Reviews in Mathematical Physics

Quantum f-divergences via Nussbaum-Szkoła and Applications to f-divergence Inequalities

--Manuscript Draft--

Manuscript Number:	RMP-D-23-00013
Full Title:	Quantum f-divergences via Nussbaum-Szkoła and Applications to f-divergence Inequalities
Article Type:	Special Issue Paper
Section/Category:	Special Issue, QMath15 – UC Davis 2022
Keywords:	Quantum f-divergence; Relative divergence; relative entropy; Relative modular operator; Nussbaum-Szkoła distributions
Abstract:	The main result in this article shows that the quantum f-divergence of two states is equal to the classical f-divergence of the corresponding Nussbaum-Szkoła distributions. This provides a general framework for studying certain properties of quantum entropic quantities using the corresponding classical entities. The usefulness of the main result is illustrated by obtaining several quantum f-divergence inequalities from their classical counterparts. All results presented here are valid in both finite and infinite dimensions and hence can be applied to continuous variable systems as well.

Quantum f-divergences via Nussbaum-Szkoła Distributions and Applications to f-divergence Inequalities

George Androulakis[1] and Tiju Cherian John[2]

[1] *University of South Carolina, Columbia, South Carolina, USA*
giorgis@math.sc.edu
[2] *University of Arizona, Arizona, USA*
tijucherian@fulbrightmail.org

January 30, 2023

Abstract

The main result in this article shows that the quantum f-divergence of two states is equal to the classical f-divergence of the corresponding Nussbaum-Szkoła distributions. This provides a general framework for studying certain properties of quantum entropic quantities using the corresponding classical entities. The usefulness of the main result is illustrated by obtaining several quantum f-divergence inequalities from their classical counterparts. All results presented here are valid in both finite and infinite dimensions and hence can be applied to continuous variable systems as well.

Keywords: Quantum f-divergence, Relative divergence, relative entropy, Relative modular operator, Nussbaum-Szkoła distributions

2020 Mathematics Subject classification: Primary 81P17; Secondary 81P99.

1 Introduction

Following the pioneering work of Shannon on information theory, Kullback and Leibler defined a divergence of two probability measures P and Q, using the formula [1]

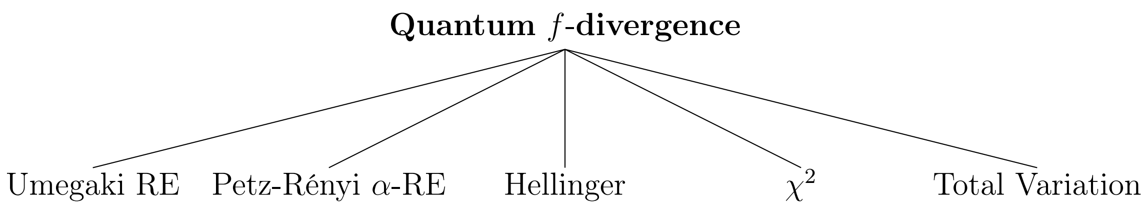
$$D(P\|Q) = \begin{cases} \sum_i P(i) \log \frac{P(i)}{Q(i)}, & \text{if } P \ll Q \text{ with the convention } 0 \log \frac{0}{0} = 0; \\ \infty, & \text{otherwise,} \end{cases}$$

This is now known as the Kullback-Leibler divergence of probability measures. Later on, Rényi [2] showed that the Kullback-Leibler divergence can be extended to a class of relative entropic quantities D_α for $\alpha \in (0,1) \cup (1,\infty)$, by defining

$$D_\alpha(P\|Q) = \begin{cases} \frac{1}{\alpha-1} \log \sum_i P(i)^\alpha Q(i)^{1-\alpha}, & \text{if } \alpha < 1, \text{ or } P \ll Q; \\ \infty, & \text{otherwise,} \end{cases}$$

with the convention $0 \cdot \infty = 0$. Then $\lim_{\alpha \to 1} D_\alpha(P||Q) = D(P||Q)$. In the same article [2, Page 561] Rényi introduced another quantity which is now known as f-divergence. This has been studied extensively in the early years in [3, 4, 5]. One may refer to [6, 7] for modern treatments. Both Kullback-Leibler and Rényi α-relative entropy, as well as many other important relative entropic quantities such as Hellinger, χ^2, and total variation distance are particular cases of f-divergences.

Quantizing these developments, Umegaki extended the notion of Kullback-Leibler divergence to the setting of von Neumann algebras [8], and Petz defined general quasi-entropies for quantum systems, (which includes quantum f-divergences) [9]. Quantum f-divergences have specific operational meaning especially because they satisfy the data processing inequality [10]. Specific cases of quantum f-divergences like Petz-Rényi α-relative entropy are useful in quantum hypothesis testing and related areas [11, 12, 13, 14, 15, 10]. One may refer to [16] and the other articles cited above to know more about different uses of relative entropy. Similar to the classical setting, a number of quantum entropic quantities arise from quantum f-divergences. Thus quantum f-divergence is a parent quantity for several useful relative entropic quantities. Some examples are pictorially represented below.



The article [11] cited above is particularly important in our context. In that article, Nussbaum and Szkoła introduced classical probability distributions, (now called Nussbaum-Szkoła distributions), corresponding to any two finite dimensional density matrices. They also proved that the Petz-Rényi α-relative entropy of two quantum states is equal to the Rényi α-relative entropy of the corresponding Nussbaum-Szkoła distributions. Furthermore, these distributions were extensively used in the literature to study various properties of Petz-Rényi α-relative entropy.

In the context of our work in this article, most important related works were produced by Hiai and Mosonyi [17], Hiai [10], and Berta, Scholz and Tomamichel [15] where quantum f-divergences are studied in general or the more specific case of Petz-Rényi relative entropy is studied. Also, Seshadreeshan, Lami and Wilde in [14] and Parthasarathy in [18] studied the Petz-Rényi relative entropy.

The structure of this article is as follows. In Section 2, we prove that the quantum f-divergence of two states on a finite or infinite dimensional Hilbert space is equal to the corresponding classical f-divergence of the Nussbaum-Szkoła distributions (see Theorem 2.8). This is a far reaching generalization of the result by Nussbaum and Szkoła that we described above. In Section 3 we show that several inequalities between various quantum relative entropic quantities follow immediately from their classical counterparts, because of our Theorem 2.8. We collect all the necessary background materials in the Appendix.

2 Quantum f-divergences via Nussbaum-Szkoła Distributions

By a state on a Hilbert space we mean a positive trace class operator with unit trace. Let ρ and σ be any two states on a Hilbert space $\mathcal{K}$. The relative modular operator $\Delta_{\rho,\sigma}$ with respect to the states ρ and σ was introduced by Araki in [19] and [20]. It is (in general) an unbounded positive selfadjoint operator defined on a dense subspace of the Hilbert space $\mathcal{B}_2(\mathcal{K})$ of Hilbert-Schmidt operators on $\mathcal{K}$, where the scalar product on $\mathcal{B}_2(\mathcal{K})$ is denoted as $\langle\cdot|\cdot\rangle_2$. Let $\xi^{\Delta_{\rho,\sigma}}$ denote the spectral measure associated with $\Delta_{\rho,\sigma}$. A detailed analysis of the relative modular operator in our setting, and its spectral decomposition is provided in Appendix B. Keeping these notations we define the quantum f-divergence of ρ, σ in Definition 2.1. Our definition of quantum f-divergences is same as that of Hiai in [10, Definition 2.1] adapted to the specific von Neumann algebra $\mathcal{B}(\mathcal{H})$.

Before defining the f-divergence, we need to fix a few notations and conventions related to a convex (or concave) function $f : (0,\infty) \to \mathbb{R}$. First of all let

$$f(0) := \lim_{t\downarrow 0} f(t), \quad f'(\infty) := \lim_{t\to\infty} \frac{f(t)}{t}. \tag{2.1}$$

Moreover, we use the conventions,

$$\begin{aligned} &0f\left(\frac{0}{0}\right) = 0, \\ &0f\left(\frac{a}{0}\right) := \lim_{t\to 0} tf\left(\frac{a}{t}\right) = a \lim_{s\to\infty} \frac{f(s)}{s} = af'(\infty), \text{ for } a > 0, \\ &0\cdot(\pm\infty) = 0 \end{aligned} \tag{2.2}$$

along with $a\cdot(\pm\infty) = \pm\infty$, for $a > 0$. In what follows, the notation "$\int\limits_{0^+}^{\infty}$" is used to denote an integral over the open interval $(0,\infty) \subseteq \mathbb{R}$. For a state η, let Π_η denote the orthogonal projection onto the support of η and $\Pi_\eta^\perp$, the projection onto $\operatorname{Ker}\eta$.

Now we define the f-divergence of two states ρ and σ as in [10, Definition 2.1]. A motivation for this definition can be seen in [17, Equations 3.9 and 3.12].

Definition 2.1. Let ρ and σ be states on a Hilbert space $\mathcal{K}$. If $f : (0,\infty) \to \mathbb{R}$ is a convex (or concave) function then the f-divergence $D_f(\rho||\sigma)$ of ρ from σ is defined as

$$D_f(\rho||\sigma) = \int_{0^+}^{\infty} f(\lambda)\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}(\mathrm{d}\lambda)\middle|\sqrt{\sigma}\right\rangle_2 + f(0)\operatorname{Tr}\left(\sigma\Pi_\rho^\perp\right) + f'(\infty)\operatorname{Tr}\left(\rho\Pi_\sigma^\perp\right). \tag{2.3}$$

The reader is warned that we use the same notation D_f for both classical and quantum f-divergence. It will be clear from the context whether we use the quantum or classical divergence each time.

Remark 2.2. It is known that, if f is convex then $D_f(\rho||\sigma)$ is well defined and takes value in $(-\infty,\infty]$ for every state ρ and σ. A proof of this fact can be seen in [10, Lemma 2.1]. In particular, we have

$$\int_{0^+}^{\infty} f(\lambda)\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}(\mathrm{d}\lambda)\middle|\sqrt{\sigma}\right\rangle_2 > -\infty, \quad (\text{when } f \text{ is convex}), \tag{2.4}$$

since the second term and the third term in (2.3) are strictly bigger than $-\infty$. If f is concave, then $-f$ is convex and we get

$$\int_{0^+}^{\infty} f(\lambda)\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}(\mathrm{d}\lambda)\middle|\sqrt{\sigma}\right\rangle_2 < \infty, \quad \text{(when } f \text{ is concave)}. \tag{2.5}$$

Examples 2.3. We can create various examples of relative entropic quantities by taking different functions f in the definition of f-divergence. The following are important examples motivated from the corresponding classical counterparts [21].

1. $f(t) = t\log t \Rightarrow D(\rho\|\sigma) := D_f(\rho\|\sigma)$, the **Umegaki Relative Entropy**.
2. For $\alpha \in (0,1)\cup(1,\infty)$, $f_\alpha(t) = t^\alpha$ we obtain $D_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log D_{f_\alpha}(\rho\|\sigma)$, the **Petz-Rényi $\boldsymbol{\alpha}$-relative entropy**.
3. For $\alpha \in (0,1)\cup(1,\infty)$, $f_\alpha(t) = \frac{t^\alpha-1}{\alpha-1}$ we obtain $\mathcal{H}_\alpha(\rho||\sigma) = D_f(\rho\|\sigma)$, the **Quantum Hellinger α-divergence**.
4. $f(t) = |t-1| \Rightarrow V(\rho||\sigma) := D_f(\rho\|\sigma)$, the **Quantum Total Variation**.
5. $f(t) = (t-1)^2 \Rightarrow \chi^2(\rho||\sigma) := D_f(\rho\|\sigma)$, the **Quantum $\boldsymbol{\chi^2}$-divergence**.

Now we define the Nussbaum-Szkoła distributions associated with the states ρ and σ, (see [11] for the original definition in the finite dimensional setting, and [12] for an infinite dimensional version in the setting of gaussian states). Our definition is valid in both finite and infinite dimensions.

Definition 2.4. (*Nussbaum-Szkoła distributions.*) Let $\mathcal{K}$ be a complex Hilbert space with $\dim\mathcal{K} = |\mathcal{I}|$, where $\mathcal{I} = \{1,2,\ldots,n\}$ for some $n \in \mathbb{N}$, or $\mathcal{I} = \mathbb{N}$. Let ρ and σ be states on $\mathcal{K}$ with spectral decomposition

$$\begin{aligned}
\rho &= \sum_{i\in\mathcal{I}} r_i\,|u_i\rangle\langle u_i|\,, \quad r_i \geq 0, \quad \sum_{i\in\mathcal{I}} r_i = 1, \quad \{u_i\}_{i\in\mathcal{I}} \text{ is an orthonormal basis of } \mathcal{K};\\
\sigma &= \sum_{j\in\mathcal{I}} s_j\,|v_j\rangle\langle v_j|\,, \quad s_j \geq 0, \quad \sum_{j\in\mathcal{I}} s_j = 1, \quad \{v_j\}_{j\in\mathcal{I}} \text{ is an orthonormal basis of } \mathcal{K}.
\end{aligned} \tag{2.6}$$

Define the Nussbaum-Szkoła distribution P and Q associated with ρ and σ on $\mathcal{I}\times\mathcal{I}$ by,

$$\begin{aligned}
P(i,j) &= r_i|\langle u_i|v_j\rangle|^2,\\
Q(i,j) &= s_j|\langle u_i|v_j\rangle|^2, \quad \forall (i,j) \in \mathcal{I}\times\mathcal{I}.
\end{aligned} \tag{2.7}$$

Remark 2.5. Observe first that

$$\sum_{i,j} P(i,j) = \sum_i r_i \sum_j |\langle u_i|v_j\rangle|^2 = \sum_i r_i = 1,$$

since $\{u_i\}_i$ and $\{v_j\}_j$ are orthonormal bases. Similarly, $\sum_{i,j} Q(i,j) = 1$. Hence P and Q are probability distributions on $\mathcal{I}\times\mathcal{I}$.

For any two probability distributions and for any convex (or concave) function f on $(0,\infty)$ the f-divergence of one distribution from the other is defined as in Definition A.1 and the discrete special case is described in Remark A.3. Under the same notations and conventions as in (2.1) and (2.2) the classical f-divergence $D_f(P||Q)$ of P from Q for a convex (or concave) f on $(0,\infty)$ is

$$D_f(P||Q) = \sum_{i,j\in\mathcal{I}} f\left(\frac{P(i,j)}{Q(i,j)}\right) Q(i,j). \tag{2.8}$$

It is known that the sum of the series defining $D_f(P||Q)$ as above is independent of any rearrangement of the series but we provide a proof of this fact in Remark A.3 in the Appendix. We refer to [6, 7] for various properties of classical f-divergences. In the next lemma, we compute a useful formula for the classical f-divergence of the Nussbaum-Szkoła distributions.

Lemma 2.6. *The f-divergence of the Nussbaum-Szkoła distributions can be computed as*

$$D_f(P||Q) = \sum_{\left\{\substack{i,j:\\ r_is_j\neq 0}\right\}} f\left(r_is_j^{-1}\right) s_j|\langle u_i|v_j\rangle|^2 + f(0)Q(P=0) + f'(\infty)P(Q=0). \tag{2.9}$$

Proof. By (A.4) in appendix, we have

$$\begin{aligned} D_f(P||Q) = &\sum_{\left\{\substack{i,j\,:r_i\neq 0,s_j\neq 0,\\ \langle u_i|v_j\rangle\neq 0}\right\}} f\left(\frac{r_i}{s_j}\right) s_j|\langle u_i|v_j\rangle|^2 + \sum_{\left\{\substack{i,j\,:r_i=0,s_j\neq 0,\\ \langle u_i|v_j\rangle\neq 0}\right\}} f(0)s_j|\langle u_i|v_j\rangle|^2 \\ &+ \sum_{\left\{\substack{i,j\,:r_i\neq 0,s_j=0,\\ \langle u_i|v_j\rangle\neq 0}\right\}} f'(\infty)r_i|\langle u_i|v_j\rangle|^2, \end{aligned}$$

which is same as (2.9). ☐

Now we analyse the measure involved in the definition of quantum f-divergence.

Lemma 2.7. *The measure $E\mapsto \langle\sqrt{\sigma}|\xi^{\Delta_{\rho,\sigma}}(E)|\sqrt{\sigma}\rangle_2$ on the Borel sigma algebra of $\mathbb{R}$ is supported on the set $\{0\}\cup\{r_is_j^{-1} : r_i\neq 0, s_j\neq 0\}$. For $r_i\neq 0$ and $s_j\neq 0$,*

$$\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}\{r_is_j^{-1}\}\middle|\sqrt{\sigma}\right\rangle_2 = \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_is_j^{-1}\}} s_{j'}|\langle u_{i'}|v_{j'}\rangle|^2, \tag{2.10}$$

and

$$\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}\{0\}\middle|\sqrt{\sigma}\right\rangle_2 = \sum_{\left\{\substack{i,j\,:\,r_i=0,s_j\neq 0\\ \langle u_i|v_j\rangle\neq 0}\right\}} s_j|\langle u_i|v_j\rangle|^2 = Q(P=0). \tag{2.11}$$

Proof. Recall from Proposition B.1 that the spectrum of the relative modular operator $\Delta_{\rho,\sigma}$ is supported on the set $\{0\}\cup\{r_is_j^{-1} : r_i\neq 0, s_j\neq 0\}$. Hence it is clear that the measure in the statement of the lemma is supported in the same set. By the spectral decomposition of the relative modular operator $\Delta_{\rho,\sigma}$ obtained in Proposition B.1, the spectral measure $\xi^{\Delta_{\rho,\sigma}}$ is supported on the eigenvalues of $\Delta_{\rho,\sigma}$, and satisfies

$$\xi^{\Delta_{\rho,\sigma}}\{\lambda\} = \begin{cases} \sum\limits_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_is_j^{-1}\}} |X_{i'j'}\rangle\langle X_{i'j'}|, & \text{for } \lambda = r_is_j^{-1}, r_i\neq 0, s_j\neq 0;\\ \sum\limits_{\{i,j\,:\,r_i=0 \text{ or } s_j=0\}} |X_{ij}\rangle\langle X_{ij}|, & \text{for } \lambda=0;\\ 0, & \text{otherwise,} \end{cases}$$

where $X_{ij} = |u_i\rangle\langle v_j| \in \mathcal{B}_2(\mathcal{K})$, for all $i, j \in \mathcal{I}$. Therefore for $r_i \neq 0$ and $s_j \neq 0$,

$$\begin{aligned}
\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}\{r_i s_j^{-1}\}\middle|\sqrt{\sigma}\right\rangle_2 &= \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} \left\langle\sqrt{\sigma}\middle|X_{i'j'}\right\rangle_2 \left\langle X_{i'j'}\middle|\sqrt{\sigma}\right\rangle_2 \\
&= \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} \left|\left\langle\sqrt{\sigma}\middle|X_{i'j'}\right\rangle_2\right|^2 \\
&= \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} \left|\mathrm{Tr}\left\{X_{i'j'}\sqrt{\sigma}\right\}\right|^2 \\
&= \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} \left|\mathrm{Tr}\left\{(|u_{i'}\rangle\langle v_{j'}|)\left(\sum_k \sqrt{s_k}\,|v_k\rangle\langle v_k|\right)\right\}\right|^2 \\
&= \sum_{\{i',j'\,:\,r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} s_{j'}|\langle u_{i'}|v_{j'}\rangle|^2.
\end{aligned}$$

A similar computation as above shows that

$$\begin{aligned}
\left\langle\sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}\{0\}\middle|\sqrt{\sigma}\right\rangle_2 &= \sum_{\{i,j\,:\,r_i=0 \text{ or } s_j=0\}} \left\langle\sqrt{\sigma}\middle|X_{ij}\right\rangle_2 \left\langle X_{ij}\middle|\sqrt{\sigma}\right\rangle_2 \\
&= \sum_{\{i,j\,:\,r_i=0 \text{ or } s_j=0\}} s_j|\langle u_i|v_j\rangle|^2 \\
&= \sum_{\left\{\substack{i,j\,:\,r_i=0, s_j\neq 0 \\ \langle u_i|v_j\rangle\neq 0}\right\}} s_j|\langle u_i|v_j\rangle|^2 \\
&= Q(P=0).
\end{aligned}$$

□

In several occasions below, we will use the following rearrangement trick for a sum of the form $\sum_k f(x_k)y_k$ with $y_k > 0$ for all k. Notice that if the sum of the negative terms in the series above is strictly bigger than $-\infty$ (or the sum of positive terms in the series is strictly less than ∞), then any rearrangement of the series produces the same sum. In particular, we have

$$\sum_k f(x_k)y_k = \sum_k f(x_k) \sum_{\{\ell\,:\,x_\ell=x_k\}} y_\ell. \tag{2.12}$$

Our next theorem is the main result in this article. It proves that the quantum f-divergence of two states is same as the classical f-divergence of corresponding Nussbaum-Szkoła distributions.

Theorem 2.8. *Let ρ, σ be as in (2.6) and P, Q denote the corresponding Nussbaum-Szkoła distributions. Let $f : (0, \infty) \to \mathbb{R}$ be a convex (or concave) function and $D_f(\rho||\sigma)$, $D_f(P||Q)$ respectively denote the quantum f-divergence of ρ from σ and the classical f-divergence of P from Q. Then*

$$D_f(\rho||\sigma) = D_f(P||Q). \tag{2.13}$$

Proof. We will show that each term on the right side of (2.3) matches with the corresponding terms on the right side of (2.9) We first evaluate the first term on the right side of (2.3) using Lemma 2.7 and (2.12)

$$\begin{aligned}\int_{0^+}^{\infty} f(\lambda)\left\langle \sqrt{\sigma}\middle|\xi^{\Delta_{\rho,\sigma}}(\mathrm{d}\lambda)\middle|\sqrt{\sigma}\right\rangle_2 &= \sum_{\left\{\substack{i,j\,:\\ r_i\neq 0, s_j\neq 0}\right\}} f(r_i s_j^{-1}) \sum_{\{i',j'\,:\, r_{i'}s_{j'}^{-1}=r_i s_j^{-1}\}} s_{j'}|\langle u_{i'}|v_{j'}\rangle|^2 \\ &= \sum_{\left\{\substack{i,j\,:\\ r_i\neq 0, s_j\neq 0}\right\}} f(r_i s_j^{-1}) s_j |\langle u_i|v_j\rangle|^2.\end{aligned}$$

This is same as the first term on the right side of (2.9). Now

$$\begin{aligned}\operatorname{Tr}\left(\sigma\Pi_\rho^\perp\right) &= \operatorname{Tr}\left(\sum_j s_j\,|v_j\rangle\langle v_j| \sum_{\{i\,:\,r_i=0\}} |u_i\rangle\langle u_i|\right) \\ &= \sum_{\left\{\substack{i,j\,:\\ r_i=0}\right\}} s_j\left|\langle u_i|v_j\rangle^2\right|.\end{aligned}$$

Therefore,

$$f(0)\operatorname{Tr}\left(\sigma\Pi_\rho^\perp\right) = f(0)Q(P=0),$$

which is same as the second term on the right side of (2.9). A similar computation as above shows that

$$f'(\infty)\operatorname{Tr}\left(\rho\Pi_\sigma^\perp\right) = f'(\infty)P(Q=0),$$

which is same as the third term on the right side of (2.9). Thus we proved (2.13). □

The correspondence between quantum states and Nussbaum-Szkoła distributions we defined in this article have the following important properties.

Proposition 2.9. *Let ρ and σ be as in (2.6) and let P and Q be as in (2.7) then*

$$P = Q \Leftrightarrow \rho = \sigma.$$

The formula for classical f-divergence simplifies if $P \ll Q$ (see (A.2)). This property for the Nussbaum-Szkoła distributions translate to the condition $\operatorname{Supp}\rho \subseteq \operatorname{Supp}\sigma$ as shown in the next proposition.

Proposition 2.10. *Let ρ and σ be as in (2.6) and let P and Q be as in (2.7), then*

$$\operatorname{Supp}\rho \subseteq \operatorname{Supp}\sigma \Leftrightarrow P \ll Q.$$

For proving Proposition 2.9, we need the following lemma.

Lemma 2.11. *Let ρ and σ be as in (2.6). Then $\operatorname{Supp}\rho \subseteq \operatorname{Supp}\sigma$ if and only if $s_j = 0$ for some j implies that for every i at least one of the two quantities $\{\langle u_i|v_j\rangle, r_i\}$ is equal to zero.*

Proof. ($\Rightarrow$) Assume $\operatorname{Supp}\rho \subseteq \operatorname{Supp}\sigma$. The assumption that $s_j = 0$ implies that $v_j \in \operatorname{Ker}\sigma$. Now the condition $\langle u_i|v_j\rangle \neq 0$ implies that $u_i \notin (\operatorname{Ker}\sigma)^\perp = \operatorname{Supp}\sigma$. Since $\operatorname{Supp}\sigma \supseteq \operatorname{Supp}\rho$ and u_i is an eigenvector of ρ we see that

$$u_i \notin \operatorname{Supp}\rho \Rightarrow u_i \in \operatorname{Ker}\rho \Rightarrow r_i = 0.$$

($\Leftarrow$) We will prove that $\operatorname{Ker}\sigma \subseteq \operatorname{Ker}\sigma$. It is enough to prove that every v_j for which $s_j = 0$ belong to $\operatorname{Ker}\rho$. Fix such j such that $s_j = 0$, by our assumption, $\langle u_i|v_j\rangle = 0$ for all i such that $r_i \neq 0$ and hence $\sum_{\{i:r_i\neq 0\}} \langle u_i|v_j\rangle\,|u_i\rangle = 0$. This means that the projection of v_j to $\operatorname{Supp}\rho$ is the zero vector which means that $v_j \in \operatorname{Ker}\rho$. □

Proof of Proposition 2.9

Proof. Clearly, if $\rho = \sigma$ then $P = Q$. Now assume $P = Q$. By definition

$$P = Q \Rightarrow r_i |\langle u_i | v_j \rangle|^2 = s_j |\langle u_i | v_j \rangle|^2, \quad \forall i, j. \tag{2.14}$$

Therefore, $s_j = 0$ for some j implies that for every i at least one of the two quantities $\{\langle u_i | v_j \rangle, r_i\}$ is equal to zero. Now by Lemma 2.11, $\operatorname{Supp} \rho \subseteq \operatorname{Supp} \sigma$. Also $\operatorname{Supp} \sigma \subseteq \operatorname{Supp} \rho$ by symmetry of the situation. Hence we have

$$\operatorname{Supp} \rho = \operatorname{Supp} \sigma.$$

Thus $\operatorname{Ker} \rho = \operatorname{Ker} \sigma$. Therefore, to prove that $\rho = \sigma$, we will show that the nonzero eigenvalues and the corresponding eigenspaces of ρ and σ are same. Now fix $i_0 \in \mathcal{I}$ such that $r_{i_0} \neq 0$. If $r_{i_0} \neq s_j$ for all j, the second equality in (2.14) shows that $\langle u_{i_0} | v_j \rangle = 0$ for all j such that $s_j \neq 0$ but this is impossible because $\{v_j | s_j \neq 0\}$ is an orthonormal basis for $\operatorname{Supp} \sigma$ and $0 \neq u_{i_0} \in \operatorname{Supp} \sigma$. Therefore, for each $i_0 \in \mathcal{I}$ such that $r_{i_0} \neq 0$, there exists $j_0 \in \mathcal{I}$ such that $r_{i_0} = s_{j_0}$. Hence $\operatorname{sp}(\rho) \subseteq \operatorname{sp}(\sigma)$, where 'sp' denotes spectrum. A similar argument shows that $\operatorname{sp}(\sigma) \subseteq \operatorname{sp}(\rho)$. Thus we have

$$\operatorname{sp}(\rho) = \operatorname{sp}(\sigma).$$

Fix i_0 and j_0 such that $r_{i_0} = s_{j_0} \neq 0$. Let R_{i_0} denote the eigenspace of ρ corresponding to the eigenvalue r_{i_0} and S_{j_0} denote the eigenspace of σ corresponding to the eigenvalue s_{j_0}. We will show that $R_{i_0} = S_{j_0}$, which will complete the proof since r_{i_0} is an arbitrary non zero eigenvalue. It is enough to show the following two claims:

1. If $r_i = r_{i_0}$ for some i then $u_i \perp v_j$ for all $j \in \mathcal{I}$ with $s_j \neq s_{j_0}$;

2. If $s_j = s_{j_0}$ for some j then $v_j \perp u_i$ for all $i \in \mathcal{I}$ with $r_i \neq r_{i_0}$.

Proof of both the claims above are similar so we will prove 1 only. Fix i such that $r_i = r_{i_0}$ and j such that $s_j \neq s_{j_0}$. Since $r_{i_0} = s_{j_0}$, we have $r_i \neq s_j$. Thus by (2.14) we have $\langle u_i | v_j \rangle = 0$, which proves 1. □

Proof of Proposition 2.10

Proof. Assume $P \not\ll Q$, we have

$$\begin{aligned} P \not\ll Q &\Leftrightarrow \exists (i,j) \in \mathcal{I} \times \mathcal{I} \text{ such that } \langle u_i | v_j \rangle \neq 0, s_j = 0, r_i \neq 0 \\ &\Leftrightarrow \exists (i,j) \in \mathcal{I} \times \mathcal{I} \text{ such that } \langle u_i | v_j \rangle \neq 0, v_j \in \operatorname{Ker} \sigma, u_i \in \operatorname{Supp} \rho \\ &\Leftrightarrow \exists i \in \mathcal{I} \text{ such that } u_i \in \operatorname{Supp} \rho, u_i \notin (\operatorname{Ker} \sigma)^\perp \\ &\Leftrightarrow \exists i \in \mathcal{I} \text{ such that } u_i \in \operatorname{Supp} \rho, u_i \notin \operatorname{Supp} \sigma \\ &\Leftrightarrow \operatorname{Supp} \rho \nsubseteq \operatorname{Supp} \sigma. \end{aligned}$$

□

3 Quantum f-divergence Inequalities

Our main Theorem 2.8 provides a framework to obtain quantum results from classical ones. To illustrate this, we list a few quantum f-divergence inequalities that follow immediately from the corresponding classical counterparts. The notations used below are as in Example 2.3. Note also, that we use same notations to denote any specific classical or quantum f-divergence. It will be clear from the context whether we are discussing quantum or classical case (we reserve the letters ρ and σ to denote quantum states and the letters P and Q for classical probability distributions). The list of inequalities provided below are not exhaustive by any means; our purpose is to show the use of our main result in obtaining such inequalities. One may refer to the the existing literature on classical f-divergences for deducing several other quantum f-divergence results from the existing classical ones. For example, the article [21] contains several f-divergence inequalities that are not presented here but are easily generalized to the quantum case using our main theorem in this article.

1. The squared Hellinger distance in classical probability satisfies the following bounds with the total variation distance [22, p. 25]

$$\mathcal{H}^2(P\|Q) \leq V(P\|Q)^2 \leq \mathcal{H}(P\|Q)\sqrt{2-\mathcal{H}^2(P\|Q)},$$

hence we have

$$\mathcal{H}^2(\rho\|\sigma) \leq V(\rho\|\sigma)^2 \leq \mathcal{H}(\rho\|\sigma)\sqrt{2-\mathcal{H}^2(\rho\|\sigma)}. \tag{3.1}$$

2. In the classical case the Kullback-Leibler divergence is bounded above by a function of χ^2-divergence as follows [23, Theorem 5],

$$D(P\|Q) \leq \log\left(1+\chi^2(P\|Q)\right),$$

where the same logarithm is used as in the definition of Kullback-Leibler divergence. Hence we have the same bound in the quantum case as well

$$D(\rho\|\sigma) \leq \log\left(1+\chi^2(\rho\|\sigma)\right). \tag{3.2}$$

3. The classical χ^2-divergence is bounded above by a function of Hellinger α-divergence for all $\alpha > 2$ [24, Corollary 5.6]

$$\chi^2(P\|Q) \leq (1+(\alpha-1)\mathcal{H}_\alpha(P\|Q))^{\frac{1}{\alpha-1}} - 1,$$

Therefore, for all $\alpha > 2$,

$$\chi^2(\rho\|\sigma) \leq (1+(\alpha-1)\mathcal{H}_\alpha(\rho\|\sigma))^{\frac{1}{\alpha-1}} - 1. \tag{3.3}$$

4. The χ^2-divergence is bounded below a function of total variation distance as follows [25, eq. (58)]

$$\chi^2(P\|Q) \geq \begin{cases} V(P\|Q)^2, & V(P\|Q) \in [0,1) \\ \frac{V(P\|Q)}{2-V(P\|Q)}, & V(P\|Q) \in [1,2] \end{cases},$$

hence we have the same bound in the quantum case as well

$$\chi^2(\rho\|\sigma) \geq \begin{cases} V(\rho\|\sigma)^2, & V(\rho\|\sigma) \in [0,1] \\ \frac{V(\rho\|\sigma)}{2-V(\rho\|\sigma)}, & V(\rho\|\sigma) \in (1,2) \end{cases}. \tag{3.4}$$

5. From the article [26] we obtain the following inequalities:

$$\frac{D^2(P\|Q)}{D(Q\|P)} \leq \frac{1}{2}\chi^2(P\|Q), \quad [26, (12) \text{ and Remark } 4],$$

$$16\mathcal{H}^4(P\|Q) \leq D(P\|Q)D(Q\|P) \leq \frac{1}{4}\chi^2(P\|Q)\chi^2(Q\|P),$$

$$[26, \text{Proposition 3 (i) and Remark } 4],$$

$$8\mathcal{H}^2(P\|Q) \leq D(P\|Q) + D(Q\|P) \leq \frac{1}{2}\left(\chi^2(P\|Q) + \chi^2(Q\|P)\right),$$

$$[26, \text{Proposition 3 (ii) and Remark } 4],$$

where the constants 16 and 4 that appear above are different than [26] because our definition of $\mathcal{H}^2$ is half of that in [26]. Moreover, the logarithm in the definition of Kullback-Leibler is taken with base e. Therefore, for the quantum case we get

$$\frac{D^2(\rho\|\sigma)}{D(\sigma\|\rho)} \leq \frac{1}{2}\chi^2(\rho\|\sigma), \tag{3.5}$$

$$16\mathcal{H}^4(\rho\|\sigma) \leq D(\rho\|\sigma)D(\sigma\|\rho) \leq \frac{1}{4}\chi^2(\rho\|\sigma)\chi^2(\sigma\|\rho), \tag{3.6}$$

$$8\mathcal{H}^2(\rho\|\sigma) \leq D(\rho\|\sigma) + D(\sigma\|\rho) \leq \frac{1}{2}\left(\chi^2(\rho\|\sigma) + \chi^2(\sigma\|\rho)\right), \tag{3.7}$$

where the logarithm in the definition of Umegaki relative entropy is taken with base e.

6. If the logarithm in the definition of Kullback-Leibler is taken with base e, we have [27, eq. (2.8)]

$$D(P\|Q) \leq \frac{1}{2}\left(V(P,Q) + \chi^2(P\|Q)\right).$$

(Notice that our definition of total variation is twice the one that appears in [27].) With the convention about the logarithm in Umegaki relative entropy, we have the following inequality in the quantum case

$$D(\rho\|\sigma) \leq \frac{1}{2}\left(V(\rho\|\sigma) + \chi^2(\rho\|\sigma)\right). \tag{3.8}$$

7. The symmetrized versions of the Kullback-Leibler and χ^2-divergences satisfy the following bounds [25, Corollary 32],

$$D(P\|Q) + D(Q\|P) \geq 2V(P\|Q)\log\left(\frac{2+V(P\|Q)}{2-V(P\|Q)}\right),$$

$$\chi^2(P\|Q) + \chi^2(Q\|P) \geq \frac{8V(P\|Q)^2}{4-V(P\|Q)^2}.$$

Therefore, the corresponding quantum versions also satisfy the same bounds

$$D(\rho\|\sigma) + D(\sigma\|\rho) \geq 2V(\rho\|\sigma)\log\left(\frac{2+V(\rho\|\sigma)}{2-V(\rho\|\sigma)}\right), \tag{3.9}$$

$$\chi^2(\rho\|\sigma) + \chi^2(\sigma\|\rho) \geq \frac{8V(\rho\|\sigma)^2}{4-V(\rho\|\sigma)^2}. \tag{3.10}$$

8. The Hellinger $\mathcal{H}_\alpha$, the Rényi α-relative entropy D_α and the Kullback-Leibler relative entropy D satisfy the following inequality [28, Proposition 2.15]:

$$\mathcal{H}_\alpha(P\|Q) \leq D_\alpha(P\|Q) \leq D(P\|Q) \leq D_\beta(P\|Q) \leq \mathcal{H}_\beta(P\|Q),$$

for $0 < \alpha < 1 < \beta < \infty$, where the logarithm used in the definitions of Kullback-Leibler and Rényi are with respect base e. Therefore, their quantum counterparts also satisfy the inequalities

$$\mathcal{H}_\alpha(\rho\|\sigma) \log e \leq D_\alpha(\rho\|\sigma) \leq D(\rho\|\sigma) \leq D_\beta(\rho\|\sigma) \leq \mathcal{H}_\beta(\rho\|\sigma), \tag{3.11}$$

for $0 < \alpha < 1 < \beta < \infty$.

9. By [29, Theorem 3.1] we get the following[1]: Let $f : (0, \infty) \to [0, \infty)$ be a strictly convex function such that $f(1) = 0$. Then there exists a real-valued function ψ_f with $\lim_{x\downarrow 0} \psi_f(x) = 0$ such that

$$V(P\|Q) \leq \psi_f\left(D_f(P\|Q)\right).$$

This implies that if

$$\lim_{n\to\infty} D_f\left(P_n\|Q_n\right) = 0 \Rightarrow \lim_{n\to\infty} V(P_n\|Q_n) = 0.$$

The assumptions on f are valid for the function giving rise to the squared Hellinger distance and χ^2-relative entropy. Therefore, if $f : (0, \infty) \to [0, \infty)$ is a strictly convex function such that $f(1) = 0$, then there exists a real-valued function ψ_f such that $\lim_{x\downarrow 0} \psi_f(x) = 0$ and

$$V(\rho\|\sigma) \leq \psi_f\left(D_f(\rho\|\sigma)\right),$$

which implies

$$\lim_{n\to\infty} D_f\left(\rho_n\|\sigma_n\right) = 0 \Rightarrow \lim_{n\to\infty} V(\rho_n\|\sigma_n) = 0. \tag{3.12}$$

4 Conclusion and Discussion

The major contribution of this article is a framework for obtaining quantum versions of several results available for classical f-divergences. This is illustrated by proving quantum versions of several important classical f-divergence inequalities. Further applications of our main theorem are provided in a subsequent article [30]. All our results work both in finite and infinite dimensional setting which is a strength of the methods adopted in this article. Hence these results are particularly useful in continuous variable quantum information theory as well. For instance, we use the results of this article in order to study the Petz-Rényi relative entropy of gaussian states in a follow-up article [31].

[1]By setting $m = 1$ and $w_1 = 1$ in [29, Theorem 3.1] and using Remark A.2 we obtain that $Q^* = P$, where Q^* is as in [29].

Appendix

A Classical f-divergence

In this section, we recall a few facts about the f-divergences in the setting of classical probability. We refer to [6] and the survey article [32] for the following definitions and results which we state in this section. The results from [32] which we use in this article are repeated here for the ease of the reader. If μ and ν are two positive measures on a measure space $(X, \mathcal{F})$, then ν is said to be absolutely continuous with respect to μ and we write $\nu \ll \mu$, if for every $E \in \mathcal{F}$ such that $\mu(E) = 0$, then $\nu(E) = 0$.

Definition A.1. [6, equation 24]. Let P, Q be probability distributions on a measure space $(X, \mathcal{F})$. Let μ be any σ-finite measure such that $P \ll \mu$ and $Q \ll \mu$. Let p and q denote the Radon-Nikodym derivatives with respect to μ, of P and Q, respectively. Let $f : (0, \infty) \to \mathbb{R}$ be a convex (or concave) function then the f-divergence $D_f(P||Q)$ is defined as

$$D_f(P||Q) = \int_X qf\left(\frac{p}{q}\right) \mathrm{d}\mu, \tag{A.1}$$

under the conventions in (2.1) and (2.2). If $P \ll Q$

$$D_f(P||Q) = \int_X f\left(\frac{p}{q}\right) \mathrm{d}Q, \tag{A.2}$$

where a proof of this fact can found in [6, Page 4398]. For completeness, we will prove it in the case of discrete probability distributions in Remark A.3.

Remark A.2. If $f(1) = 0$ then $D_f(P\|P) = 0$ for all probability distributions P. If $f(1) = 0$ and $P \ll Q$ then by Jensen's inequality it can be seen that

$$D_f(P\|Q) \geq 0.$$

Furthermore, if $f \geq 0$, then $D_f(P\|Q) \geq 0$ without any additional assumptions. Thus if $f(1) = 0$ and $f \geq 0$, then

$$\min_Q D_f(P\|Q) = D_f(P\|P) = 0.$$

Remark A.3. We refer to [6] for more details on the definition and the properties of general classical f-divergences. Here we describe briefly about the case when the measures are discrete. Let P and Q be discrete probability measures defined on a countable set $\mathcal{I}$. Let μ be the counting measure on $\mathcal{I}$, then clearly $P \ll \mu$, $Q \ll \mu$,

$$\frac{\mathrm{d}P}{\mathrm{d}\mu}(i) = P(i), \quad \text{and} \quad \frac{\mathrm{d}Q}{\mathrm{d}\mu}(i) = Q(i), \forall i \in \mathcal{I}$$

where $\frac{\mathrm{d}P}{\mathrm{d}\mu}$ and $\frac{\mathrm{d}Q}{\mathrm{d}\mu}$ are the respective Radon-Nikodym derivatives. In this case,

$$D_f(P||Q) = \sum_{i \in \mathcal{I}} Q(i) f\left(\frac{P(i)}{Q(i)}\right), \tag{A.3}$$

under the conventions in (2.1) and (2.2). We will show now that the sum in the equation above is independent of rearrangements. Assume with out loss of generality that f is a convex function. We will show in this case that the negative terms of the series above sum up to a finite number. This fact, once proved, ensures that the series is independent of rearrangements. Let

$$\mathcal{I}_- = \left\{ i \in \mathcal{I} : Q(i) f\left(\frac{P(i)}{Q(i)}\right) < 0 \right\}.$$

Since f is convex, there exists $a, b \in \mathbb{R}$ such that

$$f(x) \geq a + bx, \quad \forall x \in [0, \infty),$$

where $f(0) = \lim_{x \downarrow 0} f(x)$. For $i \in \mathcal{I}_-$ such that $Q(i) = 0$ we have by our conventions in (2.2) that

$$\begin{aligned} f\left(\frac{P(i)}{Q(i)}\right) Q(i) &= P(i) \lim_{s \to \infty} \frac{f(s)}{s} \\ &\geqslant \quad P(i) \lim_{s \to \infty} \frac{a + bs}{s} \\ &= P(i) b \\ &\geqslant -|a| Q(i) - |b| P(i). \end{aligned}$$

For $i \in \mathcal{I}_-$ such that $Q(i) \neq 0$ we have

$$\begin{aligned} f\left(\frac{P(i)}{Q(i)}\right) Q(i) &\geqslant \left(a + b \frac{P(i)}{Q(i)}\right) Q(i) \\ &= aQ(i) + bP(i) \\ &\geqslant -|a| Q(i) - |b| P(i). \end{aligned}$$

So

$$\sum_{i \in \mathcal{I}_-} f\left(\frac{P(i)}{Q(i)}\right) Q(i) \geqslant \sum_{i \in \mathcal{I}_-} (|a| Q(i) + |b| P(i)) \geqslant -|a| - |b| > -\infty.$$

Now because rearrangements are possible, we may compute an expression for $D_f(P \| Q)$ using the conventions given in (2.1) and (2.2) as follows:

$$\begin{aligned} D_f(P \| Q) &= \sum_i Q(i) f\left(\frac{P(i)}{Q(i)}\right) \\ &= \sum_{\{i : P(i) Q(i) > 0\}} Q(i) f\left(\frac{P(i)}{Q(i)}\right) + \sum_{Q(i) \neq 0, P(i) = 0} f(0) Q(i) + \sum_{Q(i) = 0, P(i) \neq 0} P(i) f'(\infty) \\ &= \sum_{\{i : P(i) Q(i) > 0\}} Q(i) f\left(\frac{P(i)}{Q(i)}\right) + f(0) Q(P = 0) + f'(\infty) P(Q = 0). \end{aligned} \tag{A.4}$$

It may be noted that, if $P \ll Q$, then the last term above is not present and we obtain (A.2).

Examples A.4. Here we list a few important examples of f-divergences. While the definitions are general, for the purpose of displaying an expression for each these quantities in our context, we assume that P and Q are discrete as in Remark A.3 and use Equation (A.4).

1. $f(t) = t \log t \Rightarrow D(P\|Q) := D_f(P\|Q)$, the **Kullback-Leibler Relative Entropy**,

$$D(P\|Q) = \begin{cases} \sum_{i\in\mathcal{I}} P(i) \log \frac{P(i)}{Q(i)}, & \text{if } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

2. For $\alpha \in (0,1) \cup (1,\infty)$, $f_\alpha(t) = t^\alpha$ we obtain $D_\alpha(P||Q) := \frac{1}{\alpha-1} \log D_{f_\alpha}(P\|Q)$, the **Rényi $\boldsymbol{\alpha}$-relative entropy**,

$$D_\alpha(P||Q) = \begin{cases} \sum_{i\in\mathcal{I}} P(i)^\alpha Q(i)^{1-\alpha}, & \text{if } \alpha < 1, \text{ or } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

3. $f(t) = \frac{1}{2}(\sqrt{t}-1)^2 \Rightarrow \mathcal{H}^2(P||Q) := D_f(P||Q)$, the **squared Hellinger distance**,

$$\mathcal{H}^2(P||Q) = \frac{1}{2} \sum_{i\in\mathcal{I}} \left(\sqrt{P(i)} - \sqrt{Q(i)}\right)^2.$$

4. For $\alpha \in (0,1) \cup (1,\infty)$, $f_\alpha(t) = \frac{t^\alpha - 1}{\alpha-1}$ we obtain $\mathcal{H}_\alpha(P||Q) := D_{f_\alpha}(P\|Q)$, the **Hellinger α-divergence**,

$$\mathcal{H}_\alpha(P||Q) = \begin{cases} \frac{1}{\alpha-1}\left(\left(\sum_i P(i)^\alpha Q(i)^{1-\alpha}\right) - 1\right), & \text{if } \alpha < 1 \text{ or } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

5. $f(t) = |t-1| \Rightarrow V(P\|Q) := D_f(P\|Q)$, the **Total Variation Distance**,

$$V(P\|Q) = \sum_{i\in\mathcal{I}} |P(i) - Q(i)|.$$

6. $f(t) = (t-1)^2 \Rightarrow \chi^2(P||Q) := D_f(P\|Q)$, the **$\boldsymbol{\chi}^2$-divergence**,

$$\chi^2(P||Q) = \begin{cases} \sum_{\{i\in\mathcal{I}|Q(i)>0\}} \frac{(P(i)-Q(i))^2}{Q(i)}, & \text{if } P \ll Q; \\ \infty, & \text{otherwise.} \end{cases}$$

B Relative Modular Operator on $\mathcal{B}(\mathcal{K})$

Let ρ and σ be two density operators on a Hilbert space $\mathcal{K}$. In this section, we analyse the relative modular operator $\Delta_{\rho,\sigma}$ and find its spectral decomposition of the same in our setting. One can refer to [20, 33, 34] for studying it in the general von Neumann algebra setting. Let $(\mathcal{B}_2(\mathcal{K}), \langle\cdot|\cdot\rangle_2)$ denote the Hilbert space of Hilbert-Schmidt operators on $\mathcal{K}$. Let Π_σ denote the orthogonal projection onto the support of σ. Define

$$D(S) = \{X\sqrt{\sigma} \,:\, X \in \mathcal{B}(\mathcal{K})\} + \{Y(I - \Pi_\sigma) \,:\, Y \in \mathcal{B}_2(\mathcal{K})\} \subseteq \mathcal{B}_2(\mathcal{K}),$$

which is a (vector space) direct sum of orthogonal linear manifolds of $\mathcal{B}_2(\mathcal{K})$. Then $D(S)$ is a dense linear submanifold. Now define the antilinear operator $S : D(S) \to \mathcal{B}_2(\mathcal{K})$ by

$$S\left(X\sqrt{\sigma} + Y(I - \Pi_\sigma)\right) = \Pi_\sigma X^\dagger \sqrt{\rho}. \tag{B.1}$$

For the densely defined antilinear operator S, the adjoint $S^\dagger$ is defined on all $a \in \mathcal{B}_2(\mathcal{K})$ such that there exists a vector $S^\dagger a \in \mathcal{B}_2(\mathcal{K})$ satisfying

$$\langle a|Sb\rangle = \overline{\langle S^\dagger a|b\rangle}, \quad \forall b \in D(S).$$

Now let $X_2 \in \mathcal{B}(\mathcal{K})$ and $Y_2 \in \mathcal{B}_2(\mathcal{K})$ be arbitrary, for the antilinear operator S defined by (B.1), we have

$$\begin{aligned}
&\left\langle \sqrt{\sigma}X_2 + (I-\Pi_\sigma)Y_2 \middle| S\left(X_1\sqrt{\sigma} + Y_1(I-\Pi_\sigma)\right)\right\rangle_2 \\
&\qquad = \left\langle \sqrt{\sigma}X_2 + (I-\Pi_\sigma)Y_2 \middle| \Pi_\sigma X_1^\dagger\sqrt{\rho}\right\rangle_2 \\
&\qquad = \operatorname{Tr}\left\{\left(X_2^\dagger\sqrt{\sigma} + Y_2^\dagger(I-\Pi_\sigma)\right)\Pi_\sigma X_1^\dagger\sqrt{\rho}\right\} \\
&\qquad = \operatorname{Tr} X_2^\dagger\sqrt{\sigma}\Pi_\sigma X_1^\dagger\sqrt{\rho} \\
&\qquad = \operatorname{Tr} X_2^\dagger\Pi_\sigma\sqrt{\sigma} X_1^\dagger\sqrt{\rho} \\
&\qquad = \operatorname{Tr} \sqrt{\sigma} X_1^\dagger\sqrt{\rho} X_2^\dagger\Pi_\sigma \\
&\qquad = \operatorname{Tr}\left\{\left(\sqrt{\sigma}X_1^\dagger + (I-\Pi_\sigma)Y_1^\dagger\right)\sqrt{\rho}X_2^\dagger\Pi_\sigma\right\} \\
&\qquad = \left\langle X_1\sqrt{\sigma} + Y_1(I-\Pi_\sigma) \middle| \sqrt{\rho}X_2^\dagger\Pi_\sigma\right\rangle_2 \\
&\qquad = \overline{\left\langle \sqrt{\rho}X_2^\dagger\Pi_\sigma \middle| X_1\sqrt{\sigma} + Y_1(I-\Pi_\sigma)\right\rangle_2}.
\end{aligned}$$

Hence $S^\dagger$ is defined of the dense set

$$\left\{\sqrt{\sigma}X_2 + (I-\Pi_\sigma)Y_2 \,:\, X_2 \in \mathcal{B}(\mathcal{K}), Y_2 \in \mathcal{B}_2(\mathcal{K})\right\},$$

and

$$S^\dagger\left(\sqrt{\sigma}X_2 + (I-\Pi_\sigma)Y_2\right) = \sqrt{\rho}X_2^\dagger\Pi_\sigma, \quad \forall X_2 \in \mathcal{B}(\mathcal{K}), Y_2 \in \mathcal{B}_2(\mathcal{K}). \tag{B.2}$$

Since S and $S^\dagger$ are densely defined, S is a closable operator. The relative modular operator $\Delta_{\rho,\sigma}$ is defined as

$$\Delta_{\rho,\sigma} = S^\dagger\overline{S} \tag{B.3}$$

where $\overline{S}$ denotes the closure of S. Furthermore, we have

$$\{X\sigma + Y(I-\Pi_\sigma) \,:\, X \in \mathcal{B}(\mathcal{K}), Y \in \mathcal{B}_2(\mathcal{K})\} \subseteq D(\Delta_{\rho,\sigma}) \tag{B.4}$$

and

$$\begin{aligned}
\Delta_{\rho,\sigma}\left(X\sigma + Y(I-\Pi_\sigma)\right) &= S^\dagger\overline{S}\left(X\sqrt{\sigma}\sqrt{\sigma} + Y(I-\Pi_\sigma)\right) \\
&= S^\dagger\left(\Pi_\sigma\sqrt{\sigma}X^\dagger\sqrt{\rho}\right) \\
&= S^\dagger\left(\sqrt{\sigma}\Pi_\sigma X^\dagger\sqrt{\rho}\right) \\
&= \sqrt{\rho}\left(\Pi_\sigma X^\dagger\sqrt{\rho}\right)^\dagger\Pi_\sigma \\
&= \rho X\Pi_\sigma, \quad \forall X \in \mathcal{B}(\mathcal{K}), Y \in \mathcal{B}_2(\mathcal{K}).
\end{aligned} \tag{B.5}$$

Proposition B.1. *Let ρ and σ be as in* (2.6). *Then the spectral decomposition of the relative modular operator $\Delta_{\rho,\sigma}$ is given by*

$$\Delta_{\rho,\sigma} = \sum_{\{i,j\,:\,r_i\neq 0, s_j\neq 0\}} r_i s_j^{-1} \left|X_{ij}\right\rangle\!\left\langle X_{ij}\right|, \tag{B.6}$$

where

$$X_{ij} = |u_i\rangle\!\langle v_j| \in \mathcal{B}_2(\mathcal{K}), \quad \forall i,j \in \mathcal{I}. \tag{B.7}$$

In particular,

$$\operatorname{Ker}\Delta_{\rho,\sigma} = \overline{\operatorname{span}}\{X_{ij} \,:\, r_i = 0 \text{ or } s_j = 0\}.$$

Proof. Since $\{u_i\}$ and $\{v_j\}$ are orthonormal bases for $\mathcal{K}$, it is easy to see that the double sequence

$$\{X_{ij}\}_{i,j\in\mathcal{I}} \text{ is an orthonormal basis of } \mathcal{B}_2(\mathcal{K}).$$

To complete the proof, we will show the following:

1. $X_{ij} \in D(\Delta_{\rho,\sigma})$ for all i,j;
2. $\Delta_{\rho,\sigma}(X_{i,j}) = r_i s_j^{-1} X_{ij}$ for all i,j such that $s_j \neq 0$;
3. if $s_j = 0$ then $X_{ij} \in \operatorname{Ker}\Delta_{\rho,\sigma}$.

To prove 1, note that for j such that $s_j \neq 0$, by (B.4),

$$X_{ij} = |u_i\rangle\!\langle v_j| = s_j^{-1} |u_i\rangle\!\langle v_j| \left(\sum_k s_k |v_k\rangle\!\langle v_k|\right) = s_j^{-1} X_{ij}\sigma \in D(\Delta_{\rho,\sigma}), \forall i. \tag{B.8}$$

Also, if $s_j = 0$ then $v_j \in \operatorname{Ran}(I - \Pi_\sigma)$ and once again from (B.4),

$$X_{ij} = |u_i\rangle\!\langle v_j| = |u_i\rangle\!\langle v_j| \,(I - \Pi_\sigma) \in D(\Delta_{\rho,\sigma}), \forall i. \tag{B.9}$$

Now by (B.5) and (B.8), for j such that $s_j \neq 0$

$$\Delta_{\rho,\sigma} X_{ij} = \Delta_{\rho,\sigma}(s_j^{-1} X_{ij}\sigma) = \rho(s_j^{-1} X_{ij})\Pi_\sigma = s_j^{-1}\left(\sum_k r_k |u_k\rangle\!\langle u_k|\right)|u_i\rangle\!\langle v_j|\,\Pi_\sigma = r_i s_j^{-1} X_{ij},$$

which proves 2. Item 3 follows from (B.5) and (B.9). □

Remark B.2. The spectral projections of general modular operator with faithful states can be seen in [35, equation 2.9] for the finite dimensional case and [10, Example 2.6] for general dimensions. Our Proposition B.1 provides the spectral decomposition of the relative modular operator even when the states are not faithful.

Acknowledgements

We thank Mark Wilde and Milán Mosonyi for their comments on an earlier version [36] of this article, and for the references that they brought to our attention.

The second author thanks The Fulbright Scholar Program and United States-India Educational Foundation for providing funding and other support to conduct this research through a Fulbright-Nehru Postdoctoral Fellowship (Award No. 2594/FNPDR/2020).

References

[1] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statistics*, vol. 22, pp. 79–86, 1951. [Online]. Available: https://doi.org/10.1214/aoms/1177729694

[2] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I.* Univ. California Press, Berkeley, Calif., 1961, pp. 547–561.

[3] T. Morimoto, "Markov processes and the h-theorem," *Journal of the Physical Society of Japan*, vol. 18, no. 3, pp. 328–331, 1963. [Online]. Available: https://doi.org/10.1143/JPSJ.18.328

[4] I. Csiszár, "Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten," *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, vol. 8, pp. 85–108, 1964.

[5] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Statist. Soc. Ser. B*, vol. 28, pp. 131–142, 1966. [Online]. Available: http://links.jstor.org/sici?sici=0035-9246(1966)28:1<131:AGCOCO>2.0.CO;2-U&origin=MSN

[6] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, 2006. [Online]. Available: https://doi.org/10.1109/TIT.2006.881731

[7] I. Csiszár and P. Shields, "Information theory and statistics: A tutorial," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, 2004. [Online]. Available: http://dx.doi.org/10.1561/0100000004

[8] H. Umegaki, "Conditional expectation in an operator algebra. IV. Entropy and information," *Kodai Math. Sem. Rep.*, vol. 14, pp. 59–85, 1962. [Online]. Available: http://projecteuclid.org/euclid.kmj/1138844604

[9] D. Petz, "Quasi-entropies for finite quantum systems," *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0034487786900674

[10] F. Hiai, "Quantum f-divergences in von Neumann algebras. I. Standard f-divergences," *J. Math. Phys.*, vol. 59, no. 10, pp. 102 202, 27, 2018. [Online]. Available: https://doi.org/10.1063/1.5039973

[11] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Statist.*, vol. 37, no. 2, pp. 1040–1057, 2009. [Online]. Available: https://doi.org/10.1214/08-AOS593

[12] M. Mosonyi, "Hypothesis testing for Gaussian states on bosonic lattices," *J. Math. Phys.*, vol. 50, no. 3, pp. 032 105, 17, 2009. [Online]. Available: https://doi.org/10.1063/1.3085759

[13] V. Jakšić, Y. Ogata, C.-A. Pillet, and R. Seiringer, "Quantum hypothesis testing and non-equilibrium statistical mechanics," *Rev. Math. Phys.*, vol. 24, no. 6, pp. 1 230 002, 67, 2012. [Online]. Available: https://doi.org/10.1142/S0129055X12300026

[14] K. P. Seshadreesan, L. Lami, and M. M. Wilde, "Rényi relative entropies of quantum gaussian states," *Journal of Mathematical Physics*, vol. 59, no. 7, p. 072204, 2018. [Online]. Available: https://doi.org/10.1063/1.5007167

[15] M. Berta, V. B. Scholz, and M. Tomamichel, "Rényi divergences as weighted non-commutative vector-valued L_p-spaces," *Ann. Henri Poincaré*, vol. 19, no. 6, pp. 1843–1867, 2018. [Online]. Available: https://doi.org/10.1007/s00023-018-0670-x

[16] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Comm. Math. Phys.*, vol. 143, no. 1, pp. 99–114, 1991. [Online]. Available: http://projecteuclid.org/euclid.cmp/1104248844

[17] F. Hiai and M. Mosonyi, "Different quantum f-divergences and the reversibility of quantum operations," *Rev. Math. Phys.*, vol. 29, no. 7, pp. 1 750 023, 80, 2017. [Online]. Available: https://doi.org/10.1142/S0129055X17500234

[18] K. R. Parthasarathy, "A pedagogical note on the computation of relative entropy of two n-mode gaussian states," *arXiv e-prints*, p. arXiv:2102.06708, Feb. 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2102.06708

[19] H. Araki, "Relative entropy of states of von Neumann algebras," *Publications of the Research Institute for Mathematical Sciences*, vol. 11, no. 3, pp. 809–833, 1976.

[20] ——, "Relative entropy for states of von Neumann algebras ii," *Publications of the Research Institute for Mathematical Sciences*, vol. 13, no. 1, pp. 173–192, 1977.

[21] I. Sason and S. Verdú, "f -divergence inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 5973–6006, 2016.

[22] L. Le Cam and G. L. Yang, *Asymptotics in statistics*, ser. Springer Series in Statistics. Springer-Verlag, New York, 1990, some basic concepts. [Online]. Available: https://doi.org/10.1007/978-1-4684-0377-0

[23] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International Statistical Review*, vol. 70, no. 3, pp. 419–435, 2002. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2002.tb00178.x

[24] A. Guntuboyina, S. Saha, and G. Schiebinger, "Sharp inequalities for f -divergences," *IEEE Trans. Inf. Theory*, vol. 60, pp. 104–121, 2013.

[25] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *Journal of Machine Learning Research*, vol. 12, no. 22, pp. 731–817, 2011. [Online]. Available: http://jmlr.org/papers/v12/reid11a.html

[26] S. Simic, "On a new moments inequality," *Statistics & Probability Letters*, vol. 78, no. 16, pp. 2671–2678, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167715208001831

[27] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains," *The Annals of Applied Probability*, vol. 6, no. 3, pp. 695 – 750, 1996. [Online]. Available: https://doi.org/10.1214/aoap/1034968224

[28] F. Liese and I. Vajda, "Convex statistical distances. teubner-texte zur mathematik [teubner texts in mathematics] 95," *BSB BG Teubner Verlagsgesellschaft, Leipzig*, vol. 656, 1987.

[29] I. Csiszár, "A class of measures of informativity of observation channels," *Periodica Mathematica Hungarica*, vol. 2, no. 1-4, pp. 191–213, 1972.

[30] G. Androulakis and T. C. John, "Petz-Rényi and umegaki relative entropy from distribution of observables," *In Preparation*, 2022.

[31] ——, "Petz-Rényi relative entropy of quantum gaussian states," *In Preparation*, 2022.

[32] T. van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014. [Online]. Available: http://doi.org/10.1109/TIT.2014.2320500

[33] M. Ohya and D. Petz, *Quantum entropy and its use*, ser. Texts and Monographs in Physics. Springer-Verlag, Berlin, 1993. [Online]. Available: https://doi.org/10.1007/978-3-642-57997-4

[34] A. Łuczak, H. Podsędkowska, and R. Wieczorek, "Relative modular operator in semifinite von Neumann algebras and its use," *arXiv preprint arXiv:1912.09633*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1912.09633

[35] N. Datta, Y. Pautrat, and C. Rouzé, "Second-order asymptotics for quantum hypothesis testing in settings beyond i.i.d.—quantum lattice systems and more," *J. Math. Phys.*, vol. 57, no. 6, pp. 062 207, 26, 2016. [Online]. Available: https://doi.org/10.1063/1.4953582

[36] G. Androulakis and T. C. John, "Quantum f-divergences via Nussbaum-Szkoła distributions with applications to Petz-Rényi and Umegaki relative entropy," *arXiv*, p. xx, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2203.01964