Exploiting Variational Formulas for Quantum Relative Entropy

Mario Berta
Institute for Quantum Information and Matter
California Institute of Technology

Omar Fawzi
Laboratoire de l'Informatique du Parallélisme
École Normale Supérieure de Lyon

Marco Tomamichel
School of Physics
The University of Sydney

***Abstract*—The relative entropy is the basic concept underlying various information measures like entropy, conditional entropy and mutual information. Here, we discuss how to make use of variational formulas for measured relative entropy and quantum relative entropy for understanding the additivity properties of various entropic quantities that appear in quantum information theory. In particular, we show that certain lower bounds on quantum conditional mutual information are superadditive.**

I. Quantum Relative Entropies

For a positive measure Q on a finite set $\mathcal{X}$ and a probability measure P on $\mathcal{X}$ that is absolutely continuous with respect to Q, denoted $P \ll Q$, the relative entropy or Kullback-Leibler divergence [1] is defined as

$$D(P\|Q) := \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}\,, \tag{1}$$

where we understand $P(x)\log\frac{P(x)}{Q(x)} = 0$ whenever $P(x) = 0$. By continuity we define it as $+\infty$ if $P \not\ll Q$. (We use log to denote the natural logarithm.) More generally, for $\alpha \in (1,\infty)$ we define the Rényi divergence [2] as

$$D_\alpha(P\|Q) := \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} P(x)\left(\frac{P(x)}{Q(x)}\right)^{\alpha-1} \tag{2}$$

if $P \ll Q$ and as $+\infty$ if $P \not\ll Q$. For $\alpha \in (0,1)$ it diverges to $+\infty$ only when P and Q are orthogonal. It is well known that the Rényi divergence converges to the Kullback-Leibler divergence when $\alpha \to 1$ and we thus set $D \equiv D_1$.

A thorough understanding of quantum generalizations of the relative entropy is of preeminent importance in quantum information theory. Let us denote the set of positive semidefinite operators acting on a finite-dimensional Hilbert space by $\mathcal{P}$ and the subset of density operators (with unit trace) by $\mathcal{S}$. For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, the *measured relative entropy* is defined as [3], [4], [5], $D^{\mathbb{M}}(\rho\|\sigma) := \sup_{(\mathcal{X},M)} D\big(P_{\rho,M}\big\|P_{\sigma,M}\big)$, where the optimization is over finite sets $\mathcal{X}$ and positive operator valued measures (POVMs) M on $\mathcal{X}$, and $P_{\rho,M}(x) = \mathrm{tr}[M(x)\rho]$. For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$ non-zero, the following identity holds [6], [7],

$$D^{\mathbb{M}}(\rho\|\sigma) = \sup_{\omega>0} \mathrm{tr}[\rho\log\omega] + 1 - \mathrm{tr}[\sigma\omega]\,. \tag{3}$$

The *measured Rényi relative entropy* for $\alpha \in (0,1)\cup(1,\infty)$ is defined as $D^{\mathbb{M}}_\alpha(\rho\|\sigma) := \sup_{(\mathcal{X},M)} D_\alpha\big(P_{\rho,M}\big\|P_{\sigma,M}\big)$. The following identities hold [7]:

$$Q^{\mathbb{M}}_\alpha(\rho\|\sigma) := \exp\big((\alpha-1)D^{\mathbb{M}}_\alpha(\rho\|\sigma)\big) \tag{4}$$
$$= \begin{cases} \inf\limits_{\omega>0} \alpha\,\mathrm{tr}\left[\rho\omega^{1-\frac{1}{\alpha}}\right] + (1-\alpha)\,\mathrm{tr}[\sigma\omega] & \text{for } \alpha\in(0,1) \\ \sup\limits_{\omega>0} \alpha\,\mathrm{tr}\left[\rho\omega^{1-\frac{1}{\alpha}}\right] + (1-\alpha)\,\mathrm{tr}[\sigma\omega] & \text{for } \alpha\in(1,\infty) \end{cases}$$
$$= \begin{cases} \inf\limits_{\omega>0} \mathrm{tr}\left[\rho\omega\right]^{\alpha}\mathrm{tr}\left[\sigma\omega^{\frac{\alpha}{\alpha-1}}\right]^{1-\alpha} & \text{for } \alpha\in(0,1) \\ \sup\limits_{\omega>0} \mathrm{tr}\left[\rho\omega^{1-\frac{1}{\alpha}}\right]^{\alpha}\mathrm{tr}[\sigma\omega]^{1-\alpha} & \text{for } \alpha\in(1,\infty) \end{cases}. \tag{5}$$

The last two expressions are a generalization of Alberti's theorem [8] for the fidelity (which corresponds to $\alpha = 1/2$) to general $\alpha \in (0,1)\cup(1,\infty)$. Since the Rényi divergences are uniformly continuous in α around one we also get for the limit $\alpha \to 1$ that $D^{\mathbb{M}} \equiv D^{\mathbb{M}}_1$.

In contrast to these measured quantities Umegaki's quantum relative entropy [9] is given as $D(\rho\|\sigma) := \mathrm{tr}\big[\rho(\log\rho - \log\sigma)\big]$ if $\sigma \gg \rho$. We recall Petz' variation expression [10],

$$D(\rho\|\sigma) = \sup_{\omega>0} \mathrm{tr}[\rho\log\omega] + 1 - \mathrm{tr}[\exp(\log\sigma + \log\omega)]\,. \tag{6}$$

We have [4], [7], $D^{\mathbb{M}}(\rho\|\sigma) \le D(\rho\|\sigma)$ with equality iff $[\rho,\sigma] = 0$. More generally, the sandwiched Rényi relative entropy of order $\alpha \in (0,1)\cup(1,\infty)$ is defined as [11], [12],

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log Q_\alpha(\rho\|\sigma), \quad \text{with} \tag{7}$$
$$Q_\alpha(\rho\|\sigma) := \mathrm{tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}\right]\,, \tag{8}$$

where the same considerations about finiteness as for the measured Rényi relative entropy apply. We also consider the limits $\alpha \to \infty$ and $\alpha \to 1$ of the above expression for which we have [11], $D_\infty(\rho\|\sigma) = \inf\{\lambda\in\mathbb{R} | \rho \le \exp(\lambda)\sigma\}$ and $D_1(\rho\|\sigma) = D(\rho\|\sigma)$, respectively. We recall the following variational expression [13],

$$Q_\alpha(\rho\|\sigma) = \begin{cases} \inf\limits_{\omega>0} \alpha\,\mathrm{tr}[\rho\omega] + (1-\alpha)\,\mathrm{tr}\left[\left(\omega^{\frac{1}{2}}\sigma^{\frac{\alpha-1}{\alpha}}\omega^{\frac{1}{2}}\right)^{\frac{\alpha}{\alpha-1}}\right] \\ \sup\limits_{\omega>0} \alpha\,\mathrm{tr}[\rho\omega] + (1-\alpha)\,\mathrm{tr}\left[\left(\omega^{\frac{1}{2}}\sigma^{\frac{\alpha-1}{\alpha}}\omega^{\frac{1}{2}}\right)^{\frac{\alpha}{\alpha-1}}\right], \end{cases} \tag{9}$$

where the first formula is valid for $\alpha \in (0,1)$ and the second one for $\alpha \in (1,\infty)$. We have $D^{\mathbb{M}}_\alpha(\rho\|\sigma) \le D_\alpha(\rho\|\sigma)$ with equality iff $[\rho,\sigma] = 0$ [7].

II. Some Optimization Problems in Quantum Information

The variational characterization of the relative entropy (6), the sandwiched Rényi relative entropy (9), and their measured counterparts (3) and (5), turn out to be useful for deriving properties of various entropic quantities that appear in quantum information theory. Here, we are interested in operational quantities of the form

$$\mathcal{M}(\rho) := \min_{\sigma \in \mathcal{C}} \mathbb{D}(\rho \| \sigma) , \tag{10}$$

where $\mathbb{D}(\cdot\|\cdot)$ stands for any relative entropy, measured relative entropy, or Rényi variant thereof, and $\mathcal{C}$ denotes some convex, compact set of states. For Umegaki's relative entropy $\mathbb{D} \equiv D$, prominent examples for $\mathcal{C}$ include the set of

- separable states, giving rise to the relative entropy of entanglement [14].
- positive partial transpose states, leading to the Rains bound on entanglement distillation [15].
- non-distillable states, leading to bounds on entanglement distillation [16].
- quantum Markov states, leading to insights about the robustness properties of these states [17].
- locally recoverable states, leading to bounds on the quantum conditional mutual information [18], [19], [20].
- k-extendible states, leading to bounds on squashed entanglement [21].

Other examples are conditional Rényi entropies which are defined by optimizing the sandwiched Rényi relative entropy over a convex set of product states with a fixed marginal, see, e.g., [22]. A central question is what properties of the underlying relative entropy $\mathbb{D}$ translate to properties of the induced measure $\mathcal{M}(\cdot)$. For example, all the relative entropies discussed in this paper are superadditive on tensor product states in the sense that

$$\mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \geq \mathbb{D}(\rho_1 \| \sigma_1) + \mathbb{D}(\rho_2 \| \sigma_2) . \tag{11}$$

We might then ask if we also have

$$\min_{\sigma_{12} \in \mathcal{C}_{12}} \mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_{12}) = \mathcal{M}(\rho_1 \otimes \rho_2) \overset{?}{\geq} \mathcal{M}(\rho_1) + \mathcal{M}(\rho_2) \tag{12}$$

$$= \min_{\sigma_1 \in \mathcal{C}_1} \mathbb{D}(\rho_1 \| \sigma_1) + \min_{\sigma_2 \in \mathcal{C}_2} \mathbb{D}(\rho_2 \| \sigma_2) . \tag{13}$$

To study questions like this we propose to make use of the variational characterizations of the form

$$\mathbb{D}(\rho \| \sigma) = \sup_{\omega > 0} f(\rho, \sigma, \omega) \quad \text{in order to write} \tag{14}$$

$$\mathcal{M}(\rho) = \min_{\sigma \in \mathcal{C}} \sup_{\omega > 0} f(\rho, \sigma, \omega) = \sup_{\omega > 0} \min_{\sigma \in \mathcal{C}} f(\rho, \sigma, \omega) , \tag{15}$$

where we made use of Sion's minimax theorem [23] for the last equality. We note that the conditions of the minimax theorem are fulfilled for all relative entropies discussed in this paper. The minimization over $\sigma \in \mathcal{C}$ then often simplifies and is a convex or even semidefinite optimization.[1] We can then use strong duality of convex optimization to rewrite this minimization as a maximization problem [24], $\min_{\sigma \in \mathcal{C}} f(\rho, \sigma, \omega) = \max_{\bar{\sigma} \in \bar{\mathcal{C}}} \bar{f}(\rho, \bar{\sigma}, \omega)$. This leads to the expression

$$\mathcal{M}(\rho) = \sup_{\omega > 0} \max_{\bar{\sigma} \in \bar{\mathcal{C}}} \bar{f}(\rho, \bar{\sigma}, \omega) , \tag{16}$$

which, in contrast to the definition of $\mathcal{M}(\rho)$ in (10), only involves maximizations. This often gives useful insights about $\mathcal{M}(\rho)$. As an example, let us come back to the question of superadditivity raised in (13). We want to argue that the following two conditions on $\bar{f}$ and $\bar{\mathcal{C}}$ imply superadditivity. First, we need that the function $\bar{f}$ is superadditive itself, i.e. we require that $\bar{f}(\rho_1 \otimes \rho_2, \bar{\sigma}_1 \otimes \bar{\sigma}_2, \omega_1 \otimes \omega_2) \geq \bar{f}(\rho_1, \bar{\sigma}_1, \omega_1) + \bar{f}(\rho_2, \bar{\sigma}_2, \omega_2)$. And second, we require that the sets $\bar{\mathcal{C}}$ are closed under tensor products in the sense that $\bar{\sigma}_1 \in \bar{\mathcal{C}}_1$ and $\bar{\sigma}_2 \in \bar{\mathcal{C}}_2$ imply that $\bar{\sigma}_1 \otimes \bar{\sigma}_2 \in \bar{\mathcal{C}}_{12}$. Using these two properties, we deduce that $\mathcal{M}(\rho_1 \otimes \rho_2) \geq \bar{f}(\rho_1 \otimes \rho_2, \bar{\sigma}_1 \otimes \bar{\sigma}_2, \omega_1 \otimes \omega_2) \geq \bar{f}(\rho_1, \bar{\sigma}_1, \omega_1) + \bar{f}(\rho_2, \bar{\sigma}_2, \omega_2)$, for any $\omega_1, \omega_2 > 0$ and any $\bar{\sigma}_1 \in \bar{\mathcal{C}}_1$, $\bar{\sigma}_2 \in \bar{\mathcal{C}}_2$. Hence, the inequalities also holds true if we maximize over these variables, implying superadditivity.

III. Relative Entropy of Recovery

In the following we denote multipartite quantum systems using capital letters, e.g., A, B, C. The set of density operators on A and B is then denoted $\mathcal{S}(AB)$, for example. We also use these letters as subscripts to indicate which systems operators act on.

As a concrete application, we study the additivity properties of the *relative entropy of recovery* defined as [20], [19], [25]

$$D^{\mathrm{rec}}(\rho_{AD} \| \sigma_{AE}) := \inf_{\Gamma_{E \to D}} D(\rho_{AD} \| (\mathcal{I}_A \otimes \Gamma_{E \to D})(\sigma_{AE})) , \tag{17}$$

where the infimum is taken over all trace-preserving completely positive maps $\Gamma_{E \to D}$. One motivation for studying the additivity properties of the relative entropy of recovery is the study of lower bounds on the quantum conditional mutual information and strengthenings of the data-processing inequality for the relative entropy [18], [20], [26], [27], [28], [29], [30], [31]. In particular, [20, Prop. 3] shows that

$$I(A : C | B) \geq \lim_{n \to \infty} \frac{1}{n} D^{\mathrm{rec}} \left(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n} \right) , \tag{18}$$

where the systems are understood as $D = BC$ and $E = B$. To obtain a lower bound that does not involve limits, the authors of [20] use a Finetti-type theorem to show that

$$\lim_{n \to \infty} \frac{1}{n} D^{\mathrm{rec}}(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n}) \geq D^{\mathbb{M},\mathrm{rec}}(\rho_{ABC} \| \rho_{AB}) . \tag{19}$$

with the *measured relative entropy of recovery* defined as [20] (see also [19, Rmk. 6]),

$$D^{\mathbb{M},\mathrm{rec}}(\rho_{AD} \| \sigma_{AE}) := \inf_{\Gamma_{E \to D}} D^{\mathbb{M}}(\rho_{AD} \| (\mathcal{I}_A \otimes \Gamma_{E \to D})(\sigma_{AE})). \tag{20}$$

[1] As an example, for the measured relative entropies the objective function becomes linear in σ.

This gives an interpretation for the conditional mutual information in terms of recoverability. The measured relative entropy is superadditive and if this property would translate to $D^{\mathbb{M},\mathrm{rec}}$ we would get an alternative proof for the step (19). Using the variational expression for the measured relative entropy (3) together with strong duality for semidefinite programming, we find the following alternative characterization for $D^{\mathbb{M},\mathrm{rec}}$.

Lemma 1. *Let $\rho_{AD} \in \mathcal{S}(AD)$ and $\sigma_{AE} \in \mathcal{S}(AE)$ with $\sigma_A > 0$, and let σ_{AEF} be a purification of σ_{AE}. Then, we have*

$$D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \begin{array}{rl} \text{maximize:} & \mathrm{tr}[\rho_{AD}\log R_{AD}] \\ \text{subject to:} & S_{AF} > 0,\ R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \\ & \mathrm{tr}[S_{AF}\sigma_{AF}] = 1\,. \end{array} \tag{21}$$

Proof. We write

$$\Gamma_{E\to D}(\sigma_{AE}) = \mathrm{tr}_F\left[\Gamma_{E\to D}\big(\sqrt{\sigma_{AE}}\Psi_{AE:F}\sqrt{\sigma_{AE}}\big)\right], \tag{22}$$

where we denote by $\Psi_{AE:F}$ the (unnormalized) maximally entangled state between $AE:F$ in the Schmidt decomposition of σ_{AEF}. With this we define the Choi-Jamiołkowski state (unnormalized) of the map $\Gamma_{E\to D}$ as

$$\tau_{ADF} = \Gamma_{E\to D}\big(\Psi_{AE:F}\big), \quad \tau_{AF} = \Pi^\sigma_{AF}\,, \tag{23}$$

where Π^σ_{AF} denotes the projector onto the support of σ_{AF}. Hence, we can write

$$\Gamma_{E\to D}(\sigma_{AE}) = \mathrm{tr}_F\left[\sqrt{\sigma_{AF}}\tau_{ADF}\sqrt{\sigma_{AF}}\right], \tag{24}$$

and thus we can express the optimization problem for $\Gamma_{E\to D}$ in terms of the Choi-Jamiołkowski state in (23). Together with the variational characterization of the measured relative entropy (3) we find

$$\begin{aligned} D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \min_{\tau_{ADF}\in\mathrm{Rec}(\sigma_{AE})} \sup_{R_{AD}>0} \mathrm{tr}[\rho_{AD}\log R_{AD}] \\ + 1 - \mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right], \end{aligned} \tag{25}$$

where $\mathrm{Rec}(\sigma_{AE}) := \{\tau_{ADF} \geq 0,\ \tau_{AF} = \Pi^\sigma_{AF}\}$. We now apply Sion's minimax theorem [23] to exchange the minimum with the supremum. The theorem is applicable as $\mathrm{Rec}(\sigma_{AE})$ is convex and compact and $\{\omega_{AD} > 0\}$ is convex. Moreover, as the logarithm is operator concave, the function

$$R_{AD} \mapsto \mathrm{tr}[\rho_{AD}\log R_{AD}] + 1 - \mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right] \tag{26}$$

is concave for any fixed τ_{ADF}. Finally, for any fixed R_{AD}, the function being optimized is linear on τ_{ADF}. As a result,

$$\begin{aligned} D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \sup_{R_{AD}>0} \mathrm{tr}[\rho_{AD}\log R_{AD}] + 1 - \\ \max_{\tau_{ADF}\in\mathrm{Rec}(\sigma_{AE})} \mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right]. \end{aligned} \tag{27}$$

Now for a fixed $R_{AD} > 0$, the inner maximization is a semidefinite program for which we can write the following programs,

$$\begin{array}{rl} \text{maximize:} & \mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right] \\ \text{subject to:} & \tau_{ADF} \geq 0 \\ & \tau_{AF} = \Pi^\sigma_{AF} \end{array} \tag{28}$$

$$\begin{array}{rl} \text{minimize:} & \mathrm{tr}[S_{AF}\sigma_{AF}] \\ \text{subject to:} & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F\,. \end{array} \tag{29}$$

As the primal problem is strictly feasible, Slater's condition (see, e.g., [24]) is satisfied and we have strong duality. This leads to the following expression for $D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE})$:

$$\begin{array}{rl} \text{maximize:} & \mathrm{tr}[\rho_{AD}\log R_{AD}] + 1 - \mathrm{tr}[S_{AF}\sigma_{AF}] \\ \text{subject to:} & S_{AF} > 0,\ R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F\,. \end{array} \tag{30}$$

Observe now that we can restrict the optimization to $\mathrm{tr}[S_{AF}\sigma_{AF}] = 1$. In fact, for arbitrary feasible R_{AD} and S_{AF}, we can define $\tilde{R}_{AD} = \frac{R_{AD}}{\mathrm{tr}[S_{AF}\sigma_{AF}]}$ and $\tilde{S}_{AF} = \frac{S_{AF}}{\mathrm{tr}[S_{AF}\sigma_{AF}]}$. The constraint $1_D \otimes \tilde{S}_{AF} \geq \tilde{R}_{AD} \otimes 1_F$ is still satisfied and the value of the objective function can only increase,

$$\mathrm{tr}[\rho_{AD}\log \tilde{R}_{AD}] \geq \mathrm{tr}[\rho_{AD}\log R_{AD}] - \mathrm{tr}[S_{AF}\sigma_{AF}] + 1\,, \tag{31}$$

where we used the inequality $\log x \leq x - 1$ and that $\mathrm{tr}[\rho_{AD}] = 1$. ☐

This readily implies that $D^{\mathbb{M},\mathrm{rec}}$ is indeed superadditive.

Proposition 2. *Let $\rho_{AD} \in \mathcal{S}(AD)$, $\tau_{A'D'} \in \mathcal{S}(A'D')$, $\sigma_{AE} \in \mathcal{S}(AE)$, and $\omega_{A'E'} \in \mathcal{S}(A'E')$ with $\sigma_A, \omega_{A'} > 0$. Then,*

$$\begin{aligned} & D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\otimes\tau_{A'D'}\|\sigma_{AE}\otimes\omega_{A'E'}) \\ & \geq D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) + D^{\mathbb{M},\mathrm{rec}}(\tau_{A'D'}\|\omega_{A'E'})\,. \end{aligned} \tag{32}$$

Proof. Given feasible operators R_{AD}, S_{AF} for the quantity $D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE})$ and feasible operators $R_{A'D'}, S_{A'F'}$ for the quantity $D^{\mathbb{M},\mathrm{rec}}(\tau_{A'D'}\|\omega_{A'E'})$, we have

$$\begin{aligned} 1_D \otimes S_{AF} \geq R_{AD}\otimes 1_F \ \wedge\ 1_{D'}\otimes S_{A'F'} \geq R_{A'D'}\otimes 1_{F'} \\ \Longrightarrow \\ 1_{DD'}\otimes S_{AF}\otimes S_{A'F'} \geq R_{AD}\otimes R_{A'D'}\otimes 1_{FF'}\,. \end{aligned} \tag{33}$$

Here, we used that $A \geq B \implies M\otimes A \geq M \otimes B$ for $M \geq 0$, which holds since taking the tensor product with M is a positive map. Moreover, we have $\mathrm{tr}\left[(S_{AF}\otimes S_{A'F'})(\sigma_{AF}\otimes \sigma_{A'F'})\right] = 1$. In other words, $(R_{AD}\otimes R_{A'D'}, S_{AF}\otimes S_{A'F'})$ is a feasible pair in the expression (21) for $D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\otimes \tau_{A'D'}\|\sigma_{AE}\otimes\omega_{A'E'})$ and we get

$$\begin{aligned} & D^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\otimes\tau_{A'D'}\|\sigma_{AE}\otimes\omega_{A'E'}) \\ & \geq \mathrm{tr}\left[\rho_{AD}\log R_{AD}\right] + \mathrm{tr}\left[\tau_{A'D'}\log R_{A'D'}\right]\,. \end{aligned} \tag{34}$$

Taking the supremum over feasible (R_{AD}, S_{AF}) and $(R_{A'D'}, S_{A'F'})$, we get the claimed superadditivity. ☐

Now, if the relative entropy of recovery would be superadditive we could strengthen (19) to

$$\lim_{n\to\infty} \frac{1}{n} D^{\mathrm{rec}}(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n}) \overset{?}{\geq} D^{\mathrm{rec}}(\rho_{ABC} \| \rho_{AB})\,. \tag{35}$$

This together with (18) would lead to a stronger lower bound on the conditional mutual information. Using strong duality for convex programming we can write D^{rec} in a dual form similar to Lemma 1.

Lemma 3. *Let $\rho_{AD} \in \mathcal{S}(AD)$ and $\sigma_{AE} \in \mathcal{S}(AE)$ with $\sigma_A > 0$, and let σ_{AEF} be a purification of σ_{AE}. Then, we have*

$$\begin{aligned} &D^{\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \\ &\quad \text{maximize:} \quad \mathrm{tr}[\rho_{AD}\log\rho_{AD}] - D^{\mathbb{M}}(\rho_{AD}\|R_{AD}) \\ &\quad \text{subject to:} \quad S_{AF} > 0,\ R_{AD} > 0 \\ &\qquad\qquad\qquad 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \\ &\qquad\qquad\qquad \mathrm{tr}[S_{AF}\sigma_{AF}] = 1\,, \end{aligned} \tag{36}$$

where R_{AD} is not normalized and the measured relative entropy term is in general negative.

We provide a proof based on strong duality for convex programming in the full version [7]. The expression (36) is to be compared to our variational expression for the measured relative entropy of recovery from Lemma 1, which has the same constraints and whose objective function can be written as $\mathrm{tr}[\rho_{AD}\log\rho_{AD}] - D(\rho_{AD}\|R_{AD})$. Unfortunately, we can not solve the open additivity question for the relative entropy of recovery using the variational expression from Lemma 3. However, we note that the argument flipped relative entropy of recovery $\bar{D}^{\mathrm{rec}}(\sigma_{AE}\|\rho_{AD}) := \inf_{\Gamma_{E\to D}} D((\mathcal{I}_A \otimes \Gamma_{E\to D})(\sigma_{AE})\|\rho_{AD})$ becomes additive.

Proposition 4. *Let $\rho_{AD} \in \mathcal{S}(AD)$ and $\sigma_{AE} \in \mathcal{S}(AE)$ with $\rho_A > 0$, and let σ_{AEF} be a purification of σ_{AE}. Then,*

$$\begin{aligned} &\bar{D}^{\mathrm{rec}}(\sigma_{AE}\|\rho_{AD}) = \\ &\quad \text{maximize:} \quad -\mathrm{tr}[S_{AF}\sigma_{AF}] \\ &\quad \text{subject to:} \quad S_{AF} > 0,\ R_{AD} > 0 \\ &\qquad\qquad\qquad 1_D \otimes S_{AF} \geq (\log\rho_{AD} - \log R_{AD}) \otimes 1_F \\ &\qquad\qquad\qquad \mathrm{tr}[R_{AD}] = 1\,. \end{aligned} \tag{37}$$

Moreover, for $\tau_{A'D'} \in \mathcal{S}(A'D')$ with $\tau_{A'} > 0$, we have $\bar{D}^{\mathrm{rec}}(\sigma_{AE} \otimes \omega_{A'E'}\|\rho_{AD} \otimes \tau_{A'D'}) = \bar{D}^{\mathrm{rec}}(\sigma_{AE}\|\rho_{AD}) + \bar{D}^{\mathrm{rec}}(\omega_{A'E'}\|\tau_{A'D'})$.

We provide a proof based on strong duality for convex programming in the full version [7].

IV. Rényi Relative Entropy of Recovery

As a generalization of the measured relative entropy of recovery (20) we define the *measured Rényi relative entropy of recovery* for $\alpha \in (1,\infty)$ as (see also [19, Rmk. 6]),

$$D_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) := \frac{1}{\alpha-1}\log Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) \tag{38}$$

with

$$Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) := \inf_{\Gamma_{E\to D}} Q_\alpha^{\mathbb{M}}(\rho_{AD}\|(\mathcal{I}_A \otimes \Gamma_{E\to D})(\sigma_{AE})) \tag{39}$$

and analogously for $\alpha \in (0,1)$ with

$$Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) := \sup_{\Gamma_{E\to D}} Q_\alpha^{\mathbb{M}}(\rho_{AD}\|(\mathcal{I}_A \otimes \Gamma_{E\to D})(\sigma_{AE})). \tag{40}$$

Using the variational expressions for the measured Rényi relative entropy (5) together with strong duality for semidefinite programming, we find the following alternative characterization for $Q_\alpha^{\mathbb{M},\mathrm{rec}}$.

Lemma 5. *Let $\rho_{AD} \in \mathcal{S}(AD)$ and $\sigma_{AE} \in \mathcal{S}(AE)$ with $\sigma_A > 0$, and let σ_{AEF} be a purification of σ_{AE}. For $\alpha \in [\frac{1}{2},1)$,*

$$\begin{aligned} &Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \\ &\quad \text{minimize:} \quad \mathrm{tr}\left[\rho_{AD}R_{AD}^{1-\frac{1}{\alpha}}\right]^\alpha \mathrm{tr}[S_{AF}\sigma_{AF}]^{1-\alpha} \\ &\quad \text{subject to:} \quad S_{AF} > 0,\ R_{AD} > 0 \\ &\qquad\qquad\qquad 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F\,. \end{aligned} \tag{41}$$

Similar dual formulas hold for $\alpha \in (0,\frac{1}{2}) \cup (1,\infty)$.

Proof. Using the variational formula (5), we get from Sion's minimax theorem [23] that

$$\begin{aligned} Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \inf_{R_{AD}>0} \max_{\gamma_{AD}\in\mathrm{Rec}(\sigma_{AE})} \alpha\,\mathrm{tr}\left[\rho_{AD}R_{AD}^{1-\frac{1}{\alpha}}\right] \\ + (1-\alpha)\,\mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right]. \end{aligned} \tag{42}$$

The set $\mathrm{Rec}(\sigma_{AE})$ is convex and compact and $\{R_{AD} > 0\}$ is convex. Moreover, the function $R_{AD} \mapsto \alpha\,\mathrm{tr}\left[\rho_{AD}R_{AD}^{1-\frac{1}{\alpha}}\right] + (1-\alpha)\,\mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right]$ is convex for any fixed τ_{ADF} because of the operator convexity of $t \mapsto t^\beta$ with $\beta = 1 - \frac{1}{\alpha}$ for $\alpha \in [\frac{1}{2},1)$. Finally, for a fixed R_{AD}, the function being optimized is linear on τ_{ADF}. As in (28) we then look at the convex dual of $\max_{\tau_{ADF}\in\mathrm{Rec}(\sigma_{AE})} \mathrm{tr}\left[\tau_{ADF}\sqrt{\sigma_{AF}}R_{AD}\sqrt{\sigma_{AF}}\right]$, and find

$$\begin{aligned} &Q_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) = \\ &\quad \text{minimize:} \quad \alpha\,\mathrm{tr}\left[\rho_{AD}R_{AD}^{1-\frac{1}{\alpha}}\right] + (1-\alpha)\,\mathrm{tr}[S_{AF}\sigma_{AF}] \\ &\quad \text{subject to:} \quad S_{AF} > 0,\ R_{AD} > 0 \\ &\qquad\qquad\qquad 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F\,. \end{aligned} \tag{43}$$

Invoking the arithmetic-geometric mean inequality establishes the claim. ☐

Using this we find that the measured Rényi entropy of recovery is superadditive for all Rényi parameters.

Proposition 6. *Let $\rho_{AD} \in \mathcal{S}(AD)$, $\tau_{A'D'} \in \mathcal{S}(A'D')$, $\sigma_{AE} \in \mathcal{S}(AE)$, and $\omega_{A'E'} \in \mathcal{S}(A'E')$ with $\sigma_A, \omega_{A'} > 0$. For $\alpha \in (0,1) \cup (1,\infty)$, we have*

$$\begin{aligned} &D_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD} \otimes \tau_{A'D'}\|\sigma_{AE} \otimes \omega_{A'E'}) \\ &\quad \geq D_\alpha^{\mathbb{M},\mathrm{rec}}(\rho_{AD}\|\sigma_{AE}) + D_\alpha^{\mathbb{M},\mathrm{rec}}(\tau_{A'D'}\|\omega_{A'E'})\,. \end{aligned} \tag{44}$$

Analogously, (17) is generalized to the *Rényi relative entropy of recovery* as (see also [19, Rmk. 6]), $D_\alpha^{\text{rec}}(\rho_{AD}\|\sigma_{AE}) := \frac{1}{\alpha-1}\log Q_\alpha^{\text{rec}}(\rho_{AD}\|\sigma_{AE})$ with

$$Q_\alpha^{\text{rec}}(\rho_{AD}\|\sigma_{AE}) := \begin{cases} \inf_{\Gamma_{E\to D}} Q_\alpha(\rho_{AD}\|(\mathcal{I}_A\otimes\Gamma_{E\to D})(\sigma_{AE})) & \text{for } \alpha\in(1,\infty) \\ \sup_{\Gamma_{E\to D}} Q_\alpha(\rho_{AD}\|(\mathcal{I}_A\otimes\Gamma_{E\to D})(\sigma_{AE})) & \text{for } \alpha\in(0,1). \end{cases} \tag{45}$$

Now, for the special case of $\alpha = 1/2$, the sandwiched Rényi relative entropy is equal to the negative logarithm of the fidelity and as the measured fidelity is the same as the fidelity (see, e.g., [32, Sec. 3.3]). That is, we have $Q_{1/2}^{\mathbb{M}} = Q_{1/2}$ as well as $\mathrm{Q}_{1/2}^{\text{rec}} = \mathrm{Q}_{1/2}^{\mathbb{M},\text{rec}}$. The same holds true at $\alpha\to\infty$ [33]. As such Proposition 6 can be seen as a generalization of the multiplicativity of the fidelity of recovery [19] that was derived by two of the authors [26]. Finally, since the Rényi divergences are uniformly continuous in α around one, we recover the superadditivity of the measured relative entropy of recovery (Proposition 2) in the limit $\alpha\to 1$.

V. Conclusion

Using variational characterizations of relative entropy we have shown that the measured Rényi relative entropy of recovery is superadditive for all $\alpha > 0$. It would be desirable to extend this to the sandwiched Rényi relative entropy of recovery, including the relative entropy of recovery corresponding to $\alpha = 1$. However, for this case we only have preliminary numerics indicating that they are not additive. Finally, it would also be interesting to use variational characterizations for relative entropy for studying the operational entropic quantities mentioned in Section II. This might lead to new applications of measured relative entropy in quantum information theory (see also [34]).

Acknowledgments: We acknowledge discussions with Fernando Brandão, Douglas Farenick and Hamza Fawzi. Most of this work was done while OF was also with the Department of Computing and Mathematical Sciences at CalTech. MT would like to thank the IQIM at CalTech and John Preskill for his hospitality during the time this project was conceived.

References

[1] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
[2] A. Rényi, "On Measures of Information and Entropy," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 1961, pp. 547–561.
[3] M. J. Donald, "On the Relative Entropy," *Communications in Mathematical Physics*, vol. 105, no. 1, pp. 13–34, 1986.
[4] D. Petz, "Sufficient Subalgebras and the Relative Entropy of States of a von Neumann Algebra," *Communications in Mathematical Physics*, vol. 105, no. 1, pp. 123–131, 1986.
[5] F. Hiai and D. Petz, "The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, 1991.
[6] D. Petz, *Quantum Information Theory and Quantum Statistics*. Springer, 2008.
[7] M. Berta, O. Fawzi, and M. Tomamichel, "On Variational Expressions for Quantum Relative Entropies," 2015. [Online]. Available: http://arxiv.org/abs/1512.02615
[8] P. M. Alberti, "A Note on the Transition Probability over C*-Algebras," *Letters in Mathematical Physics*, vol. 7, pp. 25–32, 1983.
[9] H. Umegaki, "Conditional Expectation in an Operator Algebra," *Kodai Mathematical Seminar Reports*, vol. 14, no. 2, pp. 59–85, 1962.
[10] D. Petz, "A Variational Expression for the Relative Entropy," *Communications in Mathematical Physics*, vol. 114, no. 2, pp. 345–349, 1988.
[11] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On Quantum Rényi Entropies: A New Generalization and Some Properties," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122203, 2013.
[12] M. 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," *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, 2014.
[13] R. L. Frank and E. H. Lieb, "Monotonicity of a Relative Rényi Entropy," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122201, 2013.
[14] V. Vedral and M. B. Plenio, "Entanglement Measures and Purification Procedures," *Physical Review A*, vol. 57, no. 3, pp. 1619–1633, 1998.
[15] E. Rains, "A Semidefinite Program for Distillable Entanglement," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2921–2933, 2001.
[16] V. Vedral, "On Bound Entanglement Assisted Distillation," *Physics Letters A*, vol. 262, no. 2-3, pp. 121–124, 1999.
[17] B. Ibinson, N. Linden, and A. Winter, "Robustness of Quantum Markov Chains," *Communications in Mathematical Physics*, vol. 277, no. 2, pp. 289–304, 2007.
[18] O. Fawzi and R. Renner, "Quantum Conditional Mutual Information and Approximate Markov Chains," *Communications in Mathematical Physics*, vol. 340, no. 2, pp. 575–611, 2015.
[19] K. P. Seshadreesan and M. M. Wilde, "Fidelity of Recovery, Squashed Entanglement, and Measurement Recoverability," *Physical Review A*, vol. 92, no. 4, p. 042321, 2015.
[20] F. G. S. L. Brandão, A. W. Harrow, J. Oppenheim, and S. Strelchuk, "Quantum Conditional Mutual Information, Reconstructed States, and State Redistribution," *Physical Review Letters*, vol. 115, no. 5, p. 050501, 2015.
[21] K. Li and A. Winter, "Squashed Entanglement, k-Extendibility, Quantum Markov Chains, and Recovery Maps," 2014. [Online]. Available: http://arxiv.org/abs/1410.4184
[22] M. Tomamichel, M. Berta, and M. Hayashi, "Relating Different Quantum Generalizations of the Conditional Rényi Entropy," *Journal of Mathematical Physics*, vol. 55, no. 8, p. 082206, 2014.
[23] M. Sion, "On General Minimax Theorems," *Pacific Journal of Mathematics*, vol. 8, pp. 171–176, 1958.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[25] T. Cooney, C. Hirche, C. Morgan, J. P. Olson, K. P. Seshadreesan, and M. M. Wilde, "Operational meaning of quantum measures of recovery," 2015. [Online]. Available: http://arxiv.org/abs/1512.05324
[26] M. Berta and M. Tomamichel, "The Fidelity of Recovery is Multiplicative," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1758–1763, 2016.
[27] D. Sutter, M. Tomamichel, and A. W. Harrow, "Strengthened Monotonicity of Relative Entropy via Pinched Petz Recovery Map," *accepted at IEEE Transactions on Information Theory*, 2016.
[28] D. Sutter, O. Fawzi, and R. Renner, "Universal Recovery Map for Approximate Markov Chains," *Proceedings of the Royal Society A*, vol. 472, no. 2186, feb 2016.
[29] M. M. Wilde, "Recoverability in quantum information theory," *Proceedings of the Royal Society A*, vol. 471, no. 2182, p. 20150338, 2015.
[30] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter, "Universal recovery from a decrease of quantum relative entropy," 2015. [Online]. Available: http://arxiv.org/abs/1509.07127
[31] D. Sutter, M. Berta, and M. Tomamichel, "Multivariate Trace Inequalities," 2016. [Online]. Available: http://arxiv.org/abs/1604.03023
[32] C. A. Fuchs, "Distinguishability and Accessible Information in Quantum Theory," PhD Thesis, University of New Mexico, 1996. [Online]. Available: http://arxiv.org/abs/quant-ph/9601020
[33] M. Mosonyi and T. Ogawa, "Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies," *Communications in Mathematical Physics*, vol. 334, no. 3, pp. 1617–1648, 2015.
[34] M. Piani, "Relative Entropy of Entanglement and Restricted Measurements," *Physical Review Letters*, vol. 103, no. 16, p. 160504, 2009.