MARCO TOMAMICHEL

QUANTUM RELATIVE ENTROPY: AN AXIOMATIC APPROACH

QMATH MASTERCLASS IN COPENHAGEN, AUGUST 2022

Part of the notes are transcribed by Yanglin Hu. These notes are not yet free of typos and may not be presented in the most clear way. Any comments that help reduce these deficiencies are very much appreciated.

http://www.marcotom.info/teaching/

Latest update, January 2023

Contents

1

Information-theoretic axioms and their consequences

Literature: These notes are based on joint work with Gilad Gour[1].

[1] Gilad Gour and Marco Tomamichel. Optimal extensions of resource measures and their applications. *Phys. Rev. A*, 102:062401, Dec 2020. DOI: 10.1103/PhysRevA.102.062401. URL https://link.aps.org/doi/10.1103/PhysRevA.102.062401; and Gilad Gour and Marco Tomamichel. Entropy and relative entropy from information-theoretic principles. *IEEE Transactions on Information Theory*, 67(10):6313–6327, 2021. DOI: 10.1109/TIT.2021.3078337. URL https://doi.org/10.1109/TIT.2021.3078337

1.1 *Classical relative entropies*

Let P and Q be two probability distributions on some finite alphabet $\mathcal{X}$. Arguably the most fundamental random variable in information theory is the *log-likelihood ratio*,

$$Z(X) = \log \frac{P(X)}{Q(X)}, \qquad X \leftarrow P \tag{1.1}$$

Its expectation is the Kulback-Leibler divergence, or *relative entropy*,

$$\mathbb{E}[Z] = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} =: D(P\|Q)\,. \tag{1.2}$$

It has various direct applications in statistics and information theory but is also widely used in other quantitative fields as a measure of similarity of two probability distributions. Its widespread use is a consequence of the useful mathematical properties it possesses. Amongst those are

1. The *data-processing inequality*, which states that the relative entropy cannot increase when a stochastic map is applied to both distributions.

$$D(\omega(P)\|\omega(Q)) \leq D(P\|Q)\,. \tag{1.3}$$

2. The fact that it is *additive* when evaluated on tensor products of distributions.

$$D(P_1 \times P_2\|Q_1 \times Q_2) = D(P_1\|Q_1) + D(P_2\|Q_2)\,. \tag{1.4}$$

We will formally state these properties and discuss many others in the next section, more generally for quantum states.

Are there other functionals with the above two properties? The answer is yes, Rényi relative entropies. They can be defined via the cumulant generating function of $Z(X)$, given by

$$K(t) = \log \mathbb{E}[2^{tZ}] = \log \sum_{x\in\mathbb{X}} P(x)^{1+t} Q(x)^{-t}. \quad (1.5)$$

for $t \in \mathbb{R}$. The Rényi relative entropies of order $\alpha \in [0,1) \cup (1,\infty)$ is then defined as

$$D_\alpha(P\|Q) := \frac{K(\alpha-1)}{\alpha-1} = \frac{1}{\alpha-1} \log \sum_{x\in\mathbb{X}} P(x)^\alpha Q(x)^{1-\alpha}. \quad (1.6)$$

The cases $\alpha = 1$ and $\alpha = \infty$ are defined as pointwise limits; in particular it limits to D when $\alpha \to 1$. They are a monotonically increasing family of functionals satisfying both data-processing inequality and additivity, amongst other useful properties. Rényi relative entropies have various applications in large deviation theory because of their relation to the cumulant generating function of Z. For example, they are used to describe error and strong converse exponents for many problems in information theory.

It was only very recently discovered that this picture is complete: under some weak regularity conditions it was shown that every functional that satisfies data-processing inequality and additivity is in fact a linear combination of Rényi relative entropies $D_\alpha(P\|Q)$ and $D_\alpha(Q\|P)$[2].

[2] Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. From blackwell dominance in large samples to rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021. DOI: 10.3982/ECTA17548. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17548

Can we do something similar for quantum relative entropies? We will see that this attempt fails, as there is no unique generalization of Rényi relative entropies to the quantum setting and we do not yet have a complete picture of all quantities satisfying both data-processing and additivity.

1.2 *Axioms for quantum relative entropy*

In the following we use the notation $\mathcal{S}_d$ to denote the set of quantum states[3] acting on $\mathbb{C}^d$ (as an example, $\mathcal{S}_1 = \{[1]\}$), and $\mathcal{S}_d^>$ to denote the subset with full support. We generally represent quantum states by $d \times d$ matrices. Let $0_{m\times n}$ denote an $m \times n$ zero matrix.

[3] A quantum state is a positive semi-definite matrix with unit trace.

A quantum relative entropy is a map from two quantum states to real numbers with the data-processing inequality, additivity and normalisation properties, which we will formalise as follows. We are interested in functionals of the form

$$\mathbb{D} : \bigcup_{d\in\mathbb{N}} \mathcal{S}_d \times \mathcal{S}_d \to \mathbb{R} \cup \{-\infty, \infty\}. \quad (1.7)$$

We can now formalise the data-processing inequality and additivity properties.

1. **Data-Processing Inequality**. For any pair $d, d' \in \mathbb{N}$, any pair of quantum states $\rho, \sigma \in \mathcal{S}_d$, and any quantum channel[4] $\mathcal{E} : \mathcal{S}_d \to \mathcal{S}_{d'}$, we want that

$$\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))\,. \tag{1.8}$$

[4] A quantum channel is a completely positive and trace-preserving map.

2. **Additivity**. For any pair $d_1, d_2 \in \mathbb{N}$, and any pairs of quantum states $\rho_1, \sigma_1 \in \mathcal{S}_{d_1}$ and $\rho_2, \sigma_2 \in \mathcal{S}_{d_2}$, we want that

$$\mathbb{D}(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = \mathbb{D}(\rho_1\|\sigma_1) + \mathbb{D}(\rho_2\|\sigma_2)\,. \tag{1.9}$$

Note that both 1) and 2) are completely symmetric in the two states. The normalisation condition below breaks this symmetry, however.

3. **Normalisation**. We want that

$$\mathbb{D}\left(\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\middle\|\begin{bmatrix}1/2 & 0\\ 0 & 1/2\end{bmatrix}\right) = 1\,. \tag{1.10}$$

In the following we will denote by $\mathbb{D}$ any functional of the above form that satisfies 1), 2) and 3). Can we derive some further properties of such functionals just from these three axioms? We start with the following rather direct consequence:

4. **Positivity**. We have $\mathbb{D}(\rho\|\sigma) \geq 0$ with equality if $\rho = \sigma$.

Proof. We start by noting that, by normalisation and additivity,

$$1 = \mathbb{D}\left(\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\middle\|\begin{bmatrix}1/2 & 0\\ 0 & 1/2\end{bmatrix}\right) \tag{1.11}$$

$$= \mathbb{D}\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \times [1]\middle\|\begin{bmatrix}1/2 & 0\\ 0 & 1/2\end{bmatrix} \times [1]\right) \tag{1.12}$$

$$= 1 + \mathbb{D}([1]\|[1]), \tag{1.13}$$

and hence we must have $\mathbb{D}([1]\|[1]) = 0$.

Now, on the one hand, using the data-processing inequality for the trace $\mathrm{tr} : \tau \mapsto [1]$, we can conclude that, for any pair of quantum states $\rho, \sigma \in \mathcal{S}_d$,

$$\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}([1]\|[1]) = 0\,. \tag{1.14}$$

And on the other hand, using the data-processing inequality for the preparation channel $\mathcal{E} : [1] \mapsto \rho$, we can furthermore conclude that

$$0 = \mathbb{D}([1]\|[1]) \geq \mathbb{D}(\rho\|\rho)\,. \tag{1.15}$$

This establishes the desired result. □

We have to do some more work to replace the "if" with an "if and only if". This is the topic of the next section.

1.3 Faithfulness

For this purpose, we need to exclude *trivial* quantum relative entropies. We say that a quantum relative entropy is trivial if $\mathbb{D}(\rho\|\sigma) = 0$ for any $d \in \mathbb{N}$ and any pair of quantum states $\rho, \sigma \in \mathcal{S}_d^{>}$ (with full support). Note that a trivial quantum relative entropy can have non-zero values on the boundary. A trivial quantum relative entropy (which is the only trivial quantum relative entropy I am aware of) is

$$\mathbb{D}_0(\rho\|\sigma) = -\log \operatorname{tr}(\rho^0 \sigma^1)\,. \tag{1.16}$$

Let $\operatorname{supp}(\rho)$ be the support of ρ. One can verify that $\mathbb{D}_0(\rho\|\sigma) = 0$ if $\operatorname{supp}(\rho) = \operatorname{supp}(\sigma)$ and that $\mathbb{D}_0(\rho\|\sigma) > 0$ if $\operatorname{supp}(\rho) \subsetneq \operatorname{supp}(\sigma)$.

For nontrivial relative entropies we can strengthen 4) as follows:

5. **Faithfulness**. For any non-trivial relative entropy $\mathbb{D}$ we have $\mathbb{D}(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$.

Proof. We first want to argue that for any non-trivial quantum relative entropy, there exist some constants $s_0, t_0 \in (0, \frac{1}{2})$ such that

$$\mathbb{D}\left(\begin{bmatrix} s & 0 \\ 0 & 1-s \end{bmatrix} \middle\| \begin{bmatrix} 1-t & 0 \\ 0 & t \end{bmatrix}\right) > 0 \tag{1.17}$$

for any $s \in [0, s_0]$ and $t \in [0, t_0]$. To see this, we first note that since $\mathbb{D}$ is non-trivial, there exists a $d \in \mathbb{N}$ and a pair of quantum states $\gamma, \omega \in \mathcal{S}_d^{>}$ (with full support) satisfying $\mathbb{D}(\gamma\|\omega) > 0$. We next choose s_0, t_0 such that

$$\frac{t_0}{1-s_0}\gamma \leq \omega \leq \frac{1-t_0}{s_0}\gamma\,, \tag{1.18}$$

which is always possible for sufficiently small s_0 and t_0 due to the support condition. For each s, t as above, we then construct the preparation channel

$$\mathcal{E} : \begin{bmatrix} p & \cdot \\ \cdot & 1-p \end{bmatrix} \mapsto \frac{(p-s)\omega + (1-p-t)\gamma}{1-s-t}\,. \tag{1.19}$$

It can be verified that the operator inequality enforced in Eq. (1.18) ensures that all outputs of this channel are indeed valid states. Besides, $\mathcal{E}$ also satisfies

$$\mathcal{E}\left(\begin{bmatrix} s & 0 \\ 0 & 1-s \end{bmatrix}\right) = \gamma, \quad \mathcal{E}\left(\begin{bmatrix} 1-t & 0 \\ 0 & t \end{bmatrix}\right) = \omega\,. \tag{1.20}$$

Eq. (1.17) then follows from the data-processing inequality for the preparation channel $\mathcal{E}$

$$\mathbb{D}\left(\begin{bmatrix} s & 0 \\ 0 & 1-s \end{bmatrix} \middle\| \begin{bmatrix} 1-t & 0 \\ 0 & t \end{bmatrix}\right) \geq \mathbb{D}(\gamma\|\omega) > 0\,. \tag{1.21}$$

Next, we will make use of a result in quantum hypothesis testing. We claim that for any pair of quantum states $\rho, \sigma \in \mathcal{S}_d^>$ with $\rho \neq \sigma$ and for any pair of $s, t \in (0, \frac{1}{2})$, there exists a sufficiently large n and a binary POVM $\{T, \mathbb{1} - T\}$ such that

$$\operatorname{tr}(T\rho^{\otimes n}) = s, \quad \operatorname{tr}(T\sigma^{\otimes n}) = 1 - t\,. \tag{1.22}$$

The claim can be understood intuitively as follow. For such states we know that the fidelity satisfies $F(\rho^{\otimes n}, \sigma^{\otimes n}) \to 0$ and the trace distance satisfies $\|\rho^{\otimes n} - \sigma^{\otimes n}\|_{\mathrm{tr}} \to 1$ asymptotically as $n \to \infty$. Recall that $\|\rho^{\otimes n} - \sigma^{\otimes n}\|_{\mathrm{tr}} = 1$ if and only if $\operatorname{supp}(\rho^{\otimes n}) \perp \operatorname{supp}(\sigma^{\otimes n})$. Thus there exists a projector P such that $\operatorname{tr}(P\rho^{\otimes n}) \to 1$ and $\operatorname{tr}(P\sigma^{\otimes n}) \to 0$. Consequently we may choose n large enough so that

One simple way to see this is by the Fuchs-van de Graaf inequality combined with the multiplicativity and faithfulness of fidelity.

$$\operatorname{tr}(T\rho^{\otimes n}) = s, \quad \operatorname{tr}(T\sigma^{\otimes n}) = 1 - t\,, \tag{1.23}$$

for $T = sP + (1-t)(\mathbb{1} - P)$. We thus construct the measurement channel

$$\mathcal{M} : \tau \mapsto \begin{bmatrix} \operatorname{tr}(T\tau) & 0 \\ 0 & 1 - \operatorname{tr}(T\tau) \end{bmatrix}. \tag{1.24}$$

Now one can use the additivity and the data-processing inequality for the measurement map $\mathcal{M}$ to conclude that

$$n\mathbb{D}(\rho\|\sigma) = \mathbb{D}(\rho^{\otimes n}\|\sigma^{\otimes n}) \geq \mathbb{D}\left(\begin{bmatrix} s & 0 \\ 0 & 1-s \end{bmatrix} \middle\| \begin{bmatrix} 1-t & 0 \\ 0 & t \end{bmatrix}\right) > 0\,, \tag{1.25}$$

which immediately completes the proof. □

1.4 *Bounds on quantum relative entropy*

We formally define the min-relative entropy $\mathbb{D}_{\min}$ as

$$\mathbb{D}_{\min}(\rho\|\sigma) = -\log \operatorname{tr}(\rho^0\sigma)\,, \tag{1.26}$$

where ρ^0 is the projector on $\operatorname{supp}(\rho)$. We would like to emphasise that $\mathbb{D}_{\min}$ is trivial and thus not faithful. The max-relative entropy $\mathbb{D}_{\max}$ is defined as

$$\mathbb{D}_{\max}(\rho\|\sigma) = \inf\{\lambda \in \mathbb{R} : \rho \leq 2^\lambda \sigma\}\,. \tag{1.27}$$

The nomenclature is justified by the following property.

6. **Bounds**. For any quantum relative entropy $\mathbb{D}$, any $d \in \mathcal{N}$ and any pair of quantum states $\rho, \sigma \in \mathcal{S}_d$, we have

$$\mathbb{D}_{\min}(\rho\|\sigma) \leq \mathbb{D}(\rho\|\sigma) \leq \mathbb{D}_{\max}(\rho\|\sigma)\,. \tag{1.28}$$

For the proof, we further need a lemma.

Lemma 1.1. *For any quantum relative entropy* $\mathbb{D}$*, any* $p \in [0,1]$ *and* $\sigma \in \mathcal{S}_d$*, we have*

$$f(p,\sigma) = \mathbb{D}\left(\begin{bmatrix} 1 & 0_{1\times d} \\ 0_{d\times 1} & 0_{d\times d}\end{bmatrix} \middle\| \begin{bmatrix} p & 0_{1\times d} \\ 0_{d\times 1} & (1-p)\sigma\end{bmatrix}\right) = -\log p\,. \quad (1.29)$$

Proof. Consider the quantum channel that replaces the state σ (or any other state) with ρ, i.e.

$$\mathcal{E}_\rho : \begin{bmatrix} p & \cdot \\ \cdot & (1-p)\sigma\end{bmatrix} \mapsto \begin{bmatrix} p & \cdot \\ \cdot & (1-p)\rho\end{bmatrix}. \quad (1.30)$$

By the data-processing inequality for $\mathcal{E}_\rho$ and $\mathcal{E}_\sigma$, we obtain $f(p,\sigma) = f(p)$ must in fact be independent of σ. By additivity, $f(p_1 p_2) = f(p_1) + f(p_2)$. Let the depolarising channel be

$$\mathcal{D}_{t,\sigma} : \begin{bmatrix} p & \cdot \\ \cdot & (1-p)\rho\end{bmatrix} \mapsto \begin{bmatrix} t+p-2tp & \cdot \\ \cdot & (1-t)(1-p)\rho + tp\sigma\end{bmatrix}. \quad (1.31)$$

By the data-processing inequality for $\mathcal{D}_{t,\sigma}$, we obtain $f(p_1) \geq f(p_2)$ when $p_1 \leq p_2$. This characterises additive functions and by Erdös t[5], the only function satisfying the above requirements is $f(p) = c\log p$. By normalisation, we can fix $c = -1$. ☐

[5] Paul Erdös. On the distribution function of additive functions. *Annals of Mathematics*, 47(1):1–20, 1946. ISSN 0003486X. DOI: 10.2307/1969031. URL http://www.jstor.org/stable/1969031

Now we are prepared to prove the aforementioned bounds on quantum relative entropies.

Proof. We first prove the upper bound. If $\mathbb{D}_{\max}(\rho\|\sigma) = \infty$, then we are done. Otherwise let $\lambda = 2^{-\mathbb{D}_{\max}(\rho\|\sigma)} \in (0,1]$. With the help of the previous lemma,

$$\mathbb{D}_{\max}(\rho\|\sigma) = \mathbb{D}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix} \middle\| \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda\end{bmatrix}\right) = -\log\lambda\,. \quad (1.32)$$

Let the classical-quantum preparation channel be

$$\mathcal{E} : \begin{bmatrix} p & \cdot \\ \cdot & 1-p\end{bmatrix} \mapsto p\rho + (1-p)\frac{\sigma - \lambda\rho}{1-\lambda}\,. \quad (1.33)$$

Then we apply the data-processing inequality for $\mathcal{E}$

$$\mathbb{D}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix} \middle\| \begin{bmatrix} \lambda & 0 \\ 0 & 1-\lambda\end{bmatrix}\right) \geq \mathbb{D}(\rho\|\sigma)\,, \quad (1.34)$$

which proves the upper bound together with (1.32).

As for the lower bound, let $\mu = \mathrm{tr}(\rho^0\tau) = 2^{-\mathbb{D}_{\min}(\rho\|\tau)}$ where ρ^0 is the projector on $\mathrm{supp}(\rho)$. By the previous lemma,

$$\mathbb{D}_{\min}(\rho\|\sigma) = \mathbb{D}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix} \middle\| \begin{bmatrix} \mu & 0 \\ 0 & 1-\mu\end{bmatrix}\right) = -\log\mu\,. \quad (1.35)$$

Now consider the measurement channel

$$\mathcal{M} : \tau \mapsto \begin{bmatrix} \mathrm{tr}(\rho^0 \tau) & 0 \\ 0 & 1 - \mathrm{tr}(\rho^0 \tau) \end{bmatrix} . \tag{1.36}$$

With the data-processing inequality for $\mathcal{M}$, one can find

$$\mathbb{D}(\rho \| \sigma) \geq \mathbb{D}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \middle\| \begin{bmatrix} \mu & 0 \\ 0 & 1-\mu \end{bmatrix} \right) , \tag{1.37}$$

which completes the proof together with (1.35). □

1.5 *Continuity*

With the help of bounds on quantum relative entropy, we can also prove its continuity. Here, we will start with a lemma before we proceed to the continuity.

Lemma 1.2. *We have*

$$\mathbb{D}(\rho \| \sigma) \leq \mathbb{D}(\rho \| \tau) + \mathbb{D}_{\max}(\tau \| \sigma) . \tag{1.38}$$

Proof. We may assume that $\tau \ll \sigma$ and $\rho \ll \tau$ as otherwise the inequality is trivial. Hence, also $\rho \ll \sigma$. We write $\sigma = \epsilon\tau + (1-\epsilon)\omega$ where $\omega = \tau + \frac{\sigma - \tau}{1-\epsilon}$ and $\epsilon = 2^{-\mathbb{D}_{\max}(\tau\|\sigma)}$. You may verify that $\omega \geq 0$. Recall (1.32), which implies

$$\mathbb{D}_{\max}(\tau \| \sigma) = \mathbb{D}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \middle\| \begin{bmatrix} \epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix} \right) . \tag{1.39}$$

We obtain

$$\mathbb{D}(\rho \| \tau) + \mathbb{D}_{\max}(\tau \| \sigma) = \mathbb{D}\left(\rho \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \middle\| \tau \otimes \begin{bmatrix} \epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix} \right) . \tag{1.40}$$

Define the conditional replacer quantum channel

$$\mathcal{E} : \theta \otimes \begin{bmatrix} p & \cdot \\ \cdot & 1-p \end{bmatrix} \mapsto p\theta + (1-p)\omega . \tag{1.41}$$

Now apply the data-processing inequality for $\mathcal{E}$

$$\mathbb{D}\left(\rho \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \middle\| \tau \otimes \begin{bmatrix} \epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix} \right) \geq \mathbb{D}(\rho \| \sigma) , \tag{1.42}$$

which proves the lemma combining with (1.40). □

We prove the continuity of quantum relative entropies as follows.

7. **Continuity**. Quantum relative entropies $\mathbb{D}(\,\cdot\,\|\,\cdot\,)$ is continuous on $\mathcal{S}_d^{>} \times \mathcal{S}_d^{>}$.

Proof. Consider sequences $\{(\rho_k, \sigma_k)\}_{k=1}^{\infty}$ where $(\rho_k, \sigma_k) \in \mathcal{S}_d^{>} \times \mathcal{S}_d^{>}$ converges to $(\rho, \sigma) \in \mathcal{S}_d^{>} \times \mathcal{S}_d^{>}$. Using the definition, we can verify that

$$\lim_{k\to\infty} \mathbb{D}_{\max}(\rho_k \| \sigma_k) = \mathbb{D}_{\max}(\rho\|\sigma)\,. \tag{1.43}$$

We now define a mixing quantum channel

$$\mathcal{E}_k(\omega) = (1-\epsilon_k)\omega + \rho_k - (1-\epsilon_k)\rho\,, \tag{1.44}$$

where $\epsilon_k = 1 - 2^{-\mathbb{D}_{\max}(\rho\|\rho_k)}$. As $k \to \infty$, $\epsilon_k \to 0$. Applying the data-processing inequality

$$\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}\left(\mathcal{E}_k(\rho)\|\mathcal{E}_k(\sigma)\right) = \mathbb{D}\left(\rho_k\|\mathcal{E}_k(\sigma)\right)\,. \tag{1.45}$$

Recall (1.38),

$$\mathbb{D}\left(\rho_k\|\mathcal{E}_k(\sigma)\right) \geq \mathbb{D}(\rho_k\|\sigma_k) - \mathbb{D}_{\max}(\mathcal{E}_k(\sigma)\|\sigma)\,. \tag{1.46}$$

Taking $k \to \infty$, we obtain

$$\mathbb{D}(\rho\|\sigma) \geq \limsup_{k\to\infty} \mathbb{D}(\rho_k\|\sigma_k) - \liminf_{k\to\infty} \mathbb{D}_{\max}(\mathcal{E}_k(\sigma)\|\sigma_k)\,. \tag{1.47}$$

where

$$\liminf_{k\to\infty} \mathbb{D}_{\max}(\mathcal{E}_k(\sigma)\|\sigma) = 0\,. \tag{1.48}$$

Therefore, we prove the lower bound

$$\mathbb{D}(\rho\|\sigma) \geq \limsup_{k\to\infty} \mathbb{D}(\rho_k\|\sigma_k)\,. \tag{1.49}$$

We can similarly prove the upper bound with a similar method. We again define a mixing quantum channel

$$\mathcal{E}'_k(\omega) = (1-\epsilon'_k)\omega + \rho - (1-\epsilon'_k)\rho_k\,, \tag{1.50}$$

where $\epsilon'_k = 1 - 2^{-\mathbb{D}_{\max}(\rho_k\|\rho)}$. Again apply the data-processing inequality

$$\mathbb{D}(\rho_k\|\sigma_k) \geq \mathbb{D}\left(\mathcal{E}'_k(\rho_k)\|\mathcal{E}'_k(\sigma_k)\right) = \mathbb{D}\left(\rho\|\mathcal{E}'_k(\sigma_k)\right)\,. \tag{1.51}$$

Recall (1.38),

$$\mathbb{D}\left(\rho\|\mathcal{E}'_k(\sigma_k)\right) \geq \mathbb{D}(\rho\|\sigma) - \mathbb{D}_{\max}\left(\mathcal{E}'_k(\sigma_k)\|\sigma\right)\,. \tag{1.52}$$

Taking $k \to \infty$, we again obtain

$$\mathbb{D}(\rho\|\sigma) \leq \liminf_{k\to\infty} \mathbb{D}\left(\rho_k\|\sigma_k\right) + \limsup_{k\to\infty} \mathbb{D}_{\max}\left(\mathcal{E}'_k(\sigma_k)\|\sigma\right) \tag{1.53}$$

where

$$\limsup_{k\to\infty} \mathbb{D}_{\max}\left(\mathcal{E}'_k(\sigma_k)\|\sigma\right)\,. \tag{1.54}$$

The upper bound is also obtained

$$\mathbb{D}(\rho\|\sigma) \leq \liminf_{k\to\infty} \mathbb{D}(\rho_k\|\sigma_k)\,. \tag{1.55}$$

which completes the proof. □

1.6 Recap

As a recap of previous sections, we conclude that any $\mathbb{D}$ satisfying data-processing inequality, additivity and normalisation is also continuous in interior. faithful unless trivial, and bounded by $\mathbb{D}_{\min}(\rho\|\sigma) \leq \mathbb{D}(\rho\|\sigma) \leq \mathbb{D}_{\max}(\rho\|\sigma)$.

Here are some examples of quantum relative entropies with above-mentioned properties[6]:

$$\hat{\mathbb{D}}_\alpha(\rho\|\sigma) = \frac{1}{1-\alpha} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha , \tag{1.56}$$

$$\tilde{\mathbb{D}}_\alpha(\rho\|\sigma) = \frac{1}{1-\alpha} \log \operatorname{tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha , \tag{1.57}$$

$$\bar{\mathbb{D}}_\alpha(\rho\|\sigma) = \frac{1}{1-\alpha} \log \operatorname{tr} \rho^\alpha \sigma^{1-\alpha} . \tag{1.58}$$

When α approaches 1, $\tilde{\mathbb{D}}_\alpha$ and $\bar{\mathbb{D}}_\alpha$ converge to $\mathbb{D}$ defined by

$$\mathbb{D}(\rho\|\sigma) = \operatorname{tr} \rho(\log \rho - \log \sigma) . \tag{1.59}$$

Similarly, taking the limit of α to 1, $\hat{\mathbb{D}}_\alpha$ converges to $\hat{\mathbb{D}}$

$$\hat{\mathbb{D}}(\rho\|\sigma) = \operatorname{tr} \rho \log(\rho^{\frac{1}{2}} \sigma^{-1} \rho^{\frac{1}{2}}) . \tag{1.60}$$

These quantum relative entropies satisfies

$$\tilde{\mathbb{D}}_\alpha \leq \bar{\mathbb{D}}_\alpha \leq \hat{\mathbb{D}}_\alpha . \tag{1.61}$$

and equality only holds when $[\rho, \sigma] = 0$.

Another type of quantum relative entropy is[7]

$$\mathbb{D}_{\alpha,z}(\rho\|\sigma) = \frac{1}{1-\alpha} \log \operatorname{tr} \left(\sigma^{\frac{1-\alpha}{2z}} \rho \sigma^{\frac{1-\alpha}{2z}}\right)^z , \tag{1.62}$$

where the data-processing inequality is satisfied for a certain known range of (α, z).

[6] Martin Müller-Lennert, Freédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On Quantum Rényi Entropies: A New Generalization and Some Properties. *Journal of Mathematical Physics*, 54(12):122203, jun 2013. ISSN 00222488. DOI: 10.1063/1.4838856. URL http://link.aip.org/link/JMAPAQ/v54/i12/p122203/s1&Agg=doi

[7] Koenraad M. R. Audenaert and Nilanjana Datta. α-z-Relative Renyi Entropies. *Journal of Mathematical Physics*, 56:022202, oct 2015. DOI: 10.1063/1.4906367. URL http://arxiv.org/abs/1310.7178

2

Characterising quantities satisfying the axioms

Literature: These notes are based on my book[1].

[1] Marco Tomamichel. *Quantum information processing with finite resources: mathematical foundations*. Springer Cham, 2015. DOI: 10.1007/978-3-319-21891-5. URL https://doi.org/10.1007/978-3-319-21891-5

2.1 Relation between classical and quantum relative entropies

Classical relative entropies satisfying information-theoretic axioms are complete. Quantum relative entropies are expected to reduce to classical relative entropies for commuting states. We will explore quantum relative entropies from this starting point of view.

We write $D_\alpha(\,\cdot\,\|\,\cdot\,)$ for a classical entropy and $\mathbb{D}_\alpha(\,\cdot\,\|\,\cdot\,)$ for a quantum entropy that reduces to $D_\alpha(\,\cdot\,\|\,\cdot\,)$ for $[\rho,\sigma]=0$. Two commuting quantum states (ρ,σ) can be simultaneously diagonalized with the same unitary. Let (p,q) denote eigenvalues of (ρ,σ) and λ label eigenvalues of (ρ,σ). In that case, we require that

$$\mathbb{D}_\alpha(\rho\|\sigma) = D_\alpha(p\|q) = \frac{1}{\alpha-1}\log\sum_\lambda p(\lambda)^\alpha q(\lambda)^{1-\alpha}\,. \tag{2.1}$$

Here we claim that quantum relative entropies are invariant under isometries, which justifies the above requirement. We construct the isometric channel

$$\mathcal{V}:\rho\mapsto V\rho V^\dagger \tag{2.2}$$

and the pseudo-reverse isometric channel

$$\mathcal{V}':\sigma\mapsto V^\dagger\sigma V + \xi\,\mathrm{tr}((\mathbb{I}-VV^\dagger)\sigma) \tag{2.3}$$

where V is an isometry and ξ is an arbitrary quantum state. With $V^\dagger V=\mathbb{I}$, one can verify that $\rho=\mathcal{V}'\circ\mathcal{V}(\rho)$. From the data-processing inequality with $\mathcal{V}$,

$$\mathbb{D}_\alpha(\rho\|\sigma)\geq\mathbb{D}_\alpha\left(\mathcal{V}(\rho)\|\mathcal{V}(\rho)\right)\,. \tag{2.4}$$

From the data-processing inequality with $\mathcal{V}'$,

$$\mathbb{D}_\alpha\left(\mathcal{V}(\rho)\|\mathcal{V}(\sigma)\right)\geq\mathbb{D}_\alpha(\rho\|\sigma)\,. \tag{2.5}$$

And thus we conclude

$$\mathbb{D}_\alpha(\rho\|\sigma) = \mathbb{D}_\alpha\left(\mathcal{V}(\rho)\|\mathcal{V}(\sigma)\right). \tag{2.6}$$

The quantum relative entropy thus reduce to the classical relative entropy via

$$\mathbb{D}_\alpha(\rho\|\sigma) = D(p\|q). \tag{2.7}$$

2.2 *Minimal and Maximal generalisation of Rényi relative entropies*

Here we define two quantum generalisations which are not necessarily relative entropies, the minimal and the maximal quantum generalisation of Rényi divergences, $\check{D}_\alpha(\rho\|\sigma)$ and $\hat{D}_\alpha(\rho\|\sigma)$. Their name will be justified later. They are defined as

$$\check{D}_\alpha(\rho\|\sigma) = \sup_{\mathcal{M}:\text{qc-CPTP}} D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)), \tag{2.8}$$

where qc-CPTP denotes the set of quantum-classical channels (measurements) and

$$\hat{D}_\alpha(\rho\|\sigma) = \inf_{\substack{\mathcal{P}:\text{cq-CPTP} \\ (p,q)\in\text{pdf} \\ \mathcal{P}(p)=\rho,\mathcal{P}(q)=\sigma}} D_\alpha(p\|q), \tag{2.9}$$

where pdf denotes the set of probability distributions and cq-CPTP denotes the set of classical-quantum channels (preparation) which prepares (ρ, σ) from (p, q) respectively. These two quantities satisfy the following properties.

1. $\hat{D}_\alpha$ and $\check{D}_\alpha$ satisfy

$$\hat{D}_\alpha(\rho\|\sigma) \geq \check{D}_\alpha(\rho\|\sigma). \tag{2.10}$$

Proof. Consider that we first prepare (ρ, σ) from (p, q) with $\mathcal{P}$ and obtain (p', q') from (ρ, σ) with $\mathcal{M}$. $\mathcal{M} \circ \mathcal{P}$ is equivalent to a classical channel which transforms (p, q) to (p', q') for any suitable (p, q) and (p', q') satisfying constraints. Thus by the data-processing inequality

$$\hat{D}_\alpha(\rho\|\sigma) = \inf_{\substack{p=\mathcal{M}(\rho) \\ q=\mathcal{M}(\sigma)}} D_\alpha(p\|q) \geq \sup_{\substack{\rho=\mathcal{P}(p') \\ \sigma=\mathcal{P}(q')}} D_\alpha(p'\|q') = \check{D}_\alpha(\rho\|\sigma), \tag{2.11}$$

which completes the proof. □

2. If $[\rho, \sigma] = 0$, it is guaranteed that

$$\hat{D}_\alpha(\rho\|\sigma) = \check{D}_\alpha(\rho\|\sigma) = D_\alpha(\rho\|\sigma). \tag{2.12}$$

Proof. When $[\rho,\sigma] = 0$, (ρ,σ) can be simultaneously diagonalized in common eigenbasis. Let (p,q) be eigenvalues and $|\lambda\rangle\langle\lambda|$ be common eigenbasis of (ρ,σ). Due to the isometry invariance of quantum relative entropies,

$$D_\alpha(\rho\|\sigma) = D_\alpha(p\|q)\,. \tag{2.13}$$

We define $\mathcal{M} : \tau \mapsto p_\lambda = \mathrm{tr}(|\lambda\rangle\langle\lambda|\tau)$ be measurement in common eigenbasis. Note that $\mathcal{M}(\rho) = p$ and $\mathcal{M}(\sigma) = q$. From definition of $\check{D}_\alpha$ in (2.8),

$$\check{D}_\alpha(\rho\|\sigma) \geq D_\alpha(p\|q)\,. \tag{2.14}$$

Similarly, Let $\mathcal{P} : p \mapsto \tau = \sum_\lambda p_\lambda|\lambda\rangle\langle\lambda|$ be preparation in common eigenbasis. You may verify that $\mathcal{P}(p) = \rho$ and $\mathcal{P}(q) = \sigma$. From definition of $\hat{D}_\alpha$ in (2.9),

$$\hat{D}_\alpha(\rho\|\sigma) \leq D_\alpha(p\|q)\,. \tag{2.15}$$

Combining with (2.10), one obtains

$$\hat{D}_\alpha(\rho\|\sigma) = \check{D}_\alpha(\rho\|\sigma) = D_\alpha(\rho\|\sigma) \tag{2.16}$$

□

3. $\check{D}_\alpha(\rho\|\sigma)$ and $\hat{D}_\alpha(\rho\|\sigma)$ satisfies the data-processing inequality under CPTP maps.

$$\check{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq \check{D}_\alpha(\rho\|\sigma)\,, \tag{2.17}$$

$$\hat{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq \hat{D}_\alpha(\rho\|\sigma)\,. \tag{2.18}$$

Proof. According to (2.8),

$$\check{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) = \sup_{\mathcal{M}:\text{qc-CPTP}} D_\alpha(\mathcal{M}\circ\mathcal{E}(\rho)\|\mathcal{M}\circ\mathcal{E}(\sigma))\,, \tag{2.19}$$

$$\check{D}_\alpha(\rho\|\sigma) = \sup_{\mathcal{M}':\text{qc-CPTP}} D_\alpha(\mathcal{M}'(\rho)\|\mathcal{M}'(\sigma))\,. \tag{2.20}$$

Recall that for any $\mathcal{M} \in$ qc-CPTP, $\mathcal{M}\circ\mathcal{E} \in$ qc-CPTP. The set of $\mathcal{M}\circ\mathcal{E}$ which the former maximises over is a subset of the set of $\mathcal{M}'$ which the latter maximises over. Therefore, the former is smaller than the latter,

$$\check{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq \check{D}_\alpha(\rho\|\sigma)\,. \tag{2.21}$$

Similarly, according to (2.9),

$$\hat{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) = \inf_{\substack{\mathcal{P}:\text{cq-CPTP}\\(p,q)\in\text{pdf}\\\mathcal{P}(p)=\mathcal{E}(\rho),\mathcal{P}(q)=\mathcal{E}(\sigma)}} D_\alpha(p\|q)\,, \tag{2.22}$$

$$\hat{D}_\alpha(\rho\|\sigma) = \inf_{\substack{\mathcal{P}':\text{cq-CPTP}\\(p',q')\in\text{pdf}\\\mathcal{P}'(p')=\rho,\mathcal{P}'(q')=\sigma}} D_\alpha(p'\|q')\,. \tag{2.23}$$

However, for any $(p', q') \in \text{pdf}$ and $\mathcal{P}' \in \text{cq} - \text{CPTP}$ satisfying $\mathcal{P}'(p') = \rho$ and $\mathcal{P}'(q') = \sigma$, (p', q') and $\mathcal{E} \circ \mathcal{P}'$ also satisfy $\mathcal{E} \circ \mathcal{P}'(p') = \mathcal{E}(\rho)$ and $\mathcal{E} \circ \mathcal{P}'(q') = \mathcal{E}(\rho)$. The set of (p', q') and $\mathcal{E} \circ \mathcal{P}'$ which the latter minimises is a subset of the set of (p, q) and $\mathcal{P}$ which the former minimises. Therefore, the former is smaller than the latter,

$$\hat{D}_\alpha(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq \hat{D}_\alpha(\rho \| \sigma) \,. \tag{2.24}$$

□

With property 3, we can lift property 1 into property 1*

1* For any $\mathbb{D}_\alpha$ satisfying the data-processing inequality for any (ρ, σ), it holds that

$$\check{D}_\alpha(\rho \| \sigma) \leq \mathbb{D}_\alpha(\rho \| \sigma) \leq \hat{D}_\alpha(\rho \| \sigma) \,. \tag{2.25}$$

Proof. Note that there is a bijection between classical distributions (p, q) and corresponding classical states (μ, ν),

$$D_\alpha(p|q) = \mathbb{D}_\alpha(\mu \| \nu) \,. \tag{2.26}$$

Therefore, (2.8) and (2.9) can be re-written as

$$\check{D}_\alpha(\rho \| \sigma) = \sup_{\mathcal{M}:\text{qc-CPTP}} \mathbb{D}_\alpha(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)) \,, \tag{2.27}$$

and

$$\hat{D}_\alpha(\rho \| \sigma) = \inf_{\substack{\mathcal{P}:\text{cq-CPTP} \\ (\mu,\nu) \in \text{c-state} \\ \mathcal{P}(\mu) = \rho, \mathcal{P}(\nu) = \sigma}} \mathbb{D}_\alpha(\mu \| \nu) \,. \tag{2.28}$$

where c-state denotes the set of classical states. By the data-processing inequality for $\mathcal{M}$,

$$\check{D}_\alpha(\rho \| \sigma) = \sup_{\mathcal{M}:\text{qc-CPTP}} \mathbb{D}_\alpha(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)) \leq \mathbb{D}_\alpha(\rho \| \sigma) \,. \tag{2.29}$$

Similarly, by the data-processing inequality for $\mathcal{P}$,

$$\hat{D}_\alpha(\rho \| \sigma) = \inf_{\substack{\mathcal{P}:\text{cq-CPTP} \\ (\mu,\nu) \in \text{c-state} \\ \mathcal{P}(\mu) = \rho, \mathcal{P}(\nu) = \sigma}} \mathbb{D}_\alpha(\mu \| \nu) \geq \mathbb{D}_\alpha(\rho \| \sigma) \,. \tag{2.30}$$

In conclusion of both, we obtain

$$\check{D}_\alpha(\rho \| \sigma) \leq \mathbb{D}_\alpha(\rho \| \sigma) \leq \hat{D}_\alpha(\rho \| \sigma) \,. \tag{2.31}$$

□

We still have two more comments on (2.25). One is that $\check{D}_\alpha$ and $\hat{D}_\alpha$ are not additive in general. Another is that $\check{D}_\alpha$ and $\hat{D}_\alpha$ do not collapse unless $[\rho, \sigma] = 0$.

2.3 Regularised quantum relative entropies

We have obtained the relation (2.25) between $\mathbb{D}_\alpha$, $\check{D}_\alpha$ and $\hat{D}_\alpha$. It is natural to ask whether $\check{D}_\alpha$ and $\hat{D}_\alpha$ can be made additive. Luckily, we can force them to be weakly additive by regularising them.

Here we define two regularised quantum generalisations

$$\hat{D}_\alpha^{reg}(\rho\|\sigma) = \lim_{n\to\infty} \frac{1}{n} \hat{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \,, \tag{2.32}$$

and

$$\check{D}_\alpha^{reg}(\rho\|\sigma) = \lim_{n\to\infty} \frac{1}{n} \check{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \,. \tag{2.33}$$

We say D is weakly additive if

$$D_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD(\rho\|\sigma) \tag{2.34}$$

It can be proved that any $\mathbb{D}_\alpha$ satisfying the data-processing inequality and the weak additivity is sandwiched between $\check{D}_\alpha^{reg}$ and $\hat{D}_\alpha^{reg}$. That is, for any ρ and σ,

$$\hat{D}_\alpha \geq \hat{D}_\alpha^{reg} \geq \mathbb{D}_\alpha \geq \check{D}_\alpha^{reg} \geq \check{D}_\alpha \,. \tag{2.35}$$

The properties of $\hat{D}_\alpha$, $\check{D}_\alpha$, $\hat{D}_\alpha^{reg}$ and $\check{D}_\alpha^{reg}$ are not known completely. We can only obtain concrete expressions for certain regions.

1. For any $\alpha \in [0,2]$, we have

$$\hat{D}_\alpha(\rho\|\sigma) = \hat{D}_\alpha^{reg}(\rho\|\sigma) = \frac{1}{\alpha-1} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha \,. \tag{2.36}$$

Proof. First, the expression on the right-hand-side satisfies the data-processing inequality for this range of α (we will not show this here) and is additive. Due to (2.35),

$$\hat{D}_\alpha^{reg}(\rho\|\sigma) \geq \frac{1}{\alpha-1} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right)^\alpha \,. \tag{2.37}$$

Second, we introduce the Matsumoto construction. Let

$$\Delta = \sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}} = \sum_x \lambda_x \Pi_x \,, \tag{2.38}$$

where the latter is the spectral decomposition of the former. Denoting $Q(x) = \operatorname{tr}(\sigma\Pi_x)$ and $P(x) = \lambda_x Q(x)$, we construct the preparation channel

$$\mathcal{P} : x \mapsto \frac{1}{Q(x)} \sigma^{\frac{1}{2}} \Pi_x \sigma^{\frac{1}{2}} \,. \tag{2.39}$$

One can easily verify that $\mathcal{P}$ prepares a quantum state and that $\mathcal{P}(P) = \rho$ and $\mathcal{Q} = \sigma$. By the definition of $\hat{D}_\alpha$ in (2.9),

$$\hat{D}_\alpha(\rho\|\sigma) \leq D_\alpha(P\|Q) = \frac{1}{\alpha-1} \log \sum_x Q(x)^{1-\alpha} P(x)^\alpha \,. \tag{2.40}$$

Notice that

$$\sum_x Q(x)^{1-\alpha} P(x)^\alpha = \sum_x \operatorname{tr} \sigma \Pi_x \lambda_x^\alpha = \operatorname{tr} \sigma \Delta^\alpha = \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha , \quad (2.41)$$

we obtain

$$\hat{D}_\alpha(\rho\|\sigma) \le \frac{1}{\alpha - 1} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha . \quad (2.42)$$

Combining both the first and the second,

$$\hat{D}_\alpha(\rho\|\sigma) = \hat{D}_\alpha^{reg}(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \operatorname{tr} \sigma \left(\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\right)^\alpha , \quad (2.43)$$

which completes the proof. □

We have not found an application of this quantity for any operational problem. One open question is whether we can use some other axiom to rule it out as a generalization.

2. For any $\alpha \in [\frac{1}{2}, \infty)$, we have

$$\hat{D}_\alpha^{reg}(\rho\|\sigma) = \tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha , \quad (2.44)$$

Proof. First, one again needs to verify that the expression on the right-hand-side (which is nontrivial and not done here) indeed satisfies the data-processing inequality and is additivity. Due to (2.25),

$$\check{D}_\alpha^{reg}(\rho\|\sigma) \le \tilde{D}_\alpha(\rho\|\sigma) . \quad (2.45)$$

Second, we introduce the Pinching channel. Let $\{P_x\}_x$ be a set of m orthogonal projectors that sums to identity

$$\sum_x P_x = \mathbb{1} . \quad (2.46)$$

The Pinching channel is defined as

$$\mathcal{P}_{\{P_x\}_x}(\rho) = \sum_x P_x \rho P_x . \quad (2.47)$$

Let $\{U_y\}_y$ be the Fourier transform of $\{P_x\}_x$

$$U_y = \sum_x e^{2\pi \mathrm{i} \frac{xy}{m}} P_x . \quad (2.48)$$

One can find

$$\mathcal{P}_{\{P_x\}_x}(\rho) = \frac{1}{m} \sum_y U_y \rho U_y^\dagger \rho . \quad (2.49)$$

with brute force calculation

$$\frac{1}{m} \sum_y U_u \rho U_y^\dagger = \frac{1}{m} \sum_{xx'} \sum_y e^{2\pi \mathrm{i} \frac{(x-x')y}{m}} P_x \rho P_{x'} = \sum_x P_x \rho P_x . \quad (2.50)$$

Observe that

$$U_m = \mathbb{1}\,, \tag{2.51}$$

We have

$$\rho \leq \mathcal{P}_{\{P_x\}_x}(\rho) \cdot m\,. \tag{2.52}$$

(2.52) is the pinching inequality. Consider that for any $\sigma^{\otimes n}$, there are at most $poly(n)$ distinct eigenvalues $\lambda_x^{\sigma,n}$ and distinct projectors $P_x^{\sigma,n}$ because of its permutation invariance. Therefore, denoting $\mathcal{P}_{\{P_x^{\sigma,n}\}}$ by $\mathcal{P}_{\sigma,n}$, we obtain

$$\rho^{\otimes n} \leq \mathcal{P}_{\sigma,n}(\rho^{\otimes n}) \cdot poly(n)\,. \tag{2.53}$$

It is obvious that

$$\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{n}\tilde{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n})\,. \tag{2.54}$$

Now we have to divide $(\frac{1}{2}, 1)$ into three cases to apply (2.52). When restricted to $\alpha \in (1, \infty)$, we obtain

$$\tilde{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq \tilde{D}_\alpha(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) + \log \mathrm{poly}(n)\,, \tag{2.55}$$

by using that for monotone f, if $A \leq B$ then $\operatorname{tr} f(A) \leq \operatorname{tr} f(B)$. When restricted to $\alpha = 1$, we obtain

$$\begin{aligned} D(\rho^{\otimes n}\|\sigma^{\otimes n}) &= \operatorname{tr} \rho^{\otimes n}(\log \rho^{\otimes n} - \log \sigma^{\otimes n}) \\ &\leq \operatorname{tr} \rho^{\otimes n}(\log \mathcal{P}_{\sigma,n}(\rho^{\otimes n}) - \log \sigma^{\otimes n}) + \log poly(n) \\ &= \operatorname{tr} \mathcal{P}_{\sigma,n}(\rho^{\otimes n})(\log \mathcal{P}_{\sigma,n}(\rho^{\otimes n}) - \log \sigma^{\otimes n}) + \log \mathrm{poly}(n) \\ &= D(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) + \log \mathrm{poly}(n)\,, \end{aligned}$$

or

$$D(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq D(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) + \log \mathrm{poly}(n)\,, \tag{2.56}$$

by using that for operator monotone f, if $A \leq B$ then $f(A) \leq f(B)$. When restricted to $\alpha \in (\frac{1}{2}, 1)$, more careful arguments are necessary. The result, however, does not change. In summary, for $\alpha \in (\frac{1}{2}, \infty)$, we always have

$$\tilde{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) \leq \tilde{D}(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) + \log \mathrm{poly}(n)\,. \tag{2.57}$$

By definition of $\tilde{D}_\alpha$,

$$\tilde{D}_\alpha(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) = D_\alpha(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n})\,. \tag{2.58}$$

Notice that the pinching channel is a partial measurement channel. By definition of $\check{D}_\alpha$

$$D(\mathcal{P}_{\sigma,n}(\rho^{\otimes n})\|\sigma^{\otimes n}) \leq \check{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n})\,. \tag{2.59}$$

Thus summarising all equations above,

$$\tilde{D}(\rho\|\sigma) \leq \frac{1}{n}\check{D}_\alpha(\rho^{\otimes n})\|\sigma^{\otimes n}) + \frac{\log \mathrm{poly}(n)}{n}\,. \tag{2.60}$$

Taking the limit of $n \to \infty$,

$$\tilde{D}(\rho\|\sigma) \leq \check{D}_\alpha^{reg}(\rho\|\sigma)\,. \tag{2.61}$$

Combining both the first and the second, we obtain

$$\hat{D}_\alpha^{reg}(\rho\|\sigma) = \tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha-1}\log \mathrm{tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha, \tag{2.62}$$

which completes the proof. □

Another quantum generalisation which we frequently use is due to Petz,

$$\overline{D}_\alpha = \frac{1}{\alpha-1}\log \mathrm{tr}\,\rho^\alpha\sigma^{1-\alpha}\,. \tag{2.63}$$

$\overline{D}_\alpha$ satisfies the data-processing inequality for $\alpha \in [0,2]$.

The relation between different quantum generalisations can be found in Figure 4.1 in the textbook[2]. We also present and summarise it here. $\hat{D}_\alpha$ and $\tilde{D}_\alpha$ are the maximal and minimal generalisations satisfying the data-processing inequality and the weakly additivity, respectively. $\tilde{D}_\alpha$ has optimal interpretation describing the strong converse exponent in various problem. $\overline{D}_\alpha$ is usually used to describe error exponents.

[2] Marco Tomamichel. *Quantum information processing with finite resources: mathematical foundations*. Springer Cham, 2015. DOI: 10.1007/978-3-319-21891-5. URL https://doi.org/10.1007/978-3-319-21891-5

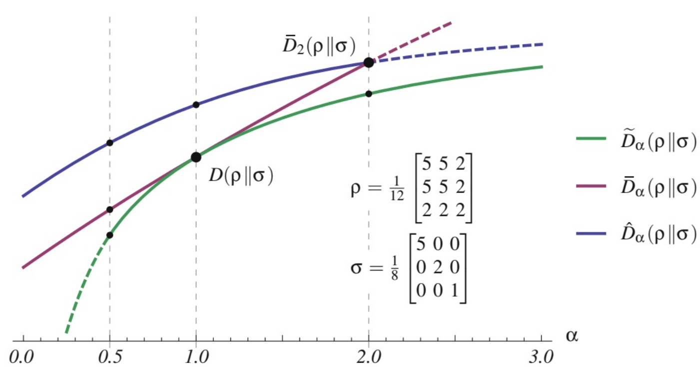

Figure 2.1: Minimal, Petz and maximal quantum Rényi entropy (for small α). These divergences are discussed in previous sections, respectively. Solid lines are used to indicate that the quantity satisfies the data-processing inequality in this range of α.

There are still open questions regarding quantum generalisations. One is whether we can also get $\overline{D}_\alpha$ from operational principles. Another is whether there is another meaningful axiom to select quantities with operational meanings.

Bibliography

Koenraad M. R. Audenaert and Nilanjana Datta. α-z-Relative Renyi Entropies. *Journal of Mathematical Physics*, 56:022202, oct 2015. DOI: 10.1063/1.4906367. URL http://arxiv.org/abs/1310.7178.

Paul Erdös. On the distribution function of additive functions. *Annals of Mathematics*, 47(1):1–20, 1946. ISSN 0003486X. DOI: 10.2307/1969031. URL http://www.jstor.org/stable/1969031.

Gilad Gour and Marco Tomamichel. Optimal extensions of resource measures and their applications. *Phys. Rev. A*, 102:062401, Dec 2020. DOI: 10.1103/PhysRevA.102.062401. URL https://link.aps.org/doi/10.1103/PhysRevA.102.062401.

Gilad Gour and Marco Tomamichel. Entropy and relative entropy from information-theoretic principles. *IEEE Transactions on Information Theory*, 67(10):6313–6327, 2021. DOI: 10.1109/TIT.2021.3078337. URL https://doi.org/10.1109/TIT.2021.3078337.

Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. From blackwell dominance in large samples to rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021. DOI: 10.3982/ECTA17548. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17548.

Martin Müller-Lennert, Freédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On Quantum Rényi Entropies: A New Generalization and Some Properties. *Journal of Mathematical Physics*, 54(12):122203, jun 2013. ISSN 00222488. DOI: 10.1063/1.4838856. URL http://link.aip.org/link/JMAPAQ/v54/i12/p122203/s1&Agg=doi.

Marco Tomamichel. *Quantum information processing with finite resources: mathematical foundations*. Springer Cham, 2015. DOI: 10.1007/978-3-319-21891-5. URL https://doi.org/10.1007/978-3-319-21891-5.