

On Variational Expressions for Quantum Relative Entropies

Mario Berta,¹ Omar Fawzi,² and Marco Tomamichel³

¹*Institute for Quantum Information and Matter, California Institute of Technology*

²*Laboratoire de l'Informatique du Parallélisme, École Normale Supérieure de Lyon*

³*School of Physics, The University of Sydney*

Distance measures between quantum states like the trace distance and the fidelity can naturally be defined by optimizing a classical distance measure over all measurement statistics that can be obtained from the respective quantum states. In contrast, Petz showed that the measured relative entropy, defined as a maximization of the Kullback-Leibler divergence over projective measurement statistics, is strictly smaller than Umegaki's quantum relative entropy whenever the states do not commute. We extend this result in two ways. First, we show that Petz' conclusion remains true if we allow general positive valued measures. To do this, we prove that maximizing the Kullback-Leibler divergence over general positive valued measures results in the measured relative entropy. Second, we extend the result to Rényi relative entropies and show that for non-commuting states the sandwiched Rényi relative entropy is strictly larger than the measured Rényi relative entropy for $\alpha \in (\frac{1}{2}, \infty)$, and strictly smaller for $\alpha \in [0, \frac{1}{2})$. The latter statement provides counterexamples for the data-processing inequality of the sandwiched Rényi relative entropy for $\alpha < \frac{1}{2}$. Our main tool is a new variational expression for the measured Rényi relative entropy, which we further exploit to show that certain lower bounds on quantum conditional mutual information are superadditive.

I. MEASURED RELATIVE ENTROPY

The relative entropy is the basic concept underlying various information measures like entropy, conditional entropy and mutual information. A thorough understanding of its quantum generalization is thus of preeminent importance in quantum information theory. We start by considering measured relative entropy, which is defined as a maximization of the Kullback-Leibler divergence over all measurement statistics that are attainable from two quantum states.

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 [25] is defined as

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

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.)

To extend this concept to quantum systems, Donald [11] as well as Hiai and Petz [19] studied measured relative entropy. In the following we restrict ourselves to a d -dimensional Hilbert space for some $d \in \mathbb{N}$. Let us denote the set of positive semidefinite operators acting on this space by \mathcal{P} and the subset of density operators (with unit trace) by \mathcal{S} . For a density operator $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, we define two variants of measured relative entropy. The general *measured relative entropy* is defined as

$$D^{\text{M}}(\rho||\sigma) := \sup_{(\mathcal{X}, M)} D(P_{\rho, M}||P_{\sigma, M}), \quad (2)$$

where the optimization is over finite sets \mathcal{X} and positive operator valued measures (POVMs) M on \mathcal{X} . (More formally, M is a map from \mathcal{X} to positive semidefinite operators and satisfies

$\sum_{x \in \mathcal{X}} M(x) = 1$, whereas $P_{\rho, M}$ is a measure on \mathcal{X} defined via the relation $P_{\rho, M}(x) = \text{tr}[M(x)\rho]$ for any $x \in \mathcal{X}$.) At first sight this definition seems cumbersome because we cannot restrict the size of \mathcal{X} that we optimize over. Therefore, following the works [11, 19, 34], let us also consider the following *projectively measured relative entropy*, which is defined as

$$D^{\mathbb{P}}(\rho \parallel \sigma) := \sup_{\{P_i\}_{i=1}^d} \left\{ \sum_{i=1}^d \text{tr}[P_i \rho] \log \frac{\text{tr}[P_i \rho]}{\text{tr}[P_i \sigma]} \right\}, \quad (3)$$

where the maximization is over all sets of mutually orthogonal projectors $\{P_i\}_{i=1}^d$ and we spelled out the Kullback-Leibler divergence for discrete measures. Note that without loss of generality we can assume that these projectors are rank-1 as any course graining of the measurement outcomes can only reduce the relative entropy due to its data-processing inequality [28, 48]. Furthermore, the quantity is finite and the supremum is achieved whenever $\rho \ll \sigma$, which here denotes that the support of ρ is contained in the support of σ . (To verify this, recall that the rank-1 projectors form a compact set and the divergence is lower semi-continuous.)

The first of the following two variational expressions for the (projectively) measured relative entropy is due to Petz [36]. Note that the second objective function is concave in ω so that the optimization problem has a particularly appealing form.

Lemma 1. *For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$ non-zero, the following identities hold:*

$$D^{\mathbb{P}}(\rho \parallel \sigma) = \sup_{\omega > 0} \text{tr}[\rho \log \omega] - \log \text{tr}[\sigma \omega] = \sup_{\omega > 0} \text{tr}[\rho \log \omega] + 1 - \text{tr}[\sigma \omega]. \quad (4)$$

Moreover, the suprema are achieved when ρ and σ are positive definite operators, i.e. $\rho > 0$ and $\sigma > 0$.

Proof. If $\rho \not\ll \sigma$ then the two expressions in the suprema of (4) are unbounded, as expected. We now assume that $\rho \ll \sigma$. Let us consider the second expression in (4). We write the supremum over $\omega > 0$ as two suprema over $\{P_i\}_{i=1}^d$ and $\{\omega_i\}_{i=1}^d$, where $\omega_i > 0$ are the eigenvalues of ω corresponding to the eigenvectors given by rank-1 projectors P_i . Using the fact that $\text{tr}[\rho] = 1$, we find

$$\sup_{\omega > 0} \text{tr}[\rho \log \omega] + 1 - \text{tr}[\sigma \omega] = \sup_{\{P_i\}_{i=1}^d} \sup_{\{\omega_i\}_{i=1}^d} \sum_{i=1}^d \text{tr}[P_i \rho] (\log \omega_i + 1) - \text{tr}[P_i \sigma] \omega_i. \quad (5)$$

For $i \in [d]$ such that $\text{tr}[P_i \sigma] = 0$, we also have $\text{tr}[P_i \rho] = 0$, and thus the corresponding term is zero. When $\text{tr}[P_i \sigma] > 0$, let us first consider $\text{tr}[P_i \rho] = 0$. In this case, the supremum of i -th term is $\sup_{\omega_i > 0} -\text{tr}[P_i \sigma] \omega_i = 0$ achieved in the limit $\omega_i \rightarrow 0$. Now in the case $\text{tr}[P_i \rho] > 0$ (which is the only possible case when $\rho > 0$), observe that the expression is concave, the inner supremum is achieved by the local maxima at $\omega_i = \text{tr}[P_i \rho] / \text{tr}[P_i \sigma]$. Plugging this into (5), we find

$$\sup_{\{P_i\}_{i=1}^d} \sum_{i=1}^d \text{tr}[P_i \rho] \log \frac{\text{tr}[P_i \rho]}{\text{tr}[P_i \sigma]} + \text{tr}[P_i \rho] - \text{tr}[P_i \rho] = \sup_{\{P_i\}_{i=1}^d} \sum_{i=1}^d \text{tr}[P_i \rho] \log \frac{\text{tr}[P_i \rho]}{\text{tr}[P_i \sigma]}. \quad (6)$$

This is the expression for the measured relative entropy in (3). The remaining supremum is achieved because the set of rank-1 projectors is compact and the divergence is lower semi-continuous.

It remains to show that the two variational expressions in (4) are equivalent. We have $\log(1+x) \leq x$ for all $x > -1$ and, thus, $-\log \text{tr}[\sigma \omega] \geq 1 - \text{tr}[\sigma \omega]$ for all $\omega > 0$. This yields

$$\sup_{\omega > 0} \text{tr}[\rho \log \omega] - \log \text{tr}[\sigma \omega] \geq \sup_{\omega > 0} \text{tr}[\rho \log \omega] + 1 - \text{tr}[\sigma \omega]. \quad (7)$$

Now note that the expression on the left hand side is invariant under the substitution $\omega \rightarrow \gamma\omega$ for $\gamma > 0$. Hence, as $\text{tr}[\sigma\omega] > 0$ for $\omega > 0$ and non-zero σ , we can add the normalization constraint $\text{tr}[\sigma\omega] = 1$ and we have

$$\sup_{\omega > 0} \text{tr}[\rho \log \omega] - \log \text{tr}[\sigma\omega] = \sup_{\substack{\omega > 0 \\ \text{tr}[\sigma\omega] = 1}} \text{tr}[\rho \log \omega] - \log \text{tr}[\sigma\omega] \quad (8)$$

$$\leq \sup_{\omega > 0} \text{tr}[\rho \log \omega] + 1 - \text{tr}[\sigma\omega], \quad (9)$$

where we used that $\log \text{tr}[\sigma\omega] = 1 - \text{tr}[\sigma\omega]$ when $\text{tr}[\sigma\omega] = 1$. \square

Using this variational expression, we are able to answer a question left open by Donald [11, Sec. 3] as well as Hiai and Petz [19, Sec. 1], namely whether the two definitions of measured relative entropy are equal.

Theorem 2. For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, we have $D^{\mathbb{P}}(\rho||\sigma) = D^{\mathbb{M}}(\rho||\sigma)$.

Proof. The direction ‘ \leq ’ holds trivially. Moreover, if $\rho \not\ll \sigma$, we can choose P_1 to be a rank-1 projector such that $\text{tr}[P_1\rho] > 0$ and $\text{tr}[P_1\sigma] = 0$ and thus $D^{\mathbb{P}}(\rho||\sigma) = D^{\mathbb{M}}(\rho||\sigma) = +\infty$.

It remains to show the direction ‘ \geq ’ when $\rho \ll \sigma$ holds. Let (\mathcal{X}, M) be a POVM. Recall that the distribution $P_{\rho, M}$ is defined by $P_{\rho, M}(x) = \text{tr}[M(x)\rho]$. Introducing $\bar{\mathcal{X}} = \{x \in \mathcal{X} : P_{\rho, M}(x)P_{\sigma, M}(x) > 0\}$, we can write

$$D(P_{\rho, M}||P_{\sigma, M}) = \sum_{x \in \bar{\mathcal{X}}} P_{\rho, M}(x) \log \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \quad (10)$$

$$= \text{tr} \left[\rho \sum_{x \in \bar{\mathcal{X}}} M(x) \log \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \right] \quad (11)$$

$$= \text{tr} \left[\rho \sum_{x \in \bar{\mathcal{X}}} \sqrt{M(x)} \log \left(\frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \cdot 1 \right) \sqrt{M(x)} \right]. \quad (12)$$

Now observe that for any $x \in \bar{\mathcal{X}}$, the spectrum of the operator $\log \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \cdot 1$ is included in $(0, \infty)$. As a result, we can apply the operator Jensen inequality for the function \log , which is operator concave on $(0, \infty)$ and get

$$D(P_{\rho, M}||P_{\sigma, M}) \leq \text{tr} \left[\rho \log \left(\sum_{x \in \bar{\mathcal{X}}} M(x) \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \right) \right]. \quad (13)$$

Now we simply choose

$$\omega = \sum_{x \in \bar{\mathcal{X}}} M(x) \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} \quad \text{such that} \quad \text{tr}[\sigma\omega] = \sum_{x \in \bar{\mathcal{X}}} P_{\sigma, M}(x) \frac{P_{\rho, M}(x)}{P_{\sigma, M}(x)} = \sum_{x \in \bar{\mathcal{X}}} P_{\rho, M}(x) = 1, \quad (14)$$

and which allows to further bound (13) by $\text{tr}[\rho \log \omega] + 1 - \text{tr}[\sigma\omega]$. Comparing this with the variational expression for the measured relative entropy in Lemma 1 yields the desired inequality. \square

Hence, the measured relative entropy, $D^{\mathbb{M}}(\rho||\sigma)$, achieves its maximum for projective rank-1 measurements and can be evaluated using the concave optimization problem in Lemma 1.

II. MEASURED RÉNYI RELATIVE ENTROPY

Here we extend the results of the previous section to Rényi divergence. Using the same notation as in the previous section, for $\alpha \in (1, \infty)$ we define the Rényi divergence [39] as

$$D_\alpha(P\|Q) := \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P(x) \left(\frac{P(x)}{Q(x)} \right)^{\alpha-1} \quad (15)$$

if $P \ll Q$ and as $+\infty$ if $P \not\ll Q$. For $\alpha \in (0, 1)$ we rewrite the sum as

$$\sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}. \quad (16)$$

Hence we see that absolute continuity $P \ll Q$ is not necessary to keep D_α finite for $\alpha < 1$. However, the Rényi divergence instead diverges to $+\infty$ when P and Q are orthogonal.¹ It is well known that the Rényi divergence converges to the Kullback-Leibler divergence when $\alpha \rightarrow 1$ and we thus set $D \equiv D_1$. Moreover, in the limit $\alpha \rightarrow \infty$ we find the max-divergence $D_\infty(P\|Q) = \sup_{x \in \mathcal{X}} \log \frac{P(x)}{Q(x)}$.

Let us now define the *measured Rényi relative entropy* as before, namely

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) := \sup_{(\mathcal{X}, M)} D_\alpha(P_{\rho, M}\|P_{\sigma, M}). \quad (17)$$

We will later show that this is equivalent to the following *projectively measured Rényi relative entropy*, which we define here for $\alpha \in (1, \infty)$ as

$$D_\alpha^{\mathbb{P}}(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q_\alpha^{\mathbb{P}}(\rho\|\sigma), \quad \text{with} \quad Q_\alpha^{\mathbb{P}}(\rho\|\sigma) := \sup_{\{P_i\}_{i=1}^d} \left\{ \sum_i \text{tr}[P_i \rho]^\alpha \text{tr}[P_i \sigma]^{1-\alpha} \right\}, \quad (18)$$

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

$$Q_\alpha^{\mathbb{P}}(\rho\|\sigma) := \min_{\{P_i\}_{i=1}^d} \left\{ \sum_i \text{tr}[P_i \rho]^\alpha \text{tr}[P_i \sigma]^{1-\alpha} \right\}. \quad (19)$$

Note that the supremum in (18) is achieved and $D_\alpha^{\mathbb{P}}(\rho\|\sigma)$ is finite whenever $\rho \ll \sigma$. Similarly, the minimum in (19) is non-zero and $D_\alpha^{\mathbb{P}}(\rho\|\sigma)$ is finite whenever $\rho \not\ll \sigma$, i.e. when the two states are not orthogonal.

Next we give variational expressions for $Q_\alpha^{\mathbb{P}}(\rho\|\sigma)$ similar to the variational characterization of the measured relative entropy in Lemma 1.

Lemma 3. *For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, the following identities hold:*

$$Q_\alpha^{\mathbb{P}}(\rho\|\sigma) = \begin{cases} \inf_{\omega > 0} \alpha \text{tr}[\rho \omega] + (1 - \alpha) \text{tr}[\sigma \omega^{\frac{\alpha}{\alpha-1}}] & \text{for } \alpha \in (0, \frac{1}{2}) \\ \inf_{\omega > 0} \alpha \text{tr}[\rho \omega^{1-\frac{1}{\alpha}}] + (1 - \alpha) \text{tr}[\sigma \omega] & \text{for } \alpha \in [\frac{1}{2}, 1) \\ \sup_{\omega > 0} \alpha \text{tr}[\rho \omega^{1-\frac{1}{\alpha}}] + (1 - \alpha) \text{tr}[\sigma \omega] & \text{for } \alpha \in (1, \infty) \end{cases} \quad (20)$$

$$= \begin{cases} \inf_{\omega > 0} \text{tr}[\rho \omega]^\alpha \text{tr}[\sigma \omega^{\frac{\alpha}{\alpha-1}}]^{1-\alpha} & \text{for } \alpha \in (0, 1) \\ \sup_{\omega > 0} \text{tr}[\rho \omega^{1-\frac{1}{\alpha}}]^\alpha \text{tr}[\sigma \omega]^{1-\alpha} & \text{for } \alpha \in (1, \infty) \end{cases}. \quad (21)$$

The infima and suprema are achieved when $\rho > 0$ and $\sigma > 0$.

¹ P and Q are orthogonal, denoted $P \perp Q$, if there exists an $A \subseteq \mathcal{X}$ such that $P(A) = 1$ and $Q(A) = 0$.

The expressions (21) can be seen as a generalization of Alberti's theorem [1] for the fidelity (which corresponds to $\alpha = 1/2$) to general $\alpha \in (0, 1) \cup (1, \infty)$.

Proof. We first show the identity (20). Let us discuss the case $\alpha \in (0, 1)$ in detail. Note that the two expressions given for $\alpha \in (0, \frac{1}{2})$ and $\alpha \in [\frac{1}{2}, 1)$ are equivalent by the transformation $\omega \mapsto \omega^{\frac{\alpha}{\alpha-1}}$ (the reason for the different ways of writing is to see that the expressions are convex in ω , which will be useful later, in particular in Theorem 4). We first write

$$\inf_{\omega > 0} \alpha \operatorname{tr} \left[\rho \omega^{1-\frac{1}{\alpha}} \right] + (1-\alpha) \operatorname{tr}[\sigma \omega] = \inf_{\{P_i\}_{i=1}^d} \inf_{\{\omega_i\}_{i=1}^d} \alpha \sum_i \omega_i^{1-\frac{1}{\alpha}} \operatorname{tr}[P_i \rho] + (1-\alpha) \sum_i \omega_i \operatorname{tr}[P_i \sigma]. \quad (22)$$

Let $i \in [d]$ be such that $\operatorname{tr}[P_i \rho]$ and $\operatorname{tr}[P_i \sigma]$ are both strictly positive (which is the case when $\rho > 0$ and $\sigma > 0$). Then a local (and thus global) minimum for ω_i is easily found at the point where

$$\alpha \left(1 - \frac{1}{\alpha}\right) \omega_i^{-\frac{1}{\alpha}} \operatorname{tr}[P_i \rho] + (1-\alpha) \operatorname{tr}[P_i \sigma] = 0 \quad \implies \quad \omega_i = \left(\frac{\operatorname{tr}[P_i \rho]}{\operatorname{tr}[P_i \sigma]} \right)^\alpha. \quad (23)$$

If both $\operatorname{tr}[P_i \rho] = \operatorname{tr}[P_i \sigma] = 0$ we can chose ω_i arbitrarily. If only $\operatorname{tr}[P_i \rho] = 0$ the infimum is achieved in the limit $\omega_i \rightarrow 0$, and if only $\operatorname{tr}[P_i \sigma] = 0$ in the limit $\omega_i \rightarrow \infty$. In all these cases the infimum of the i -th term is zero. Furthermore, it is achieved for a finite, non-zero ω_i when $\rho > 0$ and $\sigma > 0$. Plugging this solution into the above expression yields

$$\inf_{\{P_i\}_{i=1}^d} \sum_{i \in [d]} \operatorname{tr}[P_i \rho]^\alpha \operatorname{tr}[P_i \sigma]^{1-\alpha}. \quad (24)$$

This infimum is always achieved since the set we optimize over is compact. Comparing this with the definition of $D^{\mathbb{P}}(\rho||\sigma)$ yields the first equality.

For the case $\alpha \in (1, \infty)$, when $\rho \ll \sigma$, the proof is analogous to the previous argument. Otherwise, it is simple to see that the supremum is $+\infty$.

Now we show (21). Using (20) and the weighted arithmetic-geometric mean inequality we have

$$Q_\alpha^{\mathbb{P}}(\rho||\sigma) \geq \inf_{\omega > 0} \operatorname{tr} \left[\rho \omega^{1-\frac{1}{\alpha}} \right]^\alpha \operatorname{tr}[\sigma \omega]^{1-\alpha} \quad \text{for } \alpha \in (0, 1). \quad (25)$$

However, for any feasible $\omega > 0$ in (20) and $\lambda > 0$, $\lambda \omega > 0$ is also feasible and choosing $\lambda = \operatorname{tr} \left[\rho \omega^{1-\frac{1}{\alpha}} \right]^\alpha \operatorname{tr}[\sigma \omega]^{-\alpha}$ shows that (20) cannot exceed (25). Similarly, by Bernoulli's inequality,

$$Q_\alpha^{\mathbb{P}}(\rho||\sigma) \leq \sup_{\omega > 0} \operatorname{tr} \left[\rho \omega^{1-\frac{1}{\alpha}} \right]^\alpha \operatorname{tr}[\sigma \omega]^{1-\alpha} \quad \text{for } \alpha \in (1, \infty). \quad (26)$$

And the same argument as above yields the equality. \square

As for the measured relative entropy, the restriction to rank-1 projective measurements is in fact not restrictive at all.

Theorem 4. For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, we have $D_\alpha^{\mathbb{P}}(\rho||\sigma) = D_\alpha^{\mathbb{M}}(\rho||\sigma)$.

Proof. For $\alpha > 1$ we follow the steps of the proof of Theorem 2. Consider any finite set \mathcal{X} and POVM M with induced measures $P_{\rho, M}$ and $P_{\sigma, M}$. We can write

$$D_\alpha(P_{\rho, M}||P_{\sigma, M}) = \frac{1}{1-\alpha} \log \sum_{x \in \bar{\mathcal{X}}} P_{\rho, M}(x)^\alpha P_{\sigma, M}(x)^{1-\alpha}, \quad (27)$$

where we can restrict the sum over $\bar{\mathcal{X}} = \{x \in \mathcal{X} : P_{\rho,M}(x)P_{\sigma,M}(x) > 0\}$. We then find that the sum satisfies

$$\sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1} \leq \text{tr} \left[\rho \left(\sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha} \right)^{1-\frac{1}{\alpha}} \right], \quad (28)$$

where the inequality again follows by the operator Jensen inequality and the operator concavity of the function $t \mapsto t^{1-\frac{1}{\alpha}}$ on $[0, \infty)$. Now we set

$$\omega = \sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha} \quad \text{such that} \quad \text{tr}[\sigma\omega] = \sum_{x \in \mathcal{X}} P_{\rho,M}(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1}. \quad (29)$$

Thus, we can bound

$$\sum_{x \in \mathcal{X}} P_{\rho,M}(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1} \leq \alpha \text{tr}[\rho\omega^{1-\frac{1}{\alpha}}] + (1-\alpha) \text{tr}[\sigma\omega]. \quad (30)$$

Comparing this with the variational expression in Lemma 3 yields the desired inequality.

For $\alpha < 1$, we use the same notation as in (27). We further distinguish the cases $\alpha \in (0, \frac{1}{2})$ and $\alpha \in [\frac{1}{2}, 1)$. For $\alpha \in (0, \frac{1}{2})$, we define

$$\omega = \sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1}. \quad (31)$$

We can then evaluate

$$\text{tr} \left[\sigma\omega^{\frac{\alpha}{\alpha-1}} \right] = \text{tr} \left[\sigma \left(\sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1} \right)^{\frac{\alpha}{\alpha-1}} \right] \quad (32)$$

$$\leq \text{tr} \left[\sigma \sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha} \right] = \sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x)^{\alpha} P_{\sigma,M}(x)^{1-\alpha}. \quad (33)$$

where we used the operator convexity of $t \mapsto t^{\frac{\alpha}{\alpha-1}}$ on $(0, \infty)$ and the operator Jensen inequality. Moreover,

$$\text{tr}[\rho\omega] = \sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x)^{\alpha} P_{\sigma,M}(x)^{1-\alpha}. \quad (34)$$

As a result

$$\sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x)^{\alpha} P_{\sigma,M}(x)^{1-\alpha} \geq \alpha \text{tr}[\rho\omega] + (1-\alpha) \text{tr}[\sigma\omega^{\frac{\alpha}{\alpha-1}}]. \quad (35)$$

Comparing this with the variational expression in Lemma 3 yields the desired inequality.

For $\alpha \in [\frac{1}{2}, 1)$ we choose $\omega = \sum_{x \in \bar{\mathcal{X}}} M(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha}$, so that

$$\text{tr} \left[\rho\omega^{1-\frac{1}{\alpha}} \right] \leq \sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x) \left(\frac{P_{\rho,M}(x)}{P_{\sigma,M}(x)} \right)^{\alpha-1} \quad (36)$$

$$\text{tr}[\sigma\omega] = \sum_{x \in \bar{\mathcal{X}}} P_{\rho,M}(x)^{\alpha} P_{\sigma,M}(x)^{1-\alpha}, \quad (37)$$

and once again conclude using the variational expression in Lemma 3. \square

III. ACHIEVABILITY OF RELATIVE ENTROPY

A. Umegaki's Relative Entropy

Here we compare the measured relative entropy to other notions of quantum relative entropy that have been investigated in the literature and have found operational significance in quantum information theory. Umegaki's quantum relative entropy [49] has found operational significance as the threshold rate for asymmetric binary quantum hypothesis testing [19]. For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, it is defined as

$$D(\rho\|\sigma) := \text{tr}[\rho(\log \rho - \log \sigma)] \quad \text{if } \sigma \gg \rho \text{ and as } +\infty \text{ if } \sigma \not\gg \rho. \quad (38)$$

We recall the following variational expression by Petz [35] (see also [24] for another variational expression):

$$D(\rho\|\sigma) = \sup_{\omega > 0} \text{tr}[\rho \log \omega] + \log \text{tr}[\exp(\log \sigma + \log \omega)] \quad (39)$$

$$= \sup_{\omega > 0} \text{tr}[\rho \log \omega] + 1 - \text{tr}[\exp(\log \sigma + \log \omega)]. \quad (40)$$

By the data-processing inequality for the relative entropy [28, 48] we always have

$$D^{\mathbb{P}}(\rho\|\sigma) \leq D^{\mathbb{M}}(\rho\|\sigma) \leq D(\rho\|\sigma), \quad (41)$$

and moreover Petz [34] showed the inequality $D^{\mathbb{P}}(\rho\|\sigma) \leq D(\rho\|\sigma)$ is strict if ρ and σ do not commute (for $\rho > 0$ and $\sigma > 0$). Theorem 2 strengthens this to show that the strict inequality persists even when we take a supremum over POVMs. In the following we give an alternative proof of Petz' result and then extend the argument to Rényi relative entropy in Section III B.

Proposition 5. *Let $\rho \in \mathcal{S}$ with $\rho > 0$ and $\sigma \in \mathcal{P}$ with $\sigma > 0$. Then, we have*

$$D^{\mathbb{M}}(\rho\|\sigma) \leq D(\rho\|\sigma) \quad \text{with equality if and only if } [\rho, \sigma] = 0. \quad (42)$$

Our proof relies on the Golden–Thompson inequality [17, 45]. It states that for two Hermitian matrices X and Y , it holds that

$$\text{tr}[\exp(X + Y)] \leq \text{tr}[\exp(X) \exp(Y)] \quad (43)$$

with equality if and only if $[X, Y] = 0$ as shown in [42].

Proof. First, it is evident that equality holds if $[\rho, \sigma] = 0$ since then there exists a projective measurement that commutes with ρ and σ and thus does not effect the states. For the following, it is worth writing the variational expressions for the two quantities side by side. Namely, writing $H = \log \omega$, we have

$$D(\rho\|\sigma) = 1 + \sup_H \text{tr}[\rho H] - \text{tr}[\exp(\log \sigma + H)] \quad (44)$$

$$D^{\mathbb{M}}(\rho\|\sigma) = 1 + \max_H \text{tr}[\rho H] - \text{tr}[\sigma \exp(H)], \quad (45)$$

where we optimize over all Hermitian operators H . Note that, according to Lemma 1, we can write a max for (45) because we are assuming $\rho > 0$ and $\sigma > 0$. The inequality in (42) can now be seen as a direct consequence of the Golden–Thompson inequality.

It remains to show that $D^{\mathbb{M}}(\rho\|\sigma) = D(\rho\|\sigma)$ implies $[\rho, \sigma] = 0$. Let H^* be any maximizer of the variational problem in (45). Observe now that the equality $D^{\mathbb{M}}(\rho\|\sigma) = D(\rho\|\sigma)$ necessitates

$$\mathrm{tr}[\exp(\log \sigma + H^*)] = \mathrm{tr}[\sigma \exp(H^*)], \quad (46)$$

which holds only if and only if $[\sigma, H^*] = 0$ by the equality condition in (43). Now define the function

$$f(H) = \mathrm{tr}[\rho H] - \mathrm{tr}[\sigma \exp(H)], \quad (47)$$

and since H^* maximizes $f(H)$, we must have for all Hermitian Δ ,

$$0 = Df(H^*)[\Delta] = \mathrm{tr}[\rho \Delta] - \mathrm{tr}[\sigma \exp(H^*) \Delta]. \quad (48)$$

To evaluate the second summand of this Fréchet derivative we used that σ and H^* commute. Since this holds for all Δ we must in fact have $\rho = \sigma \exp(H^*)$, which means that $[\rho, \sigma] = 0$, as desired. \square

In some sense this result tells us that some quantum correlations, as measured by the relative entropy, do not survive the measurement process. This fact appears in quantum information theory in various different guises, for example in the form of locking classical correlations in quantum states [10]. (We also point to [37] for the use of measured relative entropy measures in quantum information theory.) Moreover, since Umegaki's relative entropy is the smallest quantum generalization of the Kullback-Leibler divergence that is both additive and satisfies data-processing (see, e.g., [46, Sec. 4.2.2]), the same conclusion can be drawn for any sensible quantum relative entropy. (An example being the quantum relative entropy introduced by Belavkin and Staszewski [4].)

B. Sandwiched Rényi Relative Entropy

Next we consider a family of quantum Rényi relative entropies [32, 54] that are commonly called sandwiched Rényi relative entropies and have found operational significance since they determine the strong converse exponent in asymmetric binary quantum hypothesis testing [30]. They are of particular interest here because they are, for $\alpha \geq \frac{1}{2}$, the smallest quantum generalization of the Rényi divergence that is both additive and satisfies data-processing [46, Sec. 4.2.2]. (Examples for other sensible quantum generalizations are the quantum Rényi relative entropy first studied by Petz [33] and the quantum divergences introduced by Matsumoto [29].)

For $\rho \in \mathcal{S}$ and $\sigma \in \mathcal{P}$, the sandwiched Rényi relative entropy of order $\alpha \in (0, 1) \cup (1, \infty)$ is defined as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q_\alpha(\rho\|\sigma) \quad \text{with} \quad Q_\alpha(\rho\|\sigma) := \mathrm{tr} \left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right], \quad (49)$$

where the same considerations about finiteness as for the measured Rényi relative entropy apply. We also consider the limits $\alpha \rightarrow \infty$ and $\alpha \rightarrow 1$ of the above expression for which we have [32],

$$D_\infty(\rho\|\sigma) = \inf \{ \lambda \in \mathbb{R} | \rho \leq \exp(\lambda) \sigma \} \quad \text{and} \quad D_1(\rho\|\sigma) = D(\rho\|\sigma), \quad (50)$$

respectively. We recall the following variational expression by Frank and Lieb [15]:

$$Q_\alpha(\rho\|\sigma) = \begin{cases} \inf_{\omega > 0} \alpha \mathrm{tr}[\rho \omega] + (1 - \alpha) \mathrm{tr} \left[\left(\omega^{\frac{1}{2}} \sigma^{\frac{\alpha-1}{\alpha}} \omega^{\frac{1}{2}} \right)^{\frac{\alpha}{\alpha-1}} \right] & \text{for } \alpha \in (0, 1) \\ \sup_{\omega > 0} \alpha \mathrm{tr}[\rho \omega] + (1 - \alpha) \mathrm{tr} \left[\left(\omega^{\frac{1}{2}} \sigma^{\frac{\alpha-1}{\alpha}} \omega^{\frac{1}{2}} \right)^{\frac{\alpha}{\alpha-1}} \right] & \text{for } \alpha \in (1, \infty). \end{cases} \quad (51)$$

Alternatively, we can also write

$$Q_\alpha(\rho\|\sigma) = \begin{cases} \inf_{\omega>0} \operatorname{tr}[\rho\omega]^\alpha \operatorname{tr} \left[\left(\omega^{\frac{1}{2}} \sigma^{\frac{\alpha-1}{\alpha}} \omega^{\frac{1}{2}} \right)^{\frac{\alpha}{\alpha-1}} \right]^{1-\alpha} & \text{for } \alpha \in (0, 1) \\ \sup_{\omega>0} \operatorname{tr}[\rho\omega]^\alpha \operatorname{tr} \left[\left(\omega^{\frac{1}{2}} \sigma^{\frac{\alpha-1}{\alpha}} \omega^{\frac{1}{2}} \right)^{\frac{\alpha}{\alpha-1}} \right]^{1-\alpha} & \text{for } \alpha \in (1, \infty), \end{cases} \quad (52)$$

where we have used the same arguments as in the proof of the second part of Lemma 3. By the data-processing inequality for the sandwiched Rényi relative entropy [3, 15, 32] we always have

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma) \quad \text{for } \alpha \geq \frac{1}{2}. \quad (53)$$

In the following we give an alternative proof of this fact and show that

$$[\rho, \sigma] \neq 0 \quad \implies \quad D_\alpha^{\mathbb{M}}(\rho\|\sigma) < D_\alpha(\rho\|\sigma) \quad \text{for } \alpha \in \left(\frac{1}{2}, \infty \right). \quad (54)$$

In contrast, at the boundaries $\alpha \in \{\frac{1}{2}, \infty\}$ it is known that $D_\alpha(\rho\|\sigma) = D_\alpha^{\mathbb{M}}(\rho\|\sigma)$ [16, 23, 30]. (We refer to [30, App. A] for a detailed discussion.)

Theorem 6. *Let $\rho \in \mathcal{S}$ with $\rho > 0$ and $\sigma \in \mathcal{P}$ with $\sigma > 0$. For $\alpha \in (\frac{1}{2}, \infty)$, we have*

$$D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma) \quad \text{with equality if and only if } [\rho, \sigma] = 0. \quad (55)$$

The argument is similar to the proof of Proposition 5 but with the Golden–Thompson inequality replaced by the Araki–Lieb–Thirring inequality [2, 27]. It states that for $X, Y \geq 0$ we have

$$\operatorname{tr}[YXY]^r \leq \operatorname{tr}[Y^r X^r Y^r] \quad \text{for } r \geq 1 \quad (56)$$

$$\operatorname{tr}[YXY]^r \geq \operatorname{tr}[Y^r X^r Y^r] \quad \text{for } r \in [0, 1], \quad (57)$$

with equality if and only if $[X, Y] = 0$ except for $r = 1$ as shown in [18].

Proof. We give the proof for $\alpha \in (\frac{1}{2}, 1)$ and note that the argument for $\alpha \in (1, \infty)$ is analogous. We have the following variational expressions from Lemma 3 and (51):

$$Q_\alpha(\rho\|\sigma) = \inf_{\omega>0} \alpha \operatorname{tr}[\rho\omega] + (1 - \alpha) \operatorname{tr} \left[\left(\omega^{-\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \omega^{-\frac{1}{2}} \right)^{\frac{\alpha}{1-\alpha}} \right] \quad (58)$$

$$Q_\alpha^{\mathbb{P}}(\rho\|\sigma) = \min_{\omega>0} \alpha \operatorname{tr}[\rho\omega] + (1 - \alpha) \operatorname{tr} \left[\sigma \omega^{\frac{\alpha}{\alpha-1}} \right], \quad (59)$$

where the existence of the minima relies on the fact that both operators have full support. (Note also that these two expressions are in fact equivalent for $\alpha = \frac{1}{2}$.) Since $\frac{\alpha}{1-\alpha} \geq 1$, the inequality then follows immediately by the Araki–Lieb–Thirring inequality (56):

$$Q_\alpha^{\mathbb{P}}(\rho\|\sigma) \leq Q_\alpha(\rho\|\sigma) \quad \implies \quad D_\alpha^{\mathbb{M}}(\rho\|\sigma) \leq D_\alpha(\rho\|\sigma). \quad (60)$$

Furthermore, if $[\rho, \sigma] = 0$ we have equality. To show that $Q_\alpha^{\mathbb{P}}(\rho\|\sigma) = Q_\alpha(\rho\|\sigma)$ implies $[\rho, \sigma] = 0$, we define the function

$$f_\alpha(\omega) = \alpha \operatorname{tr}[\rho\omega] + (1 - \alpha) \operatorname{tr} \left[\sigma \omega^{\frac{\alpha}{\alpha-1}} \right]. \quad (61)$$

For ω_α^* any minimizer of the variational problem in (59), we have

$$0 = Df_\alpha(\omega_\alpha^*)[\Delta] = \alpha \operatorname{tr}[\rho\Delta] - \alpha \operatorname{tr} \left[\sigma (\omega_\alpha^*)^{\frac{1}{\alpha-1}} \Delta \right], \quad (62)$$

for all Hermitian Δ . To evaluate the second summand of this Fréchet derivative we used that σ and ω_α^* commute, which holds by the equality condition for Araki–Lieb–Thirring. We thus conclude that $\rho = \sigma (\omega_\alpha^*)^{\frac{1}{\alpha-1}}$ which implies that $[\rho, \sigma] = 0$. \square

IV. VIOLATION OF DATA-PROCESSING FOR $\alpha < \frac{1}{2}$

As a further application of the variational characterization of measured Rényi relative entropy, we can show that the data-processing for the sandwiched Rényi relative entropy D_α fails for $\alpha < \frac{1}{2}$. (Numerical evidence pointed to the fact that data-processing does not hold in this regime [31].)

Theorem 7. *Let $\rho \in \mathcal{S}$ with $\rho > 0$ and $\sigma \in \mathcal{P}$ with $\sigma > 0$, and $[\rho, \sigma] \neq 0$. For $\alpha \in (0, \frac{1}{2})$, we have $D_\alpha^{\mathbb{P}}(\rho\|\sigma) > D_\alpha(\rho\|\sigma)$.*

In particular, there exists a rank-1 measurement that achieves $D_\alpha^{\mathbb{P}}(\rho\|\sigma)$ and thus violates the data-processing inequality.

Proof. First note that $[\rho, \sigma] \neq 0$ implies that the two states are not orthogonal and thus both quantities are finite. For $\alpha \in (\frac{1}{2}, 1)$ the formulas (58) and (59) still hold. However, in contrast to the proof of Theorem 6 we have $\frac{\alpha}{1-\alpha} \in [0, 1]$. Hence, we find by the Araki–Lieb–Thirring inequality (57) that

$$Q_\alpha^{\mathbb{P}}(\rho\|\sigma) \leq Q_\alpha(\rho\|\sigma) \quad \implies \quad D_\alpha^{\mathbb{P}}(\rho\|\sigma) \geq D_\alpha(\rho\|\sigma). \quad (63)$$

As in the proof of Theorem 6 we have equality if and only if $[\rho, \sigma] = 0$. This implies the claim. \square

V. EXPLOITING VARIATIONAL FORMULAS

A. Some Optimization Problems in Quantum Information

The variational characterizations of the relative entropy (39)–(40), the sandwiched Rényi relative entropy (51)–(52), and their measured counterparts (Lemma 1 and Lemma 3), can be used to derive properties of various entropic quantities that appear in quantum information theory. We are interested in operational quantities of the form

$$\mathcal{M}(\rho) := \min_{\sigma \in \mathcal{C}} \mathbb{D}(\rho\|\sigma), \quad (64)$$

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 [52].
- positive partial transpose states, leading to the Rains bound on entanglement distillation [38].
- non-distillable states, leading to bounds on entanglement distillation [51].
- quantum Markov states, leading to insights about the robustness properties of these states [21].
- locally recoverable states, leading to bounds on the quantum conditional mutual information [8, 14, 40].
- k -extendible states, leading to bounds on squashed entanglement [26, 55].

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., [47].

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

$$\mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_1 \otimes \sigma_2) \geq \mathbb{D}(\rho_1 \| \sigma_1) + \mathbb{D}(\rho_2 \| \sigma_2). \quad (65)$$

We might then ask if we also have

$$\min_{\sigma_{12} \in \mathcal{C}_{12}} \mathbb{D}(\rho_1 \otimes \rho_2 \| \sigma_{12}) = \mathcal{M}(\rho_1 \otimes \rho_2) \stackrel{?}{\geq} \mathcal{M}(\rho_1) + \mathcal{M}(\rho_2) = \min_{\sigma_1 \in \mathcal{C}_1} \mathbb{D}(\rho_1 \| \sigma_1) + \min_{\sigma_2 \in \mathcal{C}_2} \mathbb{D}(\rho_2 \| \sigma_2). \quad (66)$$

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

$$\mathbb{D}(\rho \| \sigma) = \sup_{\omega > 0} f(\rho, \sigma, \omega) \quad \text{in order to write} \quad \mathcal{M}(\rho) = \min_{\sigma \in \mathcal{C}} \sup_{\omega > 0} f(\rho, \sigma, \omega) = \sup_{\omega > 0} \min_{\sigma \in \mathcal{C}} f(\rho, \sigma, \omega), \quad (67)$$

where we made use of Sion's minimax theorem [41] 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. (As an example, for the measured relative entropies the objective function becomes linear in σ .) We can then use strong duality of convex optimization to rewrite this minimization as a maximization problem [7]:

$$\min_{\sigma \in \mathcal{C}} f(\rho, \sigma, \omega) = \max_{\bar{\sigma} \in \bar{\mathcal{C}}} \bar{f}(\rho, \bar{\sigma}, \omega). \quad (68)$$

This leads to the expression

$$\mathcal{M}(\rho) = \sup_{\omega > 0} \max_{\bar{\sigma} \in \bar{\mathcal{C}}} \bar{f}(\rho, \bar{\sigma}, \omega), \quad (69)$$

which, in contrast to the definition of $\mathcal{M}(\rho)$ in (64), 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 (66). 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

$$\bar{f}(\rho_1 \otimes \rho_2, \bar{\sigma}_1 \otimes \bar{\sigma}_2, \omega_1 \otimes \omega_2) \geq \bar{f}(\rho_1, \bar{\sigma}_1, \omega_1) + \bar{f}(\rho_2, \bar{\sigma}_2, \omega_2). \quad (70)$$

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

$$\mathcal{M}(\rho_1 \otimes \rho_2) \geq \bar{f}(\rho_1 \otimes \rho_2, \bar{\sigma}_1 \otimes \bar{\sigma}_2, \omega_1 \otimes \omega_2) \geq \bar{f}(\rho_1, \bar{\sigma}_1, \omega_1) + \bar{f}(\rho_2, \bar{\sigma}_2, \omega_2) \quad (71)$$

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 inequality also holds true if we maximize over these variables, implying superadditivity.

B. 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 [8, 20, 40]

$$D^{\text{rec}}(\rho_{AD} \| \sigma_{AE}) := \inf_{\Gamma_{E \rightarrow D}} D(\rho_{AD} \| (\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE})), \quad (72)$$

where the infimum is taken over all trace-preserving completely positive maps $\Gamma_{E \rightarrow D}$. (We restrict to $\sigma_A > 0$ such that the quantity is surely finite and the infimum becomes a minimum.) 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 [6, 8, 14, 22, 26, 40, 43, 44, 53]. In particular, [8, Prop. 3] shows that

$$I(A : C|B) \geq \lim_{n \rightarrow \infty} \frac{1}{n} D^{\text{rec}}(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n}), \quad (73)$$

where the systems are understood as $D = BC$ and $E = B$. To obtain a lower bound that does not involve limits, the authors of [8] use a Finetti-type theorem to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} D^{\text{rec}}(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n}) \geq D^{\text{M,rec}}(\rho_{ABC} \| \rho_{AB}). \quad (74)$$

with the *measured relative entropy of recovery* defined as [8] (see also [40, Rmk. 6]),

$$D^{\text{M,rec}}(\rho_{AD} \| \sigma_{AE}) := \inf_{\Gamma_{E \rightarrow D}} D^{\text{M}}(\rho_{AD} \| (\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE})). \quad (75)$$

(We restrict to $\sigma_A > 0$ such that the quantity is surely finite and the infimum becomes a minimum.) 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 $D^{\text{M,rec}}$ we would get an alternative proof for the step (74). Using the variational expression for the measured relative entropy (Lemma 1) together with strong duality for semidefinite programming, we find the following alternative characterization for $D^{\text{M,rec}}$.

Lemma 8. *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*

$$\begin{aligned} D^{\text{M,rec}}(\rho_{AD} \| \sigma_{AE}) = \text{maximize} \quad & \text{tr}[\rho_{AD} \log R_{AD}] \\ \text{subject to} \quad & S_{AF} > 0, R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \\ & \text{tr}[S_{AF} \sigma_{AF}] = 1. \end{aligned} \quad (76)$$

Proof. We write

$$\Gamma_{E \rightarrow D}(\sigma_{AE}) = \text{tr}_F [\Gamma_{E \rightarrow D}(\sigma_{AEF})] = \text{tr}_F \left[\Gamma_{E \rightarrow D}(\sqrt{\sigma_{AE}} \Psi_{AE:F} \sqrt{\sigma_{AE}}) \right], \quad (77)$$

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 $\Gamma_{E \rightarrow D}$ as

$$\tau_{ADF} = \Gamma_{E \rightarrow D}(\Psi_{AE:F}), \quad \tau_{AF} = \Pi_{AF}^\sigma, \quad (78)$$

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

$$\Gamma_{E \rightarrow D}(\sigma_{AE}) = \text{tr}_F \left[\sqrt{\sigma_{AF}} \tau_{ADF} \sqrt{\sigma_{AF}} \right], \quad (79)$$

and thus we can express the optimization problem for $\Gamma_{E \rightarrow D}$ in terms of the Choi-Jamiołkowski state in (78). Together with the variational characterization of the measured relative entropy (Lemma 1) we find

$$D^{\text{M,rec}}(\rho_{AD} \| \sigma_{AE}) = \min_{\tau_{ADF} \in \text{Rec}(\sigma_{AE})} \sup_{R_{AD} > 0} \text{tr}[\rho_{AD} \log R_{AD}] + 1 - \text{tr}[\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}], \quad (80)$$

where $\text{Rec}(\sigma_{AE}) := \{\tau_{ADF} \geq 0, \tau_{AF} = \Pi_{AF}^\sigma\}$. We now apply Sion's minimax theorem [41] to exchange the minimum with the supremum. The theorem is applicable as $\text{Rec}(\sigma_{AE})$ is convex and compact and $\{\omega_{AD} > 0\}$ is convex. Moreover, as the logarithm is operator concave, the function

$$R_{AD} \mapsto \text{tr}[\rho_{AD} \log R_{AD}] + 1 - \text{tr}[\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}] \quad (81)$$

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,

$$D^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) = \sup_{R_{AD} > 0} \text{tr}[\rho_{AD} \log R_{AD}] + 1 - \max_{\tau_{ADF} \in \text{Rec}(\sigma_{AE})} \text{tr}[\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}]. \quad (82)$$

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

$$\begin{array}{ll} \text{maximize:} & \text{tr}[\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}] \\ \text{subject to:} & \tau_{ADF} \geq 0 \\ & \tau_{AF} = \Pi_{AF}^\sigma \end{array} \quad \begin{array}{ll} \text{minimize:} & \text{tr}[S_{AF} \sigma_{AF}] \\ \text{subject to:} & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F. \end{array} \quad (83)$$

As the primal problem is strictly feasible, Slater's condition (see, e.g., [7]) is satisfied and we have strong duality. This leads to the expression:

$$\begin{aligned} D^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) = \text{maximize} & \quad \text{tr}[\rho_{AD} \log R_{AD}] + 1 - \text{tr}[S_{AF} \sigma_{AF}] \\ \text{subject to} & \quad S_{AF} > 0, R_{AD} > 0 \\ & \quad 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F. \end{aligned} \quad (84)$$

Observe now that we can restrict the optimization to $\text{tr}[S_{AF} \sigma_{AF}] = 1$. In fact, for arbitrary feasible R_{AD} and S_{AF} , we can define

$$\tilde{R}_{AD} = \frac{R_{AD}}{\text{tr}[S_{AF} \sigma_{AF}]} \quad \text{and} \quad \tilde{S}_{AF} = \frac{S_{AF}}{\text{tr}[S_{AF} \sigma_{AF}]} \quad (85)$$

The constraint $1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F$ is still satisfied and the value of the objective function can only increase,

$$\text{tr}[\rho_{AD} \log \tilde{R}_{AD}] = \text{tr}[\rho_{AD} \log R_{AD}] - \text{tr}[\rho_{AD}] \log \text{tr}[S_{AF} \sigma_{AF}] \quad (86)$$

$$\geq \text{tr}[\rho_{AD} \log R_{AD}] - \text{tr}[S_{AF} \sigma_{AF}] + 1, \quad (87)$$

where we used the inequality $\log x \leq x - 1$ and that $\text{tr}[\rho_{AD}] = 1$. This concludes the proof. \square

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

Proposition 9. *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, we have*

$$D^{\text{M,rec}}(\rho_{AD} \otimes \tau_{A'D'} \|\sigma_{AE} \otimes \omega_{A'E'}) \geq D^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) + D^{\text{M,rec}}(\tau_{A'D'} \|\omega_{A'E'}). \quad (88)$$

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

$$\begin{aligned} 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \quad \wedge \quad 1_{D'} \otimes S_{A'F'} \geq R_{A'D'} \otimes 1_{F'} \\ \implies \\ 1_{DD'} \otimes S_{AF} \otimes S_{A'F'} \geq R_{AD} \otimes R_{A'D'} \otimes 1_{FF'}. \end{aligned} \quad (89)$$

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 $\text{tr}[(S_{AF} \otimes S_{A'F'})(\sigma_{AF} \otimes \sigma_{A'F'})] = 1$. In other words, $(R_{AD} \otimes R_{A'D'}, S_{AF} \otimes S_{A'F'})$ is a feasible pair in the expression (76) for $D^{\text{M,rec}}(\rho_{AD} \otimes \tau_{A'D'} \| \sigma_{AE} \otimes \omega_{A'E'})$ and we get

$$D^{\text{M,rec}}(\rho_{AD} \otimes \tau_{A'D'} \| \sigma_{AE} \otimes \omega_{A'E'}) \geq \text{tr}[(\rho_{AD} \otimes \tau_{A'D'}) \log(R_{AD} \otimes R_{A'D'})] \quad (90)$$

$$= \text{tr}[\rho_{AD} \log R_{AD}] + \text{tr}[\tau_{A'D'} \log R_{A'D'}]. \quad (91)$$

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

Now, if the relative entropy of recovery would be superadditive we could strengthen (74) to

$$\lim_{n \rightarrow \infty} \frac{1}{n} D^{\text{rec}}(\rho_{ABC}^{\otimes n} \| \rho_{AB}^{\otimes n}) \stackrel{?}{\geq} D^{\text{rec}}(\rho_{ABC} \| \rho_{AB}). \quad (92)$$

This together with (73) 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 8.

Lemma 10. *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*

$$\begin{aligned} D^{\text{rec}}(\rho_{AD} \| \sigma_{AE}) = \text{maximize} \quad & \text{tr}[\rho_{AD} \log \rho_{AD}] - D^{\text{M}}(\rho_{AD} \| R_{AD}) \\ \text{subject to} \quad & S_{AF} > 0, R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \\ & \text{tr}[S_{AF} \sigma_{AF}] = 1, \end{aligned} \quad (93)$$

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

We provide a proof in Appendix A. This is to be compared to our expression for the measured relative entropy of recovery from Lemma 8, which can be written as

$$\begin{aligned} D^{\text{M,rec}}(\rho_{AD} \| \sigma_{AE}) = \text{maximize} \quad & \text{tr}[\rho_{AD} \log \rho_{AD}] - D(\rho_{AD} \| R_{AD}) \\ \text{subject to} \quad & S_{AF} > 0, R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F \\ & \text{tr}[S_{AF} \sigma_{AF}] = 1, \end{aligned} \quad (94)$$

where R_{AD} is not normalized and the relative entropy term is in general negative. Unfortunately, we can not solve the open additivity question for the relative entropy of recovery using the variational expression from Lemma 10. However, we note that the argument flipped relative entropy of recovery

$$\bar{D}^{\text{rec}}(\sigma_{AE} \| \rho_{AD}) := \inf_{\Gamma_{E \rightarrow D}} D((\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE}) \| \rho_{AD}) \quad (95)$$

becomes additive (restricted to $\rho_A > 0$ to assure finiteness and making the infimum a minimum).

Proposition 11. *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, we have*

$$\begin{aligned} \bar{D}^{\text{rec}}(\sigma_{AE} \| \rho_{AD}) = \text{maximize} \quad & -\text{tr}[S_{AF} \sigma_{AF}] \\ \text{subject to} \quad & S_{AF} > 0, R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq (\log \rho_{AD} - \log R_{AD}) \otimes 1_F \\ & \text{tr}[R_{AD}] = 1. \end{aligned} \quad (96)$$

Moreover, for $\tau_{A'D'} \in \mathcal{S}(A'D')$ with $\tau_{A'} > 0$, we have

$$\bar{D}^{\text{rec}}(\sigma_{AE} \otimes \omega_{A'E'} \| \rho_{AD} \otimes \tau_{A'D'}) = \bar{D}^{\text{rec}}(\sigma_{AE} \| \rho_{AD}) + \bar{D}^{\text{rec}}(\omega_{A'E'} \| \tau_{A'D'}). \quad (97)$$

We provide a proof based on strong duality for convex programming in Appendix A.

C. Rényi Relative Entropy of Recovery

As a generalization of the measured relative entropy of recovery (75) we define the *measured Rényi relative entropy of recovery* for $\alpha \in (1, \infty)$ as (see also [40, Rmk. 6]),

$$D_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}) := \frac{1}{\alpha - 1} \log Q_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}), \quad (98)$$

$$\text{with } Q_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}) := \inf_{\Gamma_{E \rightarrow D}} Q_{\alpha}^{\text{M}}(\rho_{AD} \parallel (\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE})), \quad (99)$$

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

$$Q_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}) := \sup_{\Gamma_{E \rightarrow D}} Q_{\alpha}^{\text{M}}(\rho_{AD} \parallel (\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE})), \quad (100)$$

where we restrict to $\sigma_A > 0$ such that the quantity is surely finite and the infimum/supremum is achieved. Using the variational expression for the measured Rényi relative entropy (Lemma 3) together with strong duality for semidefinite programming, we find the following alternative characterization for $Q_{\alpha}^{\text{M,rec}}$.

Lemma 12. *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 $\alpha \in [\frac{1}{2}, 1)$, we have*

$$\begin{aligned} Q_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}) = \text{minimize} \quad & \text{tr} \left[\rho_{AD} R_{AD}^{1-\frac{1}{\alpha}} \right]^{\alpha} \text{tr} [S_{AF} \sigma_{AF}]^{1-\alpha} \\ \text{subject to} \quad & S_{AF} > 0, R_{AD} > 0 \\ & 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F. \end{aligned} \quad (101)$$

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

Proof. As in the proof of Lemma 8 and using the first variational formula from Lemma 3, we get from Sion's minimax theorem [41] that

$$Q_{\alpha}^{\text{M,rec}}(\rho_{AD} \parallel \sigma_{AE}) = \max_{\tau_{ADF} \in \text{Rec}(\sigma_{AE})} \inf_{R_{AD} > 0} \alpha \text{tr} \left[\rho_{AD} R_{AD}^{1-\frac{1}{\alpha}} \right] + (1 - \alpha) \text{tr} [\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}] \quad (102)$$

$$= \inf_{R_{AD} > 0} \max_{\gamma_{AD} \in \text{Rec}(\sigma_{AE})} \alpha \text{tr} \left[\rho_{AD} R_{AD}^{1-\frac{1}{\alpha}} \right] + (1 - \alpha) \text{tr} [\gamma_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}]. \quad (103)$$

The set $\text{Rec}(\sigma_{AE})$ is convex and compact and $\{R_{AD} > 0\}$ is convex. Moreover, the function

$$R_{AD} \mapsto \alpha \text{tr} \left[\rho_{AD} R_{AD}^{1-\frac{1}{\alpha}} \right] + (1 - \alpha) \text{tr} [\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}] \quad (104)$$

is convex for any fixed τ_{ADF} because of the operator convexity of $t \mapsto t^{\beta}$ with $\beta = 1 - \frac{1}{\alpha}$ for $\alpha \in [\frac{1}{2}, 1)$ (with $\beta \in [-1, 0)$). Finally, for a fixed R_{AD} , the function being optimized is linear on τ_{ADF} . As in (83) we then look at the convex dual of

$$\max_{\tau_{ADF} \in \text{Rec}(\sigma_{AE})} \text{tr} [\tau_{ADF} \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}], \quad (105)$$

and find

$$\begin{aligned} Q_\alpha^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) &= \text{minimize} && \alpha \text{tr} \left[\rho_{AD} R_{AD}^{1-\frac{1}{\alpha}} \right] + (1-\alpha) \text{tr}[S_{AF} \sigma_{AF}] \\ &\text{subject to} && S_{AF} > 0, R_{AD} > 0 \\ &&& 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F. \end{aligned} \quad (106)$$

Invoking the arithmetic-geometric mean inequality as in the proof of Lemma 3 establishes the claim. \square

We find that the measured Rényi entropy of recovery is superadditive for all Rényi parameters.

Proposition 13. *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*

$$D_\alpha^{\text{M,rec}}(\rho_{AD} \otimes \tau_{A'D'} \|\sigma_{AE} \otimes \omega_{A'E'}) \geq D_\alpha^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) + D_\alpha^{\text{M,rec}}(\tau_{A'D'} \|\omega_{A'E'}). \quad (107)$$

Proof. We give the argument for $\alpha \in [\frac{1}{2}, 1)$ based on Lemma 12 and note that the proof for $\alpha \in (0, \frac{1}{2}) \cup (1, \infty)$ is similar. Given feasible operators R_{AD}, S_{AF} for the quantity $Q_\alpha^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE})$ and feasible operators $R_{A'D'}, S_{A'F'}$ for the quantity $Q_\alpha^{\text{M,rec}}(\tau_{A'D'} \|\omega_{A'E'})$, we apply exactly the same argument as in (89) and find

$$Q_\alpha^{\text{M,rec}}(\rho_{AD} \otimes \tau_{A'D'} \|\sigma_{AE} \otimes \omega_{A'E'}) \leq Q_\alpha^{\text{M,rec}}(\rho_{AD} \|\sigma_{AE}) + Q_\alpha^{\text{M,rec}}(\tau_{A'D'} \|\omega_{A'E'}). \quad (108)$$

This concludes the proof. \square

Analogously, the relative entropy of recovery (72) is generalized to the *Rényi relative entropy of recovery* as (see also [40, Rmk. 6]),

$$D_\alpha^{\text{rec}}(\rho_{AD} \|\sigma_{AE}) := \frac{1}{\alpha - 1} \log Q_\alpha^{\text{rec}}(\rho_{AD} \|\sigma_{AE}) \quad (109)$$

with

$$Q_\alpha^{\text{rec}}(\rho_{AD} \|\sigma_{AE}) := \begin{cases} \inf_{\Gamma_{E \rightarrow D}} Q_\alpha(\rho_{AD} \|\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE}) & \text{for } \alpha \in (1, \infty) \\ \sup_{\Gamma_{E \rightarrow D}} Q_\alpha(\rho_{AD} \|\mathcal{I}_A \otimes \Gamma_{E \rightarrow D})(\sigma_{AE}) & \text{for } \alpha \in (0, 1), \end{cases} \quad (110)$$

where we restrict to $\sigma_A > 0$ such that the quantity is surely finite and the infimum/supremum is achieved. 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., [16, Sec. 3.3]), we have

$$Q_{1/2}^{\text{M}} = Q_{1/2} \quad \text{as well as} \quad Q_{1/2}^{\text{rec}} = Q_{1/2}^{\text{M,rec}}. \quad (111)$$

The same holds true at $\alpha \rightarrow \infty$ [30]. As such Proposition 13 can be seen as a generalization of the multiplicativity of the fidelity of recovery [40] that was derived by two of the authors [6]. For the general sandwiched Rényi relative entropy of recovery, we only have preliminary numerics indicating that they are not additive. Using a non-commutative extension of the techniques from [5, Sec. 3.3.1], we can write the sandwiched Rényi relative entropy with fractional Rényi parameter as a semidefinite program. Together with the Choi state as in (79), we then get a semidefinite program for, e.g., $D_{2/3}^{\text{rec}}$. Using this, we found weak numerical evidence for additivity violations for small dimensional systems. This also challenges the possible additivity of the relative entropy of recovery D^{rec} (corresponding to the case $\alpha = 1$).

VI. CONCLUSION

We presented variational characterizations of the measured relative entropy and the measured Rényi relative entropy. Using these formulas we were able to show that these quantities can be achieved by rank-1 projective measurements instead of general POVMs. We also showed that the measured Rényi relative entropy is equal to the corresponding sandwiched Rényi relative entropy if and only if the two quantum states commute (except for the special cases with Rényi parameters $1/2$ and ∞ where they are always equal), and gave analytical counterexamples for the data-processing inequality for the sandwiched Rényi relative entropy with Rényi parameters smaller than $1/2$. Finally, we applied our variational characterizations to analyze the additivity properties of various relative entropies of recovery.

As an extension of our work it would be desirable to weaken the support conditions in Proposition 5 and Theorem 6 concerning the equality conditions for measured relative entropy and relative entropy. (We note that this is non-trivial because we make use of the equality conditions for Golden-Thompson and Araki–Lieb–Thirring). Moreover, it would be neat to extend the concept of measured relative entropy to general, continuous POVMs described by measure spaces (see, e.g., [50] for the definition of Kullback-Leibler divergence and Rényi divergence on measure spaces). Many of the steps still go through and we would like to point to a Jensen inequality for operator-valued measures [13]. It seems, however, that an extension of this inequality would be needed [12]. Finally, it would also be interesting to use the variational characterizations presented in this work for studying the operational entropic quantities mentioned in Section V A. This might lead to new applications of measured relative entropy in quantum information theory.

Acknowledgments: We acknowledge discussions with Fernando Brandão, Douglas Farenick, Hamza Fawzi, and Volkher Scholz. MB acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NFS Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028). Additional funding support was provided by the ARO grant for Research on Quantum Algorithms at the IQIM (W911NF-12-1-0521). Most of this work was done while OF was also with the Department of Computing and Mathematical Sciences, California Institute of Technology. MT would like to thank the IQIM at CalTech and John Preskill for his hospitality during the time most of the technical aspects of this project were completed. He is funded by an University of Sydney Postdoctoral Fellowship and acknowledges support from the ARC Centre of Excellence for Engineered Quantum Systems (EQUS).

Appendix A: Proofs for Results in Section V

Proof of Lemma 10. We start by writing out the relative entropy of recovery as a convex optimization program,

$$\begin{aligned}
 D^{\text{rec}}(\rho_{AD} \parallel \sigma_{AE}) = \text{tr}[\rho_{AD} \log \rho_{AD}] + \text{minimize} & \quad - \text{tr}[\rho_{AD} \log \gamma_{AD}] \\
 \text{subject to} & \quad \gamma_{AD} = \text{tr}_F [\sqrt{\sigma_{AF} \tau_{ADF} \sqrt{\sigma_{AF}}}] \\
 & \quad \tau_{ADF} \geq 0 \\
 & \quad \tau_{AF} = \Pi_{AF}^\sigma,
 \end{aligned} \tag{A1}$$

where the notation is as in the proof of Lemma 8. The Lagrangian can be written as

$$\begin{aligned} \mathcal{L}(\gamma, \tau, R, V, S) &= -\text{tr}[\rho_{AD} \log \gamma_{AD}] + \text{tr}[R_{AD} (\gamma_{AD} - \text{tr}_F [\sqrt{\sigma_{AF}} \tau_{ADF} \sqrt{\sigma_{AF}}])] \\ &\quad - \text{tr}[V_{ADF} \tau_{ADF}] + \text{tr}[S_{AF} (\tau_{AF} - \Pi_{AF}^\sigma)] \end{aligned} \quad (\text{A2})$$

$$\begin{aligned} &= -\text{tr}[\rho_{AD} \log \gamma_{AD}] + \text{tr}[R_{AD} \gamma_{AD}] - \text{tr}[\tau_{ADF} (\sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}} + V_{ADF} - S_{AF})] \\ &\quad - \text{tr}[S_{AF} \Pi_{AF}^\sigma]. \end{aligned} \quad (\text{A3})$$

In order to compute the dual objective function, we should take the infimum of this quantity over $\gamma_{AD} > 0$ and τ_{ADF} . Using the variational expression for the measured relative entropy (Lemma 1) we find

$$\inf_{\gamma_{AD} > 0, \tau_{ADF}} \mathcal{L}(\gamma, \theta, R, V, S) = -D^{\text{M}}(\rho_{AD} \| R_{AD}) + 1 - \text{tr}[S_{AF}], \quad (\text{A4})$$

when $S_{AF} \geq \sqrt{\sigma_{AF}} R_{AD} \sqrt{\sigma_{AF}}$ and $R_{AD} > 0$ (it is $-\infty$ otherwise). With Slater's strong duality and a change of variable $\bar{S}_{AF} := \sigma_{AF}^{-1/2} S_{AF} \sigma_{AF}^{-1/2}$ (but not including the bar in the following), we get

$$\begin{aligned} D^{\text{rec}}(\rho_{AD} \| \sigma_{AE}) &= \text{tr}[\rho_{AD} \log \rho_{AD}] + \text{maximize} \quad -D^{\text{M}}(\rho_{AD} \| R_{AD}) + 1 - \text{tr}[S_{AF} \sigma_{AF}] \\ &\quad \text{subject to} \quad S_{AF} > 0, R_{AD} > 0 \\ &\quad \quad \quad 1_D \otimes S_{AF} \geq R_{AD} \otimes 1_F. \end{aligned} \quad (\text{A5})$$

Adding the constraint $\text{tr}[S_{AF} \sigma_{AF}] = 1$ as in the proof of Lemma 8 concludes the proof. \square

Proof of Proposition 11. We start by writing out the argument flipped relative entropy of recovery as a convex optimization program,

$$\begin{aligned} \bar{D}^{\text{rec}}(\sigma_{AE} \| \rho_{AD}) &= \text{minimize} \quad \text{tr}[\gamma_{AD} \log \gamma_{AD}] - \text{tr}[\gamma_{AD} \log \rho_{AD}] \\ &\quad \text{subject to} \quad \gamma_{AD} = \text{tr}_F [\sqrt{\sigma_{AF}} \tau_{ADF} \sqrt{\sigma_{AF}}] \\ &\quad \quad \quad \tau_{ADF} \geq 0 \\ &\quad \quad \quad \tau_{AF} = \Pi_{AF}^\sigma, \end{aligned} \quad (\text{A6})$$

where the notation is as in the proof of Lemma 8. The Lagrangian can be written as

$$\begin{aligned} \mathcal{L}(\gamma, \tau, T, V, S) &= \text{tr}[\gamma_{AD} \log \gamma_{AD}] - \text{tr}[\gamma_{AD} \log \rho_{AD}] + \text{tr}[T_{AD} (\gamma_{AD} - \text{tr}_F [\sqrt{\sigma_{AF}} \tau_{ADF} \sqrt{\sigma_{AF}}])] \\ &\quad - \text{tr}[V_{ADF} \tau_{ADF}] + \text{tr}[S_{AF} (\tau_{AF} - \Pi_{AF}^\sigma)] \end{aligned} \quad (\text{A7})$$

$$\begin{aligned} &= \text{tr}[\gamma_{AD} \log \gamma_{AD}] + \text{tr}[\gamma_{AD} (T_{AD} - \log \rho_{AD})] \\ &\quad - \text{tr}[S_{AF} \Pi_{AF}^\sigma] - \text{tr}[\tau_{ADF} (\sqrt{\sigma_{AF}} T_{AD} \sqrt{\sigma_{AF}} + V_{ADF} - S_{AF})]. \end{aligned} \quad (\text{A8})$$

In order to compute the dual objective function, we should take the infimum of this quantity over $\gamma_{AD} > 0$ and τ_{ADF} . From the last expression we get $S_{AF} \geq \sqrt{\sigma_{AF}} T_{AD} \sqrt{\sigma_{AF}}$ and we also know how to optimize the first expression as it is an entropy maximization question:

$$\inf_{\gamma_{AD} > 0} \text{tr}[\gamma_{AD} \log T_{AD}] + \text{tr}[\gamma_{AD} (T_{AD} - \log \rho_{AD})]. \quad (\text{A9})$$

It is optimized when $\gamma_{AD} = \exp(\log \rho_{AD} - T_{AD} + \alpha 1_{AD})$ for some α [9]. This means that this infimum is given by

$$\begin{aligned} &\inf_{\alpha} \text{tr}[\exp(\log \rho_{AD} - T_{AD} + \alpha 1_{AD}) (\log \rho_{AD} - T_{AD} + \alpha 1_{AD})] \\ &\quad + \text{tr}[\exp(\log \rho_{AD} - T_{AD} + \alpha 1_{AD}) (T_{AD} - \log \rho_{AD})] \\ &= \inf_{\alpha} \alpha \text{tr}[\exp(\log \rho_{AD} - T_{AD} + \alpha 1_{AD})] \end{aligned} \quad (\text{A10})$$

$$= \inf_{\alpha} \alpha e^{\alpha} \text{tr}[\exp(\log \rho_{AD} - T_{AD} + 1_{AD})] \quad (\text{A11})$$

$$= -\text{tr}[\exp(\log \rho_{AD} - T_{AD} - 1_{AD})]. \quad (\text{A12})$$

With Slater's strong duality and a change of variable $\bar{S}_{AF} := \sigma_{AF}^{-1/2} S_{AF} \sigma_{AF}^{-1/2}$ (but not including the bar in the following), we get

$$\begin{aligned} \bar{D}^{\text{rec}}(\sigma_{AE} \|\rho_{AD}) &= \text{maximize} && -\text{tr}[S_{AF}\sigma_{AF}] - \text{tr}[\exp(\log \rho_{AD} - T_{AD} - 1_{AD})] \\ &\text{subject to} && S_{AF} > 0, T_{AD} > 0 \\ &&& 1_D \otimes S_{AF} \geq T_{AD} \otimes 1_F. \end{aligned} \quad (\text{A13})$$

We now do another change of variable (but not including the bar in the following)

$$R_{AD} := \exp(\log \rho_{AD} - T_{AD} - 1_{AD}) \quad \text{as well as} \quad 1_D \otimes \bar{S}_{AF} := 1_D \otimes S_{AF} + 1_{ADF}, \quad (\text{A14})$$

and the program becomes

$$\begin{aligned} \bar{D}^{\text{rec}}(\sigma_{AE} \|\rho_{AD}) &= \text{maximize} && -\text{tr}[S_{AF}\sigma_{AF}] + 1 - \text{tr}[R_{AD}] \\ &\text{subject to} && S_{AF} > 0, R_{AD} > 0 \\ &&& 1_D \otimes S_{AF} \geq (\log \rho_{AD} - \log R_{AD}) \otimes 1_F. \end{aligned} \quad (\text{A15})$$

Observe now that we can add the constraint $\text{tr}[R_{AD}] = 1$. In fact, let

$$\bar{R}_{AD} := \frac{R_{AD}}{\text{tr}[R_{AD}]} \quad \text{and} \quad \bar{S}_{AF} := S_{AF} + \log \text{tr}[R_{AD}]. \quad (\text{A16})$$

This solution satisfies the constraint and the objective value becomes

$$-\text{tr}[\bar{S}_{AF}\sigma_{AF}] + 1 - \text{tr}[\bar{R}_{AD}] = -\text{tr}[S_{AF}\sigma_{AF}] - \log \text{tr}[R_{AD}] \geq -\text{tr}[S_{AF}\sigma_{AF}] + 1 - \text{tr}[R_{AD}]. \quad (\text{A17})$$

This concludes the proof of (96).

To prove (97) first note that it is immediate from the definition of the argument flipped relative entropy of recovery that

$$\bar{D}^{\text{rec}}(\sigma_{AE} \otimes \omega_{A'E'} \|\rho_{AD} \otimes \tau_{A'D'}) \leq \bar{D}^{\text{rec}}(\sigma_{AE} \|\rho_{AD}) + \bar{D}^{\text{rec}}(\omega_{A'E'} \|\tau_{A'D'}), \quad (\text{A18})$$

and in the following we prove inequality in the other direction using the dual representation (96). Given feasible operators R_{AD}, S_{AF} for the quantity $\bar{D}^{\text{rec}}(\sigma_{AE} \|\rho_{AD})$ and feasible operators $R_{A'D'}, S_{A'F'}$ for the quantity $\bar{D}^{\text{rec}}(\omega_{A'E'} \|\tau_{A'D'})$, we have

$$\begin{aligned} 1_D \otimes S_{AF} &\geq (\log \rho_{AD} - \log R_{AD}) \otimes 1_F \wedge 1_{D'} \otimes S_{A'F'} \geq (\log \tau_{A'D'} - \log R_{A'D'}) \otimes 1_{F'} \\ \implies 1_{DD'} \otimes (S_{AF} \otimes 1_{A'F'} + 1_{AF} \otimes S_{A'F'}) & \\ &\geq ((\log \rho_{AD} - \log R_{AD}) \otimes 1_{A'D'} + 1_{AD} \otimes (\log \tau_{A'D'} - \log R_{A'D'})) \otimes 1_{FF'} \\ &= (\log(\rho_{AD} \otimes \tau_{A'D'}) - \log(R_{AD} \otimes R_{A'D'})) \otimes 1_{FF'}, \end{aligned} \quad (\text{A19})$$

just by multiplying with identities and adding the resulting operator inequalities. Moreover, we have $\text{tr}[R_{AD} \otimes R_{A'D'}] = 1$. Hence, $(R_{AD} \otimes R_{A'D'}, S_{AF} \otimes 1_{A'F'} + 1_{AF} \otimes S_{A'F'})$ is a feasible pair in the expression (96) for $\bar{D}^{\text{rec}}(\sigma_{AE} \otimes \omega_{A'E'} \|\rho_{AD} \otimes \tau_{A'D'})$ and we get

$$\bar{D}^{\text{rec}}(\sigma_{AE} \otimes \omega_{A'E'} \|\rho_{AD} \otimes \tau_{A'D'}) \geq -\text{tr}[(S_{AF} \otimes 1_{A'F'} + 1_{AF} \otimes S_{A'F'})(\sigma_{AF} \otimes \omega_{A'F'})] \quad (\text{A20})$$

$$= -\text{tr}[S_{AF}\sigma_{AF}] - \text{tr}[S_{A'F'}\omega_{A'F'}]. \quad (\text{A21})$$

Taking the supremum over feasible (R_{AD}, S_{AF}) and $(R_{A'D'}, S_{A'F'})$, we find the claimed additivity. \square

-
- [1] P. M. Alberti. “A Note on the Transition Probability over C^* -Algebras”. *Letters in Mathematical Physics* **7**: 25–32 (1983).
- [2] H. Araki. “On an Inequality of Lieb and Thirring”. *Letters in Mathematical Physics* **19**(2): 167–170 (1990).
- [3] S. Beigi. “Sandwiched Rényi Divergence Satisfies Data Processing Inequality”. *Journal of Mathematical Physics* **54**(12): 122202 (2013).
- [4] V. P. Belavkin and P. Staszewski. “ C^* -algebraic Generalization of Relative Entropy and Entropy”. *Annals Henri Poincaré* **37**(1): 51–58, (1982).
- [5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. *Society for Industrial and Applied Mathematics* (2001).
- [6] M. Berta and M. Tomamichel. “The Fidelity of Recovery is Multiplicative”. Preprint, [arXiv: 1502.07973](https://arxiv.org/abs/1502.07973) (2015).
- [7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press (2004).
- [8] 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* **115**(5): 050501 (2015).
- [9] V. Chandrasekaran and P. Shah. “Relative Entropy Optimization and its Applications”. Unpublished, available at [California Institute of Technology](https://caltech.edu/) (2015).
- [10] D. DiVincenzo, M. Horodecki, D. Leung, J. Smolin, and B. Terhal. “Locking Classical Correlations in Quantum States”. *Physical Review Letters* **92**(6): 067902 (2004).
- [11] M. J. Donald. “On the Relative Entropy”. *Communications in Mathematical Physics* **105**(1): 13–34 (1986).
- [12] D. R. Farenick. “Private Communication”, (2015).
- [13] D. R. Farenick and F. Zhou. “Jensen’s Inequality Relative to Matrix-Valued Measures”. *Journal of Mathematical Analysis and Applications* **327**(2): 919–929 (2007).
- [14] O. Fawzi and R. Renner. “Quantum Conditional Mutual Information and Approximate Markov Chains”. *Communications in Mathematical Physics* **340**(2): 575–611 (2015).
- [15] R. L. Frank and E. H. Lieb. “Monotonicity of a Relative Rényi Entropy”. *Journal of Mathematical Physics* **54**(12): 122201 (2013).
- [16] C. A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. Phd thesis, University of New Mexico, (1996). Available at [arXiv: quant-ph/9601020v1](https://arxiv.org/abs/quant-ph/9601020v1).
- [17] S. Golden. “Lower Bounds for the Helmholtz Function”. *Physical Review* **137**(4B): B1127–B1128 (1965).
- [18] F. Hiai. “Equality Cases in Matrix Norm Inequalities of Golden-Thompson Type”. *Linear and Multilinear Algebra* **36**(4): 239–249 (1994).
- [19] F. Hiai and D. Petz. “The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability”. *Communications in Mathematical Physics* **143**(1): 99–114 (1991).
- [20] C. Hirche, C. Morgan, K. P. Seshadreesan, and M. M. Wilde. “Unpublished Notes”, (2015).
- [21] B. Ibinson, N. Linden, and A. Winter. “Robustness of Quantum Markov Chains”. *Communications in Mathematical Physics* **277**(2): 289–304 (2007).
- [22] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. “Universal recovery from a decrease of quantum relative entropy”. Preprint, [arXiv: 1509.07127](https://arxiv.org/abs/1509.07127) (2015).
- [23] R. König, R. Renner, and C. Schaffner. “The Operational Meaning of Min- and Max-Entropy”. *IEEE Transactions on Information Theory* **55**(9): 4337–4347 (2009).
- [24] H. Kosaki. “Relative Entropy of States: A Variational Expression”. *Journal of Operator Theory* **16**(2): 335–348, (1986).
- [25] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. *Annals of Mathematical Statistics* **22**(1): 79–86 (1951).
- [26] K. Li and A. Winter. “Squashed Entanglement, k -Extendibility, Quantum Markov Chains, and Recovery Maps”. Preprint, [arXiv: 1410.4184](https://arxiv.org/abs/1410.4184) (2014).
- [27] E. H. Lieb and W. E. Thirring. “Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities”. In *The Stability of Matter: From Atoms to Stars*, chapter III, pages 205–239. *Springer* (2005).

- [28] G. Lindblad. “Completely Positive Maps and Entropy Inequalities”. *Communications in Mathematical Physics* **40**(2):147–151 (1975).
- [29] K. Matsumoto. “A New Quantum Version of f -Divergence”. Preprint, [arXiv:1311.4722](https://arxiv.org/abs/1311.4722) (2014).
- [30] M. Mosonyi and T. Ogawa. “Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies”. *Communications in Mathematical Physics* **334**(3):1617–1648 (2015).
- [31] M. Müller-Lennert. *Quantum Relative Rényi Entropies*. Master thesis, ETH Zurich,
- [32] 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* **54**(12):122203 (2013).
- [33] D. Petz. “Quasi-entropies for Finite Quantum Systems”. *Reports on Mathematical Physics* **23**(1):57–65 (1986).
- [34] D. Petz. “Sufficient Subalgebras and the Relative Entropy of States of a von Neumann Algebra”. *Communications in Mathematical Physics* **105**(1):123–131 (1986).
- [35] D. Petz. “A Variational Expression for the Relative Entropy”. *Communications in Mathematical Physics* **114**(2):345–349 (1988).
- [36] D. Petz. *Quantum Information Theory and Quantum Statistics*. Springer (2008).
- [37] M. Piani. “Relative Entropy of Entanglement and Restricted Measurements”. *Physical Review Letters* **103**(16):160504 (2009).
- [38] E. Rains. “A Semidefinite Program for Distillable Entanglement”. *IEEE Transactions on Information Theory* **47**(7):2921–2933 (2001).
- [39] A. Rényi. “On Measures of Information and Entropy”. In *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, Berkeley, California, USA(1961).
- [40] K. P. Seshadreesan and M. M. Wilde. “Fidelity of Recovery, Squashed Entanglement, and Measurement Recoverability”. *Physical Review A* **92**(4):042321 (2015).
- [41] M. Sion. “On General Minimax Theorems”. *Pacific Journal of Mathematics* **8**:171–176, (1958).
- [42] W. So. “Equality Cases in Matrix Exponential Inequalities”. *SIAM Journal on Matrix Analysis and Applications* **13**(4):1154–1158 (1992).
- [43] D. Sutter, O. Fawzi, and R. Renner. “Universal Recovery Map for Approximate Markov Chains”. Preprint, [arXiv:1504.07251](https://arxiv.org/abs/1504.07251) (2015).
- [44] D. Sutter, M. Tomamichel, and A. W. Harrow. “Strengthened Monotonicity of Relative Entropy via Pinched Petz Recovery Map”. Preprint, [arXiv:1507.00303](https://arxiv.org/abs/1507.00303) (2015).
- [45] C. J. Thompson. “Inequality with Applications in Statistical Mechanics”. *Journal of Mathematical Physics* **6**(11):1812 (1965).
- [46] M. Tomamichel. *Quantum Information Processing with Finite Resources — Mathematical Foundations*. volume 5 of *SpringerBriefs in Mathematical Physics*, Springer International Publishing (2016).
- [47] M. Tomamichel, M. Berta, and M. Hayashi. “Relating Different Quantum Generalizations of the Conditional Rényi Entropy”. *Journal of Mathematical Physics* **55**(8):082206 (2014).
- [48] A. Uhlmann. “Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory”. *Communications in Mathematical Physics* **54**(1):21–32 (1977).
- [49] H. Umegaki. “Conditional Expectation in an Operator Algebra”. *Kodai Math. Sem. Rep.* **14**:59–85, (1962).
- [50] T. van Erven and P. Harremoës. “Rényi Divergence and Kullback-Leibler Divergence”. *IEEE Transactions on Information Theory* **60**(7):3797–3820 (2014).
- [51] V. Vedral. “On Bound Entanglement Assisted Distillation”. *Physics Letters A* **262**(2-3):121–124 (1999).
- [52] V. Vedral and M. B. Plenio. “Entanglement Measures and Purification Procedures”. *Physical Review A* **57**(3):1619–1633 (1998).
- [53] M. M. Wilde. “Recoverability in quantum information theory”. *Proceedings of the Royal Society A* **471**(2182):20150338 (2015).
- [54] 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* **331**(2):593–622 (2014).
- [55] A. Winter and K. Li. “A stronger subadditivity relation? With applications to squashed entanglement, sharability and separability”, (2012). Available online: http://www.maths.bris.ac.uk/~csajw/stronger_subadditivity.pdf.