α-z-Rényi relative entropies

Koenraad M.R. Audenaert[1,2] and Nilanjana Datta[3]

[1] *Department of Mathematics, Royal Holloway University of London, Egham TW20 0EX, U.K.*

[2] *Department of Physics and Astronomy, Ghent University, S9, Krijgslaan 281, B-9000 Ghent, Belgium*

[3] *Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, U.K.*

Abstract

We consider a two-parameter family of Rényi relative entropies $D_{\alpha,z}(\rho||\sigma)$ that are quantum generalisations of the classical Rényi divergence $D_\alpha(p||q)$. This family includes many known relative entropies (or divergences) such as the quantum relative entropy, the recently defined quantum Rényi divergences, as well as the quantum Rényi relative entropies. All its members satisfy the quantum generalizations of Rényi's axioms for a divergence. We consider the range of the parameters α, z for which the data processing inequality holds. We also investigate a variety of limiting cases for the two parameters, obtaining explicit formulas for each one of them.

1 Introduction

The *quantum relative entropy* as introduced by Umegaki [1] is the proper [2] quantum generalisation of the classical Kullback-Leibler divergence and it therefore plays a central role in quantum information theory. In particular, fundamental limits on the performance of information-processing tasks in the so-called "asymptotic, memoryless (or i.i.d.) setting" is given in terms of quantities derived from the quantum relative entropy.

There are, however, several other entropic quantities and generalized relative entropies (or divergences) which are also of operational significance. One of the most important of these is the family of relative entropies called the *α-Rényi relative entropies* (α-RRE) $D_\alpha(\rho||\sigma)$, where $\alpha \in (0,1)\bigcup(1,\infty)$, which are quantum generalisations of the classical Rényi divergences. For $\alpha \in (0,1)$ these relative entropies arise in the quantum Chernoff bound [5] which characterizes the probability of error in discriminating two different quantum states in the setting of asymptotically many copies. In analogy with the operational interpretation of their classical counterparts, the α-RRE can be viewed as generalized cutoff rates in quantum binary state discrimination [6].

In the light of this plethora of different entropic quantities that arise in quantum information theory, it is desirable to find a mathematical framework that unifies as many of these quantities as possible. Recently, a non-commutative generalization of the α-RRE was defined that partially provided such a framework. Known alternatively as the α *quantum*

Rényi divergence (α-QRD) or the *"sandwiched" Rényi relative entropy*, it depends on a parameter $\alpha \in (0,1)\bigcup(1,\infty)$ [11, 12, 13, 14]. For two positive semidefinite operators ρ and σ we denote it as $\widetilde{D}_\alpha(\rho||\sigma)$. It has been proved to reduce to the min-relative entropy when $\alpha = 1/2$, to the quantum relative entropy in the limit $\alpha \to 1$, and to the max-relative entropy in the limit $\alpha \to \infty$ [7, 8]. Consequently, many properties of the min-, max- and quantum relative entropies can be inferred directly from those of the α-QRD. For example, the data-processing inequality (i.e. monotonicity under completely positive trace-preserving maps) of these relative entropies is implied by that of $\widetilde{D}_\alpha(\rho||\sigma)$ for $\alpha \geq 1/2$ [16, 17]. The fact that the min- and max-relative entropies provide lower and upper bounds to the quantum relative entropy follows directly from the fact that the function $\widetilde{D}_\alpha(\rho||\sigma)$ is monotonically increasing in α [14]. Also joint convexity of the min- and quantum relative entropies is implied by the joint convexity of $\widetilde{D}_\alpha(\rho||\sigma)$ for $1/2 \leq \alpha \leq 1$ [16].

In spite of these and various other interesting properties, which have been proved using a variety of sophisticated mathematical tools, the framework of the α-QRD family has certain limitations: (i) the data-processing inequality, which is one of the most desirable properties of any divergence-type quantity, is not satisfied for $\alpha \in (0, 1/2)$ [14, 18], and (ii) the α-QRD family is not the only quantum generalisation of the classical Rényi divergences, as it does not incorporate the previously mentioned α-RRE family.

In this paper we address both limitations by introducing a two-parameter family of quantum relative entropies that generalise the classical Rényi divergences. We refer to them as *α-z-Rényi relative entropies* (α-z-RRE), and denote them as $D_{\alpha,z}(\rho||\sigma)$, with α and z being two real parameters. For every value of the parameter z one thus obtains a different, continuously varying quantum generalisation of $D_\alpha(p||q)$. This new family satisfies the data processing inequality (DPI) for *all* values of α, with certain restrictions on the parameter z as indicated below. Furthermore, both the α-QRD and the α-RRE are included as special cases (for $z = \alpha$ and $z = 1$, respectively).

In Section 2 we define this new family of relative entropies and summarize our main results. We state how the other known relative entropies can be obtained from this family; we prove that the α-z-RRE satisfies the quantum generalizations of Rényi's axioms for a divergence, and describe the regions in the α-z plane where these entropies satisfy the data-processing inequality. We study a special case of the α-z-RRE, which we denote as $\widehat{D}_\alpha$ (and informally call the *reverse sandwiched Rényi relative entropy*) due to its similarities with the α-QRD (or sandwiched Rényi relative entropy). It satisfies the data-processing inequality for $\alpha \leq 1/2$, and we obtain an interesting closed expression for it in the limit $\alpha \to 1$. In Sections 3, 4 and 5 we study limiting cases of the α-z-RRE. We end the paper with a brief summary of our results and some open questions in Section 6.

Obtaining a single quantum generalization of the classical Rényi divergence, which would cover all possible operational scenarios in quantum information theory, is a challenging (and perhaps impossible) task. However, we believe that the α-z-RRE is currently the best candidate for such a quantity, since it unifies all known quantum relative entropies in the literature to date.

2 Definitions and Main Results

Throughout the paper $\mathcal{H}$ denotes a finite-dimensional Hilbert space. We denote by $\mathcal{P}(\mathcal{H})$ the set of positive semidefinite operators on $\mathcal{H}$ and by $\mathcal{D}(\mathcal{H})$ the set of density operators on $\mathcal{H}$, i.e. operators $\rho \in \mathcal{P}(\mathcal{H})$ with $\operatorname{Tr}\rho = 1$. Further, we denote the support of an operator ρ by $\operatorname{supp}\rho$. Logarithms are taken to base 2. We denote the ordered eigenvalues of a $d \times d$ Hermitian matrix X as $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_d(X)$.

Let us first give the definition of the α-z-Rényi relative (α-z-RRE) entropies; $\forall \rho \in \mathcal{D}(\mathcal{H}), \sigma \in \mathcal{P}(\mathcal{H})$ with $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$

$$D_{\alpha,z}(\rho||\sigma) := \frac{1}{\alpha-1}\log f_{\alpha,z}(\rho||\sigma), \tag{1}$$

where $f_{\alpha,z}(\rho||\sigma)$ is the trace functional

$$\begin{aligned} f_{\alpha,z}(\rho||\sigma) &:= \operatorname{Tr}\left(\rho^{\alpha/2z}\sigma^{(1-\alpha)/z}\rho^{\alpha/2z}\right)^z & (2)\\ &= \operatorname{Tr}\left(\sigma^{(1-\alpha)/2z}\rho^{\alpha/z}\sigma^{(1-\alpha)/2z}\right)^z. & (3) \end{aligned}$$

Here, $\alpha \in \mathbb{R}$ and the limit has to be taken for α tending to 1, and $z \in \mathbb{R}^+$ and the limit has to be taken for z tending to 0. Also, negative powers are defined in the sense of generalized inverses; that is, for negative x, $\rho^x := (\rho|_{\operatorname{supp}\rho})^x$. The above definition is easily extended to the case in which $\rho \geq 0$ but $\operatorname{Tr}\rho \neq 1$ (see (12)). The trace functional can be written alternatively as

$$f_{\alpha,z}(\rho||\sigma) = \operatorname{Tr}\left(\rho^{\alpha/z}\sigma^{(1-\alpha)/z}\right)^z. \tag{4}$$

This is because for any pair of square matrices A and B, the eigenvalues of AB and BA are the same (see, e.g. [27], exercise I.3.7). Hence, the matrix $\rho^{\alpha/z}\sigma^{(1-\alpha)/z}$ has real, non-negative eigenvalues (even though it is not in general self-adjoint), and the trace functional $\operatorname{Tr}(\cdot)^z$ in this expression is well-defined as the sum of zth powers of these eigenvalues, which are the same as those of $\rho^{\alpha/2z}\sigma^{(1-\alpha)/z}\rho^{\alpha/2z}$.

For commuting ρ and σ, $D_{\alpha,z}(\rho||\sigma)$ reduces to the classical α-Rényi divergence, for all values of z, as required.

Clearly, this family includes the α-RRE family:

$$D_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\operatorname{Tr}\left(\rho^\alpha\sigma^{1-\alpha}\right) = D_{\alpha,1}(\rho||\sigma), \tag{5}$$

and the α-QRD family:

$$\widetilde{D}_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha = D_{\alpha,\alpha}(\rho||\sigma). \tag{6}$$

Specifically, we get the known correspondences [14]

$$D_{\min} = D_{1/2,1/2}, \quad D = \lim_{\alpha\to 1} D_{\alpha,\alpha}, \quad \text{and } D_{\max} = \lim_{\alpha\to\infty} D_{\alpha,\alpha}. \tag{7}$$

Here $D_{\min}$, D and $D_{\max}$ denote the min-relative entropy [7], the quantum relative entropy and the max-relative entropy [8], respectively:

$$\begin{aligned} D_{\min}(\rho||\sigma) &:= -2\log F(\rho,\sigma), \quad \text{where} \quad F(\rho,\sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1, \\ D(\rho||\sigma) &:= \mathrm{Tr}\,\rho\log\rho - \mathrm{Tr}\,\rho\log\sigma, \\ D_{\max}(\rho||\sigma) &:= \inf\{\gamma : \rho \leq 2^\gamma \sigma\}. \end{aligned} \tag{8}$$

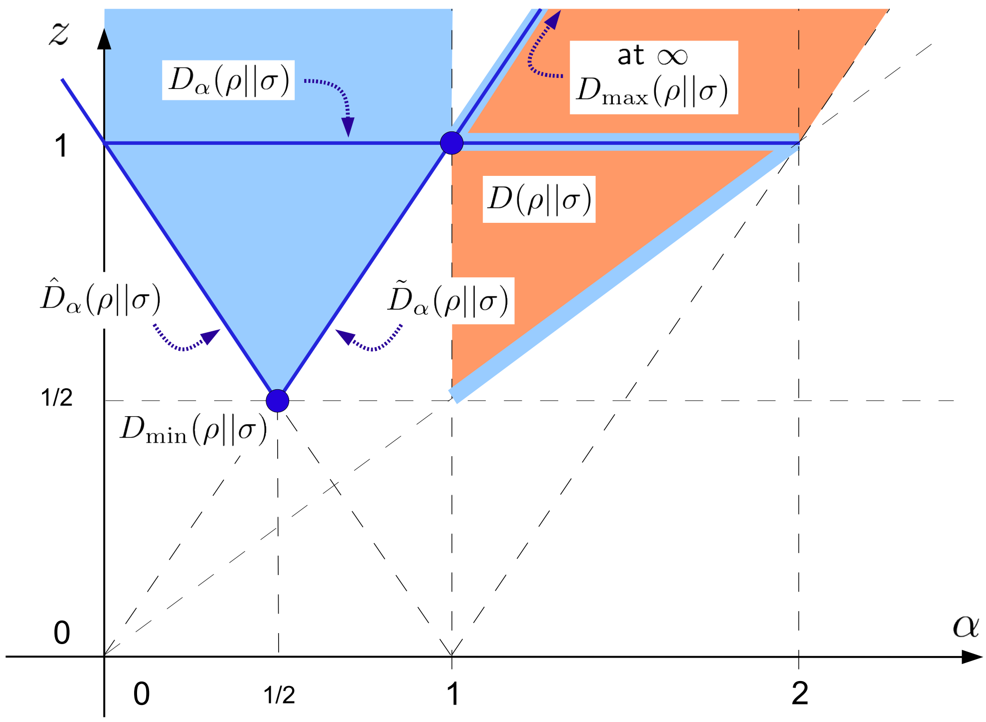

Figure 1: Schematic overview of the relative entropies that are unified by $D_{\alpha,z}$, as indicated by the dark-blue lines and dots. The region where the Data Processing Inequality (DPI) has been proven to hold has been coloured light-blue, and the orange region is where we conjecture validity of DPI. Outside these two regions DPI does not hold. For details, see Section 2.2.

These correspondences are illustrated in Figure 1. Also included in the family is a quantity defined by Hayashi in [3], which essentially is $D_{\alpha,2}$. Furthermore, as was pointed out by Lin and Tomamichel [35], the derivatives of $D_{\alpha,1}$ and $D_{\alpha,\alpha}$ with respect to α and taken at $\alpha = 1$ are both equal to one half the so-called *quantum information variance* [36, 37]

$$V(\rho||\sigma) := \mathrm{Tr}\,\rho(\log\rho - \log\sigma)^2 - D(\rho||\sigma)^2.$$

The epithet "sandwiched" in the original name of the α-QRD stems from the fact that in its formula ρ appears sandwiched between two powers of σ. Now note that one could also consider another way of sandwiching by putting σ between two powers of ρ, modifying the exponents accordingly so that the functional again coincides with D_α in the commutative setting. This new quantity $\widehat{D}_\alpha$ (which we informally call the *reverse sandwiched Rényi relative entropy*) is defined as

$$\widehat{D}_\alpha(\rho||\sigma) = \frac{1}{\alpha-1}\log \operatorname{Tr}\left(\rho^{\frac{\alpha}{2(1-\alpha)}}\sigma\rho^{\frac{\alpha}{2(1-\alpha)}}\right)^{1-\alpha} = D_{\alpha,1-\alpha}(\rho||\sigma). \tag{9}$$

From (11) we immediately obtain the symmetry relation

$$(\alpha-1)\widehat{D}_\alpha(\rho||\sigma) = (-\alpha)\widetilde{D}_{1-\alpha}(\sigma||\rho). \tag{10}$$

For $\alpha = 0$, $\widehat{D}_\alpha$ reduces to the 0-Rényi relative entropy, a quantity of particular operational relevance in one-shot information theory [9, 10]. This is in contrast to the α-quantum Rényi divergence, which does not in general reduce to the 0-Rényi relative entropy in the limit $\alpha \to 0$ [18].

Remarks.

1. For states ρ and σ with identical support, $D_{\alpha,z}$ is even in z: $D_{\alpha,z}(\rho||\sigma) = D_{\alpha,-z}(\rho||\sigma)$. This is no longer the case when the support of ρ is a proper subset of $\operatorname{supp}\sigma$. For example, one easily checks that $f_{2,-1}(\rho||\sigma) = \operatorname{Tr}\rho^2(\sigma|_{\operatorname{supp}\rho})^{-1}$, whereas $f_{2,1}(\rho||\sigma) = \operatorname{Tr}\rho^2(\sigma^{-1})|_{\operatorname{supp}\rho}$. Taking $z < 0$ might therefore complicate matters substantially, whereas there is no guarantee that the results will be interesting. We therefore have limited our considerations to $z > 0$ throughout.

2. The family obeys a symmetry condition with respect to α:

$$(\alpha-1)D_{\alpha,z}(\rho||\sigma) = (-\alpha)D_{1-\alpha,z}(\sigma||\rho). \tag{11}$$

3. The family coincides with certain quantum entropic functionals defined by Jakšić *et al.* [19] for the study of entropic fluctuations in non-equilibrium quantum statistical mechanics. These functionals were defined in the context of a dynamical system: in particular, ρ was the reference state of a dynamical system, and σ was the state ρ_t resulting from ρ due to time evolution under the action of a Hamiltonian for a time t. In contrast, we define $D_{\alpha,z}(\rho||\sigma)$ for arbitrary positive semidefinite states ρ and σ, and study its properties from a quantum information theoretic perspective.

2.1 Axiomatic properties

Following [14], we can check whether the α-z-RRE satisfies the six quantum Rényi axioms, as do the α-RRE and α-QRD. These are quantum generalizations of axioms that were put forward by Rényi in [20] as natural requirements that any classical divergence should satisfy. A quantum divergence is a functional which maps a pair of positive semidefinite operators ρ, σ, with $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$ onto $\mathbb{R}$. Its classical counterpart is obtained by replacing the operators by probability distributions.

Within this context we need to slightly redefine the α-z-RRE for non-normalized states ρ: $\forall \rho, \sigma \in \mathcal{P}(\mathcal{H})$ with $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$,

$$D_{\alpha,z}(\rho||\sigma) := \frac{1}{\alpha-1}\log\frac{f_{\alpha,z}(\rho,\sigma)}{\operatorname{Tr}\rho}. \tag{12}$$

(I) **Continuity:** For $\rho \neq 0$ and $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$, $D_{\alpha,z}(\rho||\sigma)$ is continuous in $\rho, \sigma \geq 0$ throughout the parameter space except for $\alpha \leq 0$. At $\alpha = 0$, the α-RRE is dependent on the rank of ρ and is therefore not continuous. This was actually the reason why Rényi included the continuity axiom: to exclude the cases $\alpha \leq 0$, where the relative entropy functional was not deemed a reasonable measure of information ([20], p. 558) due to its discontinuity.

The only case where it is not obvious that continuity holds for $\alpha > 0$ is the case $z = 0$. This will be considered in Section 3.

(II) **Unitary invariance:** For unitary U, $D_{\alpha,z}(U\rho U^*||U\sigma U^*) = D_{\alpha,z}(\rho||\sigma)$.

(III) **Normalization:** $D_{\alpha,z}(1||\frac{1}{2}) = 1$ (for scalar arguments, and when using base-2 logarithms), as is the case for any divergence that reduces to the classical Rényi divergence for commuting arguments.

(IV) **Order Axiom:** The axiom requires that

$$D_{\alpha,z}(\rho||\sigma)\left\{\begin{array}{c}\leq\\ \geq\end{array}\right\}0 \text{ whenever } \rho\left\{\begin{array}{c}\leq\\ \geq\end{array}\right\}\sigma.$$

Note that this axiom is a weaker version of the Data Processing Inequality (DPI) considered below, as follows from Lemma 5 in [4].

Proposition 1. *$D_{\alpha,z}$ satisfies the Order Axiom when $z \geq |\alpha - 1|$.*

Proof. Noting that $\operatorname{Tr}\rho = f_{\alpha,z}(\rho||\rho)$, we need, for $\alpha > 1$,

$$f_{\alpha,z}(\rho||\sigma)\left\{\begin{array}{c}\leq\\ \geq\end{array}\right\}f_{\alpha,z}(\rho||\rho) \text{ whenever } \rho\left\{\begin{array}{c}\leq\\ \geq\end{array}\right\}\sigma,$$

whereas, for $0 < \alpha < 1$,

$$f_{\alpha,z}(\rho||\sigma)\left\{\begin{array}{c}\leq\\ \geq\end{array}\right\}f_{\alpha,z}(\rho||\rho) \text{ whenever } \rho\left\{\begin{array}{c}\geq\\ \leq\end{array}\right\}\sigma.$$

This holds if the fractional power $(1-\alpha)/z$ that is applied to σ in (2) is operator monotone, when $0 < \alpha < 1$, and operator monotone decreasing, when $\alpha > 1$. In other words, for $0 < \alpha < 1$, $(1-\alpha)/z$ must lie between 0 and 1, i.e. $z \geq (1-\alpha)$. For $\alpha > 1$ it must lie between -1 and 0, i.e. $z \geq (\alpha - 1)$. □

In Figure 1 this corresponds to the triangular region with apex $(1, 0)$ and sides passing through the points $(0, 1)$ and $(2, 1)$, respectively.

(V) **Additivity** with respect to tensor products: clearly,

$$D_{\alpha,z}(\rho \otimes \tau||\sigma \otimes \omega) = D_{\alpha,z}(\rho||\sigma) + D_{\alpha,z}(\tau||\omega).$$

(VI) **Generalized Mean Value Axiom:** This axiom describes the behavior of $D_{\alpha,z}$ with respect to direct sums (the quantum generalization of taking the union of incomplete probability distributions). It requires the existence of a continuous, strictly increasing function g such that

$$(\operatorname{Tr}\rho + \operatorname{Tr}\tau)\; g(D_{\alpha,z}(\rho \oplus \tau || \sigma \oplus \omega)) = (\operatorname{Tr}\rho)\; g(D_{\alpha,z}(\rho||\sigma)) + (\operatorname{Tr}\tau)\; g(D_{\alpha,z}(\tau||\omega)).$$

In the classical case, if g is affine this requires that the divergence between pairs of unions of distributions is a weighted arithmetic mean of divergences, and this (along with the other axioms) limits D to be the classical relative entropy. Taking exponential g, $g(x) = \exp((\alpha - 1)x)$, we obtain the classical Rényi divergences.

Now, to see that $D_{\alpha,z}$ satisfies this axiom, it is sufficient to note that

$$f_{\alpha,z}(\rho \oplus \tau || \sigma \oplus \omega) = f_{\alpha,z}(\rho||\sigma) + f_{\alpha,z}(\tau||\omega).$$

This holds throughout the parameter space, provided we choose $g(x) = \exp((\alpha - 1)x)$, of course.

Note that in [20] only the case $\operatorname{Tr}\rho + \operatorname{Tr}\tau \leq 1$ and $\operatorname{Tr}\sigma + \operatorname{Tr}\omega \leq 1$ is considered, so that $\rho \oplus \tau$ and $\sigma \oplus \omega$ are normalized or subnormalized density matrices, the quantum generalization of generalized (i.e. complete or incomplete) probability distributions, but it turns out that even without this restriction the equality of the axiom holds.

2.2 Data Processing Inequality

A more difficult question is for which parameter range $D_{\alpha,z}$ satisfies the Data Processing Inequality (DPI). While this has not yet been established in full generality, it can be shown to hold for certain parameter ranges, indicated on Figure 1 by light-blue shading.

Theorem 1 (Data-processing Inequality)**.** *For any pair of positive semidefinite operators $\rho, \sigma \in \mathcal{P}(\mathcal{H})$, for which* $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$*, and for any CPTP map Λ acting on $\mathcal{P}(\mathcal{H})$, the Data Processing Inequality*

$$D_{\alpha,z}\left(\Lambda(\rho)||\Lambda(\sigma)\right) \leq D_{\alpha,z}\left(\rho||\sigma\right),$$

holds in each of the following cases:

- $0 < \alpha \leq 1$ *and* $z \geq \max(\alpha, 1-\alpha)$ *(Hiai),*
- $1 \leq \alpha \leq 2$ *and* $z = 1$ *(Ando),*
- $1 \leq \alpha$ *and* $z = \alpha$ *(Frank and Lieb; Beigi),*
- $1 \leq \alpha \leq 2$ *and* $z = \alpha/2$ *(Carlen, Frank and Lieb).*

It is well-known that to prove DPI for $D_{\alpha,z}$ one has to show that the trace functional $f_{\alpha,z}(\rho||\sigma)$ that lies at the heart of $D_{\alpha,z}$ is jointly concave when $\alpha \leq 1$, or jointly convex when $\alpha \geq 1$ (see, e.g. [16], its *Proof of Theorem 1 given Proposition 3*). In fact, it suffices to show that the related trace functional $f_{\alpha,z}(A; K)$, defined as

$$f_{\alpha,z}(A; K) := \operatorname{Tr}(A^{\alpha/z} K A^{(1-\alpha)/z} K^*)^{1/z}, \tag{13}$$

is concave/convex in A (for any fixed matrix K) over the set of positive semidefinite matrices. Joint concavity/convexity of the original functional $f_{\alpha,z}(\rho||\sigma)$ then follows by setting $K = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}$ and $A = \rho \oplus \sigma$.

Concavity of $f_{\alpha,z}(A;K)$ in the case $0 < \alpha \leq 1$ and $z \geq \max(\alpha, 1-\alpha)$ follows directly from a concavity theorem proven very recently by Hiai [22] (see also the older work [23]), whose proof is based on the complex analysis techniques employed by Epstein in [30]. Note that this generalises Corollary 1.1 of [15]. Epstein's paper is rather terse and uses deep results from complex analysis. A pedagogical introduction can be found, for example, in the appendix of [39]. Section 6 of [34] contains a detailed proof using similar techniques as Epstein's, but more elementary and tailored to the problem at hand.

Convexity was proven by Frank and Lieb [16] and independently by Beigi [17] for the case $1 \leq \alpha$ and $z = \alpha$, where $D_{\alpha,z}$ reduces to the QRD $\widetilde{D}_\alpha$. Convexity for $1 \leq \alpha \leq 2$ and $z = 1$ is exactly Ando's theorem [24]. Finally, after the appearance of the first version [34] of this paper, Carlen, Frank and Lieb were able to prove DPI in the case $1 \leq \alpha \leq 2$ and $z = \alpha/2$ [38].

Hiai [22] also provides necessary conditions for concavity/convexity. The regions in the parameter space where these conditions are *not* satisfied are indicated in Figure 1 as white space. About the remaining region, indicated in orange, nothing definitive is known other than that the conditions for necessity are satisfied. For this region we conjecture that the trace functional is convex, which would imply that DPI holds here as well.

When considering DPI, it is convenient to re-parameterize the trace functional $f_{\alpha,z}$ as

$$f_{p,q}(A;K) := \mathrm{Tr}(A^p K A^q K^*)^{1/(p+q)}, \tag{14}$$

where the parameters p and q are defined as $p = \alpha/z$ and $q = (1-\alpha)/z$. We obtain the original functional by setting $z = 1/(p+q)$ and $\alpha = p/(p+q)$.

Conjecture 1. *The trace functional $f_{p,q}(A;K)$ is convex on the set of positive definite $d \times d$ matrices for $-1 \leq p < 0$ and $1 \leq q \leq 2$ (or vice versa).*

Figure 2 shows the regions in (p,q)-parameter space where DPI provably holds and where we conjecture it.

Remark. One notices that whereas the α-QRD $\widetilde{D}_\alpha$ satisfies DPI only for $\alpha \geq 1/2$, the reverse α-QRD $\widehat{D}_\alpha$ satisfies DPI for $0 \leq \alpha \leq 1/2$.

2.3 Limiting cases

We study four limiting cases of the α-z-RRE: (i) limit $\alpha \to 1$ and $z \to 0$, (ii) the case of infinite α and z, (iii) fixed α and infinite z, and (iv) $z = \alpha \to 0$.

To study (i) we suitably parameterize z in terms of α as $z = r(\alpha - 1)$, where r is a non-zero finite real number, and consider the limit $\alpha \to 1$ (the case of fixed $\alpha \neq 1$ and $z \to 0$ will be studied elsewhere [25]). Note that $\alpha = 1$ is the only value of α where in the limit $z \to 0$ the Order Axiom (IV) is satisfied. For the choice $z = 1 - \alpha$, this yields the limit $\alpha \to 1$ of $\widehat{D}_\alpha(\rho||\sigma)$. In the general case in which ρ and σ do not commute, we obtain a rather

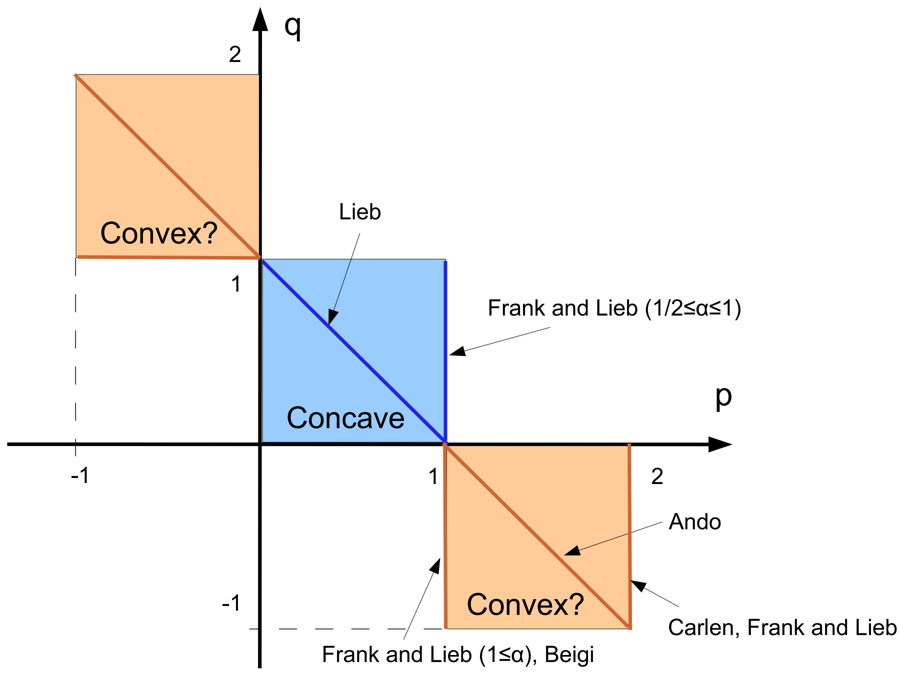

Figure 2: Regions of concavity (blue, proven in this paper) and conjectured convexity (orange) of the re-parameterized trace functional $f_{p,q}$, where $p = \alpha/z$ and $q = (1-\alpha)/z$. The dark blue and dark orange lines indicate the values of p and q for which concavity and convexity have already been proven. Note that the region where the Order Axiom is satisfied is the strip $-1 \leq q \leq 1$ and excludes the upper left orange square. The Continuity Axiom is satisfied in the region where p and $p+q$ have the same sign ($\alpha > 0$), again excluding the upper left orange square.

surprising formula for the latter: the relative entropy, not between ρ and σ, but between ρ and an operator $\widehat{\sigma}$ that is diagonal in the eigenbasis of ρ (see Theorems 2 and 3 for details). In the commuting case we recover the expected expression: the relative entropy of ρ and σ. We also prove that the α-z-RRE is continuous in ρ and σ in that limit.

To study the case (ii) of infinite α and z, we use the same parametrization of z, and take the limit $\alpha \to \infty$. In this limit the α-z-RRE is expressed in terms of a max-relative entropy (see Theorem 4 for details). In particular, our result readily yields the known [14] result that in the limit $\alpha \to \infty$, the α-QRD, $\widetilde{D}_\alpha(\rho||\sigma)$, reduces to the max-relative entropy $D_{\max}(\rho||\sigma)$.

Case (iii) concerns keeping α fixed (and finite) letting z tend to $+\infty$. Using the Lie-Trotter relation, we obtain the quantity $(1/(\alpha-1)) \log \operatorname{Tr} \exp(\alpha \log \rho + (1-\alpha) \log \sigma)$, which in the limit $\alpha \to 1$ tends to the relative entropy $D(\rho||\sigma)$.

Finally, we consider the case (iv) where α and z both tend to 0, with $z = \alpha$.

3 Limiting case $\alpha \to 1$ and $z \to 0$

In this section, we derive a closed form expression for the limit of $D_{\alpha,z}$ as z tends to 0. The most interesting point to calculate this is when $\alpha = 1$ because that is the only value where the Order Axiom remains satisfied as z goes to 0, even though DPI no longer holds. It turns out that $\lim_{z\to 0} D_{\alpha,z}$ is discontinuous in α at $\alpha = 1$ and we will have to be careful how the limit $z \to 0$ is taken. What we will consider is the limit $\alpha \to 1$ of $D_{\alpha,r(\alpha-1)}$, with fixed r, i.e. the limit along straight lines passing through the point $(1,0)$ and with slope r. This choice is particularly convenient since for $r = -1$ we recover the limit $\lim_{\alpha\to 1} \widehat{D}_\alpha$.

As we assume $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$ throughout, there is no loss of generality in only considering $\sigma > 0$; that is, all matrices will be restricted to the subspace $\operatorname{supp}\sigma$.

Lemma 1. *For $\sigma > 0$, and r a non-zero finite real number,*

$$\lim_{\alpha\to 1} \left(\rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)}\right)^{r(\alpha-1)} = \rho.$$

Proof. Since $\sigma > 0$, there exist $a, b > 0$ such that $a \leq \sigma \leq b$ (meaning that $aI \leq \sigma \leq bI$). Then, for $r > 0$, $b^{-1/r} \leq \sigma^{-1/r} \leq a^{-1/r}$ so that

$$b^{-1/r}\rho^{\alpha/r(\alpha-1)} \leq \rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)} \leq a^{-1/r}\rho^{\alpha/r(\alpha-1)}.$$

For $r < 0$ the roles of a and b get interchanged.

Raising this to the power $r(\alpha - 1)$, for $\alpha > 1$ and close enough to 1 so that this is an operator monotone operation, yields

$$b^{1-\alpha}\rho^\alpha \leq \left(\rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)}\right)^{r(\alpha-1)} \leq a^{1-\alpha}\rho^\alpha.$$

For $\alpha < 1$ and close enough to 1, a and b again have to be interchanged (as it is an operator monotone decreasing operation).

In the limit $\alpha \to 1$ we then get that $a^{1-\alpha}$ and $b^{1-\alpha}$ both tend to 1, and these inequalities become

$$\rho \leq \lim_{\alpha\to 1} \left(\rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)}\right)^{r(\alpha-1)} \leq \rho.$$

As both bounds are equal, this proves that the inequalities actually are equalities. □

A simple corollary of this lemma is that $\lim_{\alpha\to 1} f_{\alpha,r(\alpha-1)} = \operatorname{Tr}\rho = 1$. Hence, as α tends to 1, both the numerator and denominator in $D_{\alpha,r(\alpha-1)} = \log f_{\alpha,r(\alpha-1)}/(\alpha - 1)$ tend to 0. To calculate the limit it is tempting to use l'Hôpital's rule and calculate the derivative with respect to α. However, this approach did not yield any simplification. Instead, we followed a completely different approach, inspired by the power method [26] for numerically calculating eigenvalues.

We first consider the generic case in which the spectrum of ρ is non-degenerate, i.e. all its eigenvalues are distinct. Let us write the spectral decomposition of ρ as $\rho = \sum_{i=1}^d \mu_i P_i$, where the eigenvalues μ_i appear sorted in decreasing order and where P_i are the corresponding projectors $|i\rangle\langle i|$ on the (1-dimensional) eigenspaces. The main idea behind the power method is that for large positive s, ρ^s can be well-approximated by $\mu_1^s P_1$, in the sense that the sum of the remaining terms $\sum_{i=2}^d \mu_i^s P_i$ becomes much smaller in norm than μ_1^s.

Let us denote the matrix expression inside the trace of the trace functional $f_{\alpha,r(\alpha-1)}$ by $Z_{\alpha,r}(\rho||\sigma)$. Rather than applying the above approximation to the entire trace of $Z_{\alpha,r}(\rho||\sigma)$, which would be too crude, we apply it to the calculation of its largest eigenvalue λ_1 only. We get, for $z = r(\alpha-1) > 0$,

$$\begin{aligned}
\lambda_1(Z_{\alpha,r}(\rho||\sigma)) &= \lambda_1\left((\rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)})^{r(\alpha-1)}\right) \\
&\approx \mu_1^\alpha \operatorname{Tr}(P_1\sigma^{-1/r}P_1)^{r(\alpha-1)} \\
&= \mu_1^\alpha\left((\sigma^{-1/r})_{1,1}\right)^{r(\alpha-1)},
\end{aligned}$$

where $X_{1,1}$ indicates the upper left matrix element of a matrix X in the eigenbasis of ρ. This is shown in full rigor in Lemma 2 below.

As we ultimately need an expression for the trace we need approximations for all eigenvalues of $Z_{\alpha,r}$. To proceed, we will use the so-called "Weyl trick", which consists in calculating the largest eigenvalue of the kth antisymmetric tensor power of $Z_{\alpha,r}$ (see e.g. [27] Section I.5 for antisymmetric tensor powers and Section IX.2 for applications of the Weyl trick). For any given matrix X, its kth antisymmetric tensor power, denoted $X^{\wedge k}$, is defined as the restriction of its kth tensor power $X^{\otimes k}$ to the totally antisymmetric subspace. The reason for looking into this is that the largest eigenvalue of $X^{\wedge k}$ is the product of the k largest eigenvalues of X, an identity which we denote by the shorthand

$$\lambda_1(X^{\wedge k}) = \lambda_1\cdots\lambda_k(X) := \lambda_1(X)\cdots\lambda_k(X).$$

Furthermore, we have the relations $(XY)^{\wedge k} = X^{\wedge k}Y^{\wedge k}$ and $(X^s)^{\wedge k} = (X^{\wedge k})^s$.

For X of dimension d, k can take values from 1 to d. For $k = d$, the totally antisymmetric subspace is 1-dimensional and the antisymmetric tensor power $X^{\wedge d}$ is a scalar, namely the determinant of X. Analogously, the matrix elements of $X^{\wedge k}$ for $k < d$ are all possible $k \times k$ minors of X (determinants of submatrices). In particular, the "upper left" element $(X^{\wedge k})_{1,1}$ is the leading principal $k \times k$ minor of X. If we introduce the notation $X_{1:k,1:k}$ to mean the submatrix of X consisting of the first k rows and the first k columns, this element is given by

$$(X^{\wedge k})_{1,1} = \det\left(X_{1:k,1:k}\right).$$

Let us now apply the power method to $Z_{\alpha,r}^{\wedge k}$ in order to obtain an approximation for the product of the k largest eigenvalues of $Z_{\alpha,r}$. We will denote this product by $\lambda^{(k)}$, and by convention put $\lambda^{(0)} = 1$. First of all, note that $Z_{\alpha,r}(\rho||\sigma)^{\wedge k} = Z_{\alpha,r}(\rho^{\wedge k}||\sigma^{\wedge k})$. Hence, we get

$$\lambda_1(Z_{\alpha,r}(\rho||\sigma)^{\wedge k}) \approx \lambda_1(\rho^{\wedge k})^\alpha \left(\left((\sigma^{\wedge k})^{-1/r}\right)_{1,1}\right)^{r(\alpha-1)}$$

which means that

$$\lambda^{(k)} := \lambda_1\cdots\lambda_k(Z_\alpha(\rho||\sigma)) \approx (\mu_1\cdots\mu_k)^\alpha\left(\det\left((\sigma^{-1/r})_{1:k,1:k}\right)\right)^{r(\alpha-1)}. \tag{15}$$

A mathematically rigorous restatement of this approximate identity will be given below as the Approximation Lemma, Lemma 2. For $k = d$, we actually obtain an exact expression as it reduces to the well-known statement that the determinant of a product equals the product of the determinants:

$$\lambda^{(d)} = \det(Z_\alpha(\rho||\sigma)) = (\det\rho)^\alpha \left(\det\sigma^{-1/r}\right)^{r(\alpha-1)}.$$

It is now a simple matter to obtain an approximation for $\mathrm{Tr}\, Z_{\alpha,r}(\rho||\sigma)$. Indeed, by taking the quotients of successive $\lambda^{(k)}$ we get all the eigenvalues of $Z_{\alpha,r}$: $\lambda^{(k)}/\lambda^{(k-1)} = \lambda_k(Z_{\alpha,r}(\rho||\sigma))$. Summing these quotients then yields the trace of $Z_{\alpha,r}$:

$$\mathrm{Tr}\, Z_{\alpha,r}(\rho||\sigma) = \sum_{k=1}^d \lambda_k(Z_{\alpha,r}(\rho||\sigma)) = \lambda^{(1)} + \sum_{k=2}^d \frac{\lambda^{(k)}}{\lambda^{(k-1)}}.$$

Inserting the approximation (15) for $\lambda^{(k)}$ yields

$$\mathrm{Tr}\, Z_{\alpha,r}(\rho||\sigma) \approx \mu_1^\alpha \left((\sigma^{-1/r})_{1,1}\right)^{r(\alpha-1)} + \sum_{k=2}^d \mu_k^\alpha \left(\frac{\det\left((\sigma^{-1/r})_{1:k,1:k}\right)}{\det\left((\sigma^{-1/r})_{1:k-1,1:k-1}\right)}\right)^{r(\alpha-1)}. \tag{16}$$

Let us introduce the vector ν of leading principal minors of $\sigma^{-1/r}$ taken to the power $-r$, with

$$\nu_k := \det\left((\sigma^{-1/r})_{1:k,1:k}\right)^{-r}. \tag{17}$$

Note that $\nu_d = \det\sigma$. In terms of these ν_k, eq. (16) can be rewritten more succinctly as

$$\mathrm{Tr}\, Z_{\alpha,r}(\rho||\sigma) \approx \mu_1^\alpha \nu_1^{1-\alpha} + \sum_{k=2}^d \mu_k^\alpha \left(\frac{\nu_k}{\nu_{k-1}}\right)^{1-\alpha}.$$

One now recognizes the trace functional $f_{\alpha,z}$ in this formula, between the state ρ and a new positive definite matrix $\widehat{\sigma}$ that commutes with ρ and that is given by

$$\widehat{\sigma} = \mathrm{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1}). \tag{18}$$

Here, $C = \mathrm{diag}_\rho(x_1, \ldots, x_d)$ denotes a matrix C that is diagonal in the eigenbasis of ρ and has diagonal elements x_i; that is, $\mathrm{Tr}\, P_i C = x_i$.

We then finally get, for α sufficiently close to 1:

$$\mathrm{Tr}\, Z_{\alpha,r}(\rho||\sigma) \approx \mathrm{Tr}\, \rho^\alpha \widehat{\sigma}^{1-\alpha}. \tag{19}$$

The error in this approximation tends to 0 exponentially fast as $\exp(-\kappa/|r(1-\alpha|))$, where κ is a strictly positive constant depending only on the eigenvalues μ_i, as shown in Lemma 2 below. From (19) a closed form expression for the limit $\alpha \to 1$ of $D_{\alpha,r(\alpha-1)}$ can be found very easily, and it simply gives the classical relative entropy between ρ and $\widehat{\sigma}$. We have therefore proven:

Theorem 2. *Let ρ be a positive semidefinite matrix with non-degenerate spectrum and let σ be positive definite. Let r be a non-zero, finite real number. Then*

$$\begin{aligned} \lim_{\alpha\to 1} D_{\alpha,r(\alpha-1)}(\rho||\sigma) &= D(\rho||\,\mathrm{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1})), \\ \textit{with } \nu_k &= \det\left((\sigma^{-1/r})_{1:k,1:k}\right)^{-r}, \quad k = 1, \ldots, d. \end{aligned} \tag{20}$$

In particular, for $r=-1$,

$$\begin{aligned}\lim_{\alpha\to 1}\widehat{D}_\alpha(\rho||\sigma) &= D(\rho||\operatorname{diag}_\rho(\nu_1,\nu_2/\nu_1,\nu_3/\nu_2,\ldots,\nu_d/\nu_{d-1})),\\ \textit{with } \nu_k &= \det(\sigma_{1:k,1:k}), \quad k=1,\ldots,d. \end{aligned} \tag{21}$$

As a sanity check, we can consider what eq. (21) reduces to when ρ and σ commute. In that case, σ is diagonal in the eigenbasis of ρ, and its leading principal minors are just the products of its k first diagonal elements: $\nu_k=\sigma_{1,1}\cdots\sigma_{k,k}$. Hence, the successive quotients ν_k/ν_{k-1} reduce to $\sigma_{k,k}$, and $\operatorname{diag}_\rho(\nu_1,\nu_2/\nu_1,\nu_3/\nu_2,\ldots,\nu_d/\nu_{d-1})$ simply turns into σ itself. We thus find that, in the commuting case, $\lim_{\alpha\to1}D_{\alpha,r(\alpha-1)}(\rho||\sigma)=D(\rho||\sigma)$, as required.

To complete the case of non-degenerate ρ, we now provide the Approximation Lemma in full detail.

Lemma 2 (Approximation Lemma)**.** *Let A be a positive semidefinite matrix with its eigenvalues sorted in decreasing order denoted by μ_i. Let B be a positive definite matrix and let $B_{1,1}$ be the upper left matrix element expressed in the eigenbasis of A. Let $\gamma>0$. Then*

$$\lambda_1\left((A^\beta BA^\beta)^\gamma\right)=\mu_1^{2\beta\gamma}(B_{1,1})^\gamma\left(1+c(\mu_2/\mu_1)^{2\beta}\right)^\gamma,$$

for some constant value of c independent of $\gamma>0$.

In the proof of Theorem 2 we use the case $\gamma=r(1-\alpha)$, $\beta=\alpha/2\gamma$, $A=\rho$ and $B=\sigma^{-1/r}$. The limit of α going to 1 (from below, if $r<0$, or from above, if $r>0$) corresponds to the limit $\gamma\to0^+$, and we get that for α tending to 1, $\lambda_1\left((A^\beta BA^\beta)^{1-\alpha}\right)$ tends to $\mu_1^{2\beta\gamma}(B_{1,1})^\gamma=\mu_1^\alpha(B_{1,1})^{r(1-\alpha)}$ with an exponentially decreasing relative error $c\gamma\exp(-k/\gamma)$, with $k=|\log(\mu_2/\mu_1)|$, provided of course that $\mu_1>\mu_2$, strictly. This is because for $0\le x<1$ and very small γ the function $(1+cx^{1/\gamma})^\gamma$ can be approximated as

$$(1+cx^{1/\gamma})^\gamma\approx1+c\gamma\exp(-|\log x|/\gamma).$$

Proof. From the eigenvalue decomposition $A=\sum_{k=1}^d\mu_kP_k$ and the hypothesis $\mu_2<\mu_1$ we can write $A=\mu_1P_1+X$ with $0\le X\le\mu_2(I-P_1)$; note also that X is orthogonal to P_1. Thus,

$$\begin{aligned}\lambda_1\left(A^\beta BA^\beta\right) &= \lambda_1\left(B^{1/2}A^{2\beta}B^{1/2}\right)\\ &= \lambda_1\left(B^{1/2}(\mu_1^{2\beta}P_1+X^{2\beta})B^{1/2}\right).\end{aligned}$$

As the function that maps a Hermitian matrix to its largest eigenvalue is order-preserving and subadditive, this gives us

$$\begin{aligned}\lambda_1\left(A^\beta BA^\beta\right) &\ge \lambda_1\left(B^{1/2}\mu_1^{2\beta}P_1B^{1/2}\right)\\ &= \mu_1^{2\beta}B_{1,1}\end{aligned} \tag{22}$$

and

$$\begin{aligned}
\lambda_1\left(A^\beta B A^\beta\right) &\le \lambda_1\left(B^{1/2}(\mu_1^{2\beta}P_1+\mu_2^{2\beta}(I-P_1))B^{1/2}\right)\\
&\le \mu_1^{2\beta}\lambda_1\left(B^{1/2}P_1B^{1/2}\right)+\mu_2^{2\beta}\lambda_1\left(B^{1/2}(I-P_1)B^{1/2}\right)\\
&\le \mu_1^{2\beta}B_{1,1}+\mu_2^{2\beta}\lambda_1(B)\\
&= \mu_1^{2\beta}B_{1,1}\left(1+\frac{\lambda_1(B)}{B_{1,1}}(\mu_2/\mu_1)^{2\beta}\right). \qquad (23)
\end{aligned}$$

Since $B>0$, we have $B_{1,1}>0$ and the division can be done. Bracketing inequalities (22) and (23) can be combined as a single equality by introducing a constant c such that

$$\lambda_1\left(A^\beta B A^\beta\right)=\mu_1^{2\beta}B_{1,1}\left(1+c(\mu_2/\mu_1)^{2\beta}\right),$$

and imposing that c lies between 0 and $\lambda_1(B)/B_{1,1}$.

Raising all expressions to the (positive) power γ yields the equality of the lemma

$$\lambda_1\left((A^\beta B A^\beta)^\gamma\right)=\mu_1^{2\gamma\beta}(B_{1,1})^\gamma\left(1+c(\mu_2/\mu_1)^{2\beta}\right)^\gamma.$$

□

Let us now consider what happens when the spectrum of ρ is degenerate, and whether $D_{\alpha,r(\alpha-1)}(\rho||\sigma)$ is continuous in ρ and σ (with $\operatorname{supp}\rho\le\operatorname{supp}\sigma$) in the limit $\alpha\to1$. It is clear from the definition that it is continuous for all $\alpha\neq1$. Thus, if we can show that (20) has a continuous extension, one that includes degenerate ρ as well, then $D_{\alpha,r(\alpha-1)}$ is indeed continuous in the limit $\alpha\to1$.

Let us therefore consider (20) at face value (without looking back at the arguments that were used to derive it) and see whether it is even well-defined for degenerate ρ. This is not immediately clear because of the formula's non-trivial dependence on the eigenbasis of ρ: when the spectrum of ρ is degenerate, ρ has an infinity of allowed eigenbases, and the question arises whether the choice of basis affects the outcome. It turns out, however, that it does not, as the eigenvalue multiplicity 'both gives and takes', as explained below.

For the sake of concreteness, let us take a ρ for which μ_1 has multiplicity 2. Then P_1 is a 2-dimensional projector, and any pair of orthonormal vectors in the corresponding subspace can serve as basis elements. For every such basis, one gets a different matrix representation of σ. This can be recast as fixing one such representation of σ and letting a 2×2 unitary matrix U act on its upper left 2×2 block. Consequently, ν_1 depends on U whereas the other ν_i are independent from U, due to unitary invariance of the determinant. However, whereas this clearly affects the first two elements in the resulting

$$\widehat{\sigma}=\operatorname{diag}_\rho(\nu_1,\nu_2/\nu_1,\nu_3/\nu_2,\dots,\nu_d/\nu_{d-1}),$$

this is actually compensated for by the multiplicity of μ_1. The first two terms in the formula for $D(\rho||\widehat{\sigma})$ are

$$D(\rho||\widehat{\sigma})=\mu_1(\log\mu_1-\log\nu_1)+\mu_1(\log\mu_1-\log(\nu_2/\nu_1))+\cdots$$

and this simplifies to

$$D(\rho||\widehat{\sigma}) = 2\mu_1 \log \mu_1 - \mu_1 \log \nu_2 + \cdots$$

which is independent of ν_1.

One checks that this argument generalizes to all possible multiplicities. In fact, an equivalent formula for $D(\rho||\widehat{\sigma})$ is

$$D(\rho||\widehat{\sigma}) = -S(\rho) - \mu_d \log \det \sigma - \sum_{i=1}^{d-1} (\mu_i - \mu_{i+1}) \log \nu_i, \tag{24}$$

where $S(\rho) = -\operatorname{Tr} \rho \log \rho$ is the von Neumann entropy of ρ. The upshot is that $D(\rho||\widehat{\sigma})$ is independent of those elements ν_i that are dependent on a freedom of choice of basis caused by degeneracy of μ_i. This implies that $D(\rho||\widehat{\sigma})$ is continuous in ρ and σ since every term in (24) is continuous, as we now show. Indeed, the von Neumann entropy is well-known to be continuous (in the sense of Fannes), and μ_d and $\nu_d = \det \sigma$ are continuous as well since eigenvalues of a matrix depend continuously on the entries of a matrix ([28], Appendix D). The only potential problems stem from the terms $(\mu_i - \mu_{i+1}) \log \nu_i$ as they explicitly depend on the eigenprojections of ρ.

To see the problem, consider the example of a positive semidefinite matrix ρ parameterized by the variable x, $\rho(x) = \operatorname{diag}(1+x, 1-x)$, with $0 < |x| < 1$. Then for $x > 0$, $P_1 = \operatorname{diag}(1, 0)$ whereas for $x < 0$, $P_1 = \operatorname{diag}(0, 1)$. Thus for almost all σ, $\nu_1(x)$ has a discontinuity at $x = 0$. However, these discontinuities only occur at the so-called *exceptional points* of $\rho(x)$, the points where some eigenvalues coincide, a.k.a. level-crossings in physics terminology. This is because eigenprojections of Hermitian $\rho(x)$ are holomorphic functions of x ([29], Chapter II, Theorem 6.1). The discontinuities occur because the ordering of the eigenvalues changes at a level-crossing, and the eigenprojections get swapped accordingly, as in the example. The terms $(\mu_i - \mu_{i+1}) \log \nu_i$, however, remain continuous, since any level-crossing affecting ν_i occurs when the prefactor $\mu_i - \mu_{i+1}$ becomes zero, which cancels the discontinuity in ν_i (while still leaving a discontinuity in the derivative).

We have thus finally proven:

Theorem 3. *The statement from Theorem 2 still holds when the spectrum of ρ is degenerate, in the sense that (20) has to be interpreted as (24). The limit $\lim_{\alpha \to 1} D_{\alpha, r(\alpha-1)}(\rho||\sigma)$ exists as a continuous (but not necessarily smooth) function of ρ and σ.*

4 The case of infinite z

In this section we study the behaviour of $D_{\alpha,z}$ for z going to infinity. As in the previous section we first consider the parametrization $z = r(\alpha - 1)$, with $r > 0$, and take the limit of $D_{\alpha, r(\alpha-1)}$ as α tends to $+\infty$.

Noting that the operator norm is the limit of the Schatten q-norm as q tends to $+\infty$, we obtain from (4),

$$\begin{aligned}
\lim_{\alpha\to+\infty} D_{\alpha,r(\alpha-1)}(\rho||\sigma) &= \lim_{\alpha\to+\infty} \frac{1}{\alpha-1}\log\mathrm{Tr}(\rho^{\alpha/r(\alpha-1)}\sigma^{-1/r})^{r(\alpha-1)} \\
&= \lim_{\alpha\to+\infty} \log||(\rho^{\alpha/2r(\alpha-1)}\sigma^{-1/r}\rho^{\alpha/2r(\alpha-1)})^r||_{\alpha-1} \\
&= \log||(\rho^{1/2r}\sigma^{-1/r}\rho^{1/2r})^r||_\infty \\
&= \log||\rho^{1/2r}\sigma^{-1/r}\rho^{1/2r}||_\infty^r \\
&= r\log||\rho^{1/2r}\sigma^{-1/r}\rho^{1/2r}||_\infty.
\end{aligned}$$

Now the operator norm of a positive semidefinite matrix X equals the largest eigenvalue of X, which in turn is the smallest value of λ such that $X \le \lambda I$. In the present case, this condition is $\rho^{1/2r}\sigma^{-1/r}\rho^{1/2r} \le \lambda I$, which is equivalent to $\lambda\sigma^{1/r} \ge \rho^{1/r}$. Hence,

$$\log||\rho^{1/2r}\sigma^{-1/r}\rho^{1/2r}||_\infty = \log\min_\lambda\{\lambda : \lambda\sigma^{1/r} \ge \rho^{1/r}\} = D_{\max}(\rho^{1/r}||\sigma^{1/r}).$$

Thus we arrive at the following theorem:

Theorem 4. *Let ρ be a positive semidefinite matrix and let σ be positive definite. Then for a non-zero, finite real number r,*

$$\lim_{\alpha\to+\infty} D_{\alpha,r(\alpha-1)}(\rho||\sigma) = rD_{\max}(\rho^{1/r}||\sigma^{1/r}). \tag{25}$$

In particular, for $r = 1$

$$\lim_{\alpha\to+\infty} \widetilde{D}_\alpha(\rho||\sigma) = D_{\max}(\rho||\sigma).$$

For $\alpha \to -\infty$, which necessitates the stronger restriction on the supports $\mathrm{supp}\,\rho = \mathrm{supp}\,\sigma$, a similar treatment yields the result that for $r < 0$,

$$\lim_{\alpha\to-\infty} D_{\alpha,r(\alpha-1)}(\rho||\sigma) = rD_{\max}(\sigma^{-1/r}||\rho^{-1/r}) \tag{26}$$

and, for $r = -1$,

$$\lim_{\alpha\to-\infty} \widehat{D}_\alpha(\rho||\sigma) = -D_{\max}(\sigma||\rho). \tag{27}$$

Finally, we study the limit $z \to \infty$ when α is kept fixed (and finite). Let us first consider the case where $\mathrm{supp}\,\rho = \mathrm{supp}\,\sigma$. Using the well-known Lie-Trotter product formula (see, e.g. [27], Theorem IX.1.3), according to which $\lim_{m\to\infty}(\exp(A/m)\exp(B/m))^m = \exp(A+B)$ for any two matrices A and B, we easily obtain (with $A = \log\rho^\alpha$ and $B = \log\sigma^{1-\alpha}$), for $\alpha \ne 1$,

$$\lim_{z\to\infty} D_{\alpha,z}(\rho||\sigma) = \frac{1}{\alpha-1}\log\mathrm{Tr}\exp(\alpha\log\rho + (1-\alpha)\log\sigma). \tag{28}$$

In the limit $\alpha \to 1$, we use l'Hôpital's rule and the fact that $(d/d\alpha)\,\mathrm{Tr}\exp(X+\alpha Y) = \mathrm{Tr}\,Y\exp(X+\alpha Y)$ to obtain

$$\lim_{\alpha\to1}\lim_{z\to\infty} D_{\alpha,z}(\rho||\sigma) = D(\rho||\sigma). \tag{29}$$

When $\mathrm{supp}\,\rho$ is a proper subset of $\mathrm{supp}\,\sigma$, the same formulas hold except for the fact that we have to restrict $\log\sigma$ to $\mathrm{supp}\,\rho$ (more generally, both $\log\rho$ and $\log\sigma$ have to be restricted to the intersection of the supports of ρ and σ). This was proven by Hiai and Petz in [21].

After the first draft [34] of the present paper had been circulated, Lin and Tomamichel have shown [35] that the relative entropy is recovered more generally when α goes to 1 and z is taken to be $z = g(\alpha)$, for any continuously differentiable function g such that $g(1) \ne 0$.

5 Limiting case $z = \alpha \to 0^+$

In this section, we answer the question: what is the limit of $\widetilde{D}_\alpha$ as α tends to 0; that is, what is

$$\lim_{\alpha\to 0} D_{\alpha,\alpha}(\rho||\sigma) = -\log \lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma)?$$

As always, we assume that σ is full rank. We will also assume first that the spectrum of σ is non-degenerate.

The answer to this question is easy when ρ and σ commute. Choosing a basis in which both states are diagonal, with diagonal elements given by ρ_i and σ_i, respectively, the limit is given by

$$\begin{aligned}\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) &= \lim_{\alpha\to 0} \sum_{i=1}^d \rho_i^\alpha \sigma_i^{1-\alpha} \\ &= \sum_i \sigma_i : \rho_i \neq 0.\end{aligned}$$

In terms of the projector on the support of ρ, which we denote by Π_ρ, we write this as

$$\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) = \operatorname{Tr} \Pi_\rho \sigma.$$

To answer the question in the general case, we will first show that the answer does not depend on ρ itself, but only on Π_ρ, and of course also on σ. To do so, we consider the particular expression

$$\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) = \lim_{\alpha\to 0} \operatorname{Tr}(\sigma^{1/2\alpha}\rho\sigma^{1/2\alpha})^\alpha.$$

Let μ be the smallest non-zero eigenvalue of ρ. Then we have the inclusion $\mu\Pi_\rho \leq \rho \leq \Pi_\rho$. This implies

$$\mu^\alpha \operatorname{Tr}(\sigma^{1/2\alpha}\Pi_\rho\sigma^{1/2\alpha})^\alpha \leq \operatorname{Tr}(\sigma^{1/2\alpha}\rho\sigma^{1/2\alpha})^\alpha \leq \operatorname{Tr}(\sigma^{1/2\alpha}\Pi_\rho\sigma^{1/2\alpha})^\alpha.$$

In the limit of $\alpha \to 0$, μ^α of course tends to 1, so that both sides of the inclusion become equal and we have the identity

$$\lim_{\alpha\to 0} \operatorname{Tr}(\sigma^{1/2\alpha}\rho\sigma^{1/2\alpha})^\alpha = \lim_{\alpha\to 0} \operatorname{Tr}(\sigma^{1/2\alpha}\Pi_\rho\sigma^{1/2\alpha})^\alpha.$$

For the remainder of the argument, we will work in a basis in which Π_ρ is diagonal, and given by $I_r \oplus 0$, where r is the rank of ρ. Furthermore, we switch from one representation of $f_{\alpha,\alpha}$ to another, namely

$$\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) = \lim_{\alpha\to 0} \operatorname{Tr}(\Pi_\rho\sigma^{1/\alpha}\Pi_\rho)^\alpha.$$

We will also employ the spectral decomposition of σ, which we consider to be given by

$$\sigma = U\Lambda U^* = \sum_{i=1}^d \lambda_i |u_i\rangle\langle u_i|,$$

where the eigenvalues are sorted in descending order as $\lambda_1 > \lambda_2 > \cdots > \lambda_d$. To deal with the expression $\Pi_\rho \sigma^{1/\alpha} \Pi_\rho$, we will finally define the restriction of the eigenvectors to the support of ρ:

$$|u_i\rangle \mapsto |\widetilde{u}_i\rangle := \Pi_\rho |u_i\rangle.$$

With this definition, we have

$$\Pi_\rho \sigma^{1/\alpha} \Pi_\rho = \sum_{i=1}^{d} \lambda_i^{1/\alpha} |\widetilde{u}_i\rangle\langle\widetilde{u}_i|.$$

It goes without saying that the vectors $|\widetilde{u}_i\rangle$ in general no longer form an orthonormal set, and the quantities $\lambda_i^{1/\alpha}$ are not eigenvalues of $\Pi_\rho \sigma^{1/\alpha} \Pi_\rho$.

Let us first try and find an expression for the largest eigenvalue μ_1 of $Z_\alpha := (\Pi_\rho \sigma^{1/\alpha} \Pi_\rho)^\alpha$ in the limit $\alpha \to 0^+$. Given that the spectrum of σ is non-degenerate, the main contribution to $\Pi_\rho \sigma^{1/\alpha} \Pi_\rho$ as $\alpha \to 0^+$ will come from λ_1, and is given by $\lambda_1^{1/\alpha} |\widetilde{u}_1\rangle\langle\widetilde{u}_1|$. That is true, of course, only if $|\widetilde{u}_1\rangle$ is not the zero vector ($|\widetilde{u}_1\rangle = 0$ if $|u_1\rangle$ lies outside the support of ρ). We therefore have to correct our statement and say: the main contribution to $\Pi_\rho \sigma^{1/\alpha} \Pi_\rho$ will come from λ_{i_1}, and is given by $\lambda_{i_1}^{1/\alpha} |\widetilde{u}_{i_1}\rangle\langle\widetilde{u}_{i_1}|$, where i_1 is the first index value for which $|\widetilde{u}_i\rangle \neq 0$. The limit can now be calculated easily, and we get

$$\mu_1 = \lim_{\alpha \to 0} \lambda_{i_1} || \, |\widetilde{u}_{i_1}\rangle\langle\widetilde{u}_{i_1}| \, ||^\alpha = \lambda_{i_1} = \max_{i_1} \lambda_{i_1} : |\widetilde{u}_{i_1}\rangle \neq 0$$

Next, we calculate the product of the two largest eigenvalues of Z_α, $\mu_1\mu_2$, in the limit $\alpha \to 0^+$. Using the Weyl-trick, this reduces to the largest eigenvalue of the second antisymmetric tensor power, and using the formula just obtained we find

$$\mu_1\mu_2 = \max_{i_1,i_2} \lambda_{i_1}\lambda_{i_2} : |\widetilde{u}_{i_1}\rangle \wedge |\widetilde{u}_{i_2}\rangle \neq 0.$$

The latter condition amounts to the two vectors $|\widetilde{u}_{i_1}\rangle$ and $|\widetilde{u}_{i_2}\rangle$ being linearly independent. For $\mu_1\mu_2\mu_3$ we similarly obtain

$$\mu_1\mu_2\mu_3 = \max_{i_1,i_2,i_3} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} : |\widetilde{u}_{i_1}\rangle, |\widetilde{u}_{i_2}\rangle, |\widetilde{u}_{i_3}\rangle \text{ linearly independent},$$

and so on either until $\mu_1\mu_2\cdots\mu_r$ has been obtained, or no further linearly independent vectors can be added to the set. That is, the process stops at $\mu_1\mu_2\cdots\mu_s$, where s is the rank of $\Pi_\rho\sigma$ (clearly, $s \leq r$).

By successive divisions we then find the separate μ_i, for $i = 1, 2, \ldots, s$. What we are after is the sum of these μ_i, and this sum is simply given by

$$\sum_{i=1}^{s} \mu_i = \max_{i_1,i_2,\ldots,i_s} \sum_{j=1}^{s} \lambda_{i_j} : \{|\widetilde{u}_{i_j}\rangle\} \text{ linearly independent}.$$

A convenient way to find these linearly independent vectors is to use Gaussian elimination, under the guise of the Row-Echelon normal Form (REF) procedure (well-known from any introductory Linear Algebra course). The indices i_j of the formula are the column indices of those columns that contain a row-leading entry (that is, the first non-zero entry in some row) in the row-echelon normal form of the matrix $\Pi_\rho U$.

We have therefore proven:

$$\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) = \sum_{j=1}^{s} \lambda_{i_j}, \tag{30}$$

where the λ_i are the eigenvalues of σ, and the indices i_j can be found from the following procedure: calculate the row-echelon form R of the matrix $\Pi_\rho U$ (expressed in an eigenbasis of ρ). For every row of R, determine at which column the first non-zero entry appears; these column indices are the sought values of i_j and s is the number of non-zero rows in R.

The result just obtained still holds in the case when the spectrum of σ is degenerate. Suppose a certain eigenvalue of σ has multiplicity k. Let S be the subspace that is the projection of this k-dimensional eigenspace to the support of ρ. The problem is that one can choose among an infinite number of bases for S; which basis contains the highest number of vectors that are independent from the u_{i_j} that we already had? The answer is simple: that number is really basis independent and only depends on the dimension of the intersection of S with the subspace P spanned by these u_{i_j}. Thus any basis should do, and the formula remains as it stands.

We finish this section with a simple example of the procedure just described. Let ρ and σ be 4-dimensional states where σ is full rank and has non-degenerate spectrum, and ρ has rank 2. In terms of the eigenbasis of ρ, the projector Π_ρ is represented by the diagonal matrix $\Pi_\rho = \mathrm{diag}(1,1,0,0)$. Furthermore, let σ have spectral decomposition $\sigma = \sum_{i=1}^4 \lambda_i |u_i\rangle\langle u_i|$ where the eigenvectors $|u_i\rangle$ are the columns of the unitary matrix

$$U = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Thus, the matrix $\Pi_\rho U$ (after deleting the rows that are completely zero) and its REF are given by

$$\Pi_\rho U = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \text{ and } \mathrm{REF}(\Pi_\rho U) = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & -2 & -2 \end{pmatrix}.$$

The row-leader of row 1 is in column 1, and the one of row 2 is in column 3. Therefore, we put $i_1 = 1$ and $i_2 = 3$, so that the value of $\lim_{\alpha\to 0} f_{\alpha,\alpha}(\rho||\sigma) = \sum_{i=1}^s \mu_i$ is given by $\lambda_1 + \lambda_3$.

6 Discussion

In this paper we studied a two-parameter family of relative entropies, which we call the α-z-Rényi relative entropies (α-z-RRE), from which all other known relative entropies (or divergences) can be derived. This family provides a unifying framework for the analysis of properties of the different relative entropies arising in quantum information theory, such as the quantum relative entropy, the α-quantum Rényi divergences (α-QRD), and the α-quantum Rényi relative entropies. We have shown that the α-z-RRE satisfies the data-processing inequality (DPI) for suitable values of the parameters α and z.

The α-QRD (or sandwiched Rényi relative entropy), which is a special case of the α-z-RRE, has been the focus of much research of late. We have studied another special case of the α-z-RRE, which we denote as $\widehat{D}_\alpha$ (and informally call the *reverse sandwiched Rényi relative entropy*). It satisfies DPI for $\alpha \leq 1/2$, and we obtain an interesting closed expression for it in the limit $\alpha \to 1$.

Our analysis leads to some interesting open questions: (i) Does the α-z-RRE satisfy DPI in the orange regions of the α-z-plane of Figure 1? In other words, is the trace functional of the α-z-RRE convex in the orange regions of Figure 2? (ii) Operational relevance of the α-QRD for $\alpha \geq 1$ has been established in quantum hypothesis testing [32], and in the context of the second laws of quantum thermodynamics [33]. Does $\widehat{D}_\alpha$ also have operational interpretations in quantum information theory (for $0 \leq \alpha \leq 1/2$) (other than those arising through the symmetry relation (10))?

Acknowledgments

We are grateful to William Wootters for suggesting the idea of looking at "reverse sandwiching". This project was initiated during the program "Mathematical Challenges in Quantum Information" at the Isaac Newton Institute, Cambridge, UK, and was supported in part by an Odysseus grant from the Flemish Fund for Scientific Research (FWO). We thank the anonymous referee for the many comments that have greatly improved the clarity of the paper.

References

[1] H. Umegaki. Conditional expectation in an operator algebra, IV (entropy and information). *Kodai Math. Sem. Rep.* 14, 59–85 (1962).

[2] F. Hiai and D. Petz. The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability. *Commun. Math. Phys.* 143, 99–114 (1991).

[3] M. Hayashi. Large deviation analysis for quantum security via smoothing of Renyi entropy of order 2. arXiv:1202.0322 (2013).

[4] P.J. Coles, R. Colbeck, L. Yu and M. Zwolak. Uncertainty relations from simple entropic properties. *Phys. Rev. Lett.* 108, 210405 (2012).

[5] K.M.R. Audenaert, J. Calsamiglia, R. Munoz-Tapia, E. Bagan, Ll. Masanes, A. Acin and F. Verstraete. *Phys. Rev. Lett.* 98, 160501 (2007).

[6] M. Mosonyi and F. Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Trans. Inf. Th.*, 57, 2474–2487, (2011).

[7] F. Dupuis, L. Kraemer, P. Faist, J.M. Renes and R. Renner. Generalized entropies. *Proc. XVIIth Int'l Conf. on Mathematical Physics*, Aalborg (2012). See also arXiv:1211.3141.

[8] N. Datta. Min- and Max- Relative Entropies and a New Entanglement Monotone. *IEEE Trans. Inf. Th.*, 55, 2816–2826 (2009).

[9] L. Wang and R. Renner. One-Shot Classical-Quantum Capacity and Hypothesis Testing. *Phys. Rev. Lett.* 108, 200501 (2012).

[10] F. Buscemi and N. Datta. Entanglement cost in practical scenarios. *Phys. Rev. Lett.* 106, 130503 (2011).

[11] M. Tomamichel. Smooth entropies — a tutorial, with focus on applications in cryptography. Tutorial at QCRYPT 2012, September 2012.

[12] S. Fehr. On the conditional Rényi entropy. Lecture at the Beyond IID Workshop, University of Cambridge, January 2013.

[13] M.M. Wilde, A. Winter, D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels. *Commun. Math. Phys.* 331(2), 593–622 (2014).

[14] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel. On quantum Renyi entropies: a new definition and some properties. *J. Math. Phys.* 54, 122203 (2013).

[15] E.H. Lieb. Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture. *Adv. in Math* 11, 267–288 (1973).

[16] R.L. Frank and E.H. Lieb. Monotonicity of a relative Rényi entropy. *J. Math. Phys.* 54, 122201 (2013).

[17] S. Beigi. Quantum Rényi Divergence Satisfies Data Processing Inequality. *J. Math. Phys.* 54, 122202 (2013).

[18] N. Datta and F. Leditzky. A limit of the quantum Rényi divergence. *J. Phys. A: Math. Theor.* 47, 045304 (2014).

[19] V. Jakšić, Y. Ogata, Y. Pautrat, C.-A. Pillet, *Entropic fluctuations in quantum statistical mechanics. An Introduction.* In: Quantum Theory from Small to Large Scales: Lecture Notes of the Les Houches Summer School: Volume 95, August 2010, Oxford University Press (2012).

[20] A. Rényi. On measures of entropy and information. *Proc. 4th Berkeley Symp. on Math. Statist. and Probability*, Vol. 1, 547–561, University of California Press (1961). Available online at `http://projecteuclid.org/euclid.bsmsp/1200512181`.

[21] F. Hiai and D. Petz. The Golden-Thompson Trace Inequality is Complemented. *Linear Algebra Appl.* 181, 153–185 (1993).

[22] F. Hiai. Concavity of certain matrix trace and norm functions. *Linear Algebra Appl.* 439(5), 1568–1589 (2013).

[23] F. Hiai. Concavity of certain matrix trace functions. *Taiwanese J. Math.* 5, 535–554 (2001).

[24] T. Ando. Convexity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra Appl.* 26, 203–241 (1979).

[25] K.M.R. Audenaert and F. Hiai. Work in progress.

[26] G.H. Golub and C.F. Van Loan, *Matrix computations*, Johns Hopkins University Press (1983).

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

[28] R.A. Horn and C.R. Johnson. *Matrix Analysis*, Cambridge (1985).

[29] T. Kato. *Perturbation theory for linear operators*, Springer Verlag (1980).

[30] H. Epstein. Remarks on two theorems of Lieb. *Commun. Math. Phys.* 31, 317–325 (1973).

[31] W.F. Donoghue Jr. *Monotone matrix functions and analytic continuation*, Springer Verlag (1974).

[32] M. Mosonyi and T. Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. arXiv:1309.3228 (2014).

[33] F.G.S.L. Brandao, M. Horodecki, N.H.Y. Ng, J. Oppenheim and S. Wehner. The second laws of quantum thermodynamics. arXiv:1305.5278 (2014).

[34] K.M.R. Audenaert and N. Datta. arXiv:1313.7178 (2014).

[35] M.S. Lin and M. Tomamichel. Investigating Properties of a Family of Quantum Rényi Divergences. arXiv:1408.6897 (2014).

[36] M. Tomamichel and M. Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Trans. Inf. Theory* 59(11), 7693–7710 (2013).

[37] K. Li. Second-order asymptotics for quantum hypothesis testing. *Ann. Stat.* 42(1), 171–189 (2014).

[38] E.A. Carlen, R.L. Frank and E.H. Lieb. Some Operator and Trace Function Convexity Theorems. arXiv:1409.0564 (2014).

[39] M.B. Ruskai. Inequalities for quantum entropy: a review with conditions for equality. *J. Math. Phys.* 43, 4359–4376 (2002); M.B. Ruskai. Erratum. *J. Math. Phys.* 46, 019901 (2005).