\[ \alpha-z\text{-relative Rényi entropies} \]

Koenraad M.R. Audenaert\textsuperscript{1,2} and Nilanjana Datta\textsuperscript{3}

\textsuperscript{1}Department of Mathematics, Royal Holloway University of London, Egham TW20 0EX, U.K.

\textsuperscript{2}Department of Physics and Astronomy, University of Ghent, S9, Krijgslaan 281, B-9000 Gent, Belgium

\textsuperscript{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 relative Rényi entropies that unifies the study of all known relative entropies (or divergences). These include the quantum relative entropy, the recently defined quantum Rényi divergences, as well as the original quantum relative Rényi entropies. Consequently, the data-processing inequality for all these quantities follows directly from the data-processing inequality for these new relative entropies, the proof of which is one of the main technical contributions of this paper. These new relative Rényi entropies stem from the quantum entropic functionals defined by Jakšić et al. in the context of non-equilibrium statistical mechanics, and they satisfy the quantum generalizations of Rényi’s axioms for a divergence. We investigate a variety of limiting cases for the two parameters, obtaining explicit formulas for each one of them.

1 Introduction

The quantum relative entropy plays a central role in quantum information theory. In particular, fundamental limits on the performance of information-processing tasks (such as compression and transmission of information, and manipulation of entanglement) in the so-called “asymptotic, memoryless (or i.i.d.) setting\textsuperscript{1}” is given in terms of quantities derived from the quantum relative entropy. For example, the von Neumann entropy (which characterizes the data compression limit for an information source)\textsuperscript{2}, the Holevo quantity (which characterizes the classical capacity of a quantum channel)\textsuperscript{2}, and the coherent information (which characterizes the quantum capacity of a quantum channel)\textsuperscript{3}, can all be derived from the quantum relative entropy.

There are, however, several other entropic quantities and generalized relative entropies (or divergences) which are also of important operational significance in quantum information theory. One of the most important of these is the family of relative entropies called the $\alpha$-relative Rényi entropies ($\alpha$-RRE\textsuperscript{2} $D_\alpha(\rho \| \sigma)$ (with $\alpha \in (0, 1) \cup (1, \infty)$), and quantities derived

\textsuperscript{1}In this setting, the resources employed in these tasks are assumed to be memoryless, and to be available for asymptotically many independent uses.

\textsuperscript{2}We would like to point out a peculiar benefit of the acronym RRE: it treats relative Rényi entropies and Rényi relative entropies alike.
from them. For $\alpha \in (0, 1)$ these relative entropies arise in the quantum Chernoff bound [1] which characterizes the probability of error in discriminating two different quantum states in the setting of asymptotically many copies. In fact, in analogy with the operational interpretation of their classical counterparts, the $\alpha$-RRE can be viewed as generalized cutoff rates in quantum binary state discrimination [5]. The Rényi entropy of order $\alpha$, which is derived from the $\alpha$-RRE, also has various interesting operational meanings: e.g. in the context of quantum communication [6] and in determining whether a bipartite state is separable [7].

Another important class of entropies are the smooth min- and max-entropies [8] and the generalized relative entropies (namely the min- and max-relative entropies) from which they are derived [9, 10]. They play a pivotal role in “one-shot quantum information theory” which concerns the analysis of information-processing tasks in a more realistic scenario – one in which the assumption of an asymptotic i.i.d. setting is lifted. (see e.g. [11] and references therein). The 0-RRE also arises in the one-shot setting, in the context of binary hypothesis testing and transmission of information through a classical-quantum channel [12], and in the analysis of one-shot entanglement manipulation under local operations and classical communication [13]. The above list of entropic quantities is not exhaustive, quasi-entropy [14], skew divergence [15], Tsallis entropy [16] and subentropy [17] being some of the others.

In summary, there is a plethora of different entropic quantities which arise in quantum information theory, which are interesting both from the mathematical and operational points of view. It is hence desirable to have a unifying mathematical framework for the study of these different quantities.

Recently, a non-commutative generalization of the $\alpha$-relative Rényi entropy was defined that partially provided such a framework. Known alternatively as the quantum Rényi divergence (\(\alpha\)-QRD) or “sandwiched” relative Rényi entropy, it depends on a parameter $\alpha \in (0, 1) \cup (1, \infty)$ [18, 19, 20, 21]. For two positive semidefinite operators $\rho$ and $\sigma$ we denote it as $\tilde{D}_\alpha(\rho||\sigma)$. It has been proved that this quantity reduces 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 $\alpha \to \infty$. Consequently, many properties of the min-, max- and quantum relative entropies can be inferred directly from those of the $\alpha$-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 $\tilde{D}_\alpha(\rho||\sigma)$ for $\alpha \geq 1/2$ [22, 23]. 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 $\tilde{D}_\alpha(\rho||\sigma)$ is monotonically increasing in $\alpha$ [21]. Also joint convexity of the min- and quantum relative entropies is implied by the joint convexity of $\tilde{D}_\alpha(\rho||\sigma)$ for $1/2 \leq \alpha \leq 1$ [22].

In spite of these and various other interesting properties, which have been proved using a variety of sophisticated mathematical tools, the $\alpha$-QRD 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 $\alpha \in (0, 1/2)$ [21, 24], and (ii) the important $\alpha$-RRE family can only be obtained from the $\alpha$-QRD family in the special case of commuting operators. This leads us to the following question:

(Q) Can one define a more general family of relative entropies which overcomes these limitations?

In this paper we answer this question in the affirmative. The corresponding relative entropies
constitute a two-parameter family. We refer to them as \( \alpha-z \)-relative Rényi entropies (\( \alpha-z \)-RRE), and denote them as \( D_{\alpha,z}(\rho||\sigma) \), with \( \alpha \) and \( z \) being two real parameters. They stem from quantum entropic functionals defined by Jakšić et al. \[25\] 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, \( \rho \) was the reference state of a dynamical system, and \( \sigma \) was the state \( \rho_t \) resulting from \( \rho \) due to time evolution under the action of a Hamiltonian for a time \( t \). We define \( D_{\alpha,z}(\rho||\sigma) \) for arbitrary positive semidefinite states \( \rho \) and \( \sigma \), and study its properties from a quantum information theoretic perspective.

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 \( \alpha-z \)-RRE satisfies the quantum generalizations of Rényi’s axioms for a divergence, and describe the regions in the \( \alpha-z \) plane where these entropies satisfy the data-processing inequality. We study a special case of the \( \alpha-z \)-RRE, which we denote as \( \hat{D}_{\alpha} \) (and informally call the reverse sandwiched relative Rényi entropy) due to its similarities with the \( \alpha \)-QRD (or “sandwiched” relative Rényi 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 \( \alpha-z \)-RRE. In Section 6 we give a self-contained proof of the concavity/convexity of the trace functional arising in the definition of these new relative entropies. We end the paper with a brief summary of our results and some open questions in Section 7.

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

### 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 \( \text{Tr} \rho = 1 \). Further, we denote the support of an operator \( \rho \) by \( \text{supp} \rho \). 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 \( \alpha-z \)-relative Rényi (\( \alpha-z \)-RRE) entropies; \( \forall \rho \in \mathcal{D}(\mathcal{H}), \sigma \in \mathcal{P}(\mathcal{H}) \) with \( \text{supp} \rho \subseteq \text{supp} \sigma \)

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

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

\[
f_{\alpha,z}(\rho||\sigma) := \text{Tr} \left( \rho^{\alpha/z} \sigma^{(1-\alpha)/z} \right)^z = \text{Tr} \left( \rho^{\alpha/2z} \sigma^{(1-\alpha)/z} \rho^{\alpha/2z} \right)^z = \text{Tr} \left( \sigma^{(1-\alpha)/2z} \rho^{\alpha/z} \sigma^{(1-\alpha)/2z} \right)^z.
\]
Here, \( \alpha \in \mathbb{R} \) and the limit has to be taken for \( \alpha \) tending to 1, and \( z \in \mathbb{R} \) and the limit has to be taken for \( z \) tending to 0. The above definition is easily extended to the case in which \( \rho \geq 0 \) but \( \text{Tr} \rho \neq 1 \) (see (12)). Note that \( D_{\alpha,z} \) is even in \( z \): \( D_{\alpha,z}(\rho||\sigma) = D_{\alpha,-z}(\rho||\sigma) \).

For commuting \( \rho \) and \( \sigma \) this reduces to the classical \( \alpha \)-relative Rényi entropy, for all values of \( z \). Note also the symmetry

\[
(\alpha - 1)D_{\alpha,z}(\rho||\sigma) = (-\alpha)D_{1-z,1}(\sigma||\rho). \tag{5}
\]

Clearly, this family includes the \( \alpha \)-RRE family:

\[
D_{\alpha}(\rho||\sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \rho^\alpha \sigma^{1-\alpha} \right) = D_{\alpha,1}(\rho||\sigma), \tag{6}
\]

and the \( \alpha \)-QRD family:

\[
\tilde{D}_{\alpha}(\rho||\sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{1-\alpha} \rho \sigma^{1-\alpha} \right) = D_{\alpha,\alpha}(\rho||\sigma). \tag{7}
\]

Specifically, we get the known correspondences

\[
D_{\text{min}} = D_{1/2,1/2}, \quad D = \lim_{\alpha \to 1} D_{\alpha,\alpha}, \quad \text{and} \quad D_{\text{max}} = \lim_{\alpha \to \infty} D_{\alpha,\alpha}. \tag{8}
\]

Here \( D_{\text{min}} \), \( D \) and \( D_{\text{max}} \) denote the min-relative entropy \([9]\), the quantum relative entropy and the max-relative entropy \([10]\), respectively:

\[
D_{\text{min}}(\rho||\sigma) := -2 \log F(\rho, \sigma), \quad \text{where} \quad F(\rho, \sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1;
\]

\[
D(\rho||\sigma) := \text{Tr} \rho \log \rho - \text{Tr} \rho \log \sigma;
\]

\[
D_{\text{max}}(\rho||\sigma) := \inf\{\gamma : \rho \leq 2^\gamma \sigma\}. \tag{9}
\]

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

\[
\hat{D}_{\alpha}(\rho||\sigma) = \frac{1}{\alpha - 1} \log \text{Tr} \left( \rho^{\frac{\alpha}{2(1-\alpha)}} \sigma \rho^{\frac{\alpha}{2(1-\alpha)}} \right)^{1-\alpha} = D_{\alpha,1-\alpha}(\rho||\sigma). \tag{10}
\]

From (5) we immediately obtain the symmetry relation

\[
(\alpha - 1)\hat{D}_{\alpha}(\rho||\sigma) = (-\alpha)\hat{D}_{1-\alpha}(\sigma||\rho). \tag{11}
\]

For \( \alpha = 0 \), \( \hat{D}_{\alpha} \) reduces to the 0-relative Rényi entropy, a quantity of particular operational relevance in one-shot information theory \([12, 13]\). This is in contrast to the quantum Rényi divergence, which does not in general reduce to the 0-relative Rényi entropy in the limit \( \alpha \to 0 \) \([24]\).

All these correspondences are illustrated in Figure 1.
Figure 1: Schematic overview of the relative entropies that are unified by $D_{\alpha,z}$. The blue region is where the Data Processing Inequality (DPI) has been proven to hold, and the orange region is where we conjecture validity of DPI. Outside these two regions DPI does not hold.

2.1 Axiomatic properties

Following [21], we can check whether the $\alpha$-z-RRE satisfies the six quantum Rényi axioms, as does the $\alpha$-RRE and $\alpha$-QRD. These are quantum generalizations of axioms that were put forward by Rényi in [26] as natural requirements that any classical divergence should satisfy.

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

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

(I) **Continuity:** For $\rho \neq 0$ and $\text{supp} \rho \subseteq \text{supp} \sigma$, $D_{\alpha,z}(\rho||\sigma)$ is continuous in $\rho, \sigma \geq 0$ throughout the parameter space except for $\alpha \leq 0$. At $\alpha = 0$, the $\alpha$-RRE is dependent on the rank of $\rho$ and is therefore not continuous. This was actually the reason why

---

*A quantum divergence is a functional which maps a pair of positive semidefinite operators $\rho, \sigma$, with $\text{supp} \rho \subseteq \text{supp} \sigma$ onto $\mathbb{R}$. Its classical counterpart is obtained by replacing the operators by probability distributions.*
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 ([26], 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:** \( D_{\alpha,z}(I||I) = \log 2 \), as is the case for any divergence that reduces to the classical relative Rényi entropy for commuting arguments.

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

\[
D_{\alpha,z}(\rho||\sigma) \begin{cases}
\leq \quad \text{whenever } \rho \begin{cases} 
\leq \\
\geq 
\end{cases} \sigma \\

\geq \quad \text{whenever } \rho \begin{cases} 
\geq \\
\leq 
\end{cases} \sigma
\end{cases}
\]

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

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

\[
f_{\alpha,z}(\rho||\sigma) \begin{cases}
\leq \quad \text{whenever } \rho \begin{cases} 
\leq \\
\geq 
\end{cases} \sigma
\end{cases}
\]

whereas, for \( 0 < \alpha < 1 \),

\[
f_{\alpha,z}(\rho||\sigma) \begin{cases}
\geq \quad \text{whenever } \rho \begin{cases} 
\geq \\
\leq 
\end{cases} \sigma
\end{cases}
\]

This holds if the fractional power \( (1 - \alpha)/z \) that is applied to \( \sigma \) in (3) 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

\[
(\text{Tr} \rho + \text{Tr} \tau) \ g(D_{\alpha,z}(\rho \oplus \tau||\sigma \oplus \omega)) = (\text{Tr} \rho) \ g(D_{\alpha,z}(\rho||\sigma)) + (\text{Tr} \tau) \ g(D_{\alpha,z}(\tau||\omega)).
\]

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 the context of Rényi’s axioms, only the case $\text{Tr} \rho + \text{Tr} \tau \leq 1$ and $\text{Tr} \sigma + \text{Tr} \omega \leq 1$ 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 $\text{supp} \rho \subseteq \text{supp} \sigma$, and for any CPTP map $\Lambda$ acting on $\mathcal{P}(\mathcal{H})$,

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

for $0 < \alpha \leq 1$ and $z \geq \max(\alpha, 1 - \alpha)$ (Hiai), or for $1 \leq \alpha \leq 2$ and $z = 1$ (Ando), or for $1 \leq \alpha$ and $z = \alpha$ (Frank and Lieb, Beigi).

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. [22], 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

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

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 [27] (see also the older work [28]). For the sake of exposition an alternative proof of this result is given in Section 6. Convexity was proven by Frank and Lieb [22] and independently by Beigi [23] for the case $1 \leq \alpha$ and $z = \alpha$, where $D_{\alpha,z}$ reduces to the $\alpha$-QRD $\tilde{D}_{\alpha}$. Convexity for $1 \leq \alpha \leq 2$ and $z = 1$ is exactly Ando’s theorem [29].

Hiai [27] 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 there as well.

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

### 2.3 Limiting cases

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

To study (i) we suitably parameterized $z$ in terms of $\alpha$ as $z = r(\alpha - 1)$, where $r$ is a non-zero finite real number, and considered the limit $\alpha \to 1$ (the case of fixed $\alpha \neq 1$ and $z \to 0$ will be studied elsewhere [30]). Note that $\alpha = 1$ is the only point on the $\alpha$-axis where the Order Axiom (IV) is satisfied. For the choice $z = 1 - \alpha$, this yields the limit $\alpha \to 1$ of $\hat{D}_\alpha(\rho||\sigma)$. In the general case in which $\rho$ and $\sigma$ do not commute, we obtain a rather surprising formula for the latter: the relative entropy, not between $\rho$ and $\sigma$, but between $\rho$ and an operator $\hat{\sigma}$ that is diagonal in the eigenbasis of $\rho$ (see Theorems 2 and 3 for details).

In the commuting case we recover the expected expression: the relative entropy of $\rho$ and $\sigma$.

We also prove that the $\alpha$-$z$-RRE is continuous in $\rho$ and $\sigma$ in that limit.

To study the case (ii) of infinite $\alpha$ and $z$, we use the same parametrization of $z$, and take the limit $\alpha \to \infty$. In this limit the $\alpha$-$z$-RRE is expressed in terms of a max-relative entropy (see Theorem 4 for details). In particular, our result readily yields the known [21] result that in the limit $\alpha \to \infty$, the $\alpha$-QRD, $\tilde{D}_\alpha(\rho||\sigma)$, reduces to the max-relative entropy $D_{\text{max}}(\rho||\sigma)$.

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

Finally, we consider the case (iv) where $\alpha$ 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 the point $(\alpha, z) = (1, 0)$ because that is the only point on the $\alpha$-axis where the Order Axiom is satisfied. It turns out that, in this limit, $D_{\alpha,z}$ is discontinuous in $\alpha$ at that point 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 $\lim_{\alpha \to 1} \tilde{D}_\alpha$.

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

**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

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

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 $\lim_{\alpha \to 1} f_{\alpha,r}(\alpha - 1) = \text{Tr} \rho = 1$. Hence, as $\alpha$ 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 $\alpha$. However, this approach did not yield any simplification. Instead, we followed a completely different approach, inspired by the power method \cite{31} for numerically calculating eigenvalues.

We first consider the generic case in which the spectrum of $\rho$ is non-degenerate, i.e. all its eigenvalues are distinct. Let us write the spectral decomposition of $\rho$ as $\rho = \sum_{i=1}^{d} \mu_i P_i$, where the eigenvalues $\mu_i$ 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 $s$, $\rho^s$ can be well-approximated by $\mu_i^s P_i$, 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 $\mu_1^s$.

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 $\lambda_1$ only. We get (noting that $z = r(\alpha - 1)$ is always taken positive in our considerations)

$$\lambda_1(Z_{\alpha,r}(\rho \| \sigma)) = \lambda_1 \left((\rho^{\alpha/2r(\alpha-1)} \sigma^{-1/r} \rho^{\alpha/2r(\alpha-1)})^{r(\alpha-1)}\right)$$

$$\approx \mu_1^\alpha \text{Tr}(P_1 \sigma^{-1/r} P_1)^{r(\alpha-1)}$$

$$= \mu_1^\alpha \left((\sigma^{-1/r})_{1,1}\right)^{r(\alpha-1)},$$

where $X_{1,1}$ indicates the upper left matrix element of a matrix $X$ in the eigenbasis of $\rho$. This is shown in full rigor in Lemma \cite{2} below; the lemma applies to the case at hand via the substitution $A = \rho^{-1/r}$ and $B = \sigma^{-1/r}$ and taking the $(−r)^{th}$ power.

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 $k$-th antisymmetric tensor power of $Z_{α, r}$ (see e.g. [32] Section I.5 for antisymmetric tensor powers and Section IX.2 for applications of the Weyl trick). For any given matrix $X$, its $k$-th antisymmetric tensor power, denoted $X^{∧k}$, is defined as the restriction of its $k^{th}$ tensor power $X^{⊗k}$ to the totally antisymmetric subspace. The reason for looking into this is that the largest eigenvalue of $X^{∧k}$ is the product of the $k$ largest eigenvalues of $X$, an identity which we denote by the shorthand

$$
λ_1(X^{∧k}) = λ_1 \cdots λ_k(X) := λ_1(X) \cdots λ_k(X).
$$

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

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 $X^{∧d}$ is a scalar, namely the determinant of $X$. Analogously, the matrix elements of $X^{∧k}$ for $k < d$ are all possible $k \times k$ minors of $X$ (determinants of submatrices). In particular, the “upper left” element $(X^{∧k})_{1,1}$ 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

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

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

$$
λ_1(Z_{α, r}(ρ||σ)^{∧k}) \approx (ρ^{∧k})^α \left( (σ^{∧k})^{-1/r}_{1,1} \right)^{r(α-1)}
$$

which means that

$$
λ^{(k)} := λ_1 \cdots λ_k(Z_{α}(ρ||σ)) \approx (μ_1 \cdots μ_k)^α \left( \det (σ^{-1/r}_{1:k,1:k}) \right)^{r(α-1)}.
$$

A mathematically rigorous restatement of this approximate identity will be given below; see the Approximation Lemma, which is Lemma 2 at the end of this section. 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:

$$
λ^{(d)} = \det(Z_{α}(ρ||σ)) = (\det ρ)^α \left( \det σ^{-1/r} \right)^{r(α-1)}.
$$

It is now a simple matter to obtain an approximation for $\text{Tr} Z_{α, r}(ρ||σ)$. Indeed, by taking the quotients of successive $λ^{(k)}$ we get all the eigenvalues of $Z_{α, r}$: $λ^{(k)}/λ^{(k-1)} = λ_k(Z_{α, r}(ρ||σ))$.

Summing these quotients then yields the trace of $Z_{α, r}$:

$$
\text{Tr} Z_{α, r}(ρ||σ) = \sum_{k=1}^{K} λ_k(Z_{α, r}(ρ||σ)) = λ^{(1)} + \sum_{k=2}^{d} \frac{λ^{(k)}}{λ^{(k-1)}}.
$$

Inserting the approximation (14) for $λ^{(k)}$ yields

$$
\text{Tr} Z_{α, r}(ρ||σ) \approx μ_1^α \left( (σ^{-1/r}_{1,1}) \right)^{r(α-1)} + \sum_{k=2}^{d} μ_k^α \left( \frac{\det (σ^{-1/r}_{1:k,1:k})}{\det (σ^{-1/r}_{1:k-1,1:k-1})} \right)^{r(α-1)}.
$$
Let us introduce the vector $\nu$ of leading principal minors of $\sigma^{-1/r}$ taken to the power $-r$, with
$$
\nu_k := \det \left((\sigma^{-1/r})_{1:k,1:k}\right)^{-r}.
$$
(16)

Note that $\nu_d = \det \sigma$. In terms of these $\nu_k$, eq. (15) can be rewritten more succinctly as
$$
\text{Tr} Z_{\alpha,r}(\rho||\sigma) \approx \mu_1^\alpha \nu_1^{1-\alpha} + \sum_{k=2}^{d} \mu_k^\alpha \left(\frac{\nu_k}{\nu_{k-1}}\right)^{1-\alpha}.
$$

One now recognizes the trace functional of the relative Rényi entropy in this formula, between the state $\rho$ and a new positive definite matrix $\hat{\sigma}$ that commutes with $\rho$ and that is given by
$$
\hat{\sigma} = \text{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1}).
$$
(17)

We then finally get, for $\alpha$ sufficiently close to 1:
$$
\text{Tr} Z_{\alpha,r}(\rho||\sigma) \approx \text{Tr} \rho^\alpha \hat{\sigma}^{1-\alpha}.
$$
(18)

The error in this approximation tends to 0 exponentially fast as $\exp(-1/(1-\alpha))$, as shown in Lemma 2 below. From (18) a closed form expression for the limit $\alpha \to 1$ of $D_{\alpha,r}(\rho||\sigma-1)$ can be found very easily, and it simply gives the classical relative entropy between $\rho$ and $\hat{\sigma}$. We have therefore proven:

**Theorem 2.** Let $\rho$ be a positive semidefinite matrix with non-degenerate spectrum and let $\sigma$ be positive definite. Let $r$ be a non-zero, finite real number. Then
$$
\lim_{\alpha \to 1} D_{\alpha,r}(\rho||\sigma) = D(\rho||\text{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1})),
$$
with $\nu_k = \det \left((\sigma^{-1/r})_{1:k,1:k}\right)^{-r}$, $k = 1, \ldots, d$.

In particular, for $r = -1$,
$$
\lim_{\alpha \to 1} \tilde{D}_{\alpha}(\rho||\sigma) = D(\rho||\text{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1})),
$$
with $\nu_k = \det (\sigma_{1:k,1:k})$, $k = 1, \ldots, d$.

(19)

(20)

As a sanity check, we can consider what eq. (20) reduces to when $\rho$ and $\sigma$ commute. In that case, $\sigma$ is diagonal in the eigenbasis of $\rho$, 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 $\nu_k/\nu_{k-1}$ reduce to $\sigma_{k,k}$, and $\text{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1})$ simply turns into $\sigma$ itself. We thus find that, in the commuting case, $\lim_{\alpha \to 1} D_{\alpha,r}(\rho||\sigma) = D(\rho||\sigma)$, as required.

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

11
Lemma 2 (Approximation Lemma). Let $A$ be a positive semidefinite matrix with its eigenvalues sorted in decreasing order denoted by $\mu_i$. Let $\mu_1 > \mu_2$, strictly. 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 $0 < \alpha < 1$ and $\beta = \alpha/2(1 - \alpha)$. Then

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

for some constant value of $c$. Hence, in the limit $\alpha \to 1^-$, $\lambda_1\left((A^\beta BA^\beta)^{1-\alpha}\right)$ tends to $\mu_1^\alpha(B_{1,1})^{1-\alpha}$ with an exponentially decreasing relative error of the order $\exp(-1/(1 - \alpha))$.

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

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

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

$$\lambda_1\left(A^\beta BA^\beta\right) \geq \lambda_1\left(B^{1/2}\mu_1^{2\beta} P_1 B^{1/2}\right) = \mu_1^{2\beta} B_{1,1} \quad (21)$$

and

$$\lambda_1\left(A^\beta BA^\beta\right) \leq \lambda_1\left(B^{1/2}(\mu_1^{2\beta} P_1 + \mu_2^{2\beta}(I - P_1))B^{1/2}\right) \leq \mu_1^{2\beta} \lambda_1\left(B^{1/2} P_1 B^{1/2}\right) + \mu_2^{2\beta} \lambda_1\left(B^{1/2}(I - P_1)B^{1/2}\right) \leq \mu_1^{2\beta} B_{1,1} + \mu_2^{2\beta} \lambda_1(B) = \mu_1^{2\beta} B_{1,1} \left(1 + \frac{\lambda_1(B)}{B_{1,1}}(\mu_2/\mu_1)^{2\beta}\right). \quad (22)$$

Since $B > 0$, we have $B_{1,1} > 0$ and the division can be done. Bracketing inequalities (21) and (22) can be combined as a single equality by introducing a constant $c$ such that

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

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

Raising all expressions to the power $1 - \alpha$ yields the equality of the lemma

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

Since for $0 \leq x < 1$ the function $(1 + cx^{\alpha/(1-\alpha)})^{1-\alpha}$ tends to 1 exponentially fast as $\alpha$ tends to 1, the second statement of the lemma is proven too. \qed
Let us now consider what happens when the spectrum of $\rho$ is degenerate, and whether $D_{\alpha,r(\alpha-1)}(\rho||\sigma)$ is continuous in $\rho$ and $\sigma$ in the limit $\alpha \to 1$. It is clear from the definition that it is continuous for all $\alpha \neq 1$. Thus, if we can show that (19) has a continuous extension, one that includes degenerate $\rho$ as well, then $D_{\alpha,r(\alpha-1)}$ is indeed continuous in the limit $\alpha \to 1$.

Let us therefore consider (19) 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 $\rho$. This is not immediately clear because of the formula’s non-trivial dependence on the eigenbasis of $\rho$: when the spectrum of $\rho$ is degenerate, $\rho$ 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 $\rho$ for which $\mu_1$ 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 $\sigma$. This can be recast as fixing one such representation of $\sigma$ and letting a $2 \times 2$ unitary matrix $U$ act on its upper left $2 \times 2$ block. Consequently, $\nu_1$ depends on $U$ whereas the other $\nu_i$ are independent from $U$, due to unitary invariance of the determinant. However, whereas this clearly affects the first two elements in the resulting

$$\hat{\sigma} = \text{diag}_\rho(\nu_1, \nu_2/\nu_1, \nu_3/\nu_2, \ldots, \nu_d/\nu_{d-1}),$$

this is actually compensated for by the multiplicity of $\mu_1$. The first two terms in the formula for $D(\rho||\hat{\sigma})$ are

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

and this simplifies to

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

which is independent of $\nu_1$.

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

$$D(\rho||\hat{\sigma}) = -S(\rho) - \mu_d \log \det \sigma - \sum_{i=1}^{d-1} (\mu_i - \mu_{i+1}) \log \nu_i,$$  \tag{23}

where $S(\rho) = -\text{Tr} \rho \log \rho$ is the von Neumann entropy of $\rho$. The upshot is that $D(\rho||\hat{\sigma})$ is independent of those elements $\nu_i$ that are dependent on a freedom of choice of basis caused by degeneracy of $\mu_i$. This implies that $D(\rho||\hat{\sigma})$ is continuous in $\rho$ and $\sigma$ since every term in (23) is continuous, as we now show. Indeed, the von Neumann entropy is well-known to be continuous (in the sense of Fannes), and $\mu_d$ and $\nu_d = \det \sigma$ are continuous as well since eigenvalues of a matrix depend continuously on the entries of a matrix (\textit{[33]}, 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 $\rho$.

To see the problem, consider the example of a positive semidefinite matrix $\rho$ parameterized by the variable $x$, $\rho(x) = \text{diag}(1 + x, 1 - x)$, with $0 < |x| < 1$. Then for $x > 0$, $P_1 = \text{diag}(1, 0)$ whereas for $x < 0$, $P_1 = \text{diag}(0, 1)$. Thus for almost all $\sigma$, $\nu_1(x)$ 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 \( x \) (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 \( \nu_i \) occurs when the prefactor \( \mu_i - \mu_{i+1} \) becomes zero, which cancels the discontinuity in \( \nu_i \) (while still leaving a discontinuity in the derivative).

We have thus finally proven:

**Theorem 3.** The statement from Theorem 2 still holds when \( \rho \) has degenerate spectrum, in the sense that (19) has to be interpreted as (23). The limit \( \lim_{\alpha \to 1} D_{\alpha,r(\alpha-1)}(\rho||\sigma) \) exists as a continuous (but not necessarily smooth) function of \( \rho \) and \( \sigma \).

## 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 considering the parametrization \( z = r(\alpha - 1) \) and take the limit of \( D_{\alpha,r(\alpha-1)} \) as \( \alpha \) tends to \(+\infty\).

Noting that the operator norm is the limit of the Schatten \( q \)-norm as \( q \) tends to \(+\infty\), we obtain from (2),

\[
\lim_{\alpha \to +\infty} D_{\alpha,r(\alpha-1)}(\rho||\sigma) = \lim_{\alpha \to +\infty} \frac{1}{\alpha} \log \text{Tr}(\rho^{\alpha/r(\alpha-1)}\sigma^{-1/r})^{r(\alpha-1)}
= \lim_{\alpha \to +\infty} \log \left\| (\rho^{\alpha/r(\alpha-1)}\sigma^{-1/r} \rho^{1/2r(\alpha-1)})^r \right\|_{\alpha-1}
= \log \left\| (\rho^{1/2r}\sigma^{-1/r} \rho^{1/2r(\alpha-1)})^r \right\|_{\infty}
= \log \left\| \rho^{1/2r}\sigma^{-1/r} \rho^{1/2r(\alpha-1)} \right\|_{\infty}
= r \log \left\| \rho^{1/2r}\sigma^{-1/r} \rho^{1/2r(\alpha-1)} \right\|_{\infty}
\]

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

\[
\log \left\| \rho^{1/2r}\sigma^{-1/r} \rho^{1/2r(\alpha-1)} \right\|_{\infty} = \log \min_{\lambda} \{ \lambda : \lambda \sigma^{1/r} \geq \rho^{1/r} \} = D_{\max}(\rho^{1/r}||\sigma^{1/r}).
\]

Thus we arrive at the following theorem:

**Theorem 4.** Let \( \rho \) be a positive semidefinite matrix and let \( \sigma \) be positive definite. Then for a non-zero, finite real number \( r \),

\[
\lim_{\alpha \to +\infty} D_{\alpha,r(\alpha-1)}(\rho||\sigma) = rD_{\max}(\rho^{1/r}||\sigma^{1/r}). \tag{24}
\]

In particular, for \( r = 1 \)

\[
\lim_{\alpha \to +\infty} \tilde{D}_{\alpha}(\rho||\sigma) = D_{\max}(\rho||\sigma).
\]
For $\alpha \to -\infty$, which necessitates the stronger restriction on the supports $\text{supp} \rho = \text{supp} \sigma$, a similar treatment yields the result that

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

and

$$\lim_{\alpha \to -\infty} \tilde{D}_\alpha(\rho||\sigma) = -D_{\max}(\sigma||\rho).$$  (26)

Finally, we study the limit $z \to \infty$ when $\alpha$ is kept fixed (and finite). Using the well-known Lie-Trotter product formula (see, e.g. [32], 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 $\alpha \neq 1$,

$$\lim_{z \to \infty} D_{\alpha,z}(\rho||\sigma) = \frac{1}{\alpha - 1} \log \text{Tr} \exp(\alpha \log \rho + (1 - \alpha) \log \sigma).$$  (27)

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

$$\lim_{\alpha \to 1} \lim_{z \to \infty} D_{\alpha,z}(\rho||\sigma) = D(\rho||\sigma).$$  (28)

### 5 Limiting case $z = \alpha \to 0$

In this section, we answer the question: what is the limit of $\tilde{D}_\alpha$ as $\alpha$ 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 $\sigma$ is full rank. We will also assume first that the spectrum of $\sigma$ is non-degenerate.

The answer to this question is easy when $\rho$ and $\sigma$ commute. Choosing a basis in which both states are diagonal, with diagonal elements given by $\rho_i$ and $\sigma_i$, respectively, the limit is given by

$$\lim_{\alpha \to 0} f_{\alpha,\alpha}(\rho||\sigma) = \lim_{\alpha \to 0} \sum_{i=1}^d \rho_i^{\alpha} \sigma_i^{1-\alpha}$$

$$= \sum_{i} \sigma_i : \rho_i \neq 0.$$

In terms of the projector on the support of $\rho$, which we denote by $\Pi_\rho$, we write this as

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

To answer the question in the general case, we will first show that the answer does not depend on $\rho$ itself, but only on $\Pi_\rho$, and of course also on $\sigma$. To do so, we consider the particular expression

$$\lim_{\alpha \to 0} f_{\alpha,\alpha}(\rho||\sigma) = \lim_{\alpha \to 0} \text{Tr}(\sigma^{1/2\alpha} \rho^{1/2\alpha})^{\alpha}.$$  (28)
Let $\mu$ be the smallest non-zero eigenvalue of $\rho$. Then we have the inclusion $\mu \Pi \rho \leq \rho \leq \Pi \rho$. This implies
\[
\mu^\alpha \text{Tr}(\sigma^{1/2\alpha} \Pi \rho \sigma^{1/2\alpha})^\alpha \leq \text{Tr}(\sigma^{1/2\alpha} \rho \sigma^{1/2\alpha})^\alpha \leq \text{Tr}(\sigma^{1/2\alpha} \Pi \rho \sigma^{1/2\alpha})^\alpha.
\]
In the limit of $\alpha \to 0$, $\mu^\alpha$ of course tends to 1, so that both sides of the inclusion become equal and we have the identity
\[
\lim_{\alpha \to 0} \text{Tr}(\sigma^{1/2\alpha} \rho \sigma^{1/2\alpha})^\alpha = \lim_{\alpha \to 0} \text{Tr}(\sigma^{1/2\alpha} \Pi \rho \sigma^{1/2\alpha})^\alpha.
\]

For the remainder of the argument, we will work in a basis in which $\Pi \rho$ is diagonal, and given by $I_r \oplus 0$, where $r$ is the rank of $\rho$. Furthermore, we switch from one representation of $f_{\alpha,\alpha}$ to another, namely
\[
\lim_{\alpha \to 0} f_{\alpha,\alpha}(\rho||\sigma) = \lim_{\alpha \to 0} \text{Tr}(\Pi \rho \sigma^{1/\alpha} \Pi \rho)^\alpha.
\]
We will also employ the spectral decomposition of $\sigma$, which we consider 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 $\rho$:
\[
|u_i\rangle \mapsto |\tilde{u}_i\rangle := \Pi \rho |u_i\rangle.
\]
With this definition, we have
\[
\Pi \rho \sigma^{1/\alpha} \Pi \rho = \sum_{i=1}^d \lambda_i^{1/\alpha} |\tilde{u}_i\rangle \langle \tilde{u}_i|.
\]
It goes without saying that the vectors $|\tilde{u}_i\rangle$ 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 $\mu_1$ of $Z_\alpha := (\Pi \rho \sigma^{1/\alpha} \Pi \rho)^\alpha$ in the limit. Given that the spectrum of $\sigma$ is non-degenerate, the largest contribution to $\Pi \rho \sigma^{1/\alpha} \Pi \rho$ will come from $\lambda_1$, and is given by $\lambda_1^{1/\alpha} |\tilde{u}_1\rangle \langle \tilde{u}_1|$. That is true, of course, only if $|\tilde{u}_1\rangle$ is not the zero vector (which can happen if $|u_1\rangle$ lies outside the support of $\rho$). We therefore have to correct our statement and say: the largest contribution to $\Pi \rho \sigma^{1/\alpha} \Pi \rho$ will come from $\lambda_{i_1}$, and is given by $\lambda_{i_1}^{1/\alpha} |\tilde{u}_{i_1}\rangle \langle \tilde{u}_{i_1}|$, where $i_1$ is the first index value for which $|\tilde{u}_{i_1}\rangle \neq 0$. The limit can now be calculated easily, and we get
\[
\mu_1 = \lim_{\alpha \to 0} \lambda_{i_1} || |\tilde{u}_{i_1}\rangle \langle \tilde{u}_{i_1}| ||^\alpha = \lambda_{i_1} = \max_{i_1} \lambda_{i_1} : |\tilde{u}_{i_1}\rangle \neq 0
\]

Next, we calculate the product of the two largest eigenvalues of $Z$, $\mu_1 \mu_2$. Using the Weyl-trick, this reduces to the largest eigenvalue of the second antisymmetric tensor power, and using the formula just obtained we find
\[
\mu_1 \mu_2 = \max_{i_1, i_2} \lambda_{i_1} \lambda_{i_2} : |\tilde{u}_{i_1}\rangle \wedge |\tilde{u}_{i_2}\rangle \neq 0.
\]
The latter condition amounts to the two vectors $|\tilde{u}_{i_1}\rangle$ and $|\tilde{u}_{i_2}\rangle$ being linearly independent. For $\mu_1\mu_2\mu_3$ we similarly obtain
\[
\mu_1\mu_2\mu_3 = \max_{i_1, i_2, i_3} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} : |\tilde{u}_{i_1}\rangle, |\tilde{u}_{i_2}\rangle, |\tilde{u}_{i_3}\rangle \text{ linearly independent},
\]
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 $\mu_i$, for $i = 1, 2, \ldots, s$. What we are after is the sum of these $\mu_i$, and this sum is simply given by
\[
\sum_{i=1}^{s} \mu_i = \max_{i_1, i_2, \ldots, i_s} \sum_{j=1}^{s} \lambda_{i_j} : \{|\tilde{u}_{i_j}\rangle\} \text{ linearly independent}.
\]

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 (see e.g. [34]). 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:
\[
\lim_{\alpha \to 0} f_{\alpha,\alpha}(\rho||\sigma) = \sum_{j=1}^{s} \lambda_{i_j},
\]
where the $\lambda_i$ are the eigenvalues of $\sigma$, 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 $\rho$). 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 $\sigma$ is degenerate. Suppose a certain eigenvalue of $\sigma$ has multiplicity $k$. Let $S$ be the subspace that is the projection of this $k$-dimensional eigenspace to the support of $\rho$. 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 $u_{ij}$ 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 $u_{ij}$. 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 $\rho$ and $\sigma$ be 4-dimensional states where $\sigma$ is full rank and has non-degenerate spectrum, and $\rho$ has rank 2. In terms of the eigenbasis of $\rho$, the projector $\Pi_\rho$ is represented by the diagonal matrix $\Pi_\rho = \text{diag}(1, 1, 0, 0)$. Furthermore, let $\sigma$ 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

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

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 Proof of concavity of the trace functional $f_{\alpha,z}$

For any complex number $z$, and a complex matrix $C$ in $M_d(C)$ we use the notation $(C + z)$ to denote the matrix $(C + zI)$, where $I$ denotes the identity matrix. We denote the conjugate transpose of a matrix $C$ by $C^*$ and use $C^{-*}$ as a shorthand for $(C^{-1})^* = (C^*)^{-1}$. Further, we denote the spectrum of any complex matrix $C \in M_d(C)$ as $\text{Sp}(C)$.

For convenience, we will re-parameterize the trace functional $f_{\alpha,z}$ by the parameters $p = \alpha/z$ and $q = (1 - \alpha)/z$ as

\[ f_{p,q}(A; K) := \text{Tr}(A^p K A^q K^*)^{1/(p+q)}. \tag{30} \]

We obtain the original functional by setting $z = 1/(p + q)$ and $\alpha = p/(p + q)$. To lighten the notations we will henceforth just write $f_{p,q}(A)$ as we always keep $K$ fixed.

In this section we shall provide a detailed proof of the following:

**Theorem 5** (Hiai). The trace functional $f_{p,q}(A)$ is concave on the set of positive $d \times d$ matrices for $0 < p, q \leq 1$.

This theorem is a special case of Theorem 1.1 in [27] (namely the case $s = 1/(p + q)$ and replacing the positive linear maps $\Phi$ and $\Psi$ by the identity map), but as the proof of that Theorem is rather complicated, in part due to its generality, we provide a more direct proof of our own. Furthermore, the proof technique is based on Epstein’s proof [36] of Lieb’s concavity theorem, and as the paper [36] is written in a rather terse style we also felt the need to expand on the details where necessary and simplify the proof where possible. We hope that our exposition will make this powerful technique more widely accessible.

Hiai’s paper also provides necessary conditions for concavity/convexity ([27], Section 4). The region in the $(p, q)$-parameter space where these conditions are not satisfied is indicated as white space in Figure 2. Apart from the blue square where concavity provably holds, this leaves open the two orange squares, where we conjecture that convexity holds:

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

This conjecture carries some weight as the case $z = \alpha \geq 1/2$ ($p = 1$) has been proven by Frank and Lieb [22] and the case $1 \leq \alpha \leq 2$, $z = 1$ ($p + q = 1$), is Ando’s convexity theorem [29]. These two cases are indicated in Figure 1 (the blue strips within the orange region) and in Figure 2.

Before we provide the proof of Theorem 5 a number of technical tools have to be introduced.
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 before. 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 ($\alpha > 0$) in the region where $p$ and $p + q$ have the same sign, again excluding the upper left orange square.

6.1 Real and imaginary part of a matrix

The real and imaginary part of a general complex matrix $C$ are defined as

\[
\text{Re} C := \frac{C + C^*}{2}, \quad \text{Im} C := \frac{C - C^*}{2i},
\]

which are both Hermitian matrices. Conversely, any matrix $C$ can be written as $C = \text{Re} C + i \text{Im} C$, a decomposition known as the Cartesian decomposition.

We summarize a number of properties that will be needed.

Lemma 3. For any pair of matrices $C$ and $K$,

\[
\text{Re}(KCK^*) = K(\text{Re} C)K^*, \quad \text{and} \quad \text{Im}(KCK^*) = K(\text{Im} C)K^*.
\]

Proof. By the definition of the real part,

\[
\text{Re}(KCK^*) = \frac{KCK^* + (KCK^*)^*}{2} = \frac{KCK^* + K^*CK^*}{2} = \frac{K(\text{Re} C) + (\text{Re} C)^*}{2}K^* = K(\text{Re} C)K^*.
\]
The proof for the imaginary part is completely similar. □

**Lemma 4.** If \( \text{Im} \, C > 0 \), then \( C \) is invertible. If \( \text{Re} \, C > 0 \), then \( C \) is invertible.

*Proof.* Let \( A = \text{Re} \, C \) and \( B = \text{Im} \, C \), so that \( C = A + iB \). By the hypothesis, \( B > 0 \). Hence, \( B \) is invertible and its positive square root exists too. We can therefore write \( C \) as

\[
C = B^{1/2}(B^{-1/2}AB^{-1/2} + iI)B^{1/2}.
\]

As \( B^{-1/2}AB^{-1/2} \) is Hermitian, it has real eigenvalues. Any matrix commutes with the identity matrix. Thus, the eigenvalues of \( B^{-1/2}AB^{-1/2} + iI \) are of the form \( \lambda + i \) with \( \lambda \in \mathbb{R} \). No eigenvalue can therefore be zero, so that \( B^{-1/2}AB^{-1/2} + iI \) is an invertible matrix. As \( B^{1/2} \) is invertible too, this shows that \( C \) is indeed invertible.

The statement for \( \text{Re} \, C \) follows from this by noting that \( \text{Re} \, C = \text{Im}(iC) \). □

**Lemma 5.** For any invertible matrix \( C \),

\[
\begin{align*}
\text{Re} \, C^{-1} &= C^{-1}(\text{Re} \, C)C^{-*} & (32) \\
\text{Im} \, C^{-1} &= -C^{-1}(\text{Im} \, C)C^{-*} & (33)
\end{align*}
\]

*Proof.* If \( C \) is invertible, so is \( C^* \) and we can write \( C^{-1} \) as

\[
C^{-1} = C^{-1}C^*C^{-*}.
\]

Using Lemma [31] with \( K = C^{-1} \), we get

\[
\text{Re} \, C^{-1} = \text{Re}(C^{-1}C^*C^{-*}) = C^{-1} \text{Re}(C^*)C^{-*} = -C^{-1} \text{Re} \, CC^{-*}
\]

and

\[
\text{Im} \, C^{-1} = C^{-1} \text{Im}(C^*)C^{-*} = -C^{-1} \text{Im} \, CC^{-*}.
\]

□

### 6.2 Complex segments

Following [36], we denote by \( \mathcal{I}^+(\mathbb{C}) \) \( (\mathbb{C}^+ \text{ in Hiai’s notation}) \) the open half-plane of complex numbers with positive imaginary part:

\[
\mathcal{I}^+(\mathbb{C}) := \{ z : \text{Im} \, z > 0 \}. \tag{34}
\]

One can also define the open half-planes \( \mathcal{I}^- \) and \( \mathcal{R}^+ \) as the sets of complex numbers with negative imaginary part and positive real part, respectively.

These definitions generalize to complex matrices:

\[
\mathcal{I}^+(M_d(\mathbb{C})) := \{ C \in M_d(\mathbb{C}) : \text{Im} \, C > 0 \}. \tag{35}
\]

We shall drop the arguments and write \( \mathcal{I}^+ \) if it is clear from the context which set is meant.
Next, we introduce a new notation of our own, not present in \[36\]. Given two angles $\alpha$ and $\beta$, with $-\pi \leq \alpha < \beta \leq \pi$, we denote the open segment of the cut plane consisting of non-zero complex numbers whose complex argument is (strictly) between $\alpha$ and $\beta$ as

$$ S_{\alpha,\beta}(\mathbb{C}) := \{ z = r e^{i\theta} : r > 0, \alpha < \theta < \beta \}. $$

(36)

These segments come in two varieties: the convex segments, with $\beta - \alpha \leq \pi$, and the concave ones, with $\beta - \alpha > \pi$. When $\beta - \alpha = \pi$, the segment is an open half-plane and can also be defined in terms of $\mathcal{I}^+$ as

$$ S_{\alpha,\alpha+\pi}(\mathbb{C}) = \{ z : e^{-i\alpha} z \in \mathcal{I}^+ \}. $$

In particular, note that $\mathcal{I}^+ = S_{0,\pi}$, $\mathcal{I}^- = S_{-\pi,0}$ and $\mathcal{R}^+ = S_{-\pi/2,\pi/2}$.

The convex segments can equivalently be defined as the intersection between two open half-planes: for $\beta - \alpha < \pi$,

$$ S_{\alpha,\beta} = S_{\alpha,\alpha+\pi}(\mathbb{C}) \cap S_{\beta-\pi,\beta}(\mathbb{C}) $$

$$ = \{ z : e^{-i\alpha} z \in \mathcal{I}^+ \} \cap \{ z : e^{-i\beta} z \in \mathcal{I}^- \} $$

$$ = \{ z : e^{-i\alpha} z \in \mathcal{I}^+ \text{ and } e^{-i\beta} z \in \mathcal{I}^- \}. $$

(37)

(38)

Because of the latter identity, it is possible to extend the definition of a convex complex segment to matrices: for two given angles $\alpha$ and $\beta$, with $-\pi \leq \alpha < \beta \leq \pi$ and $\beta - \alpha \leq \pi$,

$$ S_{\alpha,\beta}(M_d(\mathbb{C})) := \{ C \in M_d(\mathbb{C}) : e^{-i\alpha} C \in \mathcal{I}^+ \text{ and } e^{-i\beta} C \in \mathcal{I}^- \}. $$

(39)

Again, we shall drop the arguments and write $S_{\alpha,\beta}$ if the context is clear. For concave complex segments, we are not aware of any useful generalization to matrices; several complications arise.

It is easy to see that if the angle $\phi$ is such that $-\pi \leq \alpha + \phi$ and $\beta + \phi \leq \pi$, then $C \in S_{\alpha,\beta}$ implies $e^{i\phi} C \in S_{\alpha+\phi,\beta+\phi}$. Furthermore, $C \in S_{\alpha,\beta}$ implies $C^* \in S_{-\beta,-\alpha}$.

Finally, if $C \in S_{\alpha,\beta}$ and $a > 0$ then $aC \in S_{\alpha,\beta}$. If $a < 0$, however, then $aC \in S_{\alpha-\pi,\beta-\pi}$ if $0 < \alpha < \beta$, and $aC \in S_{\alpha+\pi,\beta+\pi}$ if $\alpha < \beta < 0$. If $\alpha < 0 < \beta$ then the segment containing $aC$ necessarily has to straddle the cut, and this will not be considered.

**Lemma 6.** Let two angles $\alpha$ and $\beta$ be given, with $-\pi \leq \alpha < \beta \leq \pi$ and $\beta - \alpha \leq \pi$. Then, if $C \in S_{\alpha,\beta}$ then $\text{Sp} \ C \subset S_{\alpha,\beta}$.

**Proof.** Let us first show this for $S_{\alpha,\beta} = \mathcal{I}^+$. Clearly, if $\text{Im} \ C > 0$ and $\text{Im} \ z \leq 0$, then $\text{Im}(C-z) > 0$, which by Lemma 4 implies that $C-z$ is invertible. Hence, $C-z$ is invertible for all $z$ on the real axis or in the lower half-plane. As the spectrum of $C$ consists, by definition, of those points $z$ where $C-z$ is not invertible, this shows that $\text{Sp} \ C$ lies entirely in the upper half-plane.

To show the statement in full generality, we apply this reasoning to $e^{-i\alpha} C$ and $-e^{-i\beta} C$. Noting that $\text{Sp}(e^{-i\alpha} C) = e^{-i\alpha} \text{Sp} \ C$, we then get $e^{-i\alpha} \text{Sp} \ C \in \mathcal{I}^+$ and $e^{-i\beta} \text{Sp} \ C \in \mathcal{I}^-$. Hence, $\text{Sp} \ C \in S_{\alpha,\beta}$.

The converse of this lemma is not true, not even for normal matrices. In fact, there exist matrices that are not in $S_{\alpha,\beta}$ for any value of $\alpha, \beta$. For example, for the matrix $X =$
The Lemma ensure that \( \alpha \) other words, \( \text{Sp}(A) \). Thus, \( \text{Sp}(A) \).

This representation can be used to extend the definition of fractional powers to matrices in

\[
\begin{pmatrix}
5 - 9i & -7 - 5i \\
4 - 5i & -3 + 8i
\end{pmatrix},
\]
numerical experiments reveal that \( \text{Im} e^{-i\alpha} X \) is indefinite for any value of the angle \( \alpha \).

The main usefulness of complex matrix segments is due to the following lemma, which generalizes Lemma 2 in [36].

**Lemma 7.** Consider two convex complex segments \( S_{\alpha_i,\beta_i} \) (with \( i = 1, 2 \)), i.e. the angles satisfy \(-\pi \leq \alpha_i < \beta_i \leq \pi \) and \( \beta_i - \alpha_i \leq \pi \), such that \(-\pi \leq \alpha_1 + \alpha_2 \) and \( \beta_1 + \beta_2 \leq \pi \). Let \( A_i \in \mathcal{S}_{\alpha_i,\beta_i}(\mathbb{M}_d(\mathbb{C})) \) (\( i = 1, 2 \)). Then \( \text{Sp}(A_1 A_2) \) lies in the complex segment \( S_{\alpha_1+\alpha_2,\beta_1+\beta_2}(\mathbb{C}) \).

Note that no such conclusion can be drawn from statements about the spectra of \( A_1 \) and \( A_2 \).

The proof we give below is completely different from Epstein’s and as it does not involve deep analytical results (such as Bochner’s tube theorem) it should be much easier to understand.

**Proof.** Consider first two matrices \( A \) and \( B \) with positive real part (i.e., they are in \( \mathcal{R}^+ \)). We will show that any real eigenvalue of \( AB \) must be positive. By Lemmas 4 and 5, \( A \) and \( B \) are invertible and \( \text{Re} B^{-1} > 0 \). Let \( x \geq 0 \). Then \( \text{Re}(A + xB^{-1}) = \text{Re} A + x \text{Re} B^{-1} > 0 \), so that \( A + xB^{-1} \) is invertible (again by Lemma 4). Hence, \( AB + x \) is invertible too, and \( -x \) can not be an eigenvalue of \( AB \).

Now, the fact that \( A_i \) lies in a convex segment, \( S_{\alpha_i,\beta_i}(\mathbb{M}_d(\mathbb{C})) \), implies that there exist angles \( \gamma_i \) such that \( e^{-i\gamma_i} A_i \) is in \( \mathcal{R}^+ \). This is so if \( \gamma_i \in [-\alpha_i - \pi/2, -\beta_i + \pi/2] \). Then, by the above reasoning, no eigenvalue of \( e^{-i(\gamma_1+\gamma_2)} A_1 A_2 \) lies on the cut \( \mathbb{R}^- \), for any \( \gamma_i \) in the given interval. This corresponds to any value of \( \gamma_1 + \gamma_2 \) in the interval \( [-\pi - (\alpha_1 + \alpha_2), \pi - (\beta_1 + \beta_2)] \). Thus, \( \text{Sp}(A_1 A_2) \) lies outside the cut, and outside the segments \( S_{-\pi,\alpha_1+\alpha_2} \) and \( S_{\beta_1+\beta_2,\pi} \). In other words, \( \text{Sp}(A_1 A_2) \) lies inside the complex segment \( S_{\alpha_1+\alpha_2,\beta_1+\beta_2}(\mathbb{C}) \); the conditions of the Lemma ensure that \( \alpha_1 + \alpha_2 \) and \( \beta_1 + \beta_2 \) define a proper complex segment. \( \blacksquare \)

### 6.3 Fractional matrix powers

As Epstein’s method relies heavily on complex analysis, we will need a definition of the fractional power \( A^p \) (with \( 0 < p < 1 \)) that also applies to general complex matrices, not just the positive definite ones.

For \( 0 < p < 1 \), the complex function \( z \mapsto z^p \) is defined in the cut plane \( \mathbb{C} \setminus \mathbb{R}^- = \{ z : \text{Im} z \neq 0 \text{ or } \text{Re} z > 0 \} \) by \( \exp(p(\log |z| + i \arg z)) \), for \( \arg z \in (-\pi, \pi) \). One shows that \( z^p \) is a so-called Pick function ([37], Chapter II), a holomorphic function that maps the upper half plane into itself. Using the theory of Pick functions one can then find the following integral representation:

\[
z^p = \frac{\sin p\pi}{\pi} \int_0^\infty dt \frac{t^{p-1}}{t+z} \left( \frac{1}{t} - \frac{1}{t+z} \right). \tag{40}
\]

This representation can be used to extend the definition of fractional powers to matrices in
the obvious way. For any matrix $C$ with spectrum $\text{Sp} C \subset \mathbb{C} \setminus \mathbb{R}^-$,
\[
C^p = \frac{\sin p \pi}{\pi} \int_0^{\infty} dt \ t^p \left( \frac{1}{t} - (t + C)^{-1} \right).
\] (41)

Using this integral, we can express $\text{Im} C^p$ as an integral as well:
\[
\text{Im} C^p = \frac{\sin p \pi}{\pi} \int_0^{\infty} dt \ t^p \text{Im}(-(t + C)^{-1}).
\] (42)

However, one has to keep in mind that integral (40) can not be written as a difference of two integrals, one with integrand $t^{p-1}$ and the other with integrand $t^p/(t + z)$, as neither of these integrals converges for $0 < p < 1$. To show the validity of (42), one has to show that the integral actually converges. Since this argument is left to the reader in [36], we present it separately, in Appendix A.

Lemma 8. For $0 < p < 1$ and $C \in \mathcal{S}_{\alpha, \beta}(\mathbb{M}_d(\mathbb{C}))$, one has $C^p \in \mathcal{S}_{p\alpha, p\beta}(\mathbb{M}_d(\mathbb{C}))$ and for any $K \in \mathbb{M}_d(\mathbb{C})$, $KC^p K^* \in \mathcal{S}_{p\alpha, p\beta}(\mathbb{M}_d(\mathbb{C}))$. For $-1 < p < 0$, we get $C^p, KC^p K^* \in \mathcal{S}_{p\beta, p\alpha}(\mathbb{M}_d(\mathbb{C}))$.

Proof. The condition $|p| < 1$ ensures that the complex segment $\mathcal{S}_{p\alpha, p\beta}(\mathbb{C})$ remains convex. Consider first the case $0 < p < 1$. Again we first consider the case $C \in \mathcal{I}^+$. Then $\text{Im} C > 0$ and by Lemma 3,
\[
\text{Im}(-C^{-1}) = C^{-1} \text{Im}(C)C^{-*} > 0.
\] (43)

Combining this with the integral expression (42) we find that $\text{Im} C^p > 0$ and by Lemma 3 $\text{Im}(KC^p K^*) > 0$. To prove the general statement, we apply this to the matrices $e^{-i\alpha}C$ and $-e^{-i\beta}C$ and get $\text{Im}(e^{-i\alpha}C)^p = \text{Im} e^{-ip\alpha}C^p > 0$ and $\text{Im} e^{-ip\beta}C^p < 0$, which shows that $C^p \in \mathcal{S}_{p\alpha, p\beta}$.

For the case $-1 < p < 0$, note that $C^p$ is just $(C^{-1})^{|p|}$ and that if $C \in \mathcal{I}^+$, then by Lemma 3 $-C^{-1} \in \mathcal{I}^+$ as well. Applying the preceding argument to $-C^{-1} = e^{-i\pi}C^{-1}$ yields the statement for $-1 < p < 0$.

The restriction $|p| \leq 1$ is essential. For $p > 1$ it might seem plausible that if $C \in \mathcal{S}_{\alpha/p, \beta/p}$, with $\mathcal{S}_{\alpha, \beta}$ convex, then $C^p \in \mathcal{S}_{\alpha, \beta}$; however, this is not true. Taking $p = 2$, for example, then $C^2 \in \mathcal{S}_{-\pi/2, \pi/2} = \mathcal{R}^+$ iff $\text{Re} C^2 > 0$, which amounts to $(\text{Re} C)^2 \geq (\text{Im} C)^2$. In contrast, $C \in \mathcal{S}_{-\pi/4, \pi/4}$ iff $\text{Re} C \geq \text{Im} C$ and $\text{Re} C \geq -\text{Im} C$, which is not sufficient to ensure $(\text{Re} C)^2 \geq (\text{Im} C)^2$ (as the function $x^2$ is not operator monotone).

6.4 Proof of concavity of $f_{p,q}$ for $0 < p, q \leq 1$

After these preliminaries, we can now turn to the actual concavity proof. Epstein's technique involves the consideration of two related functions. For $A > 0$ and $\Delta = \Delta^*$, define for $\zeta$ in appropriate subsets of $\mathbb{C}$ (to be determined below)
\[
F(\zeta) := f_{p,q}(A + \zeta \Delta)
\] (44)
\[
G(\zeta) := f_{p,q}(\zeta A + \Delta),
\] (45)
where the domain of \( f_{p,q} \) has been extended to complex matrices (again, see below). Because \( f_{p,q} \) is homogeneous of order 1, these two functions are connected via the relation

\[
F(\zeta) = \zeta G(1/\zeta). \tag{46}
\]

Concavity of \( A \mapsto f_{p,q}(A) \) over the set of positive definite matrices amounts to proving concavity of \( F(x) \) for real \( x \) in the domain of \( F \). Indeed, \( f_{p,q}(xA + (1-x)A_2) \geq x f_{p,q}(A_1) + (1-x) f_{p,q}(A_2) \) for all \( x \in [0,1] \) and given \( A_1, A_2 > 0 \) is equivalent to \( F(x) \geq x F(1) + (1-x) F(0) \), with \( A = A_2 \) and \( \Delta = A_1 - A_2 \).

The essential feature of Epstein’s technique is to show that \( G \) is a so-called Pick function (a.k.a. Herglotz function): these are the functions that are holomorphic in the upper half-plane \( \mathcal{I}^+ \) and have positive imaginary part, i.e. they map \( \mathcal{I}^+ \) into itself \([37]\), and can be continued analytically by reflection (that is, \( G(\overline{z}) = \overline{G(z)} \)) to the lower half-plane \( \mathcal{I}^- \). More precisely, \( G \) is in the class \( P(\tau, +\infty) \) of Pick functions for which this analytic continuation can be achieved across an interval \( (\tau, +\infty) \), where \( \tau \) is a number depending on \( A \) and \( \Delta \) (see below). The importance of this fact is that Pick functions possess a certain integral representation, which carries over to \( F(x) \), hence to \( f_{p,q} \) itself. Proving concavity of \( f_{p,q} \) then amounts to proving concavity of the integral’s kernel, which turns out to be straightforward.

To show this, we henceforth consider the complex variable \( \zeta = x + iy \). Then \( A + \zeta \Delta = (A + x\Delta) + iy\Delta \) and \( \Delta + \zeta A = (\Delta + xA) + iyA \).

**Remark.** The class \( P(a,b) \) is very important in the theory of operator monotone functions. Loewner’s theorem ([37], Chapter IX) states that \( P(a,b) \) is exactly the class of functions that are operator monotone (that is, monotone of order \( n \) for all \( n \)) on the open interval \( (a,b) \).

### 6.4.1 Domain of holomorphy of \( F \) and \( G \)

First we need to establish the domain of holomorphy of \( F \) and \( G \). For the purposes of the proof, it is sufficient to extend the domain of \( f_{\alpha,z} \) to the set \( D \) of matrices in \( \mathbb{M}_d(\mathbb{C}) \) that are contained in one (or more) of the half-planes \( S_{\theta-\pi/2,\theta+\pi/2} \) for some \( \theta \in [-\pi/2, \pi/2] \). Thus,

\[
D := \bigcup_{\theta \in [-\pi/2, \pi/2]} S_{\theta-\pi/2,\theta+\pi/2}(\mathbb{M}_d(\mathbb{C})) = \{ A \in \mathbb{M}_d(\mathbb{C}) : \text{Re} e^{-i\theta} A > 0, \text{ for some } \theta \in [-\pi/2, \pi/2] \}.
\]

Note that

\[
\bigcup_{\theta \in [-\pi/2, \pi/2]} S_{\theta-\pi/2,\theta+\pi/2}(\mathbb{C})
\]

is just the cut complex plane \( \mathbb{C} \setminus \mathbb{R}^- \).

The function \( G(\zeta) \) is thus well-defined when \( \Delta + \zeta A = (\Delta + xA) + iyA \) is in \( D \). A sufficient (but by no means necessary) condition is that \( xA + \Delta > 0 \) (corresponding to the choice \( \theta = 0 \)). This is satisfied if \( x \) is larger than the largest eigenvalue of \( -\Delta \) divided by the smallest eigenvalue of \( A \), and this holds if (but not only if) \( x > \tau \), with \( \tau := ||\Delta||/\lambda_{\text{min}}(A) \).

Another sufficient condition is \( y \neq 0 \). Indeed, if \( y > 0 \) then \( \text{Im}(\zeta A + \Delta) = \text{Re}(-i(\zeta A + \Delta)) > 0 \), so that \( \zeta A + \Delta \in D \) (with \( \theta = \pi/2 \)); if \( y < 0 \) then \( \text{Im}(\zeta A + \Delta) = -\text{Re}(i(\zeta A + \Delta)) < 0 \), and again \( \zeta A + \Delta \in D \) (with \( \theta = -\pi/2 \)).
Thus, $\zeta A + \Delta \in D$ if (but not only if) $\zeta \in \mathbb{C} \setminus (-\infty, \tau]$. Since $G(\zeta)$ is a composition of holomorphic functions, it is holomorphic itself in this domain.

The function $F(\zeta)$ is similarly well-defined when $A + \zeta \Delta = (A + x\Delta) + iy\Delta$ is in $D$. A sufficient condition is $A + x\Delta > 0$, and this is so if (but not only if) $|x| < 1/\tau$.

By the relation $F(\zeta) = \zeta G(1/\zeta)$ this now implies that $G$ can be continued analytically to $x < -\tau$ so that $G$ is actually well-defined in and holomorphic over $\mathbb{C} \setminus [-\tau, \tau]$.

Finally, note that $G$ satisfies the reflection identity: $G(\overline{\zeta}) = \overline{G(z)}$. This is easy to verify from the definition of $G$, but it also follows from the Schwarz reflection principle: since $G$ is real and continuous on the real line (excluding the cut $[-\tau, \tau]$), so that $G$ can be continued analytically from the upper half-plane through the real interval $(\tau, +\infty)$ to the lower half-plane by reflection.

### 6.4.2 $G$ is a Pick function

To show that $G$ is a Pick function, we need to show that $\text{Im} f_{p,q}(A_1 + iA_2) > 0$ whenever $A_1 = A_1^*$ and $A_2 > 0$ (here $A_1 = \Delta + xA$ and $A_2 = yA$). From the reflection identity we then also have that $\text{Im} f_{p,q}(A_1 + iA_2) < 0$ whenever $A_2 < 0$; however, we do not actually need to invoke this identity. By homogeneity of $f_{p,q}$, we have $f_{p,q}(A) = f_{p,q}(-iA)/(-i) = f_{p,q}(iA)/i$, so that both conditions are equivalent to the one condition $\Re f_{p,q}(A) > 0$ whenever $A \in \mathcal{R}^+$.

The conditions on $p$ and $q$ are that $p, q \in (0, 1]$. Thus, if $A \in \mathcal{R}^+ = \mathcal{S}_{-\pi/2, \pi/2}$, then by Lemma 5, $A^p$ and $KA^qK^*$ are in the convex segments $\mathcal{S}_{-p\pi/2, q\pi/2}$ and $\mathcal{S}_{-q\pi/2, q\pi/2}$, respectively. Lemma 7 then shows that

$$\text{Sp}(A^pKA^qK^*) \subset \mathcal{S}_{-(p+q)\pi/2, (p+q)\pi/2}.$$  

On taking the $1/(p + q)^{th}$ power, this yields

$$\text{Sp}((A^pKA^qK^*)^{1/(p+q)}) \subset \mathcal{S}_{-\pi/2, \pi/2} = \mathcal{R}^+,$$

and, along with the analyticity properties previously discussed, this proves that $G$ is in the Pick class $\mathcal{P}(\tau, +\infty)$ with $\tau = ||\Delta||/\lambda_{\min}(A)$.

### 6.4.3 Concavity of $F$

From the theory of Pick functions it now follows that $G$ possesses the following integral representation (37, Section II.2):

$$G(\zeta) = a\zeta + b + \int_{-\tau}^{+\tau} \left( \frac{1}{t - \zeta} - \frac{t}{t^2 + 1} \right) d\mu(t).$$

We have already made use of this fact in Section 6.3 to represent fractional powers. The decisive factor is that $a$ is non-negative and $d\mu(t)$ is a positive Borel measure such that $\int_{-\tau}^{+\tau} (t^2 + 1)^{-1/2} d\mu(t)$ is finite. In fact, $a = \lim_{y \to +\infty} G(iy)/(iy)$, $b = \Re G(i)$ and the measure
is supported on the whole of $\mathbb{R}$ except where $G$ can be continued analytically from $I^+$ to $I^-$. On that subset, $\mu$ is determined by the so-called Stieltjes inversion formula:

$$\mu(b) - \mu(a) = \lim_{y \to 0^+} \frac{1}{\pi} \int_{(a,b)} \text{Im} G(x + iy) \, dx$$

for all $a < b$. From this it is clear why $a$ and $d\mu$ should be positive. Absorbing the integral $\int_{-\tau}^{+\tau} t(t^2 + 1)^{-1} d\mu(t)$ (which is finite as well) into the constant $b$ yields the simpler expression

$$G(\zeta) = a\zeta + b + \int_{-\tau}^{+\tau} \frac{1}{t - \zeta} d\mu(t).$$

From the relation $F(\zeta) = \zeta G(1/\zeta)$ it now follows that

$$F(x) = a + bx + \int_{-\tau}^{+\tau} \frac{x^2}{tx - 1} d\mu(t), \quad (47)$$

for $|x| < 1/\tau$. The second derivative of the kernel $x^2/(tx - 1)$ with respect to $x$ is $2(tx - 1)^{-3}$, which is negative for $|x| < 1/\tau < 1/|t|$. Hence the second derivative of $F$ is non-positive, which shows that $F$ is concave, as required. This ends the proof.

Remark. In fact, (47) is the canonical representation of a function that is operator concave on the interval $[-1/\tau, 1/\tau]$ (see e.g. eq. (V.31) in [32] for the canonical representation of an operator convex function; adding a minus sign yields (47)). Thus, not only is $F$ concave, it is actually operator concave, and as it is positive it is also operator monotone. More precisely, the function $Z \mapsto f_{p,q}(I \otimes A + Z \otimes \Delta; I \otimes K)$ is operator concave and operator monotone for fixed $A > 0$ and $\Delta = \Delta^*$, and with $K$ replaced by $I \otimes K$.

7 Discussion

In this paper we study a two-parameter family of relative entropies, which we call the $\alpha$-$z$-relative Rényi entropies ($\alpha$-$z$-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 quantum Rényi divergences ($\alpha$-QRD), and the quantum relative Rényi entropies. In particular, the data-processing inequality (DPI) for these quantities follows directly from DPI of the $\alpha$-$z$-RRE for suitable values of the parameters $\alpha$ and $z$.

The $\alpha$-QRD (or “sandwiched” relative Rényi entropy), which is a special case of the $\alpha$-$z$-RRE, has been the focus of much research of late. We study another special case of the $\alpha$-$z$-RRE, which we denote as $\hat{D}_\alpha$ (and informally call the reverse sandwiched relative Rényi 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 $\alpha$-$z$-RRE satisfy DPI in the orange regions of the $\alpha$-$z$-plane of Figure 1? (ii) Is the trace functional of the $\alpha$-$z$-RRE convex in the orange regions of Figure 2? (iii) Operational relevance of the $\alpha$-QRD for $\alpha \geq 1$.
has been established in quantum hypothesis testing [38], and in the context of the second
claws of quantum thermodynamics [39]. This raises the following question. Does \( \hat{\mathcal{D}}_\alpha \) also have
operational interpretations in quantum information theory (for \( 0 \leq \alpha \leq 1/2 \)) (other than
those arising through the symmetry relation (11))? 

A Proof of convergence of integral (42).

By Lemma 5,

\[
\text{Im}(-(t + C)^{-1}) = (t + C)^{-1} \text{Im}(t + C)(t + C)^{-\ast} = (t + C)^{-1} \text{Im}(C)(t + C)^{-\ast}.
\]

Now, let \( a \) and \( b \) be finite real numbers such that \( a \leq \text{Im} C \leq b \). Let \( \eta = \max(|a|, |b|) \). Then we get

\[
a(t + C)^{-1}(t + C)^{-\ast} \leq \text{Im}(-(t + C)^{-1}) \leq b(t + C)^{-1}(t + C)^{-\ast}.
\]

Since we assume that \( \text{Sp} C \) lies in the cut plane (otherwise, \( C^p \) would not be well-defined)
there exists a real positive number \( \epsilon > 0 \) such that every eigenvalue of \( C \) lies at a distance at
least \( \epsilon \) away from the cut. It follows that, for all positive \( t \), no eigenvalue of \( t + C \) lies in the
open disk around the origin of radius \( \epsilon \). Thus, \( C + t \) is invertible and, moreover, the spectrum
of \((C + t)^{-1}\) lies in the closed disk of radius \( 1/\epsilon \). Hence, the matrix \((t + C)^{-1}(t + C)^{-\ast}\), which
is a positive semidefinite matrix, is bounded above by \( 1/\epsilon^2 \), so that

\[
\|\text{Im}(-(t + C)^{-1})\| \leq \eta/\epsilon^2.
\]

For \( t \) tending to \( +\infty \), we need a bound that converges to 0. Let \( \gamma \) be the largest real
number such that \( \text{Re}(C - \gamma) \geq 0 \) (the smallest real part of the eigenvalues of \( C \)). Then for
all \( t > \gamma \), no eigenvalue of \( t + C \) is in the open disk of radius \( t - \gamma \). By the same reasoning
as before, we find that

\[
\|\text{Im}(-(t + C)^{-1})\| \leq \eta/(t - \gamma)^2.
\]

It is now easy to see that the integral (42) converges absolutely, as for \( t \) large enough, say \( t > T \), one has

\[
t^p \|\text{Im}(-(t + C)^{-1})\| \leq t^p \eta/(t - \gamma)^2 \leq c_1 t^{p-2},
\]

with \( c_1 = \eta T^2/(T - \gamma)^2 \), and for \( t < T \), one has

\[
t^p \|\text{Im}(-(t + C)^{-1})\| \leq t^p \eta/\epsilon^2 \leq c_2,
\]

with \( c_2 = T^p \eta/\epsilon^2 \). \qed

Acknowledgments

Thanks to Mr. Aidinyantz, alias Mr. A-to-Z, for inspiring the name of the new relative entropy.
We are also grateful to William Wootters for suggesting the idea of looking at “reverse
sandwiching”. This project was initiated during the program “Mathematical Challenges for
Quantum Information” at the Isaac Newton Institute, Cambridge, UK.
References

[1] B. Schumacher. Quantum Coding. Phys. Rev. A 51, 2738 (1995).

[2] A.S. Holevo. The capacity of the quantum channel with general signal states. IEEE Trans. Inf. Th. 44, (1998); B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. Phys. Rev. A 56, 131 (1997).

[3] S. Lloyd. Capacity of the noisy quantum channel. Phys. Rev. A 55, 1613 (1996); P. W. Shor. The quantum channel capacity and coherent information. MSRI Seminar, Nov. 2002 (unpublished); I. Devetak. The private classical capacity and quantum capacity of a quantum channel. IEEE Trans. Inf. Th. 51, 44 (2005).

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

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

[6] W. van Dam and P. Hayden. Rényi-entropic bounds on quantum communication. arXiv:quant-ph/0204093.

[7] K. G. H. Vollbrecht and M. M. Wolf. Conditional entropies and their relation to entanglement criteria. J. Math. Phys. 43, 4299 (2002).

[8] R. Renner. Security of quantum key distribution, PhD thesis. arXiv:quant-ph/0512258.

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

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

[11] M. Tomamichel. A Framework for Non-Asymptotic Quantum Information Theory. PhD Thesis, Department of Physics, ETH Zurich. arXiv:1203.2142.

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

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

[14] D. Petz. Quasi-entropies for finite quantum systems. Rep. Math. Phys. 23, 57 (1986).

[15] K.M.R. Audenaert. Quantum Skew Divergence: Theory and Applications. arXiv:1304.5935.

[16] C.C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. Jour. Stat. Phys. 52, 479 (1988).
[17] W. K. Wootters, *Found. Phys.* 20 1365–1378 (1990); R. Jozsa, D. Robb and W. K. Wootters. Lower bound for accessible information in quantum mechanics. *Phys. Rev. A* 49 668–677 (1994); N. Datta, T. Dorlas, R. Jozsa and F. Benatti. Properties of subentropy. arXiv:1310.1312.

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

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

[20] M.M. Wilde, A. Winter, D. Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels. arXiv:1306.1586.

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

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

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

[24] N. Datta and F. Leditzky. A limit of the quantum Rényi divergence. arXiv:1308.5961 (2013).

[25] 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).

[26] 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 on http://projecteuclid.org/euclid.bsmsp/1200512181.

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

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

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

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

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

[32] R. Bhatia. *Matrix Analysis*, Springer Verlag (1997).
[33] R.A. Horn and C.R. Johnson. *Matrix Analysis*, Cambridge (1985).
[34] C.W. Curtis, *Linear Algebra: an introductory approach*, Springer 1984; K. Hoffman and R. Kunze, *Linear Algebra*, Prentice Hall, 1971.
[35] T. Kato. *Perturbation theory for linear operators*, Springer Verlag (1980).
[36] H. Epstein. Remarks on two theorems of Lieb. *Commun. Math. Phys.* 31, 317–325 (1973).
[37] W.F. Donoghue Jr. *Monotone matrix functions and analytic continuation*, Springer Verlag (1974).
[38] M. Mosonyi and T. Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. [arXiv:1309.3228](https://arxiv.org/abs/1309.3228).
[39] 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](https://arxiv.org/abs/1305.5278) (2013).