Maximal quantum divergences and their integral representations

Filippo Girardi[1,2,*] and Ludovico Lami[1,2,3,†]

[1]QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands
[2]Korteweg–de Vries Institute for Mathematics, University of Amsterdam, Science Park 105-107, 1098 XG Amsterdam, The Netherlands
[3]Institute for Theoretical Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

In this note we review a well known integral formula for classical f–divegences and we introduce a generalization to the maximal quantum f–divergences for operator convex functions f. Then, we provide an elementary proof of the Kosaki variational formula for the Umegaki relative entropy using the Nussbaum–Szkoła distributions. We compute the dual of the optimization problem and we show that a small change in the variational formulation gives the BS entropy as the result of the minimization. After recalling Hiai's variational formula for standard f–divergences, we show that a similar result holds for any maximal f–divergence based on an operator convex function. We finally discuss DPI for these divergences and we derive some necessary and sufficient conditions for the saturation of the inequality.

I. AN INTEGRAL FORMULA FOR CLASSICAL f-DIVERGENCES

In this section we review the most elementary proof for a well known integral formula for classical f–divergences.

Definition 1. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function such that $f(1) = 0$. And let $p, q : \mathcal{X} \to [0, \infty)$ be two mutually absolutely continuous probability distributions. Then

$$D_f(p\|q) = \int_{\mathcal{X}} f\left(\frac{p(x)}{q(x)}\right) q(x)\, dx \tag{1}$$

is called f*-divergence* from p to q.

On the one hand, as an example we can consider $f_a(t) = t\log t$ and $f_b(t) = -\log t$: the quantity

$$D_{f_a}(p\|q) = D_{f_b}(q\|p) = \int_{\mathcal{X}} \log\left(\frac{p(x)}{q(x)}\right) p(x)\, dx\,, \tag{2}$$

called the *relative entropy*, is denoted as $D(p\|q)$. On the other hand, defining

$$(t)_+ = \frac{|t| + t}{2} = \begin{cases} t & t \geq 0\,, \\ 0 & t < 0\,, \end{cases} \tag{3}$$

we can consider $f_\gamma(t) = (t-\gamma)_+$ and introduce the E_γ*-divergence* as

$$E_\gamma(p\|q) = D_{f_\gamma}(p\|q) = \int_{\mathcal{X}} (p(x) - \gamma q(x))_+\, dx \tag{4}$$

The following theorem relates any f-divergence $D_f(p\|q)$ with sufficiently regular f to the E_γ-divergences

$$\{E_\gamma(p\|q),\ E_\gamma(q\|p)\}_{\gamma\geq 1} \tag{5}$$

* f.girardi@uva.nl
† ludovico.lami@gmail.com

Theorem 2. *Let $f:\mathbb{R}\to\mathbb{R}$ be a twice differentiable convex function such that $f(0)=1$. Then*

$$D_f(p\|q)=\int_1^\infty \Big(f''(\gamma)E_\gamma(p\|q)+\gamma^{-3}f''(\gamma^{-1})E_\gamma(q\|p)\Big)\,d\gamma. \tag{6}$$

Proof. By Taylor's theorem with integral reminder, for any $a,b\in\mathbb{R}$ we can write

$$f(b)=f(a)+f'(a)(b-a)+R_f(a,b)\quad\text{where}\quad R_f(a,b)=\begin{cases}\int_a^b (b-t)f''(t)\,dt & b\geq a\\ \int_b^a (t-b)f''(t)\,dt & b<a\end{cases}. \tag{7}$$

Suppose $a,b\geq 0$. The reminder can be rewritten as

$$R_f(a,b)=\begin{cases}\int_a^\infty (b-t)_+f''(t)\,dt & b\geq a\\ \int_0^a (t-b)_+f''(t)\,dt & b<a\end{cases}\overset{\text{(i)}}{=}\int_a^\infty (b-t)_+f''(t)\,dt+\int_0^a (t-b)_+f''(t)\,dt, \tag{8}$$

where in (i) we used that whenever $b\geq a$, the expression valid for $b<a$ is zero, and vice versa. Now, let us fix $a=1$ in (7) and recall that $f(1)=0$: then

$$\begin{aligned} f(b)&=f(1)+f'(1)(b-1)+R_f(1,b)\\ &=f'(1)(b-1)+\int_1^\infty (b-t)_+f''(t)\,dt+\int_0^1 (t-b)_+f''(t)\,dt\\ &=f'(1)(b-1)+\int_1^\infty (b-\gamma)_+f''(\gamma)\,d\gamma+\int_1^\infty (\gamma^{-1}-b)_+f''(\gamma^{-1})\,\gamma^{-2}\,d\gamma\\ &=f'(1)(b-1)+\int_1^\infty \Big((b-\gamma)_+f''(\gamma)+(1-\gamma b)_+f''(\gamma^{-1})\,\gamma^{-3}\Big)\,d\gamma \end{aligned} \tag{9}$$

where in the first integral we have substituted $\gamma=t$, while in the second integral we have changed variable $\gamma=\frac{1}{\gamma}$. By the definition of f-divergence,

$$\begin{aligned} D_f(p\|q)&=\int_{\mathcal{X}} f\left(\frac{p(x)}{q(x)}\right)q(x)\,dx\\ &=f'(1)\int_{\mathcal{X}}(p(x)-q(x))\,dx\\ &\quad+\int_1^\infty\left[\left(\int_{\mathcal{X}}(p(x)-\gamma q(x))_+\,dx\right)f''(\gamma)+\left(\int_{\mathcal{X}}(q(x)-\gamma p(x))_+\,dx\right)f''(\gamma^{-1})\,\gamma^{-3}\right]\,d\gamma\\ &=\int_1^\infty\Big(E_\gamma(p\|q)f''(\gamma)+E_\gamma(q\|p)f''(\gamma^{-1})\,\gamma^{-3}\Big)\,d\gamma \end{aligned} \tag{10}$$

This concludes the proof. □

As a corollary, we can write the relative entropy in the form

$$D(p\|q)=\int_1^\infty\left(\frac{1}{\gamma}E_\gamma(p\|q)+\frac{1}{\gamma^2}E_\gamma(q\|p)\right)d\gamma. \tag{11}$$

II. A NEW INTEGRAL FORMULA FOR MAXIMAL QUANTUM f-DIVERGENCES

In this section we generalize the previous result to the class of quantum maximal f–divergences. In the setting of the Umegaki relative entropy, an analogous (and technically less elementary) generalization of the classical formula was provided by Frenkel [1].

The proof provided in the previous section immediately extends to the quantum setting for the following class of quantum f-divergences.

Definition 3. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function such that $f(1) = 0$. And let $\rho, \sigma \in \mathscr{L}(\mathcal{H})$ be two quantum states such that $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$. We define

$$\tilde{D}_f(\rho\|\sigma) = \operatorname{Tr}\left[\sigma f\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\right] \tag{12}$$

which, for f operator convex function, is called *maximal quantum f-divergence* from ρ to σ [2].

The *Belavkin–Staszewski quantum relative entropy* is the maximal quantum f-divergence given by $f(t) = t\log t$:

$$\begin{aligned} D_{BS}(\rho\|\sigma) &:= \tilde{D}_f(\rho\|\sigma) \\ &= \operatorname{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\log\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\right] \\ &= \operatorname{Tr}\left[\sigma^{1/2}\rho^{-1/2}\rho\rho^{1/2}\sigma^{-1/2}\log\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\right] \\ &= \operatorname{Tr}\left[\rho\rho^{1/2}\sigma^{-1/2}\log\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\sigma^{1/2}\rho^{-1/2}\right] \\ &= \operatorname{Tr}\left[\rho\log\left(\rho^{1/2}\sigma^{-1}\rho^{1/2}\right)\right]. \end{aligned} \tag{13}$$

We define the positive part of a Hermitian matrix $A \in \mathscr{L}(\mathcal{H})$ as

$$A_+ = \frac{|A| + A}{2}, \tag{14}$$

and, considering $f_\gamma(t) = (t-\gamma)_+$, we introduce the pseudo maximal[1] E_γ-divergence:

$$\tilde{E}_\gamma(\rho\|\sigma) := \tilde{D}_{f_\gamma}(\rho\|\sigma) = \operatorname{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2} - \gamma\mathbb{1}\right)_+\right]. \tag{15}$$

Theorem 4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable convex function such that $f(0) = 1$. Then*

$$\tilde{D}_f(\rho\|\sigma) = \int_1^\infty \left(f''(\gamma)\tilde{E}_\gamma(\rho\|\sigma) + \gamma^{-3}f''(\gamma^{-1})\tilde{E}_\gamma(\rho\|\sigma)\right)d\gamma. \tag{16}$$

Proof. Let

$$\sigma^{-1/2}\rho\sigma^{-1/2} = \sum_i \lambda_i\,|i\rangle\langle i| \tag{17}$$

be the spectral decomposition of the operator $\sigma^{-1/2}\rho\sigma^{-1/2}$. Let us notice that

$$\sigma^{1/2}\rho^{-1}\sigma^{1/2} = \left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)^{-1} = \sum_i \frac{1}{\lambda_i}\,|i\rangle\langle i|\,. \tag{18}$$

[1] It does not come from a operator convex function.

Since $\sigma^{-1/2}\rho\sigma^{-1/2}$ is a positive semi-definite operator, we can introduce the following identity ($\gamma \geq 0$):

$$\lambda_i \left(\frac{1}{\lambda_i} - \gamma\right)_+ = \begin{cases} 0 & \lambda_i = 0 \\ (1-\gamma\lambda_i)_+ & \lambda_i > 0 \end{cases}, \tag{19}$$

having used the convention $0 \cdot \infty = 0$. Then, we have

$$\begin{aligned}
\tilde{D}_f(\rho\|\sigma) &= \sum_i \mathrm{Tr}\left[\sigma f\left(\lambda_i\right)|i\rangle\langle i|\right] \\
&\overset{(9}{=} f'(1)\,\mathrm{Tr}\left[\sigma \sum_i (\lambda_i - 1)\,|i\rangle\langle i|\right] \\
&\quad + \mathrm{Tr}\left[\sigma \int_1^\infty \left(\sum_i (\lambda_i-\gamma)_+ |i\rangle\langle i|\, f''(\gamma) + \sum_i \lambda_i \left(\frac{1}{\lambda_i} 1 - \gamma\right)_+ |i\rangle\langle i|\, f''(\gamma^{-1})\,\gamma^{-3}\right) d\gamma\right] \\
&= f'(1)\,\mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2} - \mathbb{1}\right)\right] \\
&\quad + \int_1^\infty \mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2} - \gamma\mathbb{1}\right)_+\right] f''(\gamma) d\gamma \\
&\quad + \int_1^\infty \mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\left(\sigma^{1/2}\rho^{-1}\sigma^{1/2} - \gamma\mathbb{1}\right)_+\right] f''(\gamma^{-1})\,\gamma^{-3}\, d\gamma
\end{aligned} \tag{20}$$

Let us notice that

$$f'(1)\,\mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2} - \mathbb{1}\right)\right] = f'(1)\,\mathrm{Tr}\left[\rho - \sigma\right] = 0 \tag{21}$$

and, defining $A := \sigma^{1/2}\rho^{-1/2}$,

$$\begin{aligned}
\sigma^{1/2}\rho^{-1}\sigma^{1/2} &= \sigma^{1/2}\rho^{-1}\sigma\sigma^{-1/2} \\
&= \left(\sigma^{1/2}\rho^{-1/2}\right)\rho^{-1/2}\sigma\rho^{-1/2}\left(\rho^{1/2}\sigma^{-1/2}\right) = A\rho^{-1/2}\sigma\rho^{-1/2}A^{-1}
\end{aligned} \tag{22}$$

whence

$$\begin{aligned}
&\mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2}\right)\left(\sigma^{1/2}\rho^{-1}\sigma^{1/2} - \gamma\mathbb{1}\right)_+\right] \\
&\quad = \mathrm{Tr}\left[\sigma^{1/2}\rho\sigma^{-1/2}\left(\sigma^{1/2}\rho^{-1}\sigma^{1/2} - \gamma\mathbb{1}\right)_+\right] \\
&\quad = \mathrm{Tr}\left[\sigma^{1/2}\rho\sigma^{-1/2}\left(A\left(\rho^{-1/2}\sigma\rho^{-1/2} - \gamma\mathbb{1}\right)A^{-1}\right)_+\right] \\
&\quad = \mathrm{Tr}\left[\sigma^{1/2}\rho\sigma^{-1/2}A\left(\rho^{-1/2}\sigma\rho^{-1/2} - \gamma\mathbb{1}\right)_+ A^{-1}\right] \\
&\quad = \mathrm{Tr}\left[(\sigma^{1/2}\rho^{-1/2})^{-1}\sigma^{1/2}\rho\sigma^{-1/2}(\sigma^{1/2}\rho^{-1/2})\left(\rho^{-1/2}\sigma\rho^{-1/2} - \gamma\mathbb{1}\right)_+\right] \\
&\quad = \mathrm{Tr}\left[\rho\left(\rho^{-1/2}\sigma\rho^{-1/2} - \gamma\mathbb{1}\right)_+\right] \\
&\quad = \tilde{E}_\gamma(\sigma\|\rho)
\end{aligned} \tag{23}$$

Therefore, (20) becomes

$$\tilde{D}_f(\rho\|\sigma) = \int_1^\infty \tilde{E}_\gamma(\rho\|\sigma) f''(\gamma) d\gamma + \int_1^\infty \tilde{E}_\gamma(\rho\|\sigma) f''(\gamma^{-1})\,\gamma^{-3}\, d\gamma. \tag{24}$$

and this concludes the proof for the quantum case. □

In particular, we can write the Belavkin–Staszewski quantum relative entropy in the form

$$D_{BS}(\rho\|\sigma) = \int_1^\infty \left(\frac{1}{\gamma}\tilde{E}_\gamma(\rho\|\sigma) + \frac{1}{\gamma^2}\tilde{E}_\gamma(\sigma\|\rho)\right) d\gamma. \tag{25}$$

Corollary 5. *Belavkin–Staszewski quantum relative entropy is larger than the Umegaki quantum relative entropy:*

$$D_U(\rho\|\sigma) \leq D_{BS}(\rho\|\sigma). \tag{26}$$

Proof. We recall that in Frenkel's formula we have

$$E_\gamma(\rho\|\sigma) = \mathrm{Tr}\left[(\rho - \gamma\sigma)_+\right] \tag{27}$$

Now we estimate

$$\begin{aligned} E_\gamma(\rho\|\sigma) &= \mathrm{Tr}\left[(\rho-\gamma\sigma)_+\right] = \sup_{0\le P\le \mathbb{1}} \mathrm{Tr}\left[P\left(\rho-\gamma\sigma\right)\right] \\ &= \sup_{0\le P\le \mathbb{1}} \mathrm{Tr}\left[\left(\sigma^{1/2}P\sigma^{1/2}\left(\sigma^{-1/2}\rho\sigma^{-1/2}-\gamma\mathbb{1}\right)\right)\right] = \sup_{0\le Q\le\sigma} \mathrm{Tr}\left[Q\left(\sigma^{-1/2}\rho\sigma^{-1/2}-\gamma\mathbb{1}\right)\right] \\ &\le \sup_{0\le Q\le\sigma} \mathrm{Tr}\left[Q\left(\sigma^{-1/2}\rho\sigma^{-1/2}-\gamma\mathbb{1}\right)_+\right] = \mathrm{Tr}\left[\sigma\left(\sigma^{-1/2}\rho\sigma^{-1/2}-\gamma\mathbb{1}\right)_+\right] = \tilde{E}_\gamma(\rho\|\sigma), \end{aligned} \tag{28}$$

which yields the thesis. □

III. KOSAKI'S FORMULA

We know that Kosaki proved the following integral representation [3]

$$D(\rho\|\sigma) = \lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\inf_X \mathrm{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right]\right\}. \tag{29}$$

It is easy to see that this equality holds by the following elementary computation.

Proof. Let $\rho = \sum_i \rho_i |\psi_i\rangle\langle\psi_i|$, $\sigma = \sum_i \sigma_i |\phi_i\rangle\langle\phi_i|$, $X = \sum_{ij} x_{ij} |\phi_i\rangle\langle\phi_j|$. We can rewrite

$$\rho = U\rho_D U^\dagger \qquad \text{where} \qquad \rho_D = \sum_i \rho_i |\phi_i\rangle\langle\phi_i| \qquad \text{and} \qquad U = \sum_i |\psi_i\rangle\langle\phi_i|. \tag{30}$$

Furthermore, in the basis $\{|\phi_i\rangle\}_i$,

$$\rho = \sum_{ij} r_{ij} |\phi_i\rangle\langle\phi_j|. \tag{31}$$

Then

$$\begin{aligned} \mathrm{Tr}\left[\rho X^\dagger X\right] &= \sum_{ijk} r_{ij} x^*_{kj} x_{ki}, \\ \mathrm{Tr}\left[\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] &= \sum_{ij} \sigma_i(\delta_{ij}-x_{ij})(\delta_{ij}-x^*_{ij}). \end{aligned} \tag{32}$$

Differentiating

$$\begin{aligned} \partial_{x^*_{ij}} \operatorname{Tr}\left[\rho X^\dagger X\right] &= \sum_k x_{ik} r_{kj} = [X\rho]_{ij}, \\ \partial_{x^*_{ij}} \operatorname{Tr}\left[\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] &= -\sigma_i(\delta_{ij}-x_{ij}) = -[\sigma(\mathbb{1}-X)]_{ij}. \end{aligned} \tag{33}$$

So the minimizer $\bar{X}$ satisfies the Sylvester's equation

$$\bar{X}\rho - \frac{1}{t}\sigma(\mathbb{1}-\bar{X}) = 0 \qquad \Longleftrightarrow \qquad \bar{X}\rho + \frac{\sigma}{t}\bar{X} = \frac{\sigma}{t} \tag{34}$$

which implies

$$\begin{aligned} \inf_X \operatorname{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] &= \operatorname{Tr}\left[\rho \bar{X}^\dagger \bar{X} + \frac{1}{t}\sigma(\mathbb{1}-\bar{X})(\mathbb{1}-\bar{X})^\dagger\right] \\ &= \operatorname{Tr}\left[\rho \bar{X}^\dagger \bar{X} + \bar{X}\rho(\mathbb{1}-\bar{X})^\dagger\right] \\ &= \operatorname{Tr}\left[\bar{X}\rho\right]. \end{aligned} \tag{35}$$

Defining $X' := \bar{X}U = \sum_{ij} x'_{ij} |\phi_i\rangle\langle\phi_j|$, (34) becomes

$$\bar{X}U\rho_D U^\dagger + \frac{\sigma}{t}\bar{X} = \frac{\sigma}{t} \qquad \Longleftrightarrow \qquad X'\rho_D + \frac{\sigma}{t}X' = \frac{\sigma}{t}U, \tag{36}$$

so we can compute the matrix element ij in the basis $\{|\phi_i\rangle\}_i$

$$x'_{ij}\rho_j + \frac{\sigma_i}{t}x'_{ij} = \frac{[\sigma U]_{ij}}{t} \qquad \Longleftrightarrow \qquad x'_{ij} = \frac{[\sigma U]_{ij}}{t\rho_j + \sigma_i} = \frac{\sigma_i \langle\phi_i|\psi_j\rangle}{t\rho_j + \sigma_i}, \tag{37}$$

so

$$X' = \sum_{ij} |\phi_i\rangle\langle\phi_i| \frac{\sigma_i}{t\rho_j + \sigma_i} |\psi_j\rangle\langle\phi_j| = \sum_j \frac{\sigma}{t\rho_j\mathbb{1} + \sigma} |\psi_j\rangle\langle\phi_j| \tag{38}$$

therefore

$$\bar{X} = X'U^\dagger = \sum_j \frac{\sigma}{t\rho_j\mathbb{1} + \sigma} |\psi_j\rangle\langle\psi_j|, \tag{39}$$

As a consequence, (35) has the analytic form

$$\operatorname{Tr}\left[\bar{X}\rho\right] = \sum_{ij} \frac{\rho_j\sigma_i}{t\rho_j + \sigma_i} |\langle\psi_j|\phi_i\rangle|^2, \tag{40}$$

Introducing the Nussbaum–Szkoła probability distributions

$$P_{ij} = \rho_j |\langle\psi_j|\phi_i\rangle|^2 \qquad \text{and} \qquad Q_{ij} = \sigma_i |\langle\psi_j|\phi_i\rangle|^2 \tag{41}$$

we can rewrite

$$\inf_X \operatorname{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] = \sum_{ij} \frac{P_{ij}Q_{ij}}{tP_{ij} + Q_{ij}} \tag{42}$$

and compute

$$
\begin{aligned}
&\lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon}-\int_\varepsilon^\infty\frac{dt}{t}\inf_X \mathrm{Tr}\left[\rho X^\dagger X+\frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right]\right\}\\
&\quad=\lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon}-\int_\varepsilon^\infty\frac{dt}{t}\sum_{ij}\frac{P_{ij}Q_{ij}}{tP_{ij}+Q_{ij}}\right\}\\
&\quad=\lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon}-\sum_{ij}\int_\varepsilon^\infty dt\left(\frac{P_{ij}}{t}-\frac{P_{ij}^2}{tP_{ij}+Q_{ij}}\right)\right\}\\
&\quad=-\lim_{\varepsilon\to 0^+}\left\{\sum_{ij}P_{ij}\log\left(\varepsilon+\frac{Q_{ij}}{P_{ij}}\right)\right\}\\
&\quad=\sum_{ij}P_{ij}\left(\log P_{ij}-\log Q_{ij}\right)\\
&\quad=D(P\|Q).
\end{aligned}
\tag{43}
$$

And this concludes the proof because $D(P\|Q)=D(\rho\|\sigma)$. □

IV. THE DUAL OF THE MINIMIZATION PROBLEM

Inspired by Kosaki's formula, we study the properties of the variational expression (46), which turns out to be the BS relative entropy (see [4]).

Let us consider again

$$
D(\rho\|\sigma)=\lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon}-\int_\varepsilon^\infty\frac{dt}{t}\inf_X \mathrm{Tr}\left[\rho X^\dagger X+\frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right]\right\}
\tag{44}
$$

and make the substitution

$$
(\mathbb{1}-X)(\mathbb{1}-X)^\dagger \quad\to\quad (\mathbb{1}-X)^\dagger(\mathbb{1}-X)
\tag{45}
$$

obtaining

$$
\hat{D}(\rho\|\sigma):=\lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon}-\int_\varepsilon^\infty\frac{dt}{t}\inf_X \mathrm{Tr}\left[\rho X^\dagger X+\frac{1}{t}\sigma(\mathbb{1}-X)^\dagger(\mathbb{1}-X)\right]\right\}.
\tag{46}
$$

We can now ask ourselves how to solve the optimizations inside the two integrals (44) and (46) using a different strategy with respect to the one discussed in the previous section. Let us handle the first minimization problem. Recalling that

$$
\begin{pmatrix}\mathbb{1} & X\\ X^\dagger & A\end{pmatrix}\geq 0 \qquad\Longleftrightarrow\qquad A\geq X^\dagger X
\tag{47}
$$

we can write

$$
\begin{aligned}
&\inf_X \operatorname{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] \\
&\qquad = \inf\left\{\operatorname{Tr}\left[\rho A + \frac{1}{t}\sigma B\right] \,:\, \begin{pmatrix}\mathbb{1} & X\\ X^\dagger & A\end{pmatrix} \geq 0,\ \begin{pmatrix}\mathbb{1} & \mathbb{1}-X^\dagger\\ X & B\end{pmatrix} \geq 0\right\} \\
&\qquad = \inf\left\{\operatorname{Tr}\left[\rho C_{22} + \frac{1}{t}\sigma D_{22}\right] \,:\, C, D \geq 0,\ C_{11} = D_{11} = \mathbb{1},\ C_{12} + D_{21} = \mathbb{1}\right\} \\
&\qquad = \inf_{C,D\geq 0}\sup_{U,V,T}\Bigg\{\operatorname{Tr}\left[\rho C_{22} + \frac{1}{t}\sigma D_{22}\right] + \operatorname{Tr}\left[U(C_{11}-\mathbb{1})\right] + \operatorname{Tr}\left[U(D_{11}-\mathbb{1})\right] \\
&\qquad\qquad\qquad + \operatorname{Tr}\left[T(C_{11}+D_{21}-\mathbb{1})\right] + \operatorname{Tr}\left[T^\dagger(C_{21}+D_{12}-\mathbb{1})\right]\Bigg\} \\
&\qquad = \sup_{U,V,T}\inf_{C,D\geq 0}\left\{\operatorname{Tr}\left[\begin{pmatrix}U & T^\dagger\\ T & \rho\end{pmatrix}\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix}\right] + \operatorname{Tr}\left[\begin{pmatrix}V & T\\ T^\dagger & \sigma/t\end{pmatrix}\begin{pmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{pmatrix}\right] - \operatorname{Tr}\left[U+V+T+T^\dagger\right]\right\} \\
&\qquad = \sup_{U,V,T}\left\{-\operatorname{Tr}\left[U+V+T+T^\dagger\right] \,:\, \begin{pmatrix}U & T^\dagger\\ T & \rho\end{pmatrix} \geq 0,\ \begin{pmatrix}V & T\\ T^\dagger & \sigma/t\end{pmatrix} \geq 0\right\} \\
&\qquad = -\inf_{U,V,T}\left\{\operatorname{Tr}\left[U+V+T+T^\dagger\right] \,:\, \begin{pmatrix}U & T^\dagger\\ T & \rho\end{pmatrix} \geq 0,\ \begin{pmatrix}V & T\\ T^\dagger & \sigma/t\end{pmatrix} \geq 0\right\} \\
&\qquad = \inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT\sigma^{-1}T^\dagger + T + T^\dagger\right]
\end{aligned} \tag{48}
$$

where in the last computation we have noticed that

$$
\begin{pmatrix}U & T^\dagger\\ T & \rho\end{pmatrix} \geq 0,\ \begin{pmatrix}V & T\\ T^\dagger & \sigma/t\end{pmatrix} \geq 0 \quad\Longleftrightarrow\quad U \geq T^\dagger\rho^{-1}T,\ V \geq Tt\sigma^{-1}T^\dagger. \tag{49}
$$

Similarly, it easy to prove that

$$
\inf_X \operatorname{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right] = \inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT^\dagger\sigma^{-1}T + T + T^\dagger\right], \tag{50}
$$

whence

$$
\begin{aligned}
D(\rho\|\sigma) &= \lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT\sigma^{-1}T^\dagger + T + T^\dagger\right]\right\} \\
\hat{D}(\rho\|\sigma) &= \lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT^\dagger\sigma^{-1}T + T + T^\dagger\right]\right\}
\end{aligned} \tag{51}
$$

The second variational problem can be solved explicitly: introducing an auxilliary operator S

$$
T =: (\rho^{-1} + t\sigma^{-1})^{-1/2} S (\rho^{-1} + t\sigma^{-1})^{-1/2}
$$

we can trace back the optimization problem to a simple quadratic form

$$
\begin{aligned}
\inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT^\dagger\sigma^{-1}T + T + T^\dagger\right] &= \inf_S \operatorname{Tr}\left[(\rho^{-1}+t\sigma^{-1})^{-1}(S^\dagger S + S + S^\dagger)\right] \\
&= \inf_S \operatorname{Tr}\left[(\rho^{-1}+t\sigma^{-1})^{-1}(S-\mathbb{1})^\dagger)(S-\mathbb{1}) - \mathbb{1})\right] \\
&= -\operatorname{Tr}\left[(\rho^{-1}+t\sigma^{-1})^{-1}\right]
\end{aligned} \tag{52}
$$

i.e.

$$\min_X \operatorname{Tr}\left[\rho X^\dagger X + \frac{1}{t}\sigma(\mathbb{1}-X)^\dagger(\mathbb{1}-X)\right] = -\operatorname{Tr}\left[(\rho^{-1}+t\sigma^{-1})^{-1}\right]. \tag{53}$$

Remark 6. Therefore, in the second optimization problem T turns out to be Hermitian. Therefore, unless $D(\rho\|\sigma) = \tilde{D}(\rho\|\sigma)$, the first optimization problem cannot be solved by a Hermitian T for each t.

Plugging the minimum inside the integral and recalling the identity

$$\log A = \lim_{\varepsilon\to 0^+}\left\{\mathbb{1}\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\frac{1}{1+tA}\right\},$$

we obtain

$$\begin{aligned}
\hat{D}(\rho\|\sigma) &= \lim_{\varepsilon\to 0^+}\left\{\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\inf_T \operatorname{Tr}\left[T^\dagger\rho^{-1}T + tT^\dagger\sigma^{-1}T + T + T^\dagger\right]\right\}\\
&= \lim_{\varepsilon\to 0^+}\operatorname{Tr}\left[\rho\left(\mathbb{1}\log\frac{1}{\varepsilon} - \int_\varepsilon^\infty \frac{dt}{t}\frac{1}{1+t\rho^{1/2}\sigma^{-1}\rho^{1/2}}\right)\right]\\
&= \operatorname{Tr}\left[\rho\log(\rho^{1/2}\sigma^{-1}\rho^{1/2})\right]\\
&= D_{BS}(\rho\|\sigma).
\end{aligned}$$

V. HIAI VARIATIONAL FORMULA

In order to generalize the previous result to the class of maximal f-divergences for f operator convex, we start from a result by Hiai [5]. Let f be an operator convex function on $(0,\infty)$ with the integral expression

$$f(t) = a + b(t-1) + c(t-1)^2 + d\frac{(t-1)^2}{t} + \int_{[0,\infty)}\frac{(t-1)^2}{t+s}d\mu(s)$$

for $a, b \in \mathbb{R}$, $c, d \geq 0$, μ positive measure. Defining

$$f_n(t) = a + b(t-1) + c\frac{n(t-1)^2}{t+n} + d\frac{(t-1)^2}{t+1/n} + \int_{[0,\infty)}\frac{(t-1)^2}{t+s}d\mu(s)$$

and

$$d\nu_n(s) = c(1+n)\delta_n + d(1+n)\delta_{1/n} + \chi_{[1/n,n]}(s)\frac{1+s}{s}d\mu(s),$$

we can rewrite

$$f_n(t) = f_n(0^+) + f_n'(\infty)t - \int_{(0,\infty)}\frac{t(1+s)}{t+s}d\nu_n(s) \tag{54}$$

with the property that

$$f_n(0^+) \nearrow f(0^+), \qquad f_n'(\infty) \nearrow f'(\infty), \qquad f_n(t) \nearrow f(t).$$

Theorem 7 (F. Hiai, [5]). *Let ρ and σ be states. Then*

$$D_f(\rho\|\sigma) = \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n} \inf_X \mathrm{Tr}\left[\sigma X^\dagger X + \frac{1}{s}\rho(\mathbb{1}-X)(\mathbb{1}-X)^\dagger\right](1+s)d\nu_n(s)\right\} \quad (55)$$

where D_f is the standard f-divergence.

Inspired by this result, let us consider

$$\hat{D}_f(\rho\|\sigma) := \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n} \inf_X \mathrm{Tr}\left[\sigma X^\dagger X + \frac{1}{s}\rho(\mathbb{1}-X)^\dagger(\mathbb{1}-X)\right](1+s)d\nu_n(s)\right\}. \quad (56)$$

Before computing this integral, let us first show that the definition of standard f–divergence given in the setting of von Neumann algebras just reduces to the classical f–divergence between the Nussbaum–Szkoła distributions introduced in (41). Replacing (42) in (55) we obtain

$$\begin{aligned}
D_f(\rho\|\sigma) &= \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n}\sum_{ij}\frac{P_{ij}Q_{ij}}{sQ_{ij}+P_{ij}}(1+s)d\nu_n(s)\right\} \\
&= \sum_{ij} Q_{ij}\sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty)\frac{P_{ij}}{Q_{ij}} - \int_{1/n}^{n}\frac{P_{ij}/Q_{ij}(1+s)}{P_{ij}/Q_{ij}+s}d\nu_n(s)\right\} \\
&= \sum_{ij} Q_{ij} f\left(\frac{P_{ij}}{Q_{ij}}\right) \\
&= D(P\|Q)
\end{aligned} \quad (57)$$

where in the last line we used the integral formula (54). Now we want to give an analytic expression to (56). As proved above, the infimum in (56) can be computed as in (53), whence

$$\begin{aligned}
\hat{D}_f(\rho\|\sigma) &= \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n}\mathrm{Tr}\left[(\sigma^{-1}+s\rho^{-1})^{-1}\right](1+s)d\nu_n(s)\right\} \\
&= \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n}\mathrm{Tr}\left[\rho(\rho^{1/2}\sigma^{-1}\rho^{1/2}+s)^{-1}\right](1+s)d\nu_n(s)\right\} \\
&= \sup_{n\in\mathbb{N}}\mathrm{Tr}\left[\rho\rho^{-1/2}\sigma\rho^{-1/2}\left\{f_n(0^+) + f_n'(\infty)\rho^{1/2}\sigma^{-1}\rho^{1/2} - \int_{1/n}^{n}\frac{\rho^{1/2}\sigma^{-1}\rho^{1/2}(1+s)}{\rho^{1/2}\sigma^{-1}\rho^{1/2}+s}d\nu_n(s)\right\}\right] \\
&= \mathrm{Tr}\left[\sigma\sigma^{1/2}\rho^{-1/2}f(\rho^{1/2}\sigma^{-1}\rho^{1/2})\rho^{1/2}\sigma^{-1/2}\right] \\
&= \mathrm{Tr}\left[\sigma f(\sigma^{-1/2}\rho\sigma^{-1/2})\right] \\
&= \tilde{D}_f(\rho\|\sigma)
\end{aligned}$$

Therefore, we have the following integral and variational representation of the maximal f–divergences for operator convex functions f written in the form (54) (similar formulations were proved in [4, 6]).

Theorem 8.

$$\tilde{D}_f(\rho\|\sigma) = \sup_{n\in\mathbb{N}}\left\{f_n(0^+) + f_n'(\infty) - \int_{1/n}^{n} \inf_X \mathrm{Tr}\left[\sigma X^\dagger X + \frac{1}{s}\rho(\mathbb{1}-X)^\dagger(\mathbb{1}-X)\right](1+s)d\nu_n(s)\right\}. \quad (58)$$

VI. DATA PROCESSING INEQUALITY

While in [1] an integral representation analogous to (25) is used to prove the DPI for the Umegaki relative entropy, in our representation we do not expect a similar strategy to be feasible: indeed, we do not think that $\tilde{E}_\gamma$ satisfies DPI because $f_\gamma(t) = (t-\gamma)_+$ is not operator convex. This makes the characterization of the states which saturate the DPI in the BS setting harder. However, the formula (58) provides a simple way of proving DPI and could be an interesting starting point to see when the inequality is saturated.

Given a channel Φ, its adjoint is a completely positive unital map $\Lambda = \Phi^\dagger$. A 2-positive unital map Λ satisfies the Schwarz inequality

$$\Lambda(A^\dagger A) \geq \Lambda(A^\dagger)\Lambda(A), \tag{59}$$

therefore

$$\begin{aligned}
\inf_X \mathrm{Tr}&\left[\Phi(\sigma)(X^\dagger X) + \frac{1}{s}\Phi(\rho)((\mathbb{1}-X)^\dagger(\mathbb{1}-X))\right] \\
&= \inf_X \mathrm{Tr}\left[\sigma\,\Lambda(X^\dagger X) + \frac{1}{s}\rho\,\Lambda((\mathbb{1}-X)^\dagger(\mathbb{1}-X))\right] \\
&\geq \inf_X \mathrm{Tr}\left[\sigma\,(\Lambda(X))^\dagger\Lambda(X) + \frac{1}{s}\rho\,(\mathbb{1}-\Lambda(X))^\dagger(\mathbb{1}-\Lambda(X))\right] \\
&= \inf_{Y=\Lambda(X)} \mathrm{Tr}\left[\sigma\,Y^\dagger Y + \frac{1}{s}\rho\,(\mathbb{1}-Y)^\dagger(\mathbb{1}-Y)\right] \\
&\geq \inf_Y \mathrm{Tr}\left[\sigma\,Y^\dagger Y + \frac{1}{s}\rho\,(\mathbb{1}-Y)^\dagger(\mathbb{1}-Y)\right].
\end{aligned} \tag{60}$$

As a consequence

$$\tilde{D}_f(\Phi(\rho)\|\Phi(\sigma)) \leq \tilde{D}_f(\rho\|\sigma)\,, \tag{61}$$

where the inequality is satisfied if and only if

$$\inf_X \mathrm{Tr}\left[\Phi(\sigma)(X^\dagger X) + \frac{1}{s}\Phi(\rho)((\mathbb{1}-X)^\dagger(\mathbb{1}-X))\right] = \inf_X \mathrm{Tr}\left[\sigma(X^\dagger X) + \frac{1}{s}\rho((\mathbb{1}-X)^\dagger(\mathbb{1}-X))\right] \tag{62}$$

for almost every $s \in \mathrm{supp}(\nu_n)$ as $n \to \infty$. It interesting to notice that the equality (62) does not depend on f. In particular, if for some Φ, ρ, σ it holds for every $s > 0$, then

$$\tilde{D}_f(\Phi(\rho)\|\Phi(\sigma)) = \tilde{D}_f(\rho\|\sigma) \tag{63}$$

for any operator convex f. Using (53), we can rewrite (62) in a different form:

$$\mathrm{Tr}\left[\left(s(\Phi(\rho))^{-1} + (\Phi(\sigma))^{-1}\right)^{-1}\right] = \mathrm{Tr}\left[\left(s\rho^{-1} + \sigma^{-1}\right)^{-1}\right]. \tag{64}$$

This condition is symmetric when we exchange ρ and σ (and s with s^{-1}). The form of both sides of (64) can be Taylor expanded

$$\mathrm{Tr}\left[\frac{1}{A^{-1}+\xi B^{-1}}\right] = \mathrm{Tr}\left[A\frac{1}{1+\xi A^{1/2}B^{-1}A^{1/2}}\right] = \sum_{n\geq 0}(-\xi)^n\,\mathrm{Tr}\left[A(A^{1/2}B^{-1}A^{1/2})^n\right], \tag{65}$$

therefore (64) can be rewritten as

$$\mathrm{Tr}\left[\Phi(\sigma)(\Phi(\sigma)^{1/2}\Phi(\rho)^{-1}\Phi(\sigma)^{1/2})^n\right] = \mathrm{Tr}\left[\sigma(\sigma^{1/2}\rho^{-1}\sigma^{1/2})^n\right] \qquad \forall n \in \mathbb{N}. \tag{66}$$

Under suitable assumptions on the support of ρ and σ, we have that $\sigma^{1/2}\rho^{-1}\sigma^{1/2}$ has a bounded spectrum, so, for any continuous function g, the previous equation implies the following condition by Stone-Weierstrass theorem

$$\mathrm{Tr}\left[\Phi(\sigma)\, g\left(\Phi(\sigma)^{1/2}\Phi(\rho)^{-1}\Phi(\sigma)^{1/2}\right)\right] = \mathrm{Tr}\left[\sigma\, g(\sigma^{1/2}\rho^{-1}\sigma^{1/2})\right]. \tag{67}$$

To recap, we have showed that the following conditions are equivalent under suitable assumptions on the support of ρ and σ: given a triple (ρ, σ, Φ), we have that

1. (ρ, σ, Φ) saturates DPI for $\tilde{D}_f$ for an operator convex f with $\lim_{n\to\infty} \operatorname{supp} \nu_n = (0, \infty)$;
2. (ρ, σ, Φ) saturates DPI for $\tilde{D}_f$ for any operator convex f;
3. (64) holds for any $s > 0$;
4. (66) holds for any $n \in \mathbb{N}$;
5. (67) holds for any g continuous.

[1] P. E. Frenkel. Integral formula for quantum relative entropy implies data processing inequality. *Quantum*, 7:1102, 2023. 3, 11

[2] K. Matsumoto. A new quantum version of f-divergence. *Preprint arXiv:1311.4722*, 2013. 3

[3] H. Kosaki. Relative entropy of states: a variational expression. *J. Operator Theory*, 16(2):335–348, 1986. 5

[4] Stefan Hollands and Alessio Ranallo. Channel divergences and complexity in algebraic QFT. *Commun. Math. Phys.*, 404(2):927–962, December 2023. 7, 10

[5] F. Hiai. Quantum f-divergences in von Neumann algebras. I. Standard f-divergences. *J. Math. Phys.*, 59(10), September 2018. 9, 10

[6] Stefan Hollands. Variational approach to relative entropies with an application to QFT. *Lett. Math. Phys.*, 111(6), December 2021. 10