8-10-2018

Optimized quantum f-divergences and data processing

Mark M. Wilde
Louisiana State University

Recommended Citation
Wilde, M. (2018). Optimized quantum f-divergences and data processing. *Journal of Physics A: Mathematical and Theoretical, 51* (37) https://doi.org/10.1088/1751-8121/aad5a1

Optimized quantum f-divergences and data processing

Mark M. Wilde*

July 25, 2018

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 paper, 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 paper 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. This paper discusses other aspects of the optimized f-divergence, such as the classical case, the classical–quantum case, and how to construct optimized f-information measures.

1 Introduction

The quantum relative entropy [Ume62] is a foundational distinguishability measure in quantum information theory. It is a function of two quantum states and measures how well one can tell the two states apart by a quantum-mechanical experiment. It is well known by now to be a parent quantity for many other information measures, such as entropy, mutual information, conditional entropy, and entanglement measures (see, e.g., [Dat11, Wil17]). One important reason for why it has found such widespread application is that it satisfies a data-processing inequality [Lin75, Uhl77]: 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 [HP91, ON00, BSS12]. Naturally, the notion of quantum relative entropy generalizes its classical counterpart [KL51], which enjoyed a rich and illustrious history prior to the development of quantum relative entropy.

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

*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

notable generalization is Rényi's relative entropy [Rén61], but this was subsequently generalized even further in the form of the f-divergence [Csi67, AS66, Mor63]. For probability distributions $\{p(x)\}_x$ and $\{q(x)\}_x$ and a convex function f, the f-divergence is defined as

$$\sum_x q(x) f(p(x)/q(x)), \tag{1.1}$$

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 $r(y|x)$ (a conditional probability distribution), 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 [Pet85, Pet86a], going under the name of "quasi-entropy" as used in [Weh79]. In [Pet85, Pet86a] and a later development [TCR09], the quantum data-processing inequality was proved in full generality for arbitrary quantum channels, whenever the underlying function f is *operator* convex. A relatively large literature on the topic of quantum f-divergence has now developed, so much that there are now many reviews and extensions of the original idea [OP93, PR98, NP05, PS09, TCR09, Sha10, Pet10b, Pet10a, HMPB11, HP12, Mat13, HM17].

Interestingly, when generalizing a notion from classical to quantum information theory, 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, and for states of many-particle quantum systems, entanglement is involved as well. 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 [Ume62] and [BS82]. 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 [HP91, ON00] accomplished this for the quantum relative entropy of [Ume62], and since then, researchers have realized more and more just how foundational the formula of [Ume62] is. As a consequence, the formula of [Ume62] 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 [Pet85, Pet86a] has been widely studied and given an operational interpretation [Nag06, Hay07], again in the context of quantum hypothesis testing (specifically, the error exponent problem). However, in recent years, the sandwiched Rényi relative entropy of [MLDS+13, WWY14] has gained prominence, due to its role in establishing strong converses for communication tasks [WWY14, GW15, TWW17, CMW16, DW15, WTB17]. The result of [MO15] solidified its fundamental meaning in quantum information theory: these authors proved that it has an operational interpretation in the strong converse exponent of quantum hypothesis testing. As such, the situation we are faced with is that there are two generalizations of Rényi relative entropy that should be considered in quantum information theory, due to their operational role mentioned above. There are further generalizations of the aforementioned quantum Rényi relative entropies [AD15], but their operational meaning (and thus their role in quantum information theory) is unclear.

The same work that introduced the Petz–Rényi relative entropy also introduced a quantum generalization of the notion of f-divergence [Pet85, Pet86a] (see also [HMPB11]), with the Petz–Rényi relative entropy being a particular example. Since then, other quantum f-divergences have appeared [PR98, HM17], now known as minimal and maximal f-divergences [Mat13, HM17]. However, hitherto it has not been known how the sandwiched Rényi relative entropy fits into the paradigm of quantum f-divergences. In fact, the authors of [HMPB11] declared in their Example 2.11 that a particular instance of the sandwiched Rényi relative entropy is not a quantum f-divergence, suggesting that it would not be possible to express it as such.

In this paper, I modify Petz's definition of quantum f-divergence [Pet85, Pet86a, HMPB11], 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 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 [HP03], which is the statement that

$$f(V^{\dagger}XV) \leq V^{\dagger}f(X)V, \tag{1.2}$$

where f is an operator convex function defined on an interval I, X is a Hermitian operator with spectrum in I, and V is an isometry. This unified approach 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 and then discuss various alternative ways of writing it, including its representation in terms of the relative modular operator formalism. In Section 3, I prove that the optimized f-divergence satisfies the quantum data processing inequality whenever the underlying function f is operator anti-monotone with domain $(0, \infty)$ and range $\mathbb{R}$. The proof of quantum data processing has two steps: I first prove that the optimized quantum f-divergence is invariant under isometric embeddings and then show that it is monotone non-increasing under the action of a partial trace. By the Stinespring dilation theorem [Sti55], these two steps establish data processing under general quantum channels. The core tool underlying both steps is the operator Jensen inequality [HP03]. The proof of monotonicity under partial trace features some novel aspects for handling non-invertible operators. In Section 4, I show how the quantum relative entropy and the sandwiched Rényi relative entropies are directly related to the optimized quantum f-divergence. Section 5 then discusses the relation between Petz's f-divergence and the optimized one. Section 6 shows how the optimized f-divergence simplifies when the operators involved have a classical or classical–quantum form. In Section 7, I discuss how to construct several information measures from the optimized f-divergence, which could potentially find application in quantum information theory or resource theories. I finally conclude in Section 8 with a summary and some open directions.

2 Optimized quantum f-divergence

Let us begin by formally defining the optimized quantum f-divergence:

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

$$\widetilde{Q}_f(X\|Y) \equiv \sup_{\tau>0,\ \mathrm{Tr}\{\tau\}\leq 1,\ \varepsilon>0} \widetilde{Q}_f(X\|Y+\varepsilon\Pi_Y^{\perp};\tau), \tag{2.1}$$

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

$$\widetilde{Q}_f(X\|Z;\tau) \equiv \langle\varphi^X|_{S\hat{S}} f(\tau_S^{-1}\otimes Z_{\hat{S}}^T)|\varphi^X\rangle_{S\hat{S}}, \tag{2.2}$$

$$|\varphi^X\rangle_{S\hat{S}} \equiv (X_S^{1/2}\otimes I_{\hat{S}})|\Gamma\rangle_{S\hat{S}}. \tag{2.3}$$

In the above, $\Pi_Y^{\perp}$ denotes the projection onto the kernel of Y, $\mathcal{H}_{\hat{S}}$ is an auxiliary Hilbert space isomorphic to $\mathcal{H}_S$,

$$|\Gamma\rangle_{S\hat{S}} \equiv \sum_{i=1}^{|S|} |i\rangle_S\, |i\rangle_{\hat{S}}\,, \tag{2.4}$$

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 $\{|i\rangle_{\hat{S}}\}_i$.

Remark 2 *Note that the expression in (2.1) simplifies considerably in the case that Y is positive definite. That is, it reduces to the following simpler expression in the case that $Y>0$:*

$$\widetilde{Q}_f(X\|Y) = \sup_{\tau>0,\ \mathrm{Tr}\{\tau\}\leq 1} \langle\varphi^X|_{S\hat{S}} f(\tau_S^{-1}\otimes Y_{\hat{S}}^T)|\varphi^X\rangle_{S\hat{S}}. \tag{2.5}$$

As such, the optimized f-divergence in (2.1) represents a modification of Petz's quantum f-divergence, a topic that I discuss in more detail in Section 5. The intention of the more general definition in (2.1) is to provide a consistent way of defining the optimized f-divergence in the case that Y is not positive semi-definite.

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., [Bha97]). This property is rather strong, but there are several functions of interest in quantum-physical applications that obey it (see Section 4). 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., [Han13]). In this case, we have the following proposition:

Proposition 3 *Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. For positive semi-definite operators X and Y acting on a Hilbert space $\mathcal{H}_S$, the following equality holds*

$$\widetilde{Q}_f(X\|Y) = \sup_{\tau>0,\,\mathrm{Tr}\{\tau\}=1} \lim_{\varepsilon\searrow 0} \widetilde{Q}_f(X\|Y+\varepsilon\Pi_Y^{\perp};\tau), \tag{2.6}$$

and furthermore, the function $\widetilde{Q}_f(X\|Y+\varepsilon\Pi_Y^{\perp};\tau)$ is concave in τ. Finally, for positive semi-definite Y_1 and Y_2 such that $Y_1\leq Y_2$, we have that

$$\widetilde{Q}_f(X\|Y_1) \geq \widetilde{Q}_f(X\|Y_2). \tag{2.7}$$

Proof. To see that we can restrict the optimization over τ to τ satisfying $\operatorname{Tr}\{\tau\} = 1$, let τ be such that $\tau > 0$ and $\operatorname{Tr}\{\tau\} < 1$. Then

$$\tau_S^{-1} \otimes Y_{\hat{S}}^T = \frac{1}{\operatorname{Tr}\{\tau_S\}} \left[\frac{\tau_S}{\operatorname{Tr}\{\tau_S\}}\right]^{-1} \otimes Y_{\hat{S}}^T \geq \left[\frac{\tau_S}{\operatorname{Tr}\{\tau_S\}}\right]^{-1} \otimes Y_{\hat{S}}^T, \tag{2.8}$$

and so

$$\widetilde{Q}_f(X\|Y + \varepsilon\Pi_Y^\perp; \tau) \leq \widetilde{Q}_f(X\|Y + \varepsilon\Pi_Y^\perp; \tau/\operatorname{Tr}\{\tau\}) \tag{2.9}$$

from the operator anti-monotonicity of f. Changing $\sup_{\varepsilon>0}$ to $\lim_{\varepsilon\searrow 0}$ follows as well from operator anti-monotonicity of f. Let $\varepsilon_2 \geq \varepsilon_1 > 0$. Then $Y + \varepsilon_1\Pi_Y^\perp \leq Y + \varepsilon_2\Pi_Y^\perp$ and so $\widetilde{Q}_f(X\|Y + \varepsilon_2\Pi_Y^\perp; \tau) \leq \widetilde{Q}_f(X\|Y + \varepsilon_1\Pi_Y^\perp; \tau)$. So then the highest value of $\widetilde{Q}_f(X\|Y + \varepsilon\Pi_Y^\perp; \tau)$ is achieved in the limit as $\varepsilon \searrow 0$, where we have also invoked the continuity of f. We note that this limit could evaluate to infinity.

Concavity in τ follows because

$$f(\tau_S^{-1} \otimes Y_{\hat{S}}^T) = f\left(\left[\tau_S \otimes \left(Y_{\hat{S}}^T\right)^{-1}\right]^{-1}\right), \tag{2.10}$$

and the function $f(x^{-1})$ is operator monotone on $(0, \infty)$, given that it is the composition of the operator anti-monotone function x^{-1} with domain $(0, \infty)$ and range $(0, \infty)$ and the function f, taken to be operator anti-monotone on $(0, \infty)$ by hypothesis. Since $f(x^{-1})$ is operator monotone on $(0, \infty)$, it is operator concave (see, e.g., [Han13]).

The dominating property in (2.7) follows from the fact that f is operator anti-monotone on $(0, \infty)$, which implies the following for a fixed $\varepsilon > 0$ and τ_S such that $\tau_S > 0$ and $\operatorname{Tr}\{\tau_S\} \leq 1$:

$$f(\tau_S^{-1} \otimes (Y_1 + \varepsilon\Pi_{Y_1}^\perp)_{\hat{S}}^T) \geq f(\tau_S^{-1} \otimes (Y_2 + \varepsilon\Pi_{Y_1}^\perp)_{\hat{S}}^T). \tag{2.11}$$

We arrive at the inequality in (2.7) after sandwiching by $|\varphi^X\rangle_{S\hat{S}}$, taking the limit as $\varepsilon \searrow 0$, and taking a supremum over τ. ■

For X positive semi-definite and Y and τ positive definite, with spectral decompositions of Y and τ given as

$$Y = \sum_y \mu_y |\phi^y\rangle\langle\phi^y|, \qquad \tau = \sum_t \nu_t |\psi^t\rangle\langle\psi^t|, \tag{2.12}$$

we can write

$$\widetilde{Q}_f(X\|Y; \tau) = \sum_{y,t} f(\mu_y \nu_t^{-1}) \operatorname{Tr}\{X^{1/2}|\phi^y\rangle\langle\phi^y|X^{1/2}|\psi^t\rangle\langle\psi^t|\} \tag{2.13}$$

$$= \sum_{y,t} f(\mu_y \nu_t^{-1}) |\langle\phi^y|X^{1/2}|\psi^t\rangle|^2, \tag{2.14}$$

by using the facts that

$$f(\tau_S^{-1} \otimes Y_{\hat{S}}^T) = \sum_{y,t} f(\mu_y \nu_t^{-1}) |\psi^t\rangle\langle\psi^t|_S \otimes |\phi^y\rangle\langle\phi^y|_{\hat{S}}^T, \tag{2.15}$$

$$(I_S \otimes Z_{\hat{S}}^T)|\Gamma\rangle_{S\hat{S}} = (Z_S \otimes I_{\hat{S}})|\Gamma\rangle_{S\hat{S}}, \tag{2.16}$$

$$\langle\Gamma|_{S\hat{S}}(Z_S \otimes I_{\hat{S}})|\Gamma\rangle_{S\hat{S}} = \operatorname{Tr}\{Z_S\}, \tag{2.17}$$

for any square operator Z acting on $\mathcal{H}_S$. The formula in (2.13) is helpful in some parts of our analysis below.

We can also phrase Definition 1 in terms of the relative modular operator formalism, which is employed in many of the works on quasi-entropy (many details of this formalism in the context of quasi-entropies are available in [HMPB11]). Let P be a positive semi-definite operator, and let R be a positive definite operator. Defining the action of the relative modular operator $\Delta(P/R)$ on an operator X as

$$\Delta(P/R)(X) = PXR^{-1}, \tag{2.18}$$

and the Hilbert–Schmidt inner product $\langle W, Z \rangle = \operatorname{Tr}\{W^\dagger Z\}$, we can write the quantity $\widetilde{Q}_f(X\|Y;\tau)$ underlying $\widetilde{Q}_f(X\|Y)$ in terms of the relative modular operator as

$$\widetilde{Q}_f(X\|Y;\tau) = \langle X^{1/2}, f(\Delta(Y/\tau))(X^{1/2})\rangle. \tag{2.19}$$

The definition of optimized quantum f-divergence following from plugging (2.19) into (2.1) can be used in more general contexts than those considered in the present paper (for example, in the context of von Neumann algebras). However, in this work, we find it more convenient to work with the expression in (2.2) (see [TCR09, Sha10] for a similar approach), and throughout this paper, we work in the setting of finite-dimensional quantum systems.

3 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 $\mathcal{N}$:

$$\widetilde{Q}_f(X\|Y) \geq \widetilde{Q}_f(\mathcal{N}(X)\|\mathcal{N}(Y)). \tag{3.1}$$

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 quantum information-theoretic contexts, the inequality in (3.1) is known as the quantum data processing inequality. According to the Stinespring dilation theorem [Sti55], every quantum channel can be realized by an isometric embedding of its input into a tensor product of the channel's output Hilbert space and an auxiliary Hilbert space, followed by a partial trace over the auxiliary Hilbert space. That is, to every quantum channel $\mathcal{N}_{S\to B}$, there exists an isometry $U^{\mathcal{N}}_{S\to BE}$ such that

$$\mathcal{N}_{S\to B}(X_S) = \operatorname{Tr}_E\{U^{\mathcal{N}}_{S\to BE} X_S \left(U^{\mathcal{N}}_{S\to BE}\right)^\dagger\}. \tag{3.2}$$

As such, we can prove the inequality in (3.1) in two steps:

1. *Isometric invariance*: First show that

$$\widetilde{Q}_f(X\|Y) = \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger) \tag{3.3}$$

for any isometry U and any positive semi-definite X and Y.[1]

[1] The importance of establishing isometric invariance of quasi-entropies has been stressed in [TCR09] and [Tom12, Appendix B].

2. *Monotonicity under partial trace*: Then show that

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A) \tag{3.4}$$

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 $Y_A = \mathrm{Tr}_B\{Y_{AB}\}$.

So we proceed and first prove isometric invariance:

Proposition 4 (Isometric invariance) *Let $U : \mathcal{H}_S \to \mathcal{H}_R$ be an isometry, let X and Y be positive semi-definite operators, and let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Then the following equality holds*

$$\widetilde{Q}_f(X\|Y) = \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger). \tag{3.5}$$

Proof. In the case that $\dim(\mathcal{H}_S) = \dim(\mathcal{H}_R)$, the statement holds trivially because U is a unitary and then $\mathcal{H}_S$ and $\mathcal{H}_R$ are isomorphic. So we focus on the case in which $\dim(\mathcal{H}_S) < \dim(\mathcal{H}_R)$. First suppose that Y is invertible when acting on $\mathcal{H}_S$. The operator X is generally not invertible, and with respect to the decomposition of $\mathcal{H}_S$ as $\mathrm{supp}(X) \oplus \ker(X)$, we can write X and each eigenprojection $|\phi^y\rangle\langle\phi^y|$ of Y respectively as

$$\begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} \phi^y_{11} & \phi^y_{12} \\ \phi^y_{21} & \phi^y_{22} \end{bmatrix}. \tag{3.6}$$

Let τ acting on $\mathcal{H}_S$ be such that $\tau > 0$ and $\mathrm{Tr}\{\tau\} = 1$. Suppose that its spectral decomposition is given by $\sum_{t=1}^{|S|} \nu_t |\psi^t\rangle\langle\psi^t|$, with each $\nu_t \in (0,1)$ and $|\psi^t\rangle$ a unit vector such that $\sum_t \nu_t = 1$. We can then write each eigenprojection $|\psi^t\rangle\langle\psi^t|$ with respect to the decomposition of $\mathcal{H}_S$ as $\mathrm{supp}(X) \oplus \ker(X)$ as

$$\begin{bmatrix} \psi^t_{11} & \psi^t_{12} \\ \psi^t_{21} & \psi^t_{22} \end{bmatrix}. \tag{3.7}$$

Applying definitions and (2.13), we then find that

$$\widetilde{Q}_f(X\|Y;\tau) = \sum_{y,t} f(\mu_y \nu_t^{-1}) \,\mathrm{Tr}\left\{ \begin{bmatrix} \sqrt{X} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \phi^y_{11} & \phi^y_{12} \\ \phi^y_{21} & \phi^y_{22} \end{bmatrix} \begin{bmatrix} \sqrt{X} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \psi^t_{11} & \psi^t_{12} \\ \psi^t_{21} & \psi^t_{22} \end{bmatrix} \right\} \tag{3.8}$$

$$= \sum_{y,t} f(\mu_y \nu_t^{-1}) \,\mathrm{Tr}\{\sqrt{X}\phi^y_{11}\sqrt{X}\psi^t_{11}\}. \tag{3.9}$$

Now consider $\widetilde{Q}_f$ for $UXU^\dagger$ and $UYU^\dagger$. Without loss of generality, we can consider the isometry U to be the trivial embedding of $\mathcal{H}_S$ into the larger Hilbert space $\mathcal{H}_R$, with it decomposed as $\mathcal{H}_R = \mathcal{H}_S \oplus \mathcal{H}_S^\perp$, so that with respect to the decomposition of $\mathcal{H}_R$ as $\mathcal{H}_R = \mathrm{supp}(X) \oplus \ker(X) \oplus \mathcal{H}_S^\perp$, we can write X and each eigenprojection $|\phi^y\rangle\langle\phi^y|$ of Y in the larger Hilbert space $\mathcal{H}_R$ as

$$\begin{bmatrix} X & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} \phi^y_{11} & \phi^y_{12} & 0 \\ \phi^y_{21} & \phi^y_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{3.10}$$

We use the notation X_R to denote the operator X embedded into $\mathcal{H}_R$. Let ω acting on $\mathcal{H}_R$ be such that $\omega > 0$ and $\mathrm{Tr}\{\omega\} = 1$. Suppose that its spectral decomposition is given by $\sum_{s=1}^{|R|} \lambda_s |\varphi^s\rangle\langle\varphi^s|$, with each $\lambda_s \in (0,1)$ and $|\varphi^s\rangle$ a unit vector such that $\sum_s \lambda_s = 1$. We can then write each eigenprojection $|\varphi^s\rangle\langle\varphi^s|$ as

$$\begin{bmatrix} \varphi^s_{11} & \varphi^s_{12} & \varphi^s_{13} \\ \varphi^s_{21} & \varphi^s_{22} & \varphi^s_{23} \\ \varphi^s_{31} & \varphi^s_{32} & \varphi^s_{33} \end{bmatrix}. \tag{3.11}$$

Since Y is no longer invertible after the embedding, we need to instead consider the operator $Y_R + \varepsilon\Pi^\perp_Y$ for some $\varepsilon \in (0,1)$, where we use the notation Y_R to denote the operator Y embedded into $\mathcal{H}_R$. Then the eigenprojections of $Y_R + \varepsilon\Pi^\perp_Y$ are now represented in this larger space as

$$\begin{bmatrix} \phi^y_{11} & \phi^y_{12} & 0 \\ \phi^y_{21} & \phi^y_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{bmatrix}. \tag{3.12}$$

Applying definitions, we then find that, in the larger Hilbert space $\mathcal{H}_R$,

$$\begin{aligned} &\widetilde{Q}_f(X_R \| Y_R + \varepsilon\Pi^\perp_Y; \omega) \\ &= \sum_{y,s} f(\mu_y \lambda_s^{-1}) \operatorname{Tr}\left\{ \begin{bmatrix} \sqrt{X} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi^y_{11} & \phi^y_{12} & 0 \\ \phi^y_{21} & \phi^y_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{X} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varphi^s_{11} & \varphi^s_{12} & \varphi^s_{13} \\ \varphi^s_{21} & \varphi^s_{22} & \varphi^s_{23} \\ \varphi^s_{31} & \varphi^s_{32} & \varphi^s_{33} \end{bmatrix} \right\} \\ &\quad + \sum_{t'} f(\varepsilon \lambda_s^{-1}) \operatorname{Tr}\left\{ \begin{bmatrix} \sqrt{X} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \sqrt{X} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varphi^s_{11} & \varphi^s_{12} & \varphi^s_{13} \\ \varphi^s_{21} & \varphi^s_{22} & \varphi^s_{23} \\ \varphi^s_{31} & \varphi^s_{32} & \varphi^s_{33} \end{bmatrix} \right\} && (3.13) \\ &= \sum_{y,s} f(\mu_y \lambda_s^{-1}) \operatorname{Tr}\{\sqrt{X}\phi^y_{11}\sqrt{X}\varphi^s_{11}\}. && (3.14) \end{aligned}$$

We now compare the expressions in (3.9) and (3.14). For a given $\tau > 0$ with spectral decomposition $\sum_{t=1}^{|S|} \nu_t |\psi^t\rangle\langle\psi^t|$ and $\delta \in (0,1)$, we can choose $\omega(\delta) > 0$ as

$$\omega(\delta) = (1-\delta) \sum_t \nu_t \begin{bmatrix} \psi^t_{11} & \psi^t_{12} & 0 \\ \psi^t_{21} & \psi^t_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \delta \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I/\dim(\mathcal{H}^\perp_S) \end{bmatrix}, \tag{3.15}$$

and then, by the above reasoning, we have that

$$\widetilde{Q}_f(X_R \| Y_R + \varepsilon\Pi^\perp_Y; \omega(\delta)) = \sum_{y,t} f(\mu_y \left[(1-\delta)\nu_t\right]^{-1}) \operatorname{Tr}\{\sqrt{X}\phi^y_{11}\sqrt{X}\psi^t_{11}\}. \tag{3.16}$$

Taking the limit $\delta \searrow 0$ and applying the continuity of f then gives

$$\lim_{\delta\searrow 0} \widetilde{Q}_f(X_R \| Y_R + \varepsilon\Pi^\perp_Y; \omega(\delta)) = \widetilde{Q}_f(X\|Y;\tau). \tag{3.17}$$

So it is clear that the following inequality holds for all τ:

$$\widetilde{Q}_f(X\|Y;\tau) \leq \sup_{\omega>0,\ \operatorname{Tr}\{\omega\}=1} \widetilde{Q}_f(X_R \| Y_R + \varepsilon\Pi^\perp_Y; \omega). \tag{3.18}$$

We can thus conclude that

$$\sup_{\tau>0,\ \operatorname{Tr}\{\tau\}=1} \widetilde{Q}_f(X\|Y;\tau) \leq \sup_{\omega>0,\ \operatorname{Tr}\{\omega\}=1} \widetilde{Q}_f(X_R \| Y_R + \varepsilon\Pi^\perp_Y; \omega), \tag{3.19}$$

which is the same as the inequality

$$\widetilde{Q}_f(X\|Y) \leq \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger). \tag{3.20}$$

This establishes the inequality $\widetilde{Q}_f(X\|Y) \leq \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger)$ in the case in which Y is invertible when acting on $\mathcal{H}_S$.

Given that the function x^{-1} is operator anti-monotone on $(0,\infty)$ and has range $(0,\infty)$, it follows that $f(x^{-1}) = g(x)$ is operator monotone on $(0,\infty)$ and thus operator concave [Han13]. Defining the embedding isometry $V \equiv V_{S\to R} \otimes V_{\hat{S}\to\hat{R}} \equiv \sum_i |i\rangle_R\langle i|_S \otimes \sum_j |j\rangle_{\hat{R}}\langle j|_{\hat{S}}$, we then have by a direct application of the operator Jensen inequality [HP03] and the fact that g is operator concave that

$$\begin{aligned}
\widetilde{Q}_f(X\|Y+\varepsilon\Pi_Y^\perp;\omega) &= \langle\varphi^X|_{R\hat{R}} f(\omega_R^{-1}\otimes\left[Y_{\hat{R}}+\varepsilon\Pi_Y^\perp\right]^T)|\varphi^X\rangle_{R\hat{R}} && (3.21)\\
&= \langle\varphi^X|_{S\hat{S}} V^\dagger f\left(\left[\omega_R\otimes\left[Y_{\hat{R}}^{-1}+\varepsilon^{-1}\Pi_Y^\perp\right]^T\right]^{-1}\right) V|\varphi^X\rangle_{S\hat{S}} && (3.22)\\
&= \langle\varphi^X|_{S\hat{S}} V^\dagger g\left(\omega_R\otimes\left[Y_{\hat{R}}^{-1}+\varepsilon^{-1}\Pi_Y^\perp\right]^T\right) V|\varphi^X\rangle_{S\hat{S}} && (3.23)\\
&\leq \langle\varphi^X|_{S\hat{S}} g\left(V^\dagger\left[\omega_R\otimes\left[Y_{\hat{R}}^{-1}+\varepsilon^{-1}\Pi_Y^\perp\right]^T\right]V\right)|\varphi^X\rangle_{S\hat{S}} && (3.24)\\
&= \langle\varphi^X|_{S\hat{S}} g\left(\omega'_S\otimes\left[Y_{\hat{S}}^{-1}\right]^T\right)|\varphi^X\rangle_{S\hat{S}} && (3.25)\\
&= \langle\varphi^X|_{S\hat{S}} f((\omega'_S)^{-1}\otimes Y_{\hat{S}}^T)|\varphi^X\rangle_{S\hat{S}} && (3.26)\\
&= \widetilde{Q}_f(X\|Y;\omega'_S) && (3.27)\\
&\leq \widetilde{Q}_f(X\|Y), && (3.28)
\end{aligned}$$

where $\omega'_S \equiv (V_{S\to R})^\dagger\omega_R V_{S\to R}$ is an operator acting on $\mathcal{H}_S$ such that $\omega'_S > 0$ and $\text{Tr}\{\omega'_S\} \leq 1$. In the above, the notation $Y_{\hat{R}}^{-1}$ indicates the inverse on the support of $Y_{\hat{R}}^{-1}$, and we have employed the facts that $\left[Y_{\hat{R}}+\varepsilon\Pi_Y^\perp\right]^{-1} = Y_{\hat{R}}^{-1}+\varepsilon^{-1}\Pi_Y^\perp$ and $(V_{\hat{S}\to\hat{R}})^\dagger\left(Y_{\hat{R}}^{-1}+\varepsilon^{-1}\Pi_Y^\perp\right)V_{\hat{S}\to\hat{R}} = Y_{\hat{R}}^{-1}$. Since the inequality holds for all $\omega > 0$ such that $\text{Tr}\{\omega\} = 1$, we conclude that

$$\widetilde{Q}_f(UXU^\dagger\|UYU^\dagger) \leq \widetilde{Q}_f(X\|Y). \tag{3.29}$$

We have now established the claim for invertible Y.

If Y is not invertible when acting on $\mathcal{H}_S$, then the definition in (2.1) applies, which in fact forces Y to become invertible when acting on $\mathcal{H}_S$. So then, in this case, we can conclude that

$$\widetilde{Q}_f(X\|Y+\varepsilon\Pi_Y^\perp) = \widetilde{Q}_f(UXU^\dagger\|U\left[Y+\varepsilon\Pi_Y^\perp\right]U^\dagger+\varepsilon\Pi_{\mathcal{H}_S^\perp}). \tag{3.30}$$

So the quantities are the same for all $\varepsilon \in (0,1)$, and then the equality follows by taking a supremum over $\varepsilon > 0$. ■

Remark 5 *The above proof establishes the inequality $\widetilde{Q}_f(X\|Y) \leq \widetilde{Q}_f(UXU^\dagger\|UYU^\dagger)$ for any continuous function f with domain $(0,\infty)$ and range $\mathbb{R}$, but for the opposite inequality $\widetilde{Q}_f(UXU^\dagger\|UYU^\dagger) \leq \widetilde{Q}_f(X\|Y)$, the proof given requires f to be operator anti-monotone with domain $(0,\infty)$ and range $\mathbb{R}$.*

We now complete the second step toward quantum data processing, as mentioned above:

Proposition 6 (Monotonicity under partial trace) *Given positive semi-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*

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A), \tag{3.31}$$

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

Proof. Throughout the proof, we take Y_{AB} to be invertible on $\mathcal{H}_A \otimes \mathcal{H}_B$. We can do so because the supremum over $\varepsilon > 0$ can be placed on the very outside, as in Definition 1, and then we can optimize over $\varepsilon > 0$ at the very end once the monotonicity inequality has been established. There are three cases to consider:

1. when $X_{AB} > 0$,
2. when $X_A > 0$, but X_{AB} is not invertible, and
3. when X_A is not invertible.

Here I show a proof for the first two cases, and the last case is shown in detail in the appendix for the interested reader.

Case $X_{AB} > 0$: We establish the claim when X_{AB} is invertible, and so X_A is as well. This is the simplest case to consider and thus has the most transparent proof (it is fruitful to understand this case well before considering the other cases). The quantities of interest are as follows:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB};\tau_{AB}) = \langle \varphi^{X_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{B}}^T)|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}}, \tag{3.32}$$

$$\widetilde{Q}_f(X_A\|Y_A;\omega_A) = \langle \varphi^{X_A}|_{A\hat{A}} f(\omega_A^{-1} \otimes Y_{\hat{A}}^T)|\varphi^{X_A}\rangle_{A\hat{A}}, \tag{3.33}$$

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

$$|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} = \left(X_{AB}^{1/2} \otimes I_{\hat{A}\hat{B}}\right)|\Gamma\rangle_{A\hat{A}} \otimes |\Gamma\rangle_{B\hat{B}}. \tag{3.34}$$

The following map, acting on an operator Z_A, is a quantum channel known as the Petz recovery channel [Pet86b, Pet88] (see also [BK02, HJPW04, LS13]):

$$Z_A \to X_{AB}^{1/2}\left(\left[X_A^{-1/2} Z_A X_A^{-1/2}\right] \otimes I_B\right) X_{AB}^{1/2}. \tag{3.35}$$

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 trace preserving because

$$\mathrm{Tr}\left\{X_{AB}^{1/2}\left(\left[X_A^{-1/2} Z_A X_A^{-1/2}\right] \otimes I_B\right) X_{AB}^{1/2}\right\} = \mathrm{Tr}\left\{X_{AB}\left(\left[X_A^{-1/2} Z_A X_A^{-1/2}\right] \otimes I_B\right)\right\} \tag{3.36}$$

$$= \mathrm{Tr}\left\{X_A\left[X_A^{-1/2} Z_A X_A^{-1/2}\right]\right\} \tag{3.37}$$

$$= \mathrm{Tr}\left\{X_A^{-1/2} X_A X_A^{-1/2} Z_A\right\} \tag{3.38}$$

$$= \mathrm{Tr}\{Z_A\}. \tag{3.39}$$

The Petz recovery channel has the property that it perfectly recovers X_{AB} if X_A is input because

$$X_A \to X_{AB}^{1/2} \left(\left[X_A^{-1/2} X_A X_A^{-1/2} \right] \otimes I_B \right) X_{AB}^{1/2} = X_{AB}. \tag{3.40}$$

Every completely positive and trace preserving map $\mathcal{N}$ has a Kraus decomposition, which is a set $\{K_i\}_i$ of operators such that

$$\mathcal{N}(\cdot) = \sum_i K_i (\cdot) K_i^\dagger, \qquad \sum_i K_i^\dagger K_i = I. \tag{3.41}$$

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

$$V = \sum_i K_i \otimes |i\rangle_E. \tag{3.42}$$

One can then readily check that $\mathcal{N}(\cdot) = \text{Tr}_E\{V(\cdot)V^\dagger\}$ and $V^\dagger V = I$. (See, e.g., [Wil17] for a review of these standard notions.) 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

$$X_{AB}^{1/2} \left(\left[X_A^{-1/2} Z_A X_A^{-1/2} \right] \otimes I_B \right) X_{AB}^{1/2} = \sum_{j=1}^{|B|} X_{AB}^{1/2} \left(\left[X_A^{-1/2} Z_A X_A^{-1/2} \right] \otimes |j\rangle\langle j|_B \right) X_{AB}^{1/2} \tag{3.43}$$

$$= \sum_{j=1}^{|B|} X_{AB}^{1/2} \left[X_A^{-1/2} \otimes |j\rangle_B \right] Z_A \left[X_A^{-1/2} \otimes \langle j|_B \right] X_{AB}^{1/2}. \tag{3.44}$$

Thus, Kraus operators for the Petz recovery channel are given by

$$\left\{ X_{AB}^{1/2} \left[X_A^{-1/2} \otimes |j\rangle_B \right] \right\}_{j=1}^{|B|}. \tag{3.45}$$

According to the standard recipe in (3.42), we can construct an isometric extension of the Petz recovery channel as

$$\sum_{j=1}^{|B|} X_{AB}^{1/2} \left[X_A^{-1/2} \otimes |j\rangle_B \right] |j\rangle_{\hat{B}} = X_{AB}^{1/2} X_A^{-1/2} \sum_{j=1}^{|B|} |j\rangle_B |j\rangle_{\hat{B}} \tag{3.46}$$

$$= X_{AB}^{1/2} X_A^{-1/2} |\Gamma\rangle_{B\hat{B}}. \tag{3.47}$$

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

$$V_{A\hat{A}\to A\hat{A}B\hat{B}} \equiv X_{AB}^{1/2} \left[X_A^{-1/2} \otimes I_{\hat{A}} \right] |\Gamma\rangle_{B\hat{B}}. \tag{3.48}$$

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 $|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}}$:

$$|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} = V_{A\hat{A}\to A\hat{A}B\hat{B}} |\varphi^{X_A}\rangle_{A\hat{A}}. \tag{3.49}$$

This can be interpreted as a generalization of (3.40) in the language of quantum information: 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:

$$\tau_{AB} = X_{AB}^{1/2}\left(\left[X_A^{-1/2}\omega_A X_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2}. \tag{3.50}$$

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

$$\begin{aligned}
&V^\dagger\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T\right)V \\
&= \left(\langle\Gamma|_{B\hat{B}}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]X_{AB}^{1/2}\right)\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T\right)\left(X_{AB}^{1/2}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]|\Gamma\rangle_{B\hat{B}}\right) &(3.51)\\
&= \langle\Gamma|_{B\hat{B}}\left(X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2}\otimes Y_{\hat{A}\hat{B}}^T\right)|\Gamma\rangle_{B\hat{B}} &(3.52)\\
&= \langle\Gamma|_{B\hat{B}}\left(\omega_A^{-1}\otimes Y_{\hat{A}\hat{B}}^T\right)|\Gamma\rangle_{B\hat{B}} &(3.53)\\
&= \omega_A^{-1}\otimes\langle\Gamma|_{B\hat{B}}Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} &(3.54)\\
&= \omega_A^{-1}\otimes Y_{\hat{A}}^T. &(3.55)
\end{aligned}$$

For the fourth equality, we used the fact that $\tau_{AB}^{-1} = X_{AB}^{-1/2}([X_A^{1/2}\omega_A^{-1}X_A^{1/2}]\otimes I_B)X_{AB}^{-1/2}$ for the choice of τ_{AB} in (3.50). With this setup, we can now readily establish the desired inequality by employing the operator Jensen inequality [HP03] and operator convexity of the function f:

$$\begin{aligned}
\widetilde{Q}_f(X_{AB}\|Y_{AB};\tau_{AB}) &= \langle\varphi^{X_{AB}}|_{AB\hat{A}\hat{B}}f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} &(3.56)\\
&= \langle\varphi^{X_A}|_{A\hat{A}}V^\dagger f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)V|\varphi^{X_A}\rangle_{A\hat{A}} &(3.57)\\
&\geq \langle\varphi^{X_A}|_{A\hat{A}}f(V^\dagger[\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T]V)|\varphi^{X_A}\rangle_{A\hat{A}} &(3.58)\\
&= \langle\varphi^{X_A}|_{A\hat{A}}f(\omega_A^{-1}\otimes Y_{\hat{A}}^T)|\varphi^{X_A}\rangle_{A\hat{A}} &(3.59)\\
&= \widetilde{Q}_f(X_A\|Y_A;\omega_A). &(3.60)
\end{aligned}$$

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:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A;\omega_A). \tag{3.61}$$

After taking a supremum over invertible states ω_A, we find that the inequality in (3.31) holds when X_{AB} is invertible.

Case $X_A > 0$, but X_{AB} not invertible: Consider the following isometry:

$$V_{A\hat{A}\to AB\hat{A}\hat{B}} \equiv X_{AB}^{1/2}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]|\Gamma\rangle_{B\hat{B}}. \tag{3.62}$$

The operator $V_{A\hat{A}\to A\hat{A}B\hat{B}}$ is indeed an isometry because

$$V^\dagger V = \left(\langle\Gamma|_{B\hat{B}}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]X_{AB}^{1/2}\right)\left(X_{AB}^{1/2}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]|\Gamma\rangle_{B\hat{B}}\right) \tag{3.63}$$

$$= \langle\Gamma|_{B\hat{B}}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]X_{AB}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right]|\Gamma\rangle_{B\hat{B}} \tag{3.64}$$

$$= \left[X_A^{-1/2}\otimes I_{\hat{A}}\right]\langle\Gamma|_{B\hat{B}}X_{AB}|\Gamma\rangle_{B\hat{B}}\left[X_A^{-1/2}\otimes I_{\hat{A}}\right] \tag{3.65}$$

$$= \left[X_A^{-1/2}\otimes I_{\hat{A}}\right]X_A\left[X_A^{-1/2}\otimes I_{\hat{A}}\right] \tag{3.66}$$

$$= X_A^{-1/2}X_AX_A^{-1/2}\otimes I_{\hat{A}} \tag{3.67}$$

$$= I_A\otimes I_{\hat{A}}. \tag{3.68}$$

Then, for $\delta\in(0,1)$ and ω_A an invertible density operator, take τ_{AB} to be the following invertible density operator:

$$\tau_{AB} = (1-\delta)\,X_{AB}^{1/2}\left(\left[X_A^{-1/2}\omega_AX_A^{-1/2}\right]\otimes I_B\right)X_{AB}^{1/2}+\delta\xi_{AB}, \tag{3.69}$$

where ξ_{AB} is some invertible density operator. We then find that

$$V^\dagger(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)V$$

$$= \langle\Gamma|_{B\hat{B}}X_A^{-1/2}X_{AB}^{1/2}(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^T)X_{AB}^{1/2}X_A^{-1/2}|\Gamma\rangle_{B\hat{B}} \tag{3.70}$$

$$= \langle\Gamma|_{B\hat{B}}X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2}\otimes Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} \tag{3.71}$$

$$= \langle\Gamma|_{B\hat{B}}\omega_A^{-1/2}\omega_A^{1/2}X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2}\omega_A^{1/2}\omega_A^{-1/2}\otimes Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} \tag{3.72}$$

$$\leq \langle\Gamma|_{B\hat{B}}\omega_A^{-1/2}\left\|\omega_A^{1/2}X_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1}X_{AB}^{1/2}X_A^{-1/2}\omega_A^{1/2}\right\|_\infty\omega_A^{-1/2}\otimes Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} \tag{3.73}$$

$$= \left\|\tau_{AB}^{-1/2}X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1/2}\right\|_\infty\langle\Gamma|_{B\hat{B}}\omega_A^{-1}\otimes Y_{\hat{A}\hat{B}}^T|\Gamma\rangle_{B\hat{B}} \tag{3.74}$$

$$\leq \frac{1}{1-\delta}\left[\omega_A^{-1}\otimes Y_{\hat{A}}^T\right]. \tag{3.75}$$

The last equality follows because $\|Z^\dagger Z\|_\infty = \|ZZ^\dagger\|_\infty$ for any operator Z (here we set $Z = \tau_{AB}^{-1/2}X_{AB}^{1/2}X_A^{-1/2}\omega_A^{1/2}$. The last inequality follows because

$$\left\|\tau_{AB}^{-1/2}X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1/2}\right\|_\infty$$

$$= \inf\left\{\mu : \tau_{AB}^{-1/2}X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}\tau_{AB}^{-1/2}\leq\mu I_{AB}\right\} \tag{3.76}$$

$$= \inf\left\{\mu : X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}\leq\mu\,\tau_{AB}\right\} \tag{3.77}$$

$$= \inf\left\{\mu : X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}\leq\mu\left[(1-\delta)\,X_{AB}^{1/2}X_A^{-1/2}\omega_AX_A^{-1/2}X_{AB}^{1/2}+\delta\xi_{AB}\right]\right\} \tag{3.78}$$

$$\leq \frac{1}{1-\delta}. \tag{3.79}$$

Then consider that

$$\langle\varphi^{X_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T})|\varphi^{X_{AB}}\rangle_{AB\hat{A}\hat{B}} = \langle\varphi^{X_A}|_{A\hat{A}} V^{\dagger} f(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}) V|\varphi^{X_A}\rangle_{A\hat{A}} \tag{3.80}$$

$$\geq \langle\varphi^{X_A}|_{A\hat{A}} f(V^{\dagger}\left[\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\right] V)|\varphi^{X_A}\rangle_{A\hat{A}} \tag{3.81}$$

$$\geq \langle\varphi^{X_A}|_{A\hat{A}} f([1-\delta]^{-1}\left[\omega_A^{-1}\otimes Y_{\hat{A}}^{T}\right])|\varphi^{X_A}\rangle_{A\hat{A}}, \tag{3.82}$$

where the first inequality is a consequence of the operator Jensen inequality and the second follows from (3.70)–(3.75) and operator anti-monotonicity of the function f. Taking a supremum over invertible density operators τ_{AB}, we then conclude that the following inequality holds for all $\delta \in (0,1)$ and for all invertible density operators ω_A:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \langle\varphi^{X_A}|_{A\hat{A}} f([1-\delta]^{-1}\left[\omega_A^{-1}\otimes Y_{\hat{A}}^{T}\right])|\varphi^{X_A}\rangle_{A\hat{A}}. \tag{3.83}$$

Since this inequality holds for all $\delta \in (0,1)$ and for all invertible density operators ω_A, we can appeal to continuity of the function f (taking the limit $\delta \searrow 0$) and then take a supremum over all invertible density operators ω_A to conclude the desired inequality for the case $X_A > 0$, but X_{AB} is not invertible:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A), \tag{3.84}$$

as claimed. ■

Remark 7 *I stress once again here that if X_{AB} and Y_{AB} are invertible, then we only require operator convexity of the function f in order to arrive at the inequality in* (3.31)*. One can examine the steps in* (3.56)*–*(3.60) *to see this.*

Based on Propositions 4 and 6 and the Stinespring dilation theorem [Sti55], we conclude the following data-processing theorem for the optimized quantum f-divergences:

Theorem 8 (Quantum data processing) *Let X_S and Y_S be positive semi-definite operators acting on a Hilbert space $\mathcal{H}_S$, and let $\mathcal{N}_{S\to B}$ be a quantum channel taking operators acting on $\mathcal{H}_S$ to operators acting on a Hilbert space $\mathcal{H}_B$. Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Then the following inequality holds*

$$\widetilde{Q}_f(X_S\|Y_S) \geq \widetilde{Q}_f(\mathcal{N}_{S\to B}(X_S)\|\mathcal{N}_{S\to B}(Y_S)). \tag{3.85}$$

4 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 [Ume62] and the sandwiched Rényi relative quasi-entropies [MLDS+13, WWY14]. The result will be that Theorem 8 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 8 and earlier work of [Pet85, Pet86a, TCR09] is a single, unified approach, based on the operator Jensen inequality [HP03], 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.

4.1 Quantum relative entropy as optimized quantum f-divergence

Let τ be an invertible state, $\varepsilon > 0$, and X and Y positive semi-definite. Let $\overline{X} = X/\operatorname{Tr}\{X\}$. Pick the function

$$f(x) = -\log x, \tag{4.1}$$

which is an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$, and we find that

$$\begin{aligned}
&\frac{1}{\operatorname{Tr}\{X\}}\langle\varphi^X|_{S\hat{S}}\left[-\log(\tau_S^{-1}\otimes\left(Y+\varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T)\right]|\varphi^X\rangle_{S\hat{S}} \\
&= \langle\varphi^{\overline{X}}|_{S\hat{S}}\left[\log(\tau_S)\otimes I_{\hat{S}} - I_S\otimes\log\left(Y+\varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T\right]|\varphi^{\overline{X}}\rangle_{S\hat{S}} && (4.2)\\
&= \langle\varphi^{\overline{X}}|_{S\hat{S}}\log(\tau_S)\otimes I_{\hat{S}}|\varphi^{\overline{X}}\rangle_{S\hat{S}} - \langle\varphi^{\overline{X}}|_{S\hat{S}} I_S\otimes\log\left(Y+\varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T|\varphi^{\overline{X}}\rangle_{S\hat{S}} && (4.3)\\
&= \operatorname{Tr}\{\overline{X}\log\tau\} - \operatorname{Tr}\{\overline{X}\log(Y+\varepsilon\Pi_Y^{\perp})\} && (4.4)\\
&\leq \operatorname{Tr}\{\overline{X}\log\overline{X}\} - \operatorname{Tr}\{\overline{X}\log(Y+\varepsilon\Pi_Y^{\perp})\} && (4.5)\\
&= D(\overline{X}\|Y+\varepsilon\Pi_Y^{\perp}). && (4.6)
\end{aligned}$$

The inequality is a consequence of Klein's inequality [Kle31] (see also [Rus02]), establishing that the optimal τ is set to $\overline{X}$.[2] Now taking a supremum over $\varepsilon > 0$, we find that

$$\widetilde{Q}_{-\log(\cdot)}(X\|Y) = \operatorname{Tr}\{X\}D(\overline{X}\|Y), \tag{4.7}$$

where the quantum relative entropy $D(\overline{X}\|Y)$ is defined as [Ume62]

$$D(\overline{X}\|Y) = \operatorname{Tr}\{\overline{X}\left[\log\overline{X} - \log Y\right]\} \tag{4.8}$$

if $\operatorname{supp}(X) \subseteq \operatorname{supp}(Y)$ and $D(\overline{X}\|Y) = +\infty$ otherwise.

4.2 Sandwiched Rényi relative quasi-entropy as optimized quantum f-divergence

Take τ, ε, X, and Y as defined in Section 4.1. For $\alpha \in [1/2, 1)$, pick the function

$$f(x) = -x^{(1-\alpha)/\alpha}, \tag{4.9}$$

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

$$\widetilde{Q}_{-(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = -\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \tag{4.10}$$

[2]Technically, we would require an invertible τ that approximates $\overline{X}$ arbitrarily well in order to achieve equality in Klein's inequality. One can alternatively establish the inequality $\operatorname{Tr}\{\overline{X}\log\overline{X}\} \geq \operatorname{Tr}\{\overline{X}\log\tau\}$ by employing the non-negativity of quantum relative entropy $D(\overline{X}\|\tau) \geq 0$ for quantum states.

which is the known expression for sandwiched Rényi relative quasi-entropy for $\alpha \in [1/2, 1)$ [MLDS+13, WWY14]. To see this, consider that

$$
\begin{aligned}
& - \langle\varphi^X|_{S\hat{S}} \left[\tau_S^{-1} \otimes \left(Y + \varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T\right]^{(1-\alpha)/\alpha} |\varphi^X\rangle_{S\hat{S}} \\
&= -\langle\varphi^X|_{S\hat{S}} \tau_S^{(\alpha-1)/\alpha} \otimes \left(\left(Y + \varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T\right)^{(1-\alpha)/\alpha} |\varphi^X\rangle_{S\hat{S}} && (4.11) \\
&= -\langle\Gamma|_{S\hat{S}} X_S^{1/2} \tau_S^{(\alpha-1)/\alpha} X_S^{1/2} \otimes \left(\left(Y + \varepsilon\Pi_Y^{\perp}\right)_{\hat{S}}^T\right)^{(1-\alpha)/\alpha} |\Gamma\rangle_{S\hat{S}} && (4.12) \\
&= -\operatorname{Tr}\left\{X^{1/2}\tau^{(\alpha-1)/\alpha} X^{1/2} (Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha}\right\} && (4.13) \\
&= -\operatorname{Tr}\left\{X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}. && (4.14)
\end{aligned}
$$

The formulas in the above development are related to those given in the proof of [MLDS+13, Lemma 19]. Now optimizing over invertible states τ and employing Hölder duality [Bha97], in the form of the reverse Hölder inequality and as observed in [MLDS+13], we find that

$$\sup_{\substack{\tau>0,\\ \operatorname{Tr}\{\tau\}=1}} \left[-\operatorname{Tr}\left\{X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}\right] = -\left\|X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2}\right\|_\alpha, \tag{4.15}$$

where for positive semi-definite Z, we define

$$\|Z\|_\alpha = [\operatorname{Tr}\{Z^\alpha\}]^{1/\alpha}. \tag{4.16}$$

Now taking the limit $\varepsilon \searrow 0$, we get that

$$\widetilde{Q}_{-(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = -\left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_\alpha = -\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha, \tag{4.17}$$

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 [MLDS+13, WWY14]

$$\widetilde{D}_\alpha(X\|Y) = \frac{\alpha}{\alpha - 1}\log\left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha. \tag{4.18}$$

Thus, Theorem 8 implies quantum data processing for the sandwiched Rényi relative entropy

$$\widetilde{D}_\alpha(X\|Y) \geq \widetilde{D}_\alpha(\mathcal{N}(X)\|\mathcal{N}(Y)), \tag{4.19}$$

for the parameter range $\alpha \in [1/2, 1)$, which is a result previously established in [FL13].

For $\alpha \in (1, \infty]$, pick the function

$$f(x) = x^{(1-\alpha)/\alpha}, \tag{4.20}$$

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

$$\widetilde{Q}_{(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = \begin{cases} \left\|Y^{(1-\alpha)/2\alpha}XY^{(1-\alpha)/2\alpha}\right\|_\alpha & \text{if } \operatorname{supp}(X) \subseteq \operatorname{supp}(Y) \\ +\infty & \text{else} \end{cases}, \tag{4.21}$$

which is the known expression for sandwiched Rényi relative quasi-entropy for $\alpha \in (1,\infty]$ [MLDS+13, WWY14]. To see this, consider that the same development as above gives that

$$\langle \varphi^X|_{S\hat{S}}(\tau_S^{-1} \otimes \left(Y + \varepsilon \Pi_Y^{\perp}\right)^T_{\hat{S}})^{(1-\alpha)/\alpha} |\varphi^X\rangle_{S\hat{S}} = \operatorname{Tr}\left\{X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2} \tau^{(\alpha-1)/\alpha}\right\}. \quad (4.22)$$

Again employing Hölder duality, as observed in [MLDS+13], we find that

$$\sup_{\tau>0,\operatorname{Tr}\{\tau\}=1} \operatorname{Tr}\left\{X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2} \tau^{(\alpha-1)/\alpha}\right\} = \left\|X^{1/2}(Y + \varepsilon\Pi_Y^{\perp})^{(1-\alpha)/\alpha} X^{1/2}\right\|_\alpha, \quad (4.23)$$

Now taking the limit $\varepsilon \searrow 0$, we get that

$$\widetilde{Q}_{(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = \left\|X^{1/2} Y^{(1-\alpha)/\alpha} X^{1/2}\right\|_\alpha = \left\|Y^{(1-\alpha)/2\alpha} X Y^{(1-\alpha)/2\alpha}\right\|_\alpha, \quad (4.24)$$

where the equalities hold if $\operatorname{supp}(X) \subseteq \operatorname{supp}(Y)$ and otherwise $\widetilde{Q}_{(\cdot)^{(1-\alpha)/\alpha}}(X\|Y) = +\infty$, as observed in [MLDS+13]. The sandwiched Rényi relative entropy itself is defined up to a normalization factor as in (4.18) if $\operatorname{supp}(X) \subseteq \operatorname{supp}(Y)$ and otherwise $\widetilde{D}_\alpha(X\|Y) = +\infty$ for $\alpha \in (1,\infty]$ [MLDS+13, WWY14]. Thus, Theorem 8 implies quantum data processing for the sandwiched Rényi relative entropy

$$\widetilde{D}_\alpha(X\|Y) \geq \widetilde{D}_\alpha(\mathcal{N}(X)\|\mathcal{N}(Y)), \quad (4.25)$$

for the parameter range $\alpha \in (1,\infty]$, which is a result previously established in full by [FL13, Bei13, MO15] and for $\alpha \in (1,2]$ by [MLDS+13, WWY14].

4.3 Optimized α-divergence: monotonicity under partial trace for invertible density operators

Interestingly, for $\alpha \in [1/3, 1/2]$, the function

$$f(x) = x^{(1-\alpha)/\alpha} \quad (4.26)$$

is operator convex on the domain $(0,\infty)$ and with range $\mathbb{R}$. Note that this is a reparametrization of x^β for $\beta \in [1,2]$. Thus, by following the same development as before, for positive definite X and Y we find that

$$\langle \varphi^X|_{S\hat{S}}(\tau_S^{-1} \otimes Y_{\hat{S}}^T)^\beta |\varphi^X\rangle_{S\hat{S}} = \operatorname{Tr}\left\{X^{1/2} Y^\beta X^{1/2} \tau^{-\beta}\right\}. \quad (4.27)$$

Now optimizing over τ, we find that the following function

$$\widetilde{Q}_{(\cdot)^\beta}(X\|Y) = \sup_{\tau>0,\operatorname{Tr}\{\tau\}=1} \operatorname{Tr}\left\{X^{1/2} Y^\beta X^{1/2} \tau^{-\beta}\right\} \quad (4.28)$$

is monotone with respect to partial trace for $\beta \in [1,2]$. That is, the inequality

$$\widetilde{Q}_{(\cdot)^\beta}(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_{(\cdot)^\beta}(X_A\|Y_A) \quad (4.29)$$

holds for $\beta \in [1,2]$ and positive definite X_{AB} and Y_{AB}, by applying Remark 7.

Take note that $\widetilde{Q}_{(\cdot)^\beta}(X\|Y)$ for $\beta \in [1,2]$ is not a sandwiched Rényi relative quasi-entropy because the optimization over τ goes the opposite way when compared to that for the sandwiched Rényi relative entropy for $\alpha \in [1/3, 1/2]$. This is consistent with the fact that data processing is known not to hold for the sandwiched Rényi relative entropy for $\alpha \in (0, 1/2)$ [MLDS+13, DL14, BFT17].

5 On Petz's quantum f-divergence

I now discuss in more detail the relation between the optimized quantum f-divergence and the Petz quantum f-divergence from [Pet85, Pet86a]. In brief, we find that the Petz f-divergence can be recovered by replacing τ in Definition 1 with $X + \delta\pi_X^\perp$.

Definition 9 (Petz quantum f-divergence) *Let f be a continuous function with domain $(0,\infty)$ and range $\mathbb{R}$. For positive semi-definite operators X and Y acting on a Hilbert space $\mathcal{H}_S$, the Petz quantum f-divergence is defined as*

$$Q_f(X\|Y) \equiv \sup_{\varepsilon>0} \lim_{\delta\searrow 0} \langle\varphi^X|_{S\hat{S}} f\left(\left[X_S + \delta\pi_X^\perp\right]^{-1} \otimes \left[Y_{\hat{S}} + \varepsilon\Pi_Y^\perp\right]^T\right)|\varphi^X\rangle_{S\hat{S}}, \tag{5.1}$$

where $\pi_X^\perp = \Pi_X^\perp/\operatorname{Tr}\{\Pi_X^\perp\}$ is the maximally mixed state on the kernel of X and the rest of the notation is the same as in Definition 1. If the kernel of X is equal to zero, then we set $\pi_X^\perp = 0$.

Let spectral decompositions of positive semi-definite X and positive definite Y be given as

$$X = \sum_x \lambda_x |\psi^x\rangle\langle\psi^x|, \qquad Y = \sum_y \mu_y |\phi^y\rangle\langle\phi^y|. \tag{5.2}$$

By following the same development needed to arrive at (2.13), we see that $Q_f(X\|Y)$ can be written for non-invertible X and invertible Y as

$$\begin{aligned} Q_f(X\|Y) &= \lim_{\delta\searrow 0} \sum_y \Bigg[\sum_{x:\lambda_x\neq 0} f(\mu_y\lambda_x^{-1}) \operatorname{Tr}\{X^{1/2}|\psi^x\rangle\langle\psi^x|X^{1/2}|\phi^y\rangle\langle\phi^y|\} \\ &\qquad\qquad + f(\mu_y \operatorname{Tr}\{\Pi_X^\perp\}\delta^{-1}) \operatorname{Tr}\{X^{1/2}\Pi_X^\perp X^{1/2}|\phi^y\rangle\langle\phi^y|\}\Bigg] \quad (5.3) \\ &= \lim_{\delta\searrow 0} \sum_y \sum_{x:\lambda_x\neq 0} f(\mu_y\lambda_x^{-1}) \operatorname{Tr}\{X^{1/2}|\psi^x\rangle\langle\psi^x|X^{1/2}|\phi^y\rangle\langle\phi^y|\} \quad (5.4) \\ &= \sum_y \sum_{x:\lambda_x\neq 0} \lambda_x f(\mu_y\lambda_x^{-1}) \left|\langle\psi^x|\phi^y\rangle\right|^2. \quad (5.5) \end{aligned}$$

Note that we get the same formula for $Q_f(X\|Y)$ if X is invertible. For non-invertible Y, we just substitute $Y + \varepsilon\Pi_Y^\perp$ and take the supremum over $\varepsilon > 0$ at the end.

The next concern is about quantum data processing with the Petz f-divergence as defined above. 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 3, 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 Proposition 4. Monotonicity of $Q_f(X_{AB}\|Y_{AB})$ under partial trace breaks down into three cases depending on invertibility of X_{AB} or X_A, as discussed in the proof of Proposition 6. For the proof, we assume as previously done that Y_{AB} is invertible throughout. If it is not, then Definition 9 forces it to be invertible and then a supremum over $\varepsilon > 0$ is finally taken at the end.

1. The case when X_{AB} is invertible is already handled by Petz's proof from [Pet85, Pet86a], which relies on the operator Jensen inequality [HP03]. In this case, the operator $X_{AB} + \delta\pi_{AB}^\perp$ reduces to X_{AB} because $\Pi_{AB}^\perp = 0$.

2. The case when X_{AB} is not invertible but X_A is can be understood as an appeal to continuity, as discussed in Remark 15. For this case, we take the operator τ_{AB} for some $\delta_1 \in (0,1)$ to be $(1-\delta_1) X_{AB} + \delta_1 \pi^{\perp}_{X_{AB}}$, which is a positive definite operator. The rest of the proof proceeds the same and then the monotonicity under partial trace holds for this case.

3. As far as I can tell, the case when X_A is not invertible was not discussed in several papers of Petz *et al.* [Pet85, Pet86a, PS09, Pet10b, Pet10a], and it was only considered recently in [HM17, Proposition 3.12]. However, the method I have given in the proof of Proposition 6 appears to be different. Also, Remark 15 in the appendix discusses how this approach arguably extends beyond a mere appeal to continuity. For this case, we take the channel in (A.1) to be

$$Z_A \to X_{AB}^{1/2} \left(\left[X_A^{-1/2} Z_A X_A^{-1/2} \right] \otimes I_B \right) X_{AB}^{1/2} + \mathrm{Tr}\{\Pi^{\perp}_{X_A} Z_A\} \pi^{\perp}_{X_{AB}}. \tag{5.6}$$

Inputting $X_A + \delta \pi^{\perp}_A$ then leads to the output $X_{AB} + \delta \pi^{\perp}_{AB}$, which is a positive definite operator. The rest of the proof proceeds the same and then the monotonicity under partial trace holds for this case.

Special and interesting cases of the Petz f-divergence are found by taking

$$f(x) = -\log x, \tag{5.7}$$

$$f(x) = -x^{\beta} \qquad \text{for } \beta \in (0,1], \tag{5.8}$$

$$f(x) = x^{\beta} \qquad \text{for } \beta \in [-1,0). \tag{5.9}$$

Each of these functions are operator anti-monotone with domain $(0,\infty)$ and range $\mathbb{R}$. By following similar reasoning as in Section 4 to simplify Q_f and by applying the above arguments for data processing, we find that all of the following quantities obey the data processing inequality:

$$Q_{-\log(\cdot)}(X\|Y) = \mathrm{Tr}\{X\} D(\overline{X}\|Y), \tag{5.10}$$

$$Q_{-(\cdot)^{\beta}}(X\|Y) = -\mathrm{Tr}\{X^{1-\beta} Y^{\beta}\}, \text{ for } \beta \in (0,1], \tag{5.11}$$

$$Q_{(\cdot)^{\beta}}(X\|Y) = \begin{cases} \mathrm{Tr}\{X^{1-\beta} Y^{\beta}\} & \text{if } \mathrm{supp}(X) \subseteq \mathrm{supp}(Y) \\ +\infty & \text{else} \end{cases}, \text{ for } \beta \in [-1,0), \tag{5.12}$$

where again $\overline{X} = X/\mathrm{Tr}\{X\}$. By a reparametrization $\alpha = 1-\beta$, we find that the latter two quantities are directly related to the Petz Rényi relative entropy, defined as

$$D_{\alpha}(X\|Y) \equiv \begin{cases} \frac{1}{\alpha-1} \log \mathrm{Tr}\{X^{\alpha} Y^{1-\alpha}\} & \text{if } \mathrm{supp}(X) \subseteq \mathrm{supp}(Y) \text{ and } \alpha > 1 \\ +\infty & \text{else} \end{cases}. \tag{5.13}$$

Thus, the data processing proof establishes the data processing inequality for $D_{\alpha}(X\|Y)$ for $\alpha \in [0,1) \cup (1,2]$, which is the range for which it was already known to hold from prior work [Pet86a, TCR09].

Remark 10 *One beneficial aspect of the present paper is that we now see that there is a single, unified approach, based on the operator Jensen inequality, for establishing the data processing inequality for both the Petz–Rényi relative entropy for $\alpha \in [0,1) \cup (1,2]$ and the sandwiched Rényi relative entropy for $\alpha \in [1/2,1) \cup (1,\infty]$, the full ranges of α for which the data processing inequality* *is already known from [Pet85, Pet86a, TCR09, MLDS+13, WWY14, FL13, Bei13, MO15] to hold for these quantities. Prior to the present paper, there were a variety of different ways for establishing the data processing inequality for the sandwiched Rényi relative entropy, which can be found in [MLDS+13, WWY14, FL13, Bei13, MO15].*

Interestingly, for $\beta \in [1, 2]$, the function

$$f(x) = x^{\beta} \tag{5.14}$$

is operator convex on the domain $(0, \infty)$ and with range $\mathbb{R}$. Thus, for positive definite X and Y we find that

$$Q_{(\cdot)^{\beta}}(X\|Y) = \operatorname{Tr}\left\{X^{1-\beta}Y^{\beta}\right\} \tag{5.15}$$

is monotone with respect to partial trace for $\beta \in [1, 2]$. That is, the inequality

$$Q_{(\cdot)^{\beta}}(X_{AB}\|Y_{AB}) \geq Q_{(\cdot)^{\beta}}(X_A\|Y_A) \tag{5.16}$$

holds for $\beta \in [1, 2]$ and positive definite X_{AB} and Y_{AB}, by applying Remark 7. By reparametrizing with $\alpha = 1 - \beta$, we find that the following inequality holds for positive definite X_{AB} and Y_{AB} and $\alpha \in [-1, 0]$:

$$D_{\alpha}(X_{AB}\|Y_{AB}) \leq D_{\alpha}(X_A\|Y_A). \tag{5.17}$$

Note that there is trivially an equality when $\alpha = 0$, under the assumption that X_{AB} and Y_{AB} are positive definite, because

$$D_0(X_{AB}\|Y_{AB}) = -\log \operatorname{Tr}\{Y_{AB}\} = -\log \operatorname{Tr}\{Y_A\} = D_0(X_A\|Y_A). \tag{5.18}$$

5.1 Inequality for sandwiched and Petz–Rényi relative entropies

The development above motivates the following inequality relating the sandwiched and Petz–Rényi relative entropies. The same inequality was shown in [Jen18] when X and Y are normal states of an arbitrary von Neumann algebra and for $\alpha > 1$, whereas the following proposition considers the case when X and Y are positive semi-definite operators acting on a finite-dimensional Hilbert space and the range $\alpha \in [1/2, 1) \cup (1, \infty)$.

Proposition 11 *Let X and Y be positive semi-definite operators such that $X, Y \neq 0$. Then the following inequality holds for $\alpha \in [1/2, 1) \cup (1, \infty)$:*

$$\widetilde{D}_{\alpha}(X\|Y) \geq D_{(2\alpha-1)/\alpha}(X\|Y) - \log \operatorname{Tr}\{X\}. \tag{5.19}$$

Proof. Without loss of generality, let us assume that X and Y are invertible. The above inequality follows simply by picking the state $\tau = X/\operatorname{Tr}\{X\}$. Indeed, let $\alpha \in [1/2, 1)$. Consider that

$$-\left\|X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\right\|_{\alpha} = \sup_{\substack{\tau>0,\\ \operatorname{Tr}\{\tau\}=1}} \left[-\operatorname{Tr}\left\{X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\tau^{(\alpha-1)/\alpha}\right\}\right] \tag{5.20}$$

$$\geq -\operatorname{Tr}\left\{X^{1/2}Y^{(1-\alpha)/\alpha}X^{1/2}\left(X/\operatorname{Tr}\{X\}\right)^{(\alpha-1)/\alpha}\right\} \tag{5.21}$$

$$= -\operatorname{Tr}\left\{Y^{(1-\alpha)/\alpha}X^{(2\alpha-1)/\alpha}\right\}\left[\operatorname{Tr}\{X\}\right]^{(1-\alpha)/\alpha}. \tag{5.22}$$

This inequality implies that

$$\log \left\| X^{1/2} Y^{(1-\alpha)/\alpha} X^{1/2} \right\|_\alpha \leq \log \operatorname{Tr}\left\{ X^{(2\alpha-1)/\alpha} Y^{(1-\alpha)/\alpha} \right\} + \frac{1-\alpha}{\alpha} \log \operatorname{Tr}\{X\}. \tag{5.23}$$

Multiplying by $\frac{\alpha}{\alpha-1}$ leads to

$$\widetilde{D}_\alpha(X\|Y) = \frac{\alpha}{\alpha-1} \log \left\| X^{1/2} Y^{(1-\alpha)/\alpha} X^{1/2} \right\|_\alpha \tag{5.24}$$

$$\geq \frac{\alpha}{\alpha-1} \log \operatorname{Tr}\left\{ X^{(2\alpha-1)/\alpha} Y^{(1-\alpha)/\alpha} \right\} - \log \operatorname{Tr}\{X\} \tag{5.25}$$

$$= \frac{1}{\left[\frac{2\alpha-1}{\alpha}\right] - 1} \log \operatorname{Tr}\left\{ X^{(2\alpha-1)/\alpha} Y^{(1-\alpha)/\alpha} \right\} - \log \operatorname{Tr}\{X\} \tag{5.26}$$

$$= D_{\frac{2\alpha-1}{\alpha}}(X\|Y) - \log \operatorname{Tr}\{X\}. \tag{5.27}$$

This establishes the claim for $\alpha \in [1/2, 1)$. The proof for $\alpha \in (1, \infty)$ is very similar. ■

6 Classical and classical–quantum cases

When the operators X and Y commute, the optimized f-divergence takes on a simpler form, as stated in the following proposition:

Proposition 12 (Classical case) *Let f be an operator anti-monotone function with domain $(0, \infty)$ and range $\mathbb{R}$. Let X and Y be positive semi-definite operators that commute, having spectral decompositions*

$$X = \sum_z \lambda_z |z\rangle\langle z|, \qquad Y = \sum_z \mu_z |z\rangle\langle z|, \tag{6.1}$$

for a common eigenbasis $\{|z\rangle\}_z$. Then

$$\widetilde{Q}_f(X\|Y) = \sup_{\{\tau_z\}_z,\ \varepsilon>0} \left\{ \sum_{z\,:\,\mu_z\neq 0} \lambda_z f(\mu_z/\tau_z) + \sum_{z\,:\,\mu_z=0} \lambda_z f(\varepsilon/\tau_z) : \tau_z > 0\ \forall z,\ \sum_z \tau_z = 1 \right\}. \tag{6.2}$$

Proof. For simplicity, we prove the statement for the case in which Y is invertible, and then the extension to non-invertible Y is straightforward. For a spectral decomposition of τ as $\tau = \sum_t \nu_t |\phi_t\rangle\langle\phi_t|$ and by applying (2.13), we find that

$$\widetilde{Q}_f(X\|Y;\tau) = \sum_{z,t} f(\mu_z/\nu_t) |\langle z|X^{1/2}|\phi_t\rangle|^2 \tag{6.3}$$

$$= \sum_{z,t} f((\nu_t/\mu_z)^{-1}) \lambda_z |\langle z|\phi_t\rangle|^2 \tag{6.4}$$

$$\leq \sum_z f((\tau_z/\mu_z)^{-1}) \lambda_z, \tag{6.5}$$

where $\tau_z = \sum_t |\langle z|\phi_t\rangle|^2 \nu_t = \langle z|\tau|z\rangle$. The inequality follows because the function $f(x^{-1})$ is concave, due to the assumption that f is operator anti-monotone with domain $(0, \infty)$ and range $\mathbb{R}$. ■

If X and Y have a classical–quantum form, as follows

$$X = \sum_z |z\rangle\langle z| \otimes X^z, \qquad Y = \sum_z |z\rangle\langle z| \otimes Y^z, \tag{6.6}$$

where $\{|z\rangle\}_z$ is an orthonormal basis and $\{X^z\}_z$ and $\{Y^z\}_z$ are sets of positive semi-definite operators, then the optimized f-divergence simplifies as well, generalizing Proposition 12. That is, it suffices to optimize over positive definite states τ respecting the same classical–quantum form:

Proposition 13 (Classical–quantum case) *Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Let X and Y be positive semi-definite, having the classical–quantum form in (6.6). Then*

$$\widetilde{Q}_f(X\|Y) = \sup_{\{\hat{\tau}^z\}_z,\ \varepsilon>0} \sum_z \widetilde{Q}_f(X^z\|Y^z + \varepsilon\Pi^{\perp}_{Y_z}; \hat{\tau}^z), \tag{6.7}$$

where each $\hat{\tau}^z$ is positive definite such that $\sum_z \operatorname{Tr}\{\hat{\tau}^z\} = 1$.

Proof. The main idea here is to show that the optimal τ takes on a classical–quantum form as well, as $\tau = \sum_z |z\rangle\langle z| \otimes \tau^z$. This follows from an application of the operator Jensen inequality [HP03], as shown below. We focus on the case in which each Y^z is invertible. We adopt system labels Z for the classical system and A for the quantum system. For a given positive definite τ with $\operatorname{Tr}\{\tau\} = 1$, we have that

$$\widetilde{Q}_f(X\|Y;\tau) = \langle\varphi^X|_{ZA\hat{Z}\hat{A}} f(\tau_{ZA}^{-1} \otimes Y^T_{\hat{Z}\hat{A}})|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \tag{6.8}$$

$$= \langle\varphi^X|_{ZA\hat{Z}\hat{A}} f\left(\tau_{ZA}^{-1} \otimes \sum_z |z\rangle\langle z|_{\hat{Z}} \otimes (Y^z_{\hat{A}})^T\right)|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \tag{6.9}$$

$$= \langle\varphi^X|_{ZA\hat{Z}\hat{A}} \sum_z |z\rangle\langle z|_{\hat{Z}} \otimes f\left(\tau_{ZA}^{-1} \otimes (Y^z_{\hat{A}})^T\right)|\varphi^X\rangle_{ZA\hat{Z}\hat{A}}. \tag{6.10}$$

Consider that $|z\rangle\langle z|_{\hat{Z}}$ is invariant under the action of the decoherence or "pinching" channel

$$(\cdot) \to \mathcal{D}_{\hat{Z}}(\cdot) = \sum_z |z\rangle\langle z|_{\hat{Z}}(\cdot)|z\rangle\langle z|_{\hat{Z}}. \tag{6.11}$$

This implies that

$$\langle\varphi^X|_{ZA\hat{Z}\hat{A}} f(\tau_{ZA}^{-1} \otimes Y^T_{\hat{Z}\hat{A}})|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} = \langle\varphi^X|_{ZA\hat{Z}\hat{A}} \mathcal{D}_{\hat{Z}}\left[f(\tau_{ZA}^{-1} \otimes Y^T_{\hat{Z}\hat{A}})\right]|\varphi^X\rangle_{ZA\hat{Z}\hat{A}}. \tag{6.12}$$

By (2.16), the fact that $X_{ZA} = \mathcal{D}_Z(X_{ZA})$, and defining $g(x) = f(x^{-1})$, we find that

$$\langle\varphi^X|_{ZA\hat{Z}\hat{A}} \mathcal{D}_{\hat{Z}}\left[f(\tau_{ZA}^{-1} \otimes Y^T_{\hat{Z}\hat{A}})\right]|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} = \langle\varphi^X|_{ZA\hat{Z}\hat{A}} \mathcal{D}_Z\left[f(\tau_{ZA}^{-1} \otimes Y^T_{\hat{Z}\hat{A}})\right]|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \tag{6.13}$$

$$= \langle\varphi^X|_{ZA\hat{Z}\hat{A}} \mathcal{D}_Z\left[g(\tau_{ZA} \otimes (Y^T_{\hat{Z}\hat{A}})^{-1})\right]|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \tag{6.14}$$

$$\leq \langle\varphi^X|_{ZA\hat{Z}\hat{A}} \left[g(\mathcal{D}_Z(\tau_{ZA}) \otimes (Y^T_{\hat{Z}\hat{A}})^{-1})\right]|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \tag{6.15}$$

$$= \langle\varphi^X|_{ZA\hat{Z}\hat{A}} f([\mathcal{D}_Z(\tau_{ZA})]^{-1} \otimes Y^T_{\hat{Z}\hat{A}})|\varphi^X\rangle_{ZA\hat{Z}\hat{A}}. \tag{6.16}$$

The inequality follows from the operator Jensen inequality [HP03] and the fact that $g(x)$ is operator concave with domain $(0,\infty)$ and range $\mathbb{R}$. Consider that

$$\mathcal{D}_Z(\tau_{ZA}) = \sum_z |z\rangle\langle z|_Z \otimes \hat{\tau}^z_{\hat{A}}, \tag{6.17}$$

for some $\{\tau^z\}_z$, where each $\hat{\tau}^z_{\hat{A}}$ is positive definite and $\sum_z \mathrm{Tr}\{\hat{\tau}^z_{\hat{A}}\} = 1$. Now consider that

$$\begin{aligned}
&\langle \varphi^X|_{ZA\hat{Z}\hat{A}} f([\mathcal{D}_Z(\tau_{ZA})]^{-1} \otimes Y^T_{\hat{Z}\hat{A}})|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} \\
&= \sum_{z,z'} \langle \varphi^X|_{ZA\hat{Z}\hat{A}} \otimes |z\rangle\langle z|_Z \otimes |z'\rangle\langle z'|_{\hat{Z}} \otimes f\left((\hat{\tau}^z_{\hat{A}})^{-1} \otimes (Y^{z'}_{\hat{A}})^T\right)|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} &(6.18)\\
&= \sum_z \langle \varphi^{X^z}|_{A\hat{A}} f\left((\hat{\tau}^z_{\hat{A}})^{-1} \otimes (Y^z_{\hat{A}})^T\right)|\varphi^{X^z}\rangle_{A\hat{A}} &(6.19)\\
&= \sum_z \widetilde{Q}_f(X^z\|Y^z;\hat{\tau}^z), &(6.20)
\end{aligned}$$

where the second-to-last line follows because

$$\begin{aligned}
|\varphi^X\rangle_{ZA\hat{Z}\hat{A}} &= (X^{1/2}_{ZA} \otimes I_{\hat{Z}\hat{A}})|\Gamma\rangle_{Z\hat{Z}}|\Gamma\rangle_{A\hat{A}} &(6.21)\\
&= \sum_z |z\rangle\langle z|_Z \otimes (X^z_A)^{1/2}|\Gamma\rangle_{Z\hat{Z}}|\Gamma\rangle_{A\hat{A}} &(6.22)\\
&= \sum_z |z\rangle_Z|z\rangle_{\hat{Z}}(X^z_A)^{1/2}|\Gamma\rangle_{A\hat{A}} &(6.23)\\
&= \sum_z |z\rangle_Z|z\rangle_{\hat{Z}}|\varphi^{X^z}\rangle_{A\hat{A}}. &(6.24)
\end{aligned}$$

This completes the proof after optimizing over $\{\hat{\tau}^z\}_z$ satisfying $\sum_z \mathrm{Tr}\{\hat{\tau}^z\} = 1$. We handle the case of non-invertible Y by taking a supremum over $\varepsilon > 0$ at the end. ■

7 Optimized quantum f-information measures

It is well known that the quantum relative entropy is a parent quantity for many information measures used in quantum information theory (see, e.g., [Dat11] or [Wil17, Chapter 11]). As such, once one has a base relative entropy or divergence to work with, there is now a relatively standard recipe for generating other information measures, such as entropy, conditional entropy, coherent information, mutual information, entanglement measures, and more generally resource measures. This method has been used in many works now [VP98, Dat09, Sha10, MH11, WWY14, MLDS+13, Bei13, GW15, TWW17, WTB17, KW17]. Each of the resulting quantities then satisfies a particular kind of quantum data processing inequality, which follows as a consequence of the monotonocity of the underlying relative entropy.

With the above in mind, we now mention some different information measures that can be derived from the optimized f-divergence and we state the data processing inequality that they satisfy. In what follows, ρ_{AB} is a density operator acting on a tensor-product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, $\rho_A = \mathrm{Tr}_B\{\rho_{AB}\}$, and f is an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. Let $\mathcal{W}_{A\to A'}$ denote a subunital channel, satisfying $\mathcal{W}_{A\to A'}(I_A) \leq I_{A'}$, and let $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ be quantum channels. All statements about data processing follow from Theorem 8 and some slight extra reasoning (see, e.g., [Wil17, Section 11.9]). One can find various operational interpretations of entropic quantities discussed in [Wil17, Hay06, Hol12].

1. The optimized f-entropy is defined as

$$\widetilde{S}_f(A)_\rho \equiv \widetilde{S}_f(\rho_A) \equiv -\widetilde{Q}_f(\rho_A \| I_A). \tag{7.1}$$

It does not decrease under the action of a subunital channel $\mathcal{W}_{A\to A'}$, in the sense that

$$\widetilde{S}_f(A)_\rho \leq \widetilde{S}_f(A')_{\mathcal{W}(\rho)}. \tag{7.2}$$

2. The optimized f-mutual information is defined as

$$\widetilde{I}_f(A;B)_\rho \equiv \inf_{\sigma_B} \widetilde{Q}_f(\rho_{AB} \| \rho_A \otimes \sigma_B). \tag{7.3}$$

It does not increase under the action of the product channel $\mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'}$, in the sense that

$$\widetilde{I}_f(A;B)_\rho \geq \widetilde{I}_f(A';B')_{(\mathcal{N}\otimes\mathcal{M})(\rho)}. \tag{7.4}$$

3. The optimized conditional f-entropy is defined as

$$\widetilde{S}_f(A|B)_\rho \equiv -\inf_{\sigma_B} \widetilde{Q}_f(\rho_{AB} \| I_A \otimes \sigma_B). \tag{7.5}$$

It does not decrease under the action of the product channel $\mathcal{W}_{A\to A'} \otimes \mathcal{M}_{B\to B'}$:

$$\widetilde{S}_f(A|B)_\rho \leq \widetilde{S}_f(A|B)_{(\mathcal{W}\otimes\mathcal{M})(\rho)}. \tag{7.6}$$

4. Related to the above, the optimized f-coherent information is defined as

$$\widetilde{I}_f(A\rangle B)_\rho \equiv -\widetilde{S}_f(A|B)_\rho, \tag{7.7}$$

and we have that

$$\widetilde{I}_f(A\rangle B)_\rho \geq \widetilde{I}_f(A\rangle B)_{(\mathcal{W}\otimes\mathcal{M})(\rho)}. \tag{7.8}$$

5. In recent years, there has been much activity surrounding quantum resource theories [BG15, Fri15, dRKR15, KdR16, CG18]. Such a resource theory consists of a few basic elements. There is a set $\mathcal{F}$ of free quantum states, i.e., those that the players involved are allowed to access without any cost. Related to these, there is a set of free channels, and they should have the property that a free state remains free after a free channel acts on it. Once these are defined, it follows that any state that is not free is considered resourceful, i.e., useful in the context of the resource theory. We can also then define a measure of the resourcefulness of a quantum state, and some fundamental properties that it should satisfy are that 1) it should be monotone non-increasing under the action of a free channel and 2) it should be equal to zero when evaluated on a free state. A typical choice of a resourcefulness measure of a state ρ satisfying these requirements is the relative entropy of resourcefulness, defined in terms of relative entropy as $\inf_{\sigma\in F} D(\rho\|\sigma)$. We can thus consider an optimized f-relative entropy of resourcefulness as

$$\widetilde{R}_f(\rho) \equiv \inf_{\sigma\in\mathcal{F}} \widetilde{Q}_f(\rho\|\sigma), \tag{7.9}$$

and it thus satisfies the following data processing inequality

$$\widetilde{R}_f(\rho) \geq \widetilde{R}_f(\mathcal{N}(\rho)), \tag{7.10}$$

whenever $\mathcal{N}$ is a free channel as described above.

6. We can extend all of the above measures to quantum channel measures by optimizing over inputs to the channel. For example, optimized f-mutual information of a channel $\mathcal{N}_{A\to B}$ is defined as

$$\sup_{\psi_{RA}} \widetilde{I}_f(R;B)_\omega, \tag{7.11}$$

where $\omega_{RB} = \mathcal{N}_{A\to B}(\psi_{RA})$ and ψ_{RA} is a pure bipartite state. Due to the Schmidt decomposition theorem and data processing, it suffices to optimize over pure bipartite states ψ_{RA} with the reference system R isomorphic to the channel input system A.

7.1 Duality of optimized conditional f-entropy

This paper's final contribution is the following proposition, which generalizes a well known duality relation for conditional quantum entropy:

Proposition 14 (Duality) *Let f be an operator anti-monotone function with domain $(0,\infty)$ and range $\mathbb{R}$. For a pure state $|\psi\rangle\langle\psi|_{ABC}$, we have that*

$$\widetilde{S}_f(A|B)_\psi = -\widetilde{S}_k(A|C)_\psi, \tag{7.12}$$

where $k(x) = -f(x^{-1})$.

Proof. The method of proof is related to that given in [Bei13, MLDS+13]. Set $\rho_{AB} = \operatorname{Tr}_C\{|\psi\rangle\langle\psi|_{ABC}\}$ and consider that

$$\begin{aligned}
\widetilde{S}_f(A|B)_\psi &= -\inf_{\sigma_B} \widetilde{Q}_f(\rho_{AB}\|I_A\otimes\sigma_B) && (7.13)\\
&= -\inf_{\sigma_B}\sup_{\tau_{AB}} \langle\varphi^{\rho_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes\sigma_{\hat{B}}^T)|\varphi^{\rho_{AB}}\rangle_{AB\hat{A}\hat{B}} && (7.14)\\
&= -\sup_{\tau_{AB}}\inf_{\sigma_B} \langle\varphi^{\rho_{AB}}|_{AB\hat{A}\hat{B}} f(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes\sigma_{\hat{B}}^T)|\varphi^{\rho_{AB}}\rangle_{AB\hat{A}\hat{B}} && (7.15)\\
&= -\sup_{\tau_{AB}}\inf_{\sigma_B} \langle\varphi^{\rho_{AB}}|_{AB\hat{A}\hat{B}} f\left(I_A\otimes\sigma_B\otimes\left(\tau_{\hat{A}\hat{B}}^{-1}\right)^T\right)|\varphi^{\rho_{AB}}\rangle_{AB\hat{A}\hat{B}} && (7.16)\\
&= -\sup_{\tau_C}\inf_{\sigma_B} \langle\psi|_{ABC} f\left(I_A\otimes\sigma_B\otimes\left(\tau_C^{-1}\right)^T\right)|\psi\rangle_{ABC} && (7.17)
\end{aligned}$$

The first two equalities follow by definition. For simplicity, we consider σ_B to be an invertible density operator. The third equality follows from an application of the minimax theorem [Sio58], considering that $f(x)$ is operator convex and $f(x^{-1})$ is operator concave. The fourth equality follows by applying (2.13)–(2.17). The fifth equality follows because $|\varphi^{\rho_{AB}}\rangle_{AB\hat{A}\hat{B}}$ and $|\psi\rangle_{ABC}$ are purifications of ρ_{AB} and all purifications are related by an isometry (see, e.g., [Wil17]). Furthermore, we have isometric invariance of the optimized $\langle\varphi^{\rho_{AB}}|_{AB\hat{A}\hat{B}} f(I_A\otimes\sigma_B\otimes(\tau_{\hat{A}\hat{B}}^{-1})^T)|\varphi^{\rho_{AB}}\rangle_{AB\hat{A}\hat{B}}$ by the same reasoning as given in the proof of Proposition 4. Continuing,

$$\begin{aligned}
&= -\sup_{\tau_C}\inf_{\sigma_B}\left[-\langle\psi|_{ABC}\ k(I_A\otimes\sigma_B^{-1}\otimes\tau_C^T)|\psi\rangle_{ABC}\right] && (7.18)\\
&= \inf_{\tau_C}\sup_{\sigma_B}\langle\psi|_{ABC}\ k(I_A\otimes\sigma_B^{-1}\otimes\tau_C^T)|\psi\rangle_{ABC} && (7.19)\\
&= \inf_{\tau_C}\widetilde{Q}_k(\psi_{AC}\|I_A\otimes\tau_C) && (7.20)\\
&= -\widetilde{S}_k(A|C)_\psi. && (7.21)
\end{aligned}$$

The first equality follows from the definition of the function k. The second equality follows from propagating the inside minus sign to the outside. The last equalities follow from applying similar steps as in the beginning of the proof and then the definitions of $\widetilde{Q}_k(\rho_{AC}\|I_A \otimes \tau_C)$ and $\widetilde{S}_k(A|C)_\psi$. ■

When $f(x) = x^{(1-\alpha)/\alpha}$ for $\alpha \in (1, \infty]$, we recover a duality relation similar to that for sandwiched Rényi relative entropy [Bei13, MLDS+13]. Duality relations for conditional entropy and the data processing inequality are known to be closely related to entropic uncertainty relations [CCYZ12], so there could be interesting new ones to develop by choosing more general operator anti-monotone functions.

8 Conclusion

The main contribution of the present paper 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 [HP03], and it bears some similarities to the original approach from [Pet85, Pet86a, TCR09]. 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 [HP03], 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. In the remainder of the paper, I considered other aspects such as the classical case, the classical–quantum case, and information measures that one could construct from the optimized f-divergence.

There are several directions that one could pursue going forward. Equation (2.19) represents the function underlying the optimized f-divergence in the relative modular operator formalism—this should be helpful in understanding the optimized f-divergence in more general contexts. Combined with the methods of [Pet85, Pet86a] and the approach in this paper, it is clear that the data processing inequality will hold in more general contexts. It would also be interesting to show that the data processing inequality holds for maps beyond quantum channels, such as the Schwarz and stochastic maps considered in [HMPB11]. I suspect that the methods of [HMPB11] and the present paper could be used to establish the data processing inequality for more general classes of maps.

Acknowledgements. I thank Milan Mosonyi and Anna Vershynina for discussions related to the topic of this paper, and I acknowledge support from the National Science Foundation under grant no. 1714215. I also thank the anonymous referees for many suggestions that helped to improve the paper.

A Final case for Proposition 6

Case X_A not invertible: We now discuss how to extend the proof detailed in the main text to the case in which X_A is not invertible (and thus X_{AB} is not invertible either). In this case, with $X_A^{-1/2}$ understood as a square-root inverse of X_A on its support, the Petz recovery map in (3.35) is no longer a quantum channel, but it is instead a completely positive and trace non-increasing map. A standard method for producing a quantum channel from the map in (3.35) is to specify an additional action on the kernel of X_A, as

$$Z_A \to X_{AB}^{1/2}\left(\left[X_A^{-1/2} Z_A X_A^{-1/2}\right] \otimes I_B\right) X_{AB}^{1/2} + \text{Tr}\{\Pi_{X_A}^{\perp} Z_A\}\xi_{AB}, \tag{A.1}$$

where we take ξ_{AB} to be an invertible density operator (see, e.g., [Wil17, Chapter 12] for this standard construction). One can check that the map in (A.1) is completely positive and trace preserving, and furthermore, an invertible input state leads to an invertible output state. Our goal is now to find a Kraus decomposition for the above quantum channel, so that we can work with its isometric extension as we did previously. To begin with, suppose that the invertible state ξ_{AB} has a spectral decomposition as

$$\xi_{AB} = \sum_{l=1}^{|A||B|} p_l |\phi_l\rangle\langle\phi_l|_{AB}, \tag{A.2}$$

where $\{p_l\}_l$ is a probability distribution and $\{|\phi_l\rangle_{AB}\}_l$ is an orthonormal basis. Then we can write the channel in (A.1) as

$$\begin{aligned}
&\sum_{j=1}^{|B|} X_{AB}^{1/2}\left(\left[X_A^{-1/2} Z_A X_A^{-1/2}\right] \otimes |j\rangle\langle j|_B\right) X_{AB}^{1/2} + \sum_{k=1}^{|A|}\langle k|_A \Pi_{X_A}^{\perp} Z_A \Pi_{X_A}^{\perp}|k\rangle_A \xi_{AB} \\
&= \sum_{j=1}^{|B|} X_{AB}^{1/2}\left[X_A^{-1/2} \otimes |j\rangle_B\right] Z_A \left[X_A^{-1/2} \otimes \langle j|_B\right] X_{AB}^{1/2} \\
&\quad + \sum_{k=1}^{|A|}\sum_{l=1}^{|A||B|} \sqrt{p_l}|\phi_l\rangle_{AB}\langle k|_A \Pi_{X_A}^{\perp} Z_A \Pi_{X_A}^{\perp}|k\rangle_A\langle\phi_l|_{AB}\sqrt{p_l}.
\end{aligned} \tag{A.3}$$

Thus, Kraus operators for it are as follows:

$$\left\{\left\{X_{AB}^{1/2}\left[X_A^{-1/2} \otimes |j\rangle_B\right]\right\}_{j=1}^{|B|}, \left\{\sqrt{p_l}|\phi_l\rangle_{AB}\langle k|_A \Pi_{X_A}^{\perp}\right\}_{l\in\{1,\ldots,|A||B|\}, k\in\{1,\ldots,|A|\}}\right\}. \tag{A.4}$$

We now define an enlarged Hilbert space $\hat{C}$ to be the direct sum of $\hat{B}$ and $\hat{A}$, and thus with dimension $|\hat{B}| + |\hat{A}|$, and an orthonormal basis for it as

$$\{|1\rangle_{\hat{C}}, \ldots, |j\rangle_{\hat{C}}, \ldots, ||\hat{B}|\rangle_{\hat{C}}, ||\hat{B}|+1\rangle_{\hat{C}}, \ldots, ||\hat{B}|+k\rangle_{\hat{C}}, \ldots, ||\hat{B}|+|\hat{A}|\rangle_{\hat{C}}\}. \tag{A.5}$$

We also define an auxiliary Hilbert space E with orthonormal basis

$$\{|e\rangle_E, |1\rangle_E, \ldots, ||\hat{A}||\hat{B}|\rangle_E\}, \tag{A.6}$$

and we represent a purification $|\varphi^{\xi_{AB}}\rangle_{ABE}$ of the state ξ_{AB} as

$$|\varphi^{\xi_{AB}}\rangle_{ABE} = \sum_{l=1}^{|\hat{A}||\hat{B}|} \sqrt{p_l}|\phi_l\rangle_{AB}|l\rangle_E. \tag{A.7}$$

Thus, an isometric extension of the Petz recovery channel in (A.1), according to the standard recipe in (3.42), is given by

$$\begin{aligned}
&\sum_{j=1}^{|B|} X_{AB}^{1/2}\left[X_A^{-1/2} \otimes |j\rangle_B\right]|j\rangle_{\hat{C}}|e\rangle_E + \sum_{k=1}^{|\hat{A}|}\sum_{l=1}^{|\hat{A}||\hat{B}|}\sqrt{p_l}|\phi_l\rangle_{AB}\langle k|_A\Pi_{X_A}^{\perp} \otimes |k+|B|\rangle_{\hat{C}}|l\rangle_E \\
&= X_{AB}^{1/2} X_A^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_E + |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}}\Pi_{X_A}^{\perp},
\end{aligned} \tag{A.8}$$

where we set

$$|\Gamma\rangle_{B\hat{C}} \equiv \sum_{j=1}^{|B|} |j\rangle_B |j\rangle_{\hat{C}}, \tag{A.9}$$

and we define the embedding map

$$U_{A\to\hat{C}} \equiv \sum_{k=1}^{|\hat{A}|} |k+|B|\rangle_{\hat{C}} \langle k|_A. \tag{A.10}$$

So we set the isometry $V_{A\to B\hat{C}E}$ as

$$V_{A\to B\hat{C}E} \equiv X_{AB}^{1/2} X_A^{-1/2} |\Gamma\rangle_{B\hat{C}} |e\rangle_E + |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp}. \tag{A.11}$$

By construction, the operator $V_{A\to B\hat{C}E}$ is an isometry, but we can verify by the following alternative calculation that this operator is indeed an isometry:

$$\begin{aligned} V^\dagger V &= \left(\langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE}\right) \times \\ &\qquad \left(X_{AB}^{1/2} X_A^{-1/2} |\Gamma\rangle_{B\hat{C}} |e\rangle_E + |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp}\right) &\text{(A.12)} \\ &= \langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} X_{AB}^{1/2} X_A^{-1/2} |\Gamma\rangle_{B\hat{C}} |e\rangle_E \\ &\qquad + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE} X_{AB}^{1/2} X_A^{-1/2} |\Gamma\rangle_{B\hat{C}} |e\rangle_E \\ &\qquad + \langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp} \\ &\qquad + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE} |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp} &\text{(A.13)} \\ &= \langle e|_E X_A^{-1/2} \langle\Gamma|_{B\hat{C}} X_{AB} |\Gamma\rangle_{B\hat{C}} X_A^{-1/2} |e\rangle_E + \Pi_{X_A}^{\perp} &\text{(A.14)} \\ &= X_A^{-1/2} X_A X_A^{-1/2} + \Pi_{X_A}^{\perp} &\text{(A.15)} \\ &= \Pi_{X_A} + \Pi_{X_A}^{\perp} &\text{(A.16)} \\ &= I_A. &\text{(A.17)} \end{aligned}$$

In the above, we have used the fact that $\langle e|_E |\varphi^{\xi_{AB}}\rangle_{ABE} = 0$. Now we extend V to

$$V_{A\hat{A}\to B\hat{C}E} \equiv V_{A\to B\hat{C}E} \otimes I_{\hat{A}}, \tag{A.18}$$

and observe that

$$V_{A\hat{A}\to B\hat{C}E} |\varphi^{X_A}\rangle_{A\hat{A}} = |\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{C}} |e\rangle_E = (X_{AB}^{1/2} \otimes I_{\hat{A}\hat{C}}) |\Gamma\rangle_{A\hat{A}} |\Gamma\rangle_{B\hat{C}} |e\rangle_E. \tag{A.19}$$

Let τ_{AB} be the output of the Petz recovery channel when the invertible state ω_A is input:

$$\tau_{AB} = X_{AB}^{1/2} \left(\left[X_A^{-1/2} \omega_A X_A^{-1/2}\right] \otimes I_B\right) X_{AB}^{1/2} + \mathrm{Tr}\{\Pi_{X_A}^{\perp} \omega_A\} \xi_{AB}. \tag{A.20}$$

Note that τ_{AB} is invertible because we chose ξ_{AB} to be invertible. Consider the positive definite operator $Y_{\hat{A}\hat{B}}$, whose $\hat{B}$ system we embed into $\mathrm{span}\{|1\rangle_{\hat{C}}, \ldots, |j\rangle_{\hat{C}}, \ldots, ||\hat{B}|\rangle_{\hat{C}}\}$ of system $\hat{C}$, calling the embedded operator $Y_{\hat{A}\hat{C}}$. We then have that $Y_{\hat{A}\hat{C}}^T U_{A\to\hat{C}} = 0$, and so we find that

$$\begin{aligned}
&V^\dagger \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) V \\
&= \left[\left(\langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE}\right) \otimes I_{\hat{A}}\right] \times \\
&\quad \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) \left[\left(X_{AB}^{1/2} X_A^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_E + |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp}\right) \otimes I_{\hat{A}}\right] \qquad \text{(A.21)} \\
&= \langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) X_{AB}^{1/2} X_A^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_E \\
&\quad + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) X_{AB}^{1/2} X_A^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_E \\
&\quad + \langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp} \\
&\quad + \Pi_{X_A}^{\perp} \left(U_{A\to\hat{C}}\right)^\dagger \langle\varphi^{\xi_{AB}}|_{ABE} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) |\varphi^{\xi_{AB}}\rangle_{ABE} U_{A\to\hat{C}} \Pi_{X_A}^{\perp}. \qquad \text{(A.22)}
\end{aligned}$$

The last three terms are equal to zero because $\langle\varphi^{\xi_{AB}}|_{ABE}|e\rangle_E = 0$ and $Y_{\hat{A}\hat{C}}^T U_{A\to\hat{C}} = 0$. Continuing, the last expression above is equal to

$$\begin{aligned}
&\langle\Gamma|_{B\hat{C}}\langle e|_E X_A^{-1/2} X_{AB}^{1/2} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T \otimes I_E\right) X_{AB}^{1/2} X_A^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_E \\
&= \langle\Gamma|_{B\hat{C}} \omega_A^{-1/2} \omega_A^{1/2} X_A^{-1/2} X_{AB}^{1/2} \left(\tau_{AB}^{-1} \otimes Y_{\hat{A}\hat{C}}^T\right) X_{AB}^{1/2} X_A^{-1/2} \omega_A^{1/2} \omega_A^{-1/2} |\Gamma\rangle_{B\hat{C}} \qquad \text{(A.23)} \\
&= \langle\Gamma|_{B\hat{C}} \left(\omega_A^{-1/2} \left[\omega_A^{1/2} X_A^{-1/2} X_{AB}^{1/2} \tau_{AB}^{-1} X_{AB}^{1/2} X_A^{-1/2} \omega_A^{1/2}\right] \omega_A^{-1/2} \otimes Y_{\hat{A}\hat{C}}^T\right) |\Gamma\rangle_{B\hat{C}} \qquad \text{(A.24)} \\
&\leq \left\|\omega_A^{1/2} X_A^{-1/2} X_{AB}^{1/2} \tau_{AB}^{-1} X_{AB}^{1/2} X_A^{-1/2} \omega_A^{1/2}\right\|_\infty \langle\Gamma|_{B\hat{C}} \left(\omega_A^{-1/2}\omega_A^{-1/2} \otimes Y_{\hat{A}\hat{C}}^T\right) |\Gamma\rangle_{B\hat{C}} \qquad \text{(A.25)} \\
&= \left\|\tau_{AB}^{-1/2} X_{AB}^{1/2} X_A^{-1/2} \omega_A X_A^{-1/2} X_{AB}^{1/2} \tau_{AB}^{-1/2}\right\|_\infty \langle\Gamma|_{B\hat{C}} \left(\omega_A^{-1} \otimes Y_{\hat{A}\hat{C}}^T\right) |\Gamma\rangle_{B\hat{C}} \qquad \text{(A.26)} \\
&\leq \langle\Gamma|_{B\hat{C}} \left(\omega_A^{-1} \otimes Y_{\hat{A}\hat{C}}^T\right) |\Gamma\rangle_{B\hat{C}} \qquad \text{(A.27)} \\
&= \omega_A^{-1} \otimes Y_{\hat{A}}^T. \qquad \text{(A.28)}
\end{aligned}$$

The third equality follows because $\left\|Z^\dagger Z\right\|_\infty = \left\|Z Z^\dagger\right\|_\infty$ for an operator Z. The last inequality follows because for a positive semi-definite operator W, we have that $\|W\|_\infty = \inf\{\mu : W \leq \mu I\}$. Applying this, we find that

$$\begin{aligned}
&\inf\left\{\mu : \tau_{AB}^{-1/2} X_{AB}^{1/2} X_A^{-1/2} \omega_A X_A^{-1/2} X_{AB}^{1/2} \tau_{AB}^{-1/2} \leq \mu I_{AB}\right\} \\
&= \inf\left\{\mu : X_{AB}^{1/2} X_A^{-1/2} \omega_A X_A^{-1/2} X_{AB}^{1/2} \leq \mu\, \tau_{AB}\right\} \qquad \text{(A.29)} \\
&= \inf\left\{\mu : X_{AB}^{1/2} X_A^{-1/2} \omega_A X_A^{-1/2} X_{AB}^{1/2} \leq \mu \left[X_{AB}^{1/2} X_A^{-1/2} \omega_A X_A^{-1/2} X_{AB}^{1/2} + \mathrm{Tr}\{\Pi_{X_A}^{\perp}\omega_A\}\xi_{AB}\right]\right\} \qquad \text{(A.30)} \\
&\leq 1. \qquad \text{(A.31)}
\end{aligned}$$

Let $P_{\hat{C}}$ be the embedding of the identity operator I_A into the subspace of $\hat{C}$ spanned by

$$\{||\hat{B}|+1\rangle_{\hat{C}}, \ldots, ||\hat{B}|+k\rangle_{\hat{C}}, \ldots, ||\hat{B}|+|\hat{A}|\rangle_{\hat{C}}\}. \qquad \text{(A.32)}$$

That is, $P_{\hat{C}} \equiv \sum_{k=1}^{|\hat{A}|} ||\hat{B}|+k\rangle_{\hat{C}}\langle|\hat{B}|+k|_{\hat{C}}$. Consider that, due to $P_{\hat{C}}|\Gamma\rangle_{B\hat{C}} = 0$, we have that

$$
\begin{aligned}
&V^{\dagger}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)V \\
&=\left[\left(\langle\Gamma|_{B\hat{C}}\langle e|_{E}X_{A}^{-1/2}X_{AB}^{1/2}+\Pi_{X_A}^{\perp}\left(U_{A\to\hat{C}}\right)^{\dagger}\langle\varphi^{\xi_{AB}}|_{ABE}\right)\otimes I_{\hat{A}}\right]\times \\
&\qquad\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)\left[\left(X_{AB}^{1/2}X_{A}^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_{E}+|\varphi^{\xi_{AB}}\rangle_{ABE}U_{A\to\hat{C}}\Pi_{X_A}^{\perp}\right)\otimes I_{\hat{A}}\right] &&\text{(A.33)}\\
&=\langle\Gamma|_{B\hat{C}}\langle e|_{E}X_{A}^{-1/2}X_{AB}^{1/2}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)X_{AB}^{1/2}X_{A}^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_{E} \\
&\qquad+\Pi_{X_A}^{\perp}\left(U_{A\to\hat{C}}\right)^{\dagger}\langle\varphi^{\xi_{AB}}|_{ABE}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)X_{AB}^{1/2}X_{A}^{-1/2}|\Gamma\rangle_{B\hat{C}}|e\rangle_{E} \\
&\qquad+\langle\Gamma|_{B\hat{C}}\langle e|_{E}X_{A}^{-1/2}X_{AB}^{1/2}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)|\varphi^{\xi_{AB}}\rangle_{ABE}U_{A\to\hat{C}}\Pi_{X_A}^{\perp} \\
&\qquad+\Pi_{X_A}^{\perp}\left(U_{A\to\hat{C}}\right)^{\dagger}\langle\varphi^{\xi_{AB}}|_{ABE}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)|\varphi^{\xi_{AB}}\rangle_{ABE}U_{A\to\hat{C}}\Pi_{X_A}^{\perp} &&\text{(A.34)}\\
&=\Pi_{X_A}^{\perp}\left(U_{A\to\hat{C}}\right)^{\dagger}\langle\varphi^{\xi_{AB}}|_{ABE}\left(\tau_{AB}^{-1}\otimes I_{\hat{A}}\otimes P_{\hat{C}}\otimes I_{E}\right)|\varphi^{\xi_{AB}}\rangle_{ABE}U_{A\to\hat{C}}\Pi_{X_A}^{\perp} &&\text{(A.35)}\\
&=\Pi_{X_A}^{\perp}\otimes I_{\hat{A}}\ \langle\varphi^{\xi_{AB}}|_{ABE}\tau_{AB}^{-1}|\varphi^{\xi_{AB}}\rangle_{ABE} &&\text{(A.36)}\\
&=\Pi_{X_A}^{\perp}\otimes I_{\hat{A}}\ \operatorname{Tr}\{\xi_{AB}\tau_{AB}^{-1}\}. &&\text{(A.37)}
\end{aligned}
$$

Observe that

$$
\begin{aligned}
f\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\right)&=\langle e|_{E}\left[f\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\right)\otimes I_{E}\right]|e\rangle_{E} &&\text{(A.38)}\\
&=\langle e|_{E}\left[f\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\otimes I_{E}\right)\right]|e\rangle_{E}. &&\text{(A.39)}
\end{aligned}
$$

Furthermore, consider that for an $\varepsilon\in(0,1)$, we have that

$$
\begin{aligned}
&\widetilde{Q}_{f}(X_{AB}\|Y_{AB};\tau_{AB}) \\
&=\langle\varphi^{X_{AB}}|_{A\hat{A}B\hat{B}}f\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\right)|\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{B}} &&\text{(A.40)}\\
&=\langle\varphi^{X_{AB}}|_{A\hat{A}B\hat{B}}\langle e|_{E}\left[f\left(\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{B}}^{T}\otimes I_{E}\right)\right]|\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{B}}|e\rangle_{E} &&\text{(A.41)}\\
&=\langle\varphi^{X_{AB}}|_{A\hat{A}B\hat{C}}\langle e|_{E}\left[f\left(\tau_{AB}^{-1}\otimes\left(Y_{\hat{A}\hat{C}}^{T}+\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]\right)\otimes I_{E}\right)\right]|\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{C}}|e\rangle_{E} &&\text{(A.42)}\\
&=\langle\varphi^{X_{A}}|_{A\hat{A}}V^{\dagger}\left[f\left(\tau_{AB}^{-1}\otimes\left(Y_{\hat{A}\hat{C}}^{T}+\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]\right)\otimes I_{E}\right)\right]V|\varphi^{X_{A}}\rangle_{A\hat{A}} &&\text{(A.43)}\\
&\geq\langle\varphi^{X_{A}}|_{A\hat{A}}f\left(V^{\dagger}\left[\tau_{AB}^{-1}\otimes\left(Y_{\hat{A}\hat{C}}^{T}+\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]\right)\otimes I_{E}\right]V\right)|\varphi^{X_{A}}\rangle_{A\hat{A}} &&\text{(A.44)}\\
&\geq\langle\varphi^{X_{A}}|_{A\hat{A}}f\left(\omega_{A}^{-1}\otimes Y_{\hat{A}}^{T}+\varepsilon\left[\Pi_{X_A}^{\perp}\otimes I_{\hat{A}}\ \operatorname{Tr}\{\xi_{AB}\tau_{AB}^{-1}\}\right]\right)|\varphi^{X_{A}}\rangle_{A\hat{A}} &&\text{(A.45)}
\end{aligned}
$$

The third equality follows because the term $\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]$ gets zeroed out due to the sandwich by $|\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{C}}$, given that $|\varphi^{X_{AB}}\rangle_{A\hat{A}B\hat{C}}$ only has support in $\operatorname{span}\{|1\rangle_{\hat{C}},\ldots,|j\rangle_{\hat{C}},\ldots,||\hat{B}|\rangle_{\hat{C}}\}$ (this can be seen explicitly by examining the proof of Proposition 4). Furthermore, note that the operator $Y_{\hat{A}\hat{C}}^{T}+\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]$ is invertible. The first inequality follows from the operator Jensen inequality [HP03]. The next inequality follows because

$$
\begin{aligned}
&V^{\dagger}\left[\tau_{AB}^{-1}\otimes\left(Y_{\hat{A}\hat{C}}^{T}+\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]\right)\otimes I_{E}\right]V \\
&=V^{\dagger}\left[\tau_{AB}^{-1}\otimes Y_{\hat{A}\hat{C}}^{T}\otimes I_{E}\right]V+V^{\dagger}\left[\tau_{AB}^{-1}\otimes\varepsilon\left[I_{\hat{A}}\otimes P_{\hat{C}}\right]\otimes I_{E}\right]V &&\text{(A.46)}\\
&\leq\omega_{A}^{-1}\otimes Y_{\hat{A}}^{T}+\varepsilon\Pi_{X_A}^{\perp}\otimes I_{\hat{A}}\ \operatorname{Tr}\{\xi_{AB}\tau_{AB}^{-1}\}, &&\text{(A.47)}
\end{aligned}
$$

and by applying operator anti-monotonicity of f. This establishes the inequality for all $\varepsilon \in (0, 1)$. Thus, we can apply continuity of f and take the limit $\varepsilon \searrow 0$ to find that

$$\widetilde{Q}_f(X_{AB}\|Y_{AB};\tau_{AB}) \geq \langle\varphi^{X_A}|_{A\hat{A}} f\left(\omega_A^{-1} \otimes Y_{\hat{A}}^T\right)|\varphi^{X_A}\rangle_{A\hat{A}} = \widetilde{Q}_f(X_A\|Y_A;\omega_A). \tag{A.48}$$

We can now take the supremum over all invertible states τ_{AB} to get the following inequality holding for all invertible states ω_A:

$$\widetilde{Q}_f(X_{AB}\|Y_{AB}) \geq \widetilde{Q}_f(X_A\|Y_A;\omega_A). \tag{A.49}$$

After taking a supremum over invertible states ω_A, we find that the inequality in (3.31) holds when X_A is not invertible.

Remark 15 *Several of the works [Pet86a, PS09, Pet10a, Sha10] on quantum f-divergence consider only invertible density operators and then appeal to continuity in order to extend proofs to the whole set of density operators. This is often understood as simply adding εI to a density operator and then taking the limit $\varepsilon \to 0$ later. In the second case given in the proof of Theorem 8, in which X_{AB} is not invertible but X_A is, the method can be understood as falling under an appeal to continuity. However, in the last case detailed above, when X_A is not invertible, the method arguably goes beyond a mere appeal to continuity, given the construction of the channel in (A.1), the corresponding isometric extension in (A.11), and the ensuing analysis.*

References

[AD15] Koenraad M. R. Audenaert and Nilanjana Datta. α-z-Rényi relative entropies. *Journal of Mathematical Physics*, 56(2):022202, February 2015. arXiv:1310.7178.

[AS66] 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)*, 28(1):131–142, 1966.

[Bei13] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, December 2013. arXiv:1306.5920.

[BFT17] Mario Berta, Omar Fawzi, and Marco Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107(12):2239–2265, December 2017. arXiv:1512.02615.

[BG15] Fernando G. S. L. Brandao and Gilad Gour. Reversible framework for quantum resource theories. *Physical Review Letters*, 115(7):070503, August 2015. arXiv:1502.03149.

[Bha97] Rajendra Bhatia. *Matrix Analysis*. Springer, 1997.

[BK02] Howard Barnum and Emanuel Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, May 2002. arXiv:quant-ph/0004088.

[BS82] V. P. Belavkin and P. Staszewski. C*-algebraic generalization of relative entropy and entropy. *Annales de l'I.H.P. Physique théorique*, 37(1):51–58, 1982.

[BSS12] Igor Bjelakovic and Rainer Siegmund-Schultze. Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb. July 2012. arXiv:quant-ph/0307170.

[CCYZ12] Patrick J. Coles, Roger Colbeck, Li Yu, and Michael Zwolak. Uncertainty relations from simple entropic properties. *Physical Review Letters*, 108(21):210405, May 2012. arXiv:1112.0543.

[CG18] Eric Chitambar and Gilad Gour. Quantum resource theories. June 2018. arXiv:1806.06107.

[CMW16] Tom Cooney, Milan Mosonyi, and Mark M. Wilde. Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *Communications in Mathematical Physics*, 344(3):797–829, June 2016. arXiv:1408.3373.

[Csi67] Imre Csiszár. Information type measure of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

[Dat09] Nilanjana Datta. Min- and max-relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, June 2009. arXiv:0803.2770.

[Dat11] Nilanjana Datta. Relative entropies and entanglement monotones. In *Mathematical Results In Quantum Physics: Proceedings of the QMath11 Conference*, pages 3–20. World Scientific Publishing Company, May 2011.

[DL14] Nilanjana Datta and Felix Leditzky. A limit of the quantum Rényi divergence. *Journal of Physics A: Mathematical and Theoretical*, 47(4):045304, January 2014. arXiv:1308.5961.

[dRKR15] Lidia del Rio, Lea Kraemer, and Renato Renner. Resource theories of knowledge. November 2015. arXiv:1511.08818.

[DW15] David Ding and Mark M. Wilde. Strong converse exponents for the feedback-assisted classical capacity of entanglement-breaking channels. June 2015. arXiv:1506.02228.

[FL13] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12):122201, December 2013. arXiv:1306.5358.

[Fri15] Tobias Fritz. Resource convertibility and ordered commutative monoids. *Mathematical Structures in Computer Science*, page 189, 2015. arXiv:1504.03661.

[GW15] Manish Gupta and Mark M. Wilde. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. *Communications in Mathematical Physics*, 334(2):867–887, March 2015. arXiv:1310.7028.

[Han13] Frank Hansen. The fast track to Löwner's theorem. *Linear Algebra and its Applications*, 438(11):4557–4571, June 2013. arXiv:1112.0098.

[Hay06] Masahito Hayashi. *Quantum Information: An Introduction*. Springer, 2006.

[Hay07] Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76(6):062301, December 2007. arXiv:quant-ph/0611013.

[HJPW04] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246(2):359–374, April 2004. arXiv:quant-ph/0304007.

[HM17] Fumio Hiai and Milán Mosonyi. Different quantum f-divergences and the reversibility of quantum operations. *Reviews in Mathematical Physics*, 29(7):1750023, August 2017. arXiv:1604.03089.

[HMPB11] Fumio Hiai, Milán Mosonyi, Dénes Petz, and Cedric Beny. Quantum f-divergences and error correction. *Reviews in Mathematical Physics*, 23(7):691–747, August 2011. arXiv:1008.2529.

[Hol12] Alexander S. Holevo. *Quantum Systems, Channels, Information*. de Gruyter Studies in Mathematical Physics (Book 16). de Gruyter, November 2012.

[HP91] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99–114, December 1991.

[HP03] Frank Hansen and Gert K. Pedersen. Jensen's operator inequality. *Bulletin of the London Mathematical Society*, 35(4):553–564, July 2003. arXiv:math/0204049.

[HP12] Fumio Hiai and Dénes Petz. From quasi-entropy to various quantum information quantities. *Publ. RIMS Kyoto University*, 48:525–542, 2012.

[Jen18] Anna Jenčová. Rényi relative entropies and noncommutative L_p-spaces. *to appear in Annales Henri Poincaré*, 2018. arXiv:1609.08462.

[KdR16] Lea Kraemer and Lidia del Rio. Currencies in resource theories. May 2016. arXiv:1605.01064.

[KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[Kle31] O. Klein. Zur Quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre. *Z. Physik*, 72:767–775, 1931.

[KW17] Eneet Kaur and Mark M. Wilde. Relative entropy of steering: on its definition and properties. *Journal of Physics A: Mathematical and Theoretical*, 50(46):465301, November 2017. arXiv:1612.07152.

[Lin75] Göran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, June 1975.

[LS13] Matthew S. Leifer and Robert W. Spekkens. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Physical Review A*, 88(5):052130, November 2013. arXiv:1107.5849.

[Mat13] Keiji Matsumoto. A new quantum version of f-divergence. 2013. arXiv:1311.4722.

[MH11] Milán Mosonyi and Fumio Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, April 2011. arXiv:0912.1286.

[MLDS+13] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new definition and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv:1306.3142.

[MO15] Milán Mosonyi and Tomohiro Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Communications in Mathematical Physics*, 334(3):1617–1648, March 2015. arXiv:1309.3228.

[Mor63] Tetsuzo Morimoto. Markov processes and the h-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963.

[Nag06] Hiroshi Nagaoka. The converse part of the theorem for quantum Hoeffding bound. November 2006. arXiv:quant-ph/0611289.

[NP05] Michael A. Nielsen and Denés Petz. A simple proof of the strong subadditivity inequality. *Quantum Information and Computation*, 5(6):507–513, September 2005. arXiv:quant-ph/0408130.

[ON00] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, November 2000. arXiv:quant-ph/9906090.

[OP93] Masanori Ohya and Dénes Petz. *Quantum Entropy and Its Use*. Springer-Verlag, 1993.

[Pet85] Dénes Petz. Quasi-entropies for states of a von Neumann algebra. *Publ. RIMS, Kyoto University*, 21:787–800, 1985.

[Pet86a] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports in Mathematical Physics*, 23:57–65, 1986.

[Pet86b] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105(1):123–131, March 1986.

[Pet88] Dénes Petz. Sufficiency of channels over von Neumann algebras. *Quarterly Journal of Mathematics*, 39(1):97–108, 1988.

[Pet10a] Dénes Petz. From f-divergence to quantum quasi-entropies and their use. *Entropy*, 12(3):304–325, 2010.

[Pet10b] Dénes Petz. From quasi-entropy. September 2010. arXiv:1009.2679.

[PR98] Dénes Petz and Mary Beth Ruskai. Contraction of generalized relative entropy under stochastic mappings on matrices. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1(1):83–89, January 1998.

[PS09] Dénes Petz and V.E. Sandor Szabo. From quasi-entropy to skew information. *International Journal of Mathematics*, 20(11):1421–1430, November 2009. arXiv:0712.2881.

[Rén61] Alfréd Rényi. On measures of entropy and information. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1:547–561, 1961. held at the Statistical Laboratory, University of California, 1960, edited by J. Neyman (University of California Press, Berkeley).

[Rus02] Mary Beth Ruskai. Inequalities for quantum entropy: A review with conditions for equality. *Journal of Mathematical Physics*, 43(9):4358–4375, September 2002. arXiv:quant-ph/0205064.

[Sha10] Naresh Sharma. Equality conditions for the quantum f-relative entropy and generalized data processing inequalities. In *2010 IEEE International Symposium on Information Theory*, pages 2698–2702, June 2010. arXiv:0906.4755.

[Sio58] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[Sti55] William F. Stinespring. Positive functions on C*-algebras. *Proceedings of the American Mathematical Society*, 6:211–216, 1955.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, December 2009. arXiv:0811.1221.

[Tom12] Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012. arXiv:1203.2142.

[TWW17] Marco Tomamichel, Mark M. Wilde, and Andreas Winter. Strong converse rates for quantum communication. *IEEE Transactions on Information Theory*, 63(1):715–727, January 2017. arXiv:1406.2946.

[Uhl77] Armin Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, 1977.

[Ume62] Hisaharu Umegaki. Conditional expectations in an operator algebra IV (entropy and information). *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962.

[VP98] Vlatko Vedral and Martin B. Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57(3):1619–1633, March 1998. arXiv:quant-ph/9707035.

[Weh79] Alfred Wehrl. A remark on the concavity of entropy. *Foundations of Physics*, 9(11):939–946, December 1979.

[Wil17] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, second edition, 2017. Available as arXiv:1106.1445.

[WTB17] Mark M. Wilde, Marco Tomamichel, and Mario Berta. Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory*, 63(3):1792–1817, March 2017. arXiv:1602.08898.

[WWY14] Mark M. Wilde, Andreas Winter, and Dong 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, October 2014. arXiv:1306.1586.