginals of the state ρ_{ABC}, we use the standard notation such as $\rho_A = \mathrm{tr}_{BC}(\rho_{ABC})$.

$$H(A)_\rho = -\mathrm{tr}(\rho_A \log \rho_A) \tag{1}$$

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho \tag{2}$$

$$I(A:C)_\rho = H(A)_\rho - H(A|C)_\rho \tag{3}$$

$$I(A:C|B)_\rho = H(A|B)_\rho - H(A|BC)_\rho \ . \tag{4}$$

Let us now try to give some justification for the naming of these quantities, in particular the conditioning. If we have a qc-state $\rho_{AB} = \sum_b p(b)\rho_{A,b} \otimes |b\rangle\langle b|$, then one can verify that

$$H(A|B)_\rho = \sum_b p(b) H(A)_{\rho_{A,b}} \ , \tag{5}$$

and this is a justification for calling it conditional entropy. When the system B is quantum, the entropy $H(A|B)$ cannot be written as an average of unconditional von Neuman entropies. In fact $H(A|B)$ can even be negative when ρ_{AB} is entangled. If $\rho_{AB} = |\Phi\rangle\langle\Phi|$ with $|\Phi\rangle = \frac{1}{\sqrt{d_A}} \sum_{i \in [d_A]} |i\rangle_A \otimes |i\rangle_B$, then $H(A|B)_\rho = -\log d_A$, and this is the smallest it can get as shown in the following:

$$-\log d_A \leq H(A|B) \leq \log d_A \ . \tag{6}$$

It is worth mentioning that $H(A|B)$ has an operational interpretation in terms of state merging. ψ_{ABR} shared between Alice and Bob. Alice wants to transmit her part to Bob. Suppose classical communication is free, what is the minimum number of ebits needed to do that? If it is a product state $|\psi\rangle_A \otimes |\psi\rangle_B$, then Bob can prepare it locally. If we have a maximally correlated classical state, then gives $\log d_A$. If Alice and Bob start with a maximally entangled state: then the reference is product and Bob can reproduce it locally but we have gained one ebit of entanglement in the story, this is where the negative conditional entropy means.

Let us now move to the mutual information. To understand properties of the mutual information, it is often useful to write it using a quantum relative entropy:

$$D(\rho \| \sigma) = \mathrm{tr}(\rho(\log \rho - \log \sigma)) \ . \tag{7}$$

Note that this quantity is infinite if the support of σ is not included in the support of ρ. It is simple to see that

$$I(A:C) = D(\rho_{AC} \| \rho_A \otimes \rho_C) \ . \tag{8}$$

Theorem 1.2. *For any density operators ρ, σ acting on A,*

$$D(\rho \| \sigma) \geq 0 \ , \tag{9}$$

with equality if and only if $\rho = \sigma$. This implies that $I(A:C)_\rho = H(A) - H(A|C) \geq 0$.

This is a direct consequence of Klein's inequality. See [6] for a proof.

Note that this property is quite important. Having the uncertainty decrease if we know more is a very much desirable property. Such a property is sometimes called a data processing inequality: if I forget about some information, then the uncertainty I have cannot decrease. Without such a property, it would be quite difficult to call it a entropic quantity.

In terms of data processing inequality, we would expect something stronger to hold: if one holds a system BC and discards the C part, then the entropy should only increase. In the case where B is classical, this is easy to prove. As for the conditional entropy, when the system B we are conditioning on is classical $\rho_{ABC} = \sum_b p(b) |b\rangle\langle b|_B \otimes \rho_{AC,b}$, the conditional mutual information can be written as an average of unconditional mutual information quantities:

$$I(A:C|B) = \sum_b p(b) I(A:C)_{\rho_{AC,b}} \ . \tag{10}$$

From this, it follows that the conditional mutual information is always non-negative. However, when B is quantum, we cannot write the conditional mutual information as an average of mutual information quantities. This is in fact true but it is much more difficult to prove than Theorem 1.2. This property can be formulated in terms of a simple mathematical property of the relative entropy: joint convexity.

Theorem 1.3. *The relative entropy is jointly convex, i.e., for any states $\rho_0, \rho_1, \sigma_0, \sigma_1$ and $p \in [0,1]$, we have*

$$D(p\rho_0 + (1-p)\rho_1 \| p\sigma_0 + (1-p)\sigma_1) \leq pD(\rho_0\|\sigma_0) + (1-p)D(\rho_1\|\sigma_1) \ . \tag{11}$$

This joint convexity of a related function was proved by Lieb [17]. A very operational property follows from this mathematical property of the relative entropy: the monotonicity of relative entropy under completely positive trace preserving maps.

Theorem 1.4. *Let ρ, σ be density operators on A and $\mathcal{W}_{A\to B}$ be a completely positive trace-preserving map. Then*

$$D(\mathcal{W}(\rho) \| \mathcal{W}(\sigma)) \leq D(\rho\|\sigma) \ . \tag{12}$$

Proof. To obtain this from joint convexity, we first consider an isometry $W_{A\to BE}$ that is Stinespring dilation of the map $\mathcal{W}$, i.e., $\mathcal{W}(\rho) = \mathrm{tr}_E(W\rho W^\dagger)$. Then we take the family of states $V_x W \rho W^\dagger V_x^\dagger$, where V_x for $x \in [m]$ are Weyl-Heisenberg operators on the space E. Then

$$D(\frac{1}{m}\sum_x V_x W\rho W^\dagger V_x^\dagger \| \frac{1}{m}\sum_x V_x W \sigma W^\dagger V_x^\dagger) = D(\mathcal{W}(\rho)\otimes \pi_E \| \mathcal{W}(\sigma)\otimes\pi_E) = D(\mathcal{W}(\rho)\|\mathcal{W}(\sigma)) \ . \tag{13}$$

On the other hand, for any x, we have $D(V_x W\rho W^\dagger V_x^\dagger \| V_x W\sigma W^\dagger V_x^\dagger) = D(\rho\|\sigma)$. □

Now we can apply it to the map $\mathcal{W}$ being the partial trace to get the famous strong subadditivity theorem first proved by [16].

Theorem 1.5. *For any state ρ_{ABC} acting on $A \otimes B \otimes C$, we have*

$$I(A : C|B)_\rho = H(A|B)_\rho - H(A|BC)_\rho \geq 0 \ . \tag{14}$$

Written explicitly in terms of unconditional von Neuman entropies, we get

$$H(AB) + H(BC) \geq H(B) + H(ABC) \ . \tag{15}$$

Proof. We just apply the monotonicity theorem to the states ρ_{ABC} and

$$D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) = \mathrm{tr}(\rho_{ABC} \log \rho_{ABC}) - \mathrm{tr}(\rho_{ABC} \log(\rho_A \otimes \rho_{BC})) \tag{16}$$

$$= -H(ABC)_\rho + H(A)_\rho + H(BC)_\rho \ . \tag{17}$$

Moreover,

$$D(\rho_{AB} \| \rho_A \otimes \rho_B) = -H(AB)_\rho + H(A)_\rho + H(B)_\rho \ . \tag{18}$$

Taking $\rho = \rho_{ABC}$, $\sigma = \rho_A \otimes \rho_{BC}$ and $\mathcal{W} = \mathrm{tr}_C$, we get

$$-H(AB)_\rho + H(A)_\rho + H(B)_\rho \leq -H(ABC)_\rho + H(A)_\rho + H(BC)_\rho \ , \tag{19}$$

which gives the desired inequality. □

1.1 Motivation for studying von Neumann entropy quantities

The von Neumann entropy quantities are "average case" entropies. They usually have an operational meaning only when we have iid copies of a resource or when we look at some average cost. In more general one-shot setting, there are other entropic quantities that are more relevant. In particular, in cryptography, one usually uses a worst-case kind of entropy called min-entropy to quantify randomness.

1. Characterises the optimal rates at which operational tasks can be done. Example: state merging. Compression. Channel coding. Randomness extraction. Properties like strong subadditivity are essential is proofs of converse results in particular.

2. It properties make it a useful tool for proofs. The main reason that makes it so useful is

$$I(A_1 \dots A_n : C|B) = \sum_i I(A_i : C|BA_1 \dots A_{i-1}) \ . \tag{20}$$

2 States (approximately) saturating strong subadditivity

We would now like to understand the structure of states satisfying $I(A : C|B)_\rho = 0$. In the classical case, this is easy to determine such distributions P_{ABC}. In fact, we have for any b, $I(A : C)_{P_{|b}} = 0$, which implies that $P_{AC|B=b} = P_{A|B=b} \times P_{C|B=b}$. In other words, A and C are independent conditioned on B. This means that $A \leftrightarrow B \leftrightarrow C$ form a short Markov chain. A useful way of stating this is that there exists a mapping $\mathcal{R} : B \to BC$, namely $\mathcal{R}(\delta_b) = \delta_b \times P_{C|B=b}$, such that $\mathcal{I}_A \otimes \mathcal{R}_{B\to BC}(P_{AB}) = P_{ABC}$.

A quantum analogue of this characterization was proved in [19, 11]. It has been found that a zero conditional mutual information corresponds to states ρ_{ABC} whose C system can be reconstructed just by acting on B. More precisely:

Theorem 2.1. *The following conditions are equivalent*

1. $I(A : C|B)_\rho = 0$

2. *There exists a quantum channel $\mathcal{T}_{B\to BC}$ such that*

$$\mathcal{I}_A \otimes \mathcal{T}_{B\to BC}(\rho_{AB}) = \rho_{ABC} \tag{21}$$

A state satisfying this property is called a quantum Markov chain. Petz [19] showed that the map $\mathcal{T}_{B\to BC}$ can be taken to be $\mathcal{T}_{B\to BC}(X_B) = \rho_{BC}^{1/2}(\rho_B^{-1/2} X_B \rho_B^{-1/2} \otimes \mathrm{id}_C)\rho_{BC}^{1/2}$.

3. *The Hilbert space B can be decomposed into $B = \bigoplus_j b_j^L \otimes b_j^R$ such that*

$$\rho_{ABC} = \bigoplus_j q_j \rho^j_{Ab_j^L} \otimes \rho^j_{b_j^R C} , \tag{22}$$

where q_j is a probability distribution and $\rho^j_{Ab_j^L}$ and $\rho^j_{b_j^R C}$ are density operators on $A \otimes b_j^L$ and $b_j^R \otimes C$.

Corollary 2.2. *If $I(A:C|B)_\rho = 0$, then ρ_{AC} is separable. Conversely, if ρ_{AC} is separable, then there exists an extension ρ_{ABC} such that $I(A:C|B) = 0$.*

Proof.

$$\mathrm{tr}_B \left(\bigoplus_j q_j \rho^j_{Ab_j^L} \otimes \rho^j_{b_j^R C} \right) = \sum_j q_j \mathrm{tr}_B (\rho^j_{Ab_j^L} \otimes \rho^j_{b_j^R C}) = \sum_j q_j \rho^j_A \otimes \rho^j_C . \tag{23}$$

For the converse, write $\rho_{AC} = \sum_j q_j \rho^j_A \otimes \rho^j_C$. Then $\rho_{ACJ} = \sum_j q_j \rho^j_A \otimes \rho^j_C \otimes |j\rangle\langle j|$ satisfies $I(A:C|J) = 0$. □

2.1 Approximate Markov chains

A natural question that is very relevant for applications is to characterise states for which the conditional mutual information is approximately zero, i.e., for which it is guaranteed that $I(A:C|B) \leq \epsilon$ for some $\epsilon > 0$. In applications involving n systems $A_1, \ldots, A_n$, such a guarantee is often obtained from an upper bound on the total conditional mutual information $I(A_1 \ldots A_n : C|B) \leq c$ (which can even be the trivial bound $2\log_2 \dim C$). The chain rule mentioned above then implies that, on average over i, we have $I(A_i : C|BA_1 \ldots A_{i-1}) \leq c/n$.

One first candidate conjecture is the following

$$I(A:C|B)_\rho \leq \epsilon \quad \Rightarrow \quad \rho_{ABC} \approx_{f(\epsilon)} \sigma_{ABC} \text{ with } \sigma \text{ a quantum Markov chain} . \tag{24}$$

The authors of [12] gave evidence for the difficulty of characterising such states in the quantum setting by finding states for which the conditional mutual information is small whereas their distance to any Markov chain is large (see also [8] for more extreme examples). We will see such an example when we mention applications to quantifying entanglement.

Recent works by [26, 13, 27] made the important observation that instead of considering the distance to a (perfect) Markov chain, another possibly more appropriate measure would be the accuracy with which Eq. 21 is satisfied.

$$I(A:C|B)_\rho \leq \epsilon \quad \Rightarrow \quad \rho_{ABC} \approx_{f(\epsilon)} \mathcal{I}_A \otimes \mathcal{T}_{B\to BC}(\rho_{AB}) . \tag{25}$$

It was conjectured in [13] that the conditional mutual information is lower bounded by the trace distance between the two sides of Eq. 21 for a specific form for the map $\mathcal{T}_{B\to BC}$ known sometimes as the Petz map, which is defined as $\mathcal{T}(X_B) = \rho_{BC}^{1/2}(\rho_B^{-1/2} X_B \rho_B^{-1/2} \otimes \mathrm{id}_C)\rho_{BC}^{1/2}$. Later, in the context of studying Rényi generalisations of the conditional mutual information, the authors of [2] refined this conjecture by replacing the trace distance with the negative logarithm of the fidelity (see also [20]). The fidelity between two states is defined as

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 . \tag{26}$$

Such an inequality was recently shown in [10].

Theorem 2.3. *For any state ρ_{ABC}, there exists a recovery map $\mathcal{T} : B \to BC$*

$$I(A:C|B)_\rho \geq -2\log F(\rho_{ABC}, \mathcal{T}_{B\to BC}(\rho_{AB})) . \tag{27}$$

Trace distance In terms of the trace distance, it can be written as

$$\frac{1}{\ln 2}\Delta(\rho_{ABC},\sigma_{ABC})^2 \leq I(A:C|B)_\rho \ . \tag{28}$$

It is tight up to a logarithmic factor in the dimension of A.

$$I(A:C|B)_\rho \leq 7\log_2(\dim A)\sqrt{\Delta(\rho_{ABC},\sigma_{ABC})} \ . \tag{29}$$

Recovery map The map $\mathcal{T}_{B\to BC}$ is not fully explicit, we know that the best map works, but we don't know about the Petz map. We can rewrite the inequality as

$$I(A:C|B)_\rho \geq \min_{\mathcal{T}:B\to BC} -2\log F(\rho_{ABC},\mathcal{T}_{B\to BC}(\rho_{AB})) \ . \tag{30}$$

The quantity on the right is often called fidelity of recovery and seems like a quantity of interest. We actually do have some structure on the map, we can assume it has the form of a rotated Petz map:

$$\mathcal{T}_{B\to BC}(X_B) = U_{BC}\rho_{BC}^{1/2}\rho_B^{-1/2}U_B X_B U_B^\dagger \rho_B^{-1/2}\rho_{BC}^{1/2}U_{BC}^\dagger \ . \tag{31}$$

Combining with subsequent work, one can even show that the unitaries commute with the corresponding marginal [21], see also [24].

Strenghtenings There are multiple strengthenings of this result by now: either giving a better lower bound

$$I(A:C|B)_\rho \geq \min_{\mathcal{T}:B\to BC} D_{\mathbb{M}}(\rho_{ABC}\|\mathcal{T}_{B\to BC}(\rho_{AB})) \ , \tag{32}$$

or also having a more general statement as a remainder term for the monotonicity of the relative entropy

$$D(\rho\|\sigma) - D(\mathcal{W}(\rho)\|\mathcal{W}(\sigma)) \geq -2\log F(\rho,\mathcal{R}_{\sigma,\mathcal{W}}(\mathcal{W}(\rho))) \ . \tag{33}$$

See [1, 24, 22].

Proofs All the currently known proofs [10, 4, 3, 22] have the following rough pattern of considering n copies of the state $\rho_{ABC}^{\otimes n}$, except for one [24]. The intuition behind the usefulness of this is that states of the form $\rho_{ABC}^{\otimes n}$ have marginals that are close to flat. This parts usually takes the following form.

Lemma 2.4.

$$\begin{aligned} I(A:C|B)_\rho &\geq \lim_{n\to\infty}\frac{1}{n}\min_{\mathcal{T}:B^n\to B^nC^n} D(\rho_{ABC}^{\otimes n}\|\mathcal{T}(\rho_{AB}^{\otimes n})) &(34)\\ &\geq \lim_{n\to\infty}\frac{1}{n}\min_{\mathcal{T}:B^n\to B^nC^n} -2\log F(\rho_{ABC}^{\otimes n}\|\mathcal{T}(\rho_{AB}^{\otimes n})) \ . &(35)\end{aligned}$$

The second inequality follows from the fact that $-2\log F = D_{1/2}$ is a sandwiched Renyi divergence [18, 25] of order $1/2$ and we know that D_α is an increasing function of α.

The second step is to obtain from $\mathcal{T}_{B^n\to B^nC^n}$ a map on just one copy.

Lemma 2.5.

$$\lim_{n\to\infty}\frac{1}{n}\min_{\mathcal{T}:B^n\to B^nC^n} -2\log F(\rho_{ABC}^{\otimes n}\|\mathcal{T}(\rho_{AB}^{\otimes n})) = \min_{\mathcal{T}:B\to BC} -2\log F(\rho_{ABC}\|\mathcal{T}(\rho_{AB})) \ . \tag{36}$$

Go to the board and draw diagram of the different proof strategies.

Let us now proceed to the proof of some of these claims.

Proof of Lemma 2.4. This proof is due to Sutter, Tomamichel and Harrow [22]. We are going to construct a particular map $\mathcal{T}_{B^n\to B^nC^n}$ that satisfies the inequality with a correction factor that vanishes as n grows. To define this map, we first define the picking map $\mathcal{P}_\sigma$ for a state σ. Let $\sigma = \sum_{\lambda\in\mathrm{spec}(\sigma)} \lambda\Pi_\lambda$, with Π_λ the projector on the eigenspace λ, then

$$\mathcal{P}_\sigma(X) = \sum_{\lambda\in\mathrm{spec}(\sigma)} \Pi_\lambda X \Pi_\lambda \ . \tag{37}$$

The map $\mathcal{P}$ is clearly CPTP. The reason the pinching map is so useful is that for any X we have that $\mathcal{P}_\sigma(X)$ commutes with σ. In fact

$$\sum_{\lambda\in\mathrm{spec}(\sigma)} \Pi_\lambda X \Pi_\lambda \sigma = \sum_{\lambda\in\mathrm{spec}(\sigma)} \Pi_\lambda X \Pi_\lambda \lambda \Pi_\lambda = \sum_{\lambda\in\mathrm{spec}(\sigma)} \sigma \Pi_\lambda X \Pi_\lambda \ . \tag{38}$$

Moreover, the map $\mathcal{P}$ conserves some of the properties of X in the following sense: for any $X \geq 0$, $\mathcal{P}_\sigma(X) \geq \frac{1}{|\mathrm{spec}(\sigma)|}X$. To see this, let $m = |\mathrm{spec}(\sigma)|$ and we label the projectors P_λ arbitrarily from $x = 0$ to $m-1$ and define $U_y = \sum_{x\in[m]} e^{2\pi i xy/m} P_x$. Then we have

$$\frac{1}{m}\sum_{y\in[m]} U_y X U_y^\dagger = \frac{1}{m}\sum_y \sum_{x,x'} e^{2\pi(x-x')y/m} P_x X P_{x'} = \mathcal{P}_\sigma(X) \ . \tag{39}$$

But now given that $U_0 = \mathrm{id}$ and that $U_y X U_y^\dagger \geq 0$, we have $\mathcal{P}_\sigma(X) \geq X/m$ (this argument is from [23]).

Let us now define our recovery map:

$$\mathcal{T}_{B^n\to B^nC^n}(X_{B^n}) = \mathcal{P}_{\rho_{BC}^{\otimes n}}\left((\rho_{BC}^{\otimes n})^{1/2}(\rho_B^{\otimes n})^{-1/2}\mathcal{P}_{\rho_B^{\otimes n}}(X_{B^n})(\rho_B^{\otimes n})^{-1/2}(\rho_{BC}^{\otimes n})^{1/2}\right) \ . \tag{40}$$

This is clearly a CPTP map as a composition of CPTP maps.

Now, we write

$$\begin{aligned}
&D(\rho^{\otimes n}\|\mathcal{T}_{B^n\to B^nC^n}(\rho_{AB}^{\otimes n})) \qquad (41)\\
&= -nH(ABC)_\rho - \mathrm{tr}(\rho_{ABC}^{\otimes n}\log \mathrm{id}_{A^n}\otimes\rho_{BC}^{\otimes n}) - \mathrm{tr}\left(\rho_{ABC}^{\otimes n}\log\mathcal{P}_{\rho_{BC}^{\otimes n}}\left((\rho_B^{\otimes n})^{-1/2}\mathcal{P}_{\rho_B^{\otimes n}}(\rho_{AB}^{\otimes n})(\rho_B^{\otimes n})^{-1/2}\right)\right) \qquad (42)\\
&= -nH(ABC)_\rho + nH(BC)_\rho - \mathrm{tr}\left(\rho_{ABC}^{\otimes n}\log\mathcal{P}_{\rho_{BC}^{\otimes n}}\left((\rho_B^{\otimes n})^{-1/2}\mathcal{P}_{\rho_B^{\otimes n}}(\rho_{AB}^{\otimes n})(\rho_B^{\otimes n})^{-1/2}\right)\right) \ . \qquad (43)
\end{aligned}$$

We now use the second property of the pinching map. We have

$$\mathcal{P}_{\rho_{BC}^{\otimes n}}(X_{BC}) \geq \frac{1}{\mathrm{spec}(\rho_{BC}^{\otimes n})} X_{BC} \geq \frac{1}{n^{d_{BC}}} X_{BC} \ . \tag{44}$$

Combining this with the operator monotonicity of the log function, we have

$$\begin{aligned}
&-\mathrm{tr}\left(\rho_{ABC}^{\otimes n}\log\mathcal{P}_{\rho_{BC}^{\otimes n}}\left((\rho_B^{\otimes n})^{-1/2}\mathcal{P}_{\rho_B^{\otimes n}}(\rho_{AB}^{\otimes n})(\rho_B^{\otimes n})^{-1/2}\right)\right) \qquad (45)\\
&\leq -\mathrm{tr}\left(\rho_{ABC}^{\otimes n}\log\left((\rho_B^{\otimes n})^{-1/2}\mathcal{P}_{\rho_B^{\otimes n}}(\rho_{AB}^{\otimes n})(\rho_B^{\otimes n})^{-1/2}\right)\right) + O(\log n) \qquad (46)\\
&= -\mathrm{tr}(\rho_{ABC}^{\otimes n}\log(\rho_B^{\otimes n})^{-1}) - \mathrm{tr}(\rho_{ABC}^{\otimes n}\log\mathcal{P}_{\rho_B^{\otimes n}}(\rho_{AB}^{\otimes n})) + O(\log n) \qquad (47)\\
&\leq -nH(B)_\rho - \mathrm{tr}(\rho_{ABC}^{\otimes n}\log\rho_{AB}^{\otimes n}) + O(\log n) \ . \qquad (48)
\end{aligned}$$

Putting everything together, we get

$$D(\rho^{\otimes n}\|\mathcal{T}_{B^n\to B^nC^n}(\rho_{AB}^{\otimes n})) \leq nI(A:C|B) + O(\log n) \ . \tag{49}$$

□

For the second part, where we get back again to just one copy of the state, there are two approaches. The first one (historically) was to use some specific structure of the map $\mathcal{T}_{B^n \to B^n C^n}$ that is constructed, namely that it is invariant under permutation of the systems. Then one uses a de Finetti type theorem to say that it is not too far from an iid channel.

Corollary 2.6. *Let D and E be Hilbert spaces. Then there exists a probability measure $\mathrm{d}\tau$ on the set of completely positive trace-preserving maps $\tau_{D\to E}$ such that*[1]

$$\mathcal{W}_{D^n \to E^n} \leq (n+1)^{d^2-1} \int \tau_{D\to E}^{\otimes n} \mathrm{d}\tau \tag{50}$$

holds for any $n \in \mathbb{N}$, any completely positive trace-preserving map $\mathcal{W}_{D^n\to E^n}$ that is permutation-invariant (i.e., $\mathcal{W} \circ \pi = \pi \circ \mathcal{W}$ for all permutations π), and $d = \dim(D)\dim(E)^2$.

The proof of this is based on some Schur-Weyl duality. Then it takes a little bit of work to get to the desired statement, purify everything, then the fidelity is nice and can then take the best map. In fact, this was not exactly the way the first argument was done, in some sense the de Finetti was applied to directly replace the type projectors by product unitaries instead of just globally to the map.

Later, it was realised by Berta and Tomamichel [3] that in fact one does not need to use any structure of the map. In fact, for the fidelity, the optimal map $\mathcal{T}_{B^n \to B^n C^n}$ has product form. More precisely

Theorem 2.7. *For any $\rho^1_{A_1B_1C_1}, \rho^2_{A_2B_2C_2}$, we have*

$$\min_{\mathcal{T}:B_1B_2\to B_1B_2C_1C_2} -2\log F(\rho^1_{A_1B_1C_1} \otimes \rho^2_{A_2B_2C_2}, \mathcal{T}(\rho^1_{A_1B_1} \otimes \rho^2_{A_2B_2})) \tag{51}$$

$$= \min_{\mathcal{T}:B_1\to B_1C_1} -2\log F(\rho^1_{A_1B_1C_1}, \mathcal{T}(\rho^1_{A_1B_1})) + \min_{\mathcal{T}:B_2\to B_2C_2} -2\log F(\rho^2_{A_2B_2C_2}, \mathcal{T}(\rho^2_{A_2B_2}))\ . \tag{52}$$

Proof sketch. First, the inequality $\leq$ is clear as we can just take $\mathcal{T}_{B_1B_2\to B_1B_2C_1C_2} = \mathcal{T}_{B_1\to B_1C_1} \otimes \mathcal{T}_{B_2 \to B_2C_1}$. For the other inequality, we use semidefinite programming duality. We can write the fidelity of recovery as a semidefinite program: this is just optimizing a linear function over the intersection of the positive semidefinite cone and a affine subspace. In particular, this one can be written as

$$\begin{aligned} F_{\mathrm{rec}}(\rho_{ABC}) = \underset{X}{\text{maximize}} \quad & \frac{1}{2}\mathrm{tr}(X) + \mathrm{tr}(X^\dagger) \\ \text{subject to} \quad & \begin{pmatrix} \rho_{ABC} & X \\ X^\dagger & \omega_{ABC} \end{pmatrix} \geq 0 \\ & \omega_{ABC} = \mathrm{tr}_{B'}((\mathrm{id}_A \otimes \theta_{B'BC})(\mathrm{id}_{BC} \otimes \rho'_{AB'})) \\ & \theta_{B'BC} \geq 0 \end{aligned} \tag{53}$$

where $\rho^T_{AB'}$ is just a copy of ρ_{AB} with a partial transpose applied on B. This can be solved efficiently on a computer. Here, the usefulness in writing it as an SDP is that we can use duality theory. In particular, using strong duality, we may write $F_{\mathrm{rec}}(\rho_{ABC}) = \min_\lambda \mathrm{dual}(\rho_{ABC}, \lambda)$ as a minimization problem instead.

$$F_{\mathrm{rec}}(\rho^1_{A_1B_1C_1} \otimes \rho^2_{A_2B_2C_2}) = \min_{\lambda_{12}} \mathrm{dual}(\rho^1_{A_1B_1C_1} \otimes \rho^2_{A_2B_2C_2}, \lambda_{12}) \tag{54}$$

$$\leq \min_{\lambda_1,\lambda_2} \mathrm{dual}(\rho^1_{A_1B_1C_1} \otimes \rho^2_{A_2B_2C_2}, \lambda_1 \otimes \lambda_2) \tag{55}$$

$$= \min_{\lambda_1,\lambda_2} \mathrm{dual}(\rho^1_{A_1B_1C_1}, \lambda_1) \cdot \mathrm{dual}(\rho^2_{A_2B_2C_2}, \lambda_2) \tag{56}$$

$$= F_{\mathrm{rec}}(\rho^1_{A_1B_1C_1}) \cdot F_{\mathrm{rec}}(\rho^2_{A_2B_2C_2})\ . \tag{57}$$

The key here is in showing that if λ_1 and λ_2 are dual feasible for ρ^1 and ρ^2, then $\lambda_1 \otimes \lambda_2$ is dual feasible for $\rho^1 \otimes \rho^2$ and also that the objective value is the product. This is a very useful technique to show additivity kind of properties. □

The same technique can be used with now convex duality to prove the lower bound with the measured relative entropy.

[1] The inequality means that the difference between the right hand side and the left hand side is a completely positive map.

3 Applications

3.1 Squashed entanglement

As an example for how our result can be applied, we present here an argument proposed by Li and Winter [26]. *Squashed entanglement* is a measure of entanglement defined for any bipartite state ρ_{AC} as

$$E_{\mathrm{sq}}(\rho_{AC}) = \frac{1}{2} \inf_{\rho_{ACE}} I(A:C|E)_{\rho} \ , \tag{58}$$

where the infimum ranges over all non-negative extensions ρ_{ACE} of ρ_{AC} [9].

As an illustration, and also to get the promised counter example to (25), consider the antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$. This is the space spanned by $|aa'\rangle - |a'a\rangle$ for $a \neq a'$. The dimension is $d_{as} = \frac{d(d-1)}{2}$. Let Π_{as} be the projector onto the antisymmetric subspace and $\rho_{as} = \Pi_{as}/d_{as}$.

$$E_{\mathrm{sq}}\left(\rho_{as}\right) = O(1/d) \ . \tag{59}$$

In fact, consider the totally antisymmetric subspace on $(\mathbb{C}^d)^m$. For each subset $\{a_1, \dots, a_m\} \subseteq [d]$, we can define the state $\frac{1}{\sqrt{m!}} \sum_{\pi \in \mathcal{S}_m} sgn(\pi) |\pi(a_1) \dots \pi(a_m)\rangle$. Let Π^m_{as} be the projector onto the antisymmetric state on m copies and the dimension is $d_{as,m} = \binom{d}{m}$. One can verify that $\mathrm{tr}_1(\rho_{as,m}) = \rho_{as,m}$. Now consider ρ_{ACE} acting on $\mathbb{C}^d \otimes \mathbb{C}^d \otimes (\mathbb{C}^d)^{\otimes m}$ where A and C are the first and the second copy of $\mathbb{C}^d$ and E is the remaining copies m copies. Then

$$\begin{aligned} I(A:C|E)_\rho &= H(AE) + H(CE) - H(ACE) - H(E) && (60) \\ &= \log d_{as,1+m} + \log d_{as,1+m} - \log d_{as,2+m} - \log d_{as,m} && (61) \\ &= \log \frac{\binom{d}{1+m}^2}{\binom{d}{m} \cdot \binom{d}{2+m}} \ . && (62) \end{aligned}$$

We now assume that d is even and set $m = d/2 - 1$. Then the quantity becomes

$$I(A:C|E)_{\rho_{as,1+d/2}} = 2 \log \left(\frac{\binom{d}{d/2}}{\binom{d}{d/2-1}} \right) = 2 \log \frac{d+2}{d} = O(1/d) \ . \tag{63}$$

The state $\rho_{as,1+d/2}$ has a small conditional mutual information. We are now going to show that it is far from Markov chains. For that recall than any Markov state ρ_{ABC}, the marginal ρ_{AC} is separable. So it suffices to show that $\rho_{as,2}$ is far from separable states. To see this, let $\sigma_{AC} = \sum_i p_i \sigma^i_A \otimes \sigma^i_C$. Then we have

$$\mathrm{tr}\left(\Pi_{as,2} \left(\rho_{as,2} - \sum_i p_i \sigma^i_A \otimes \sigma^i_C \right) \right) = 1 - \sum_i p_i \mathrm{tr}(\Pi_{as,2} \sigma^i_A \otimes \sigma^i_C) \geq \frac{1}{2} \ . \tag{64}$$

So even though this state has a conditional mutual information of $O(1/d)$, it is a constant away from any Markov chain. This antisymmetric state is a counterexample for many things you might conjecture, so good to keep in mind. See [8] for more properties of the antisymmetric state.

Let's now get back to the properties of squashed entanglement. We are mainly interested in faithfulness. It is known that squashed entanglement is *faithful*, i.e., strictly positive for any entangled state [5, 14]. In other words, $E_{\mathrm{sq}}(\rho_{AC}) = 0$ if and only if the state ρ_{AC} is separable. Theorem 2.3 implies a quantitative version of this claim. The main idea is to relate $E_{\mathrm{sq}}(\rho_{AC})$ to the distance between ρ_{AC} and the closest state that is k-extendible (a state ρ_{AC} is k-extendible is there exists a state $\rho_{AC_1 \dots C_k}$ such that $\rho_{AC_i} = \rho_{AC}$ for all $1 \leq i \leq k$.)

Theorem 3.1 ([15]). *For any density operator ρ_{AC} on $A \otimes C$ and any $k \in \mathbb{N}$ there exists a k-extendible density operator ω_{AC} such that*

$$\Delta(\rho_{AC}, \omega_{AC}) \leq (k-1) \sqrt{\frac{\ln 2}{2} E_{\mathrm{sq}}(\rho_{AC})} \ . \tag{65}$$

See [15] for a proof.

Lemma 3.2. *For any k-extendible density operator ω_{AC} on $A \otimes C$*

$$\inf_{\sigma_{AC} \in S_{A:C}} \Delta(\omega_{AC}, \sigma_{AC}) \leq 2\frac{(\dim C)^2}{k} \ . \tag{66}$$

Proof. By definition, there exists a density operator $\bar{\omega}_{AC_1 \cdots C_k}$ such that $\omega_{AC} = \bar{\omega}_{AC_i}$ for $i = 1, \ldots, k$. Because this condition still holds if the order of the subsystems $C_1, \ldots, C_n$ is permuted, one can assume without loss of generality that $\bar{\omega}_{AC_1 \ldots C_k}$ is invariant under such permutations. The claim then follows immediately from Theorem II.7′ of [7]. □

Corollary 3.3 ([15])**.** *For any density operator ρ_{AC} on $A \otimes C$*

$$\inf_{\sigma_{AC} \in S_{A:C}} \Delta(\rho_{AC}, \sigma_{AC}) \leq 2 \dim C \sqrt[4]{2 \ln(2) E_{\mathrm{sq}}(\rho_{AC})} \ . \tag{67}$$

4 Some open questions

1. Can we say something more on the map. For example, does the Petz map satisfy the inequality? We know that the map can be chosen to be universal, i.e., depending only on the marginal ρ_{BC} and not on the particular correlations we wish to recover [21].

2. Can we strengthen the inequality? We know that

$$I(A:C|B)_\rho \geq \min_{\mathcal{T}:B \to BC} D_{\mathbb{M}}(\rho_{ABC} \| \mathcal{T}_{B \to BC}(\rho_{AB})) \ . \tag{68}$$

But do we have

$$I(A:C|B)_\rho \geq \min_{\mathcal{T}:B \to BC} D(\rho_{ABC} \| \mathcal{T}_{B \to BC}(\rho_{AB})) \ . \tag{69}$$

3. Can these new inequalities tell us more about states the entanglement in states with small conditional mutual information. Does $I(A:C|B)_\rho \leq \epsilon$ imply that $\rho \approx_{f(\epsilon)} \sigma$ where $\approx$ would be measured in some other norms (for example restricted norms, SEP, etc...)?

References

[1] M. Berta, M. Lemm, and M. M. Wilde. Monotonicity of quantum relative entropy and recoverability. *arXiv preprint arXiv:1412.4067*, 2014.

[2] M. Berta, K. Seshadreesan, and M. Wilde. Renyi generalizations of the conditional quantum mutual information. 2014. arXiv:1403.6102.

[3] M. Berta and M. Tomamichel. The fidelity of recovery is multiplicative. *arXiv preprint arXiv:1502.07973*, 2015.

[4] F. Brandão, A. Harrow, J. Oppenheim, and S. Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. 2014. arXiv:1411.4921.

[5] F. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Comm. Math. Phys.*, 306(3):805–830, 2011. arXiv:1010.1750.

[6] E. Carlen. Trace inequalities and quantum entropy: an introductory course. 2010.

[7] M. Christandl, R. König, G. Mitchison, and R. Renner. One-and-a-half quantum de Finetti theorems. *Comm. Math. Phys.*, 273(2):473–498, 2007. arXiv:quant-ph/0602130.

[8] M. Christandl, N. Schuch, and A. Winter. Entanglement of the antisymmetric state. *Comm. Math. Phys.*, 311(2):397–422, 2012. arXiv:0910.4151.

[9] M. Christandl and A. Winter. "Squashed entanglement": An additive entanglement measure. *J. Math. Phys.*, 45(3):829–840, 2004.

[10] O. Fawzi and R. Renner. Quantum Conditional Mutual Information and Approximate Markov Chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.

[11] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Comm. Math. Phys.*, 246(2):359–374, 2004. arXiv:quant-ph/0304007.

[12] B. Ibinson, N. Linden, and A. Winter. Robustness of Quantum Markov Chains. *Comm. Math. Phys.*, 277(2):289–304, 2008. arXiv:quant-ph/0611057.

[13] I. Kim. Application of conditional independence to gapped quantum many-body systems, 2013. http://www.physics.usyd.edu.au/quantum/Coogee2013/Presentations/Kim.pdf.

[14] K. Li and A. Winter. Relative entropy and squashed entanglement. *Comm. Math. Phys.*, 326(1):63–80, 2014. arXiv:1210.3181.

[15] K. Li and A. Winter. Squashed entanglement, k-extendibility, quantum Markov chains, and recovery maps. 2014. arXiv:1410.4184.

[16] E. Lieb and M. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.*, 14(12):1938–1941, 1973.

[17] E. H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11(3):267 – 288, 1973.

[18] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *J. Math. Phys.*, 54(12):122203, 2013. arXiv:1306.3142.

[19] D. Petz. Sufficiency of channels over von Neumann algebras. *Q. J. Math.*, 39(1):97–108, 1988.

[20] K. Seshadreesan and M. Wilde. Fidelity of recovery and geometric squashed entanglement. 2014. arXiv:1410.1441.

[21] D. Sutter, O. Fawzi, and R. Renner. Universal recovery map for approximate markov chains. *arXiv preprint arXiv:1504.07251*, 2015.

[22] D. Sutter, M. Tomamichel, and A. W. Harrow. Strengthened Monotonicity of Relative Entropy via Pinched Petz Recovery Map. *arXiv preprint arXiv:1507.00303*, 2015.

[23] M. Tomamichel. *Quantum Information Processing with Finite Resources-Mathematical Foundations.* 2015.

[24] M. Wilde. Recoverability in quantum information theory, 2015. *arXiv preprint arXiv:1505.04661.*

[25] M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Comm. Math. Phys.*, 331(2):593–622, 2014. arXiv:1306.1586.

[26] A. Winter and K. Li. A stronger subadditivity relation? with applications to squashed entanglement, shareability and separability. available at http://www.maths.bris.ac.uk/˜csajw/stronger_subadditivity.pdf, 2012.

[27] L. Zhang. Conditional mutual information and commutator. *Int. J. of Theor. Phys.*, 52(6):2112–2117, 2013. arXiv:1212.5023.