Optimized quantum f-divergences

Mark M. Wilde
Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA, Email: mwilde@lsu.edu

***Abstract*—The quantum relative entropy is a measure of the distinguishability of two quantum states, and it is a unifying concept in quantum information theory: many information measures such as entropy, conditional entropy, mutual information, and entanglement measures can be realized from it. As such, there has been broad interest in generalizing the notion to further understand its most basic properties, one of which is the data processing inequality. The quantum f-divergence of Petz is one generalization of the quantum relative entropy, and it also leads to other relative entropies, such as the Petz–Rényi relative entropies. In this contribution, I introduce the optimized quantum f-divergence as a related generalization of quantum relative entropy. I prove that it satisfies the data processing inequality, and the method of proof relies upon the operator Jensen inequality, similar to Petz's original approach. Interestingly, the sandwiched Rényi relative entropies are particular examples of the optimized f-divergence. Thus, one benefit of this approach is that there is now a single, unified approach for establishing the data processing inequality for both the Petz–Rényi and sandwiched Rényi relative entropies, for the full range of parameters for which it is known to hold.**

Full version of this paper is accessible at arXiv:1710.10252

I. Introduction

The quantum relative entropy [1] is a foundational distinguishability measure in quantum information theory (QIT). It is a function of two quantum states and measures how well one can tell the two states apart by a quantum-mechanical experiment. One important reason for why it has found such widespread application is that it satisfies a data-processing inequality [2], [3]: it does not increase under the action of a quantum channel on the two states. This can be interpreted as saying that two quantum states do not become more distinguishable if the same quantum channel is applied to them, and a precise interpretation of this statement in terms of quantum hypothesis testing is available in [4]–[6]. Quantum relative entropy generalizes its classical counterpart [7].

The wide interest in relative entropy sparked various researchers to generalize and study it further, in an attempt to elucidate the fundamental properties that govern its behavior. One notable generalization is Rényi's relative entropy [8], but this was subsequently generalized even further in the form of the f-divergence [9]–[11]. For probability distributions $\{p(x)\}_x$ and $\{q(x)\}_x$ and a convex function f, the f-divergence is defined as $\sum_x q(x) f(p(x)/q(x))$, in the case that $p(x) = 0$ for all x such that $q(x) = 0$. The resulting quantity is then non-increasing under the action of a classical channel $r(y|x)$ that produces the output distributions $\sum_x r(y|x)p(x)$ and $\sum_x r(y|x)q(x)$. Some years after these developments, a quantum generalization of f-divergence appeared in [12], [13] In [12], [13] and a later development [14], the quantum data-processing inequality was proved in full generality for arbitrary quantum channels, whenever the underlying function f is *operator* convex.

Interestingly, when generalizing a notion from classical to QIT, there is often more than one way to do so, and sometimes there could even be an infinite number of ways to do so. This has to do with the non-commutativity of quantum states. For example, there are several different ways that one could generalize the relative entropy to the quantum case, and two prominent formulas were put forward in [1] and [15]. This added complexity for the quantum case could potentially be problematic, but the typical way of determining on which generalizations we should focus is to show that a given formula is the answer to a meaningful operational task. The papers [4], [5] accomplished this for the formula from [1], and since then, researchers have realized more and more just how foundational the formula of [1] is. As a consequence, the formula of [1] is now known as quantum relative entropy.

The situation becomes more intricate when it comes to quantum generalizations of Rényi relative entropy. For many years, the Petz–Rényi relative entropy of [12], [13] has been widely studied and given an operational interpretation [16], [17], again in the context of quantum hypothesis testing. However, in recent years, the sandwiched Rényi relative entropy of [18], [19] has gained prominence, due to its role in establishing strong converses for communication tasks (see, e.g., [19], [20]). The result of [21] solidified its fundamental meaning in QIT, proving that it has an operational interpretation in the strong converse exponent of quantum hypothesis testing. As such, the situation is that there are two generalizations of Rényi relative entropy that should be considered in QIT, due to their operational role mentioned above.

The same work that introduced the Petz–Rényi relative entropy also introduced a quantum generalization of the notion of f-divergence [12], [13] (see also [22]), with the Petz–Rényi relative entropy being a particular example. Since then, other quantum f-divergences have appeared [23], [24], now known as minimal and maximal f-divergences [24], [25]. However, it has not been known how the sandwiched Rényi relative entropy fits into the paradigm of quantum f-divergences.

In this paper, I modify Petz's definition of quantum f-divergence [12], [13], [22], by allowing for a particular optimization (see Definition 1 for details of the modification). As such, I call the resulting quantity the *optimized* quantum f-divergence. I prove that it obeys a quantum data processing inequality, and as such, my perspective is that it deserves to

978-1-5386-4780-6/18/$31.00©2018 IEEE

be considered as another variant of the quantum f-divergence, in addition to the original, the minimal, and the maximal. Interestingly, the sandwiched Rényi relative entropy is directly related to the optimized quantum f-divergence, thus bringing the sandwiched quantity into the f-divergence formalism.

One benefit of the results of this paper is that there is now a single, unified approach for establishing the data-processing inequality for both the Petz–Rényi relative entropy and the sandwiched Rényi relative entropy, for the full Rényi parameter ranges for which it is known to hold. This unified approach is based on Petz's original approach that employed the operator Jensen inequality [26], and it is useful for presenting a succint proof of the data processing inequality for both quantum Rényi relative entropy families.

In the rest of the paper, I begin by defining the optimized quantum f-divergence in the next section. In Section III, I prove that the optimized f-divergence satisfies the quantum data processing inequality under partial trace whenever the underlying function f is operator anti-monotone with domain $(0,\infty)$ and range $\mathbb{R}$. The core tool underlying this proof is the operator Jensen inequality [26]. In Section IV, I show how the quantum relative entropy and the sandwiched Rényi relative entropies are directly related to the optimized quantum f-divergence. Section V then discusses the relation between Petz's f-divergence and the optimized one. I finally conclude in Section VI with a summary.

II. Optimized quantum f-divergence

Let us begin by defining the optimized quantum f-divergence. Here I focus exclusively on the case of positive definite operators, and the full version provides details for the more general case of positive semi-definite operators.

Definition 1 (Optimized quantum f-divergence): Let f be a function with domain $(0,\infty)$ and range $\mathbb{R}$. For positive definite operators X and Y acting on a Hilbert space $\mathcal{H}_S$, we define the *optimized* quantum f-divergence as

$$\widetilde{Q}_f(X\|Y) \equiv \sup_{\tau>0,\ \mathrm{Tr}\{\tau\}\leq 1} \widetilde{Q}_f(X\|Y;\tau), \tag{1}$$

where $\widetilde{Q}_f(X\|Y;\tau)$ is defined for positive definite Y and τ acting on $\mathcal{H}_S$ as

$$\widetilde{Q}_f(X\|Y;\tau) \equiv \langle\varphi^X|_{S\hat{S}} f(\tau_S^{-1}\otimes Y_{\hat{S}}^T)|\varphi^X\rangle_{S\hat{S}}, \tag{2}$$

$$|\varphi^X\rangle_{S\hat{S}} \equiv (X_S^{1/2}\otimes I_{\hat{S}})|\Gamma\rangle_{S\hat{S}}. \tag{3}$$

In the above, $\mathcal{H}_{\hat{S}}$ is an auxiliary Hilbert space isomorphic to $\mathcal{H}_S$, $|\Gamma\rangle_{S\hat{S}} \equiv \sum_{i=1}^{|S|} |i\rangle_S |i\rangle_{\hat{S}}$, for orthonormal bases $\{|i\rangle_S\}_{i=1}^{|S|}$ and $\{|i\rangle_{\hat{S}}\}_{i=1}^{|\hat{S}|}$, and the T superscript indicates transpose with respect to the basis $\{|i\rangle_{\hat{S}}\}_i$.

The case of greatest interest for us here is when the underlying function f is operator anti-monotone; i.e., for Hermitian operators A and B, the function f is such that $A \leq B \Rightarrow f(B) \leq f(A)$ (see, e.g., [27]). This property is rather strong, but there are several functions of interest in quantum physical applications that obey it (see Section IV). One critical property of an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$ is that it is also operator convex and continuous (see, e.g., [28]). In this case, we have the following simple proposition, proved in the full version:

Proposition 2: Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. For positive definite operators X and Y acting on a Hilbert space $\mathcal{H}_S$,

$$\widetilde{Q}_f(X\|Y) = \sup_{\tau>0,\ \mathrm{Tr}\{\tau\}=1} \widetilde{Q}_f(X\|Y;\tau), \tag{4}$$

and the function $\widetilde{Q}_f(X\|Y;\tau)$ is concave in τ.

III. Quantum data processing

Our first main objective is to prove that $\widetilde{Q}_f(X\|Y)$ deserves the name "f-divergence" or "f-relative entropy," i.e., that it is monotone non-increasing under the action of a completely positive trace-preserving map $\mathcal{N}$:

$$\widetilde{Q}_f(X\|Y) \geq \widetilde{Q}_f(\mathcal{N}(X)\|\mathcal{N}(Y)). \tag{5}$$

Such a map $\mathcal{N}$ is also called a quantum channel, due to its purpose in quantum physics as modeling the physical evolution of the state of a quantum system. In QIT contexts, the inequality in (5) is known as the quantum data processing inequality. According to the Stinespring dilation theorem [29], to every quantum channel $\mathcal{N}_{S\to B}$, there exists an isometry $U_{S\to BE}^{\mathcal{N}}$ such that

$$\mathcal{N}_{S\to B}(X_S) = \mathrm{Tr}_E\{U_{S\to BE}^{\mathcal{N}} X_S \left(U_{S\to BE}^{\mathcal{N}}\right)^\dagger\}. \tag{6}$$

As such, we can prove the inequality in (5) in two steps. *Isometric invariance*: First show that

$$\widetilde{Q}_f(X\|Y) = \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger) \tag{7}$$

for any isometry U and any positive semi-definite X and Y. This is done in the full version of this work, using the general definition given there. *Monotonicity under partial trace*: Then show that

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A) \tag{8}$$

for positive semi-definite operators X_{AB} and Y_{AB} acting on the tensor-product Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B$, with $X_A = \mathrm{Tr}_B\{X_{AB}\}$ and $Y_A = \mathrm{Tr}_B\{Y_{AB}\}$.

We now discuss the second step toward quantum data processing, mentioned above, and here we focus exclusively on positive definite operators:

Theorem 3 (Monotonicity under partial trace): Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Given positive definite operators X_{AB} and Y_{AB} acting on the tensor-product Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B$, the optimized quantum f-divergence does not increase under the action of a partial trace, in the sense that

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A), \tag{9}$$

where $X_A = \mathrm{Tr}_B\{X_{AB}\}$ and $Y_A = \mathrm{Tr}_B\{Y_{AB}\}$.

Proof. The quantities of interest are as follows:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB};\tau_{AB}) = \langle\varphi^{X_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}}, \tag{10}$$

$$\widetilde{Q}_f(X_A\|Y_A;\omega_A) = \langle\varphi^{X_A}|_{A\hat{A}} f(\omega_A^{-1}\otimes Y_{\hat{A}}^T)|\varphi^{X_A}\rangle_{A\hat{A}}, \quad (11)$$

where τ_{AB} and ω_A are invertible density operators and, by definition,

$$|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} = \left(X_{AB}^{1/2}\otimes I_{\hat{A}\hat{B}}\right)|\Gamma\rangle_{A\hat{A}}\otimes|\Gamma\rangle_{B\hat{B}}. \quad (12)$$

The following map, acting on an operator Z_A, is a quantum channel known as the Petz recovery channel [30], [31]:

$$Z_A \to X_{AB}^{1/2}\left(\left[X_A^{-1/2}Z_A X_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2}. \quad (13)$$

It is completely positive because it consists of the serial concatenation of three completely positive maps: sandwiching by $X_A^{-1/2}$, tensoring in the identity I_B, and sandwiching by $X_{AB}^{1/2}$. It is also trace preserving. The Petz recovery channel has the property that it perfectly recovers X_{AB} if X_A is input because

$$X_A \to X_{AB}^{1/2}\left(\left[X_A^{-1/2}X_A X_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2} = X_{AB}. \quad (14)$$

Every completely positive and trace preserving map $\mathcal{N}$ has a Kraus decomposition, which is a set $\{K_i\}_i$ of operators such that $\mathcal{N}(\cdot) = \sum_i K_i(\cdot)K_i^\dagger$ and $\sum_i K_i^\dagger K_i = I$. A standard construction for an isometric extension of a channel is then to pick an orthonormal basis $\{|i\rangle_E\}_i$ for an auxiliary Hilbert space $\mathcal{H}_E$ and define

$$V = \sum_i K_i\otimes|i\rangle_E. \quad (15)$$

One can then readily check that $\mathcal{N}(\cdot) = \mathrm{Tr}_E\{V(\cdot)V^\dagger\}$ and $V^\dagger V = I$. For the Petz recovery channel, we can figure out a Kraus decomposition by expanding the identity operator $I_B = \sum_{j=1}^{|B|}|j\rangle\langle j|_B$, with respect to some orthonormal basis $\{|j\rangle_B\}_j$, so that

$$\begin{aligned}
&X_{AB}^{1/2}\left(\left[X_A^{-1/2}\omega_A X_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2}\\
&= \sum_{j=1}^{|B|} X_{AB}^{1/2}\left(\left[X_A^{-1/2}\omega_A X_A^{-1/2}\right]\otimes|j\rangle\langle j|_B\right)X_{AB}^{1/2}\\
&= \sum_{j=1}^{|B|} X_{AB}^{1/2}\left[X_A^{-1/2}\otimes|j\rangle_B\right]\omega_A\left[X_A^{-1/2}\otimes\langle j|_B\right]X_{AB}^{1/2}.
\end{aligned}$$

Thus, Kraus operators for the Petz recovery channel are $\left\{X_{AB}^{1/2}\left[X_A^{-1/2}\otimes|j\rangle_B\right]\right\}_{j=1}^{|B|}$. According to the standard recipe in (15), we can construct an isometric extension of the Petz recovery channel as

$$\begin{aligned}
\sum_{j=1}^{|B|} X_{AB}^{1/2}\left[X_A^{-1/2}\otimes|j\rangle_B\right]|j\rangle_{\hat{B}} &= X_{AB}^{1/2}X_A^{-1/2}\sum_{j=1}^{|B|}|j\rangle_B|j\rangle_{\hat{B}}\\
&= X_{AB}^{1/2}X_A^{-1/2}|\Gamma\rangle_{B\hat{B}}. \quad (16)
\end{aligned}$$

We can then extend this isometry to act as an isometry on a larger space by tensoring it with the identity operator $I_{\hat{A}}$, and so we define

$$V_{A\hat{A}\to A\hat{A}B\hat{B}} \equiv X_{AB}^{1/2}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]|\Gamma\rangle_{B\hat{B}}. \quad (17)$$

We can also see that $V_{A\hat{A}\to A\hat{A}B\hat{B}}$ acting on $|\varphi^{X_A}\rangle_{A\hat{A}}$ generates $|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}}$: $|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} = V_{A\hat{A}\to A\hat{A}B\hat{B}}|\varphi^{X_A}\rangle_{A\hat{A}}$. This can be interpreted as a generalization of (14) in the language of QIT: an isometric extension of the Petz recovery channel perfectly recovers a purification $|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}}$ of X_{AB} from a purification $|\varphi^{X_A}\rangle_{A\hat{A}}$ of X_A. Since the Petz recovery channel is indeed a channel, we can pick τ_{AB} as the output state of the Petz recovery channel acting on an invertible state ω_A:

$$\tau_{AB} = X_{AB}^{1/2}\left(\left[X_A^{-1/2}\omega_A X_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2}. \quad (18)$$

Observe that τ_{AB} is invertible. Then consider that

$$\begin{aligned}
&V^\dagger\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T\right)V\\
&= \langle\Gamma|_{B\hat{B}}\left(X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2}\otimes Y_{\hat{A}\hat{B}}^T\right)|\Gamma\rangle_{B\hat{B}} \quad (19)\\
&= \langle\Gamma|_{B\hat{B}}\left(\omega_A^{-1}\otimes I_B\otimes Y_{\hat{A}\hat{B}}^T\right)|\Gamma\rangle_{B\hat{B}} \quad (20)\\
&= \omega_A^{-1}\otimes\langle\Gamma|_{B\hat{B}}Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} \quad (21)\\
&= \omega_A^{-1}\otimes Y_{\hat{A}}^T. \quad (22)
\end{aligned}$$

For the second equality, we used the fact that $X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2} = \omega_A^{-1}\otimes I_B$ for the choice of τ_{AB} in (18). With this setup, we can now readily establish the desired inequality by employing the operator Jensen inequality [26] and operator convexity of the function f:

$$\begin{aligned}
&\widetilde{Q}_f(X_{AB}\|Y_{AB};\tau_{AB})\\
&= \langle\varphi^{X_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} \quad (23)\\
&= \langle\varphi^{X_A}|_{A\hat{A}} V^\dagger f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)V|\varphi^{X_A}\rangle_{A\hat{A}} \quad (24)\\
&\geq \langle\varphi^{X_A}|_{A\hat{A}} f(V^\dagger[\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T]V)|\varphi^{X_A}\rangle_{A\hat{A}} \quad (25)\\
&= \langle\varphi^{X_A}|_{A\hat{A}} f(\omega_A^{-1}\otimes Y_{\hat{A}}^T)|\varphi^{X_A}\rangle_{A\hat{A}} \quad (26)\\
&= \widetilde{Q}_f(X_A\|Y_A;\omega_A). \quad (27)
\end{aligned}$$

Taking a supremum over τ_{AB} such that $\tau_{AB} > 0$ and $\mathrm{Tr}\{\tau_{AB}\} = 1$, we conclude that the following inequality holds for all invertible states ω_A:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A;\omega_A). \quad (28)$$

After taking a supremum over invertible states ω_A, we find that the inequality in (9) holds when X_{AB} is invertible. ■

IV. Examples of optimized quantum f-divergences

I now show how several known quantum divergences are particular examples of an optimized quantum f-divergence, including the quantum relative entropy [1] and the sandwiched Rényi relative quasi-entropies [18], [19]. The result will be that Theorem 3 recovers quantum data processing for the sandwiched Rényi relative entropies for the full range of parameters for which it is known to hold. Thus, one benefit of Theorem 3 and earlier work of [12]–[14] is a single, unified approach, based on the operator Jensen inequality [26], for establishing quantum data processing for all of the Petz– and sandwiched Rényi relative entropies for the full parameter ranges for which data processing is known to hold.

A. Quantum relative entropy as optimized quantum f-divergence

Let τ be an invertible state and X and Y positive definite. Let $\overline{X} = X/\operatorname{Tr}\{X\}$. Pick the function $f(x) = -\log x$, which is an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$, and we find that

$$\begin{aligned}
&\frac{1}{\operatorname{Tr}\{X\}}\langle\varphi^X|_{S\hat{S}}\left[-\log(\tau_S^{-1}\otimes Y_{\hat{S}}^T)\right]|\varphi^X\rangle_{S\hat{S}}\\
&= \langle\varphi^{\overline{X}}|_{S\hat{S}}\left[\log(\tau_S)\otimes I_{\hat{S}} - I_S\otimes\log Y_{\hat{S}}^T\right]|\varphi^{\overline{X}}\rangle_{S\hat{S}} \quad (29)\\
&= \langle\varphi^{\overline{X}}|_{S\hat{S}}\log(\tau_S)\otimes I_{\hat{S}}|\varphi^{\overline{X}}\rangle_{S\hat{S}}\\
&\quad - \langle\varphi^{\overline{X}}|_{S\hat{S}} I_S\otimes\log\left(Y_{\hat{S}}^T\right)|\varphi^{\overline{X}}\rangle_{S\hat{S}} \quad (30)\\
&= \operatorname{Tr}\{\overline{X}\log\tau\} - \operatorname{Tr}\{\overline{X}\log Y\} \quad (31)\\
&\leq \operatorname{Tr}\{\overline{X}\log\overline{X}\} - \operatorname{Tr}\{\overline{X}\log Y\} = D(\overline{X}\|Y). \quad (32)
\end{aligned}$$

The inequality is a consequence of Klein's inequality [32] (see also [33]), establishing that the optimal τ is set to $\overline{X}$. So we find that $\widetilde{Q}_{-\log(\cdot)}(X\|Y) = \operatorname{Tr}\{X\}D(\overline{X}\|Y)$, where the quantum relative entropy $D(\overline{X}\|Y)$ is defined as [1] $D(\overline{X}\|Y) = \operatorname{Tr}\{\overline{X}\left[\log\overline{X} - \log Y\right]\}$.

B. Sandwiched Rényi relative quasi-entropy as optimized quantum f-divergence

Take τ, X, and Y as defined in Section IV-A. For $\alpha \in [1/2,1)$, pick the function $f(x) = -x^{(1-\alpha)/\alpha}$, which is an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Note that this is a reparametrization of $-x^\beta$ for $\beta \in (0,1]$. I now show that

$$\widetilde{Q}_{-(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = -\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \quad (33)$$

which is the known expression for sandwiched quasi-entropy for $\alpha \in [1/2,1)$ [18], [19]. To see this, consider that

$$\begin{aligned}
&-\langle\varphi^X|_{S\hat{S}}\left[\tau_S^{-1}\otimes Y_{\hat{S}}^T\right]^{(1-\alpha)/\alpha}|\varphi^X\rangle_{S\hat{S}}\\
&= -\langle\varphi^X|_{S\hat{S}}\tau_S^{(\alpha-1)/\alpha}\otimes\left(Y_{\hat{S}}^T\right)^{(1-\alpha)/\alpha}|\varphi^X\rangle_{S\hat{S}}\\
&= -\langle\Gamma|_{S\hat{S}}X_S^{1/2}\tau_S^{(\alpha-1)/\alpha}X_S^{1/2}\otimes\left(Y_{\hat{S}}^T\right)^{(1-\alpha)/\alpha}|\Gamma\rangle_{S\hat{S}}\\
&= -\operatorname{Tr}\left\{X^{1/2}\tau^{(\alpha-1)/\alpha}X^{1/2}Y^{(1-\alpha)/\alpha}\right\}\\
&= -\operatorname{Tr}\left\{X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}. \quad (34)
\end{aligned}$$

Now optimizing over invertible states τ and employing Hölder duality, in the form of the reverse Hölder inequality and as observed in [18], we find that

$$\begin{aligned}
&\sup_{\substack{\tau>0,\\ \operatorname{Tr}\{\tau\}=1}}\left[-\operatorname{Tr}\left\{X^{1/2}Y^{\frac{1-\alpha}{\alpha}}X^{1/2}\tau^{\frac{\alpha-1}{\alpha}}\right\}\right]\\
&\qquad = -\left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_\alpha, \quad (35)
\end{aligned}$$

where for positive semi-definite Z, we define $\|Z\|_\alpha = \left[\operatorname{Tr}\{Z^\alpha\}\right]^{1/\alpha}$. We then get that

$$\begin{aligned}
\widetilde{Q}_{-(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) &= -\left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_\alpha \quad (36)\\
&= -\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \quad (37)
\end{aligned}$$

which is the sandwiched Rényi relative quasi-entropy for the range $\alpha \in [1/2,1)$. The sandwiched Rényi relative entropy itself is defined up to a normalization factor as [18], [19]

$$\widetilde{D}_\alpha(X\|Y) = \frac{\alpha}{\alpha-1}\log\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha. \quad (38)$$

Thus, Theorem 3 implies quantum data processing for the sandwiched Rényi relative entropy $\widetilde{D}_\alpha(X_{AB}\|Y_{AB}) \geq \widetilde{D}_\alpha(X_A\|Y_A)$, for the parameter range $\alpha \in [1/2,1)$, which is a result previously established in [34].

For $\alpha \in (1,\infty]$, pick the function $f(x) = x^{(1-\alpha)/\alpha}$, which is an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Note that this is a reparametrization of x^β for $\beta \in [-1,0)$. I now show that

$$\widetilde{Q}_{(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = \left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \quad (39)$$

which is the known expression for sandwiched Rényi relative quasi-entropy for $\alpha \in (1,\infty]$ [18], [19]. To see this, consider that the same development as above gives that

$$\begin{aligned}
&\langle\varphi^X|_{S\hat{S}}(\tau_S^{-1}\otimes Y_{\hat{S}}^T)^{(1-\alpha)/\alpha}|\varphi^X\rangle_{S\hat{S}}\\
&\qquad = \operatorname{Tr}\left\{X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}. \quad (40)
\end{aligned}$$

Again employing Hölder duality, as observed in [18], we find

$$\begin{aligned}
&\sup_{\tau>0,\operatorname{Tr}\{\tau\}=1}\operatorname{Tr}\left\{X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}\\
&\qquad = \left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_\alpha, \quad (41)
\end{aligned}$$

We then get that

$$\begin{aligned}
\widetilde{Q}_{(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) &= \left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_\alpha \quad (42)\\
&= \left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \quad (43)
\end{aligned}$$

where the equalities hold as observed in [18]. The sandwiched Rényi relative entropy itself is defined up to a normalization factor as in (38) [18], [19]. Thus, Theorem 3 implies quantum data processing for the sandwiched Rényi relative entropy $\widetilde{D}_\alpha(X_{AB}\|Y_{AB}) \geq \widetilde{D}_\alpha(X_A\|Y_A)$, for the parameter range $\alpha \in (1,\infty]$, which is a result previously established in full by [21], [34], [35] and for $\alpha \in (1,2]$ by [18], [19].

V. On Petz's quantum f-divergence

I now discuss in more detail the relation between the optimized quantum f-divergence and Petz's f-divergence from [12], [13]. In brief, we find that the Petz f-divergence can be recovered by replacing τ in Definition 1 with X.

Definition 4 (Petz quantum f-divergence): Let f be a continuous function with domain $(0,\infty)$ and range $\mathbb{R}$. For positive definite operators X and Y acting on a Hilbert space $\mathcal{H}_S$, the Petz quantum f-divergence is defined as

$$Q_f(X\|Y) \equiv \langle\varphi^X|_{S\hat{S}}f\left(X_S^{-1}\otimes Y_{\hat{S}}^T\right)|\varphi^X\rangle_{S\hat{S}}, \quad (44)$$

where the notation is the same as in Definition 1.

One main concern is about quantum data processing with the Petz f-divergence. To show this, we take f to be an

operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. As discussed in Section III, one can establish data processing by showing isometric invariance and monotonicity under partial trace. Isometric invariance of $Q_f(X\|Y)$ follows from the same proof as given in the full version and was also shown in [14]. Monotonicity of $Q_f(X_{AB}\|Y_{AB})$ under partial trace for positive definite X_{AB} and Y_{AB} follows from the operator Jensen inequality [12], [13].

Special and interesting cases of the Petz f-divergence are found by taking $f(x) = -\log x$, $f(x) = -x^\beta$ for $\beta \in (0,1]$, and $f(x) = x^\beta$ for $\beta \in [-1,0)$. Each of these functions are operator anti-monotone with domain $(0,\infty)$ and range $\mathbb{R}$. As shown in [12], [13], all of the following quantities obey the data processing inequality:

$$Q_{-\log(\cdot)}(X\|Y) = \operatorname{Tr}\{X\}D(\overline{X}\|Y), \tag{45}$$

$$Q_{-(\cdot)^\beta}(X\|Y) = -\operatorname{Tr}\{X^{1-\beta}Y^\beta\}, \text{ for } \beta \in (0,1], \tag{46}$$

$$Q_{(\cdot)^\beta}(X\|Y) = \operatorname{Tr}\{X^{1-\beta}Y^\beta\}, \text{ for } \beta \in [-1,0). \tag{47}$$

By a reparametrization $\alpha = 1-\beta$, the latter two quantities are directly related to the Petz Rényi relative entropy $D_\alpha(X\|Y) \equiv \frac{1}{\alpha-1}\log \operatorname{Tr}\{X^\alpha Y^{1-\alpha}\}$. Thus, the data processing inequality holds for $D_\alpha(X\|Y)$ for $\alpha \in [0,1)\cup(1,2]$ [13], [14].

VI. Conclusion

The main contribution of the present work is the definition of the optimized quantum f-divergence and the proof that the data processing inequality holds for it whenever the function f is operator anti-monotone with domain $(0,\infty)$ and range $\mathbb{R}$. The proof of the data processing inequality relies on the operator Jensen inequality [26], and it bears some similarities to the original approach from [12]–[14]. Furthermore, I showed how the sandwiched Rényi relative entropies are particular examples of the optimized quantum f-divergence. As such, one benefit of this paper is that there is now a single, unified approach, based on the operator Jensen inequality [26], for establishing the data processing inequality for the Petz–Rényi and sandwiched Rényi relative entropies, for the full range of parameters for which it is known to hold.

Acknowledgements. I thank Anna Vershynina for discussions related to the topic of this paper, and I acknowledge support from the NSF under grant no. 1714215.

References

[1] H. Umegaki, "Conditional expectations in an operator algebra IV," *Kodai Math. Sem. Rep.*, vol. 14, no. 2, pp. 59–85, 1962.

[2] G. Lindblad, "Completely positive maps and entropy inequalities," *Comm. Math. Phys.*, vol. 40, no. 2, pp. 147–151, June 1975.

[3] A. Uhlmann, "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory," *Communications in Mathematical Physics*, vol. 54, no. 1, pp. 21–32, 1977.

[4] 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, December 1991.

[5] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428–2433, November 2000.

[6] I. Bjelakovic and R. Siegmund-Schultze, "Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb," July 2012, arXiv:quant-ph/0307170.

[7] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, March 1951.

[8] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[9] I. Csiszár, "Information type measure of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299—-318, 1967.

[10] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 28, no. 1, pp. 131–142, 1966.

[11] T. Morimoto, "Markov processes and the h-theorem," *Journal of the Physical Society of Japan*, vol. 18, no. 3, pp. 328–331, 1963.

[12] D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publ. RIMS, Kyoto University*, vol. 21, pp. 787–800, 1985.

[13] ——, "Quasi-entropies for finite quantum systems," *Reports in Mathematical Physics*, vol. 23, pp. 57–65, 1986.

[14] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5840–5847, December 2009.

[15] V. P. Belavkin and P. Staszewski, "C*-algebraic generalization of relative entropy and entropy," *Annales de l'I.H.P. Physique théorique*, vol. 37, no. 1, pp. 51–58, 1982.

[16] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," November 2006, arXiv:quant-ph/0611289.

[17] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding," *Physical Review A*, vol. 76, no. 6, p. 062301, December 2007.

[18] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: a new definition and some properties," *J. Math. Phys.*, vol. 54, no. 12, p. 122203, December 2013.

[19] 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, October 2014.

[20] M. M. Wilde, M. Tomamichel, and M. Berta, "Converse bounds for private communication over quantum channels," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1792–1817, March 2017.

[21] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," *Comm. Math. Phys.*, vol. 334, no. 3, pp. 1617–1648, March 2015.

[22] F. Hiai, M. Mosonyi, D. Petz, and C. Beny, "Quantum f-divergences and error correction," *Reviews in Mathematical Physics*, vol. 23, no. 7, pp. 691–747, August 2011.

[23] D. Petz and M. B. Ruskai, "Contraction of generalized relative entropy under stochastic mappings on matrices," *Inf. Dim. Ana., Quantum Prob. and Related Topics*, vol. 1, no. 1, pp. 83–89, January 1998.

[24] F. Hiai and M. Mosonyi, "Different quantum f-divergences and the reversibility of quantum operations," *Reviews in Mathematical Physics*, vol. 29, no. 7, p. 1750023, August 2017.

[25] K. Matsumoto, "A new quantum version of f-divergence," 2013, arXiv:1311.4722.

[26] F. Hansen and G. K. Pedersen, "Jensen's operator inequality," *Bulletin London Math. Soc.*, vol. 35, no. 4, pp. 553–564, July 2003.

[27] R. Bhatia, *Matrix Analysis*. Springer, 1997.

[28] F. Hansen, "The fast track to Löwner's theorem," *Linear Algebra and its Applications*, vol. 438, no. 11, pp. 4557–4571, June 2013.

[29] W. F. Stinespring, "Positive functions on C*-algebras," *Proceedings of the American Mathematical Society*, vol. 6, pp. 211–216, 1955.

[30] 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, March 1986.

[31] ——, "Sufficiency of channels over von Neumann algebras," *Quarterly Journal of Mathematics*, vol. 39, no. 1, pp. 97–108, 1988.

[32] O. Klein, "Zur Quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre," *Z. Physik*, vol. 72, pp. 767–775, 1931.

[33] M. B. Ruskai, "Inequalities for quantum entropy: A review with conditions for equality," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4358–4375, September 2002.

[34] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," *J. Math. Phys.*, vol. 54, no. 12, p. 122201, December 2013.

[35] S. Beigi, "Sandwiched Rényi divergence satisfies data processing inequality," *J. Math. Phys.*, vol. 54, no. 12, p. 122202, December 2013.