Inequalities for the quantum Rényi divergences 
with applications to compound coding problems

Milán Mosonyi

Física Teòrica: Informació i Fenomenes Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain.
Mathematical Institute, Budapest University of Technology and Economics
Egry József u 1., Budapest, 1111 Hungary

Abstract

We show two-sided bounds between the traditional quantum Rényi divergences and the new notion of Rényi divergences introduced recently in Müller-Lennert, Dupuis, Szehr, Fehr and Tomamichel, J. Math. Phys. 54, 122203, (2013), and Wilde, Winter, Yang, arXiv:1306.1586. The bounds imply that the two versions can be used interchangeably near \( \alpha = 1 \), and hence one can benefit from the best properties of both when proving coding theorems in the case of asymptotically vanishing error. We illustrate this by giving short and simple proofs of the quantum Stein’s lemma with composite null-hypothesis, universal source compression, and the achievability part of the classical capacity of compound quantum channels. Apart from the above interchangeability, we benefit from a weak quasi-concavity property of the new Rényi divergences that we also establish here.

1 Introduction

Rényi introduced a generalization of the Kullback-Leibler divergence (relative entropy) in [49]. According to his definition, the \( \alpha \)-divergence of two probability distributions \( p \) and \( q \) on a finite set \( \mathcal{X} \) for a parameter \( \alpha \in [0, +\infty) \setminus \{1\} \) is given by

\[
D_{\alpha} (p\|q) := \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} p(x)^\alpha q(x)^{1-\alpha}.
\]  

The limit \( \alpha \to 1 \) yields the standard relative entropy. These quantities turned out to play a central role in information theory and statistics; indeed, the Rényi divergences quantify the trade-off between the exponents of the relevant quantities in many information-theoretic tasks, including hypothesis testing, source coding and noisy channel coding; see, e.g. [34] for an overview of these results. It was also shown in [34] that the Rényi relative entropies, and other related quantities, like the Rényi entropies and the Rényi capacities, have direct operational interpretations as so-called generalized cutoff rates in the corresponding information-theoretic tasks.
In quantum theory, the state of a system is described by a density operator instead of a probability distribution, and the definition (1) can be extended for pairs of density operators in various inequivalent ways, due to the non-commutativity of operators. The traditional way to define the Rényi divergence of two density operators is

\[
D^{(\text{old})}_\alpha (\rho\|\sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \rho^n \sigma^{1-\alpha}.
\]

(2)

It has been shown in [38] that, similarly to the classical case, the Rényi \(\alpha\)-divergences \(D^{(\text{old})}_\alpha\) with \(\alpha \in (0, 1)\) have a direct operational interpretation as generalized cutoff rates in the so-called direct domain of binary state discrimination. This is a consequence of another, indirect, operational interpretation in the setting of the quantum Hoeffding bound [5, 21, 23, 42].

Recently, a new quantum extension of the Rényi \(\alpha\)-divergences has been proposed in [40, 56], defined as

\[
D^{(\text{new})}_\alpha (\rho\|\sigma) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha.
\]

(3)

This definition was introduced in [40] as a parametric family that connects the min- and max-relative entropies [16, 48] and Umegaki’s relative entropy [55]. In [56], the corresponding generalized Holevo capacities were used to establish the strong converse property for the classical capacities of entanglement-breaking and Hadamard channels. It was shown in [39] that these new Rényi divergences play the same role in the (strong) converse problem of binary state discrimination as the traditional Rényi divergences in the direct problem. In particular, the strong converse exponent was expressed as a function of the new Rényi divergences, and from that a direct operational interpretation was derived for them as generalized cutoff rates in the sense of [14].

The above results suggest that, somewhat surprisingly, one should use two different quantum extensions of the classical Rényi divergences: for the direct part, corresponding to \(\alpha \in (0, 1)\), the “right” definition is the one given in (2), while for the converse part, corresponding to \(\alpha > 1\), the “right” definition is the one in (3). Although coding theorems supporting this separation have only been shown for binary state discrimination so far, it seems reasonable to expect the same separation in the case of other information-theoretic tasks. We remark that, in line with this expectation, lower bounds on the classical capacity of quantum channels can be obtained in terms of the traditional Rényi divergences [37], while upper bounds were found in terms of the new Rényi divergences in [56].

On the other hand, the above two quantum Rényi divergences have different mathematical properties, which might make them better or worse suited for certain mathematical manipulations, and therefore it might be beneficial to use the new Rényi divergences in the direct part of coding problems, and the traditional ones in converse parts, despite the “real” quantities being the opposite. The problem that one faces then is how to arrive back to the natural quantity of the given problem. As it turns out, this is possible, at least if one’s aim is to study the case of asymptotically vanishing error, corresponding to \(\alpha \to 1\); this is thanks to the well-known Araki-Lieb-Thirring inequality, and its complement due to Audenaert [6]. We explain this in detail in Section 3.1.
Convexity properties of these divergences are of particular importance for applications. As it was shown in [18, 56], both versions of the Rényi divergences are jointly quasi-convex around $\alpha = 1$. In Section 3.2 we show a certain converse to this quasi-convexity in the form of a weak partial quasi-concavity (Corollary 3.15 and Proposition 3.17), which is still strong enough to be useful for applications, as we illustrate on various examples in Section 4.

Coding theorems for the problems considered in Section 4 have been established in [9, 44] for Stein's lemma with composite null-hypothesis, in [30] for universal source compression, and in [11, 15] for the classical capacity of compound and averaged channels. Here we provide alternative proofs for these coding theorems, using the following general approach:

1. We take a single-shot coding theorem that bounds the relevant error probability in terms of a Rényi divergence. In the case of Stein's lemma and source compression, this is Audenaert's inequality [4], while in the case of channel coding, we use the random coding theorem due to Hayashi and Nagaoka [19]. The bounds are given in terms of $D^{(\alpha)}_{\text{old}}$.

2. We use lemma 3.3 to switch from the old to the new Rényi divergences in the upper bound to the error probability, and then we use the weak partial quasi-concavity properties of the Rényi divergences, given in Corollary 3.15 and Proposition 3.17 to decouple the upper bound into a sum of individual Rényi divergences.

3. If necessary, we use again lemma 3.3 to return to $D^{(\alpha)}_{\text{old}}$ in the upper bound.

4. We use the additivity of the relevant Rényi quantities (divergences, entropies, generalized Holevo quantities) to obtain the asymptotics.

The advantage the above approach is that it only uses very general arguments that are largely independent of the concrete model in consideration. Once the single-shot coding theorems are available, the coding theorems for the composite cases follow essentially by the same amount of effort as for the simple cases (simple null-hypothesis, single source, single channel), using only very general properties of the Rényi divergences. This makes the proofs considerably shorter and simpler than e.g., in [9, 11, 15]. Moreover, this approach is very easy to generalize to non-i.i.d. compound problems, unlike the methods of [30, 44], which are based on the method of types.

2 Notations

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})_+$ denote the set of all non-zero positive semidefinite operators on $\mathcal{H}$, and let $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H})_+ ; \text{Tr } \rho = 1\}$ be the set of all density operators (states) on $\mathcal{H}$.

We define the powers of a positive semidefinite operator $A$ only on its support; that is, if $\lambda_1, \ldots, \lambda_r$ are the strictly positive eigenvalues of $A$, with corresponding spectral projections $P_1, \ldots, P_r$, then we define $A^\alpha := \sum_{i=1}^r \lambda_i^\alpha P_i$ for all $\alpha \in \mathbb{R}$. In particular, $A^0 = \sum_{i=1}^r P_i$ is the projection onto the support of $A$.

We will use the convention $\log 0 := -\infty$ and $\log +\infty := +\infty$. 
3 Rényi divergences

3.1 Two definitions

For non-zero positive semidefinite operators \( \rho, \sigma \), the Rényi \( \alpha \)-divergence \( D_{\alpha} \) of \( \rho \) w.r.t. \( \sigma \) with parameter \( \alpha \in (0, +\infty) \setminus \{1\} \) is traditionally defined as

\[
D_{\alpha}^{(\text{old})} (\rho \| \sigma) := \begin{cases} 
\frac{1}{1-\alpha} \log \text{Tr} \rho^\alpha \sigma^{1-\alpha} - \frac{1}{1-\alpha} \log \text{Tr} \rho, & \alpha \in (0, 1) \text{ or } \text{supp } \rho \subseteq \text{supp } \sigma, \\
+\infty, & \text{otherwise}.
\end{cases}
\]

For the mathematical properties of \( D_{\alpha}^{(\text{old})} \), see, e.g. \[32, 38, 47\]. Recently, a new notion of Rényi divergence has been introduced in \[40, 56\], defined as

\[
D_{\alpha}^{(\text{new})} (\rho \| \sigma) := \begin{cases} 
\frac{1}{1-\alpha} \log \left( \sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right) - \frac{1}{1-\alpha} \log \text{Tr} \rho, & \alpha \in (0, 1) \text{ or } \text{supp } \rho \subseteq \text{supp } \sigma, \\
+\infty, & \text{otherwise}.
\end{cases}
\]

For the mathematical properties of \( D_{\alpha}^{(\text{new})} \), see, e.g. \[8, 18, 39, 40, 56\].

Remark 3.1. It is easy to see that for non-zero \( \rho \), we have \( \lim_{\sigma \to 0} D_{\alpha}^{(\text{old})} (\rho \| \sigma) = \lim_{\rho \to 0} D_{\alpha}^{(\text{new})} (\rho \| \sigma) = +\infty \), and hence we define \( D_{\alpha}^{(\text{old})} (\rho \| 0) := D_{\alpha}^{(\text{new})} (\rho \| 0) := +\infty \) when \( \rho \neq 0 \). On the other hand, for non-zero \( \sigma \), the limits \( \lim_{\rho \to 0} D_{\alpha}^{(\text{old})} (\rho \| \sigma) \) and \( \lim_{\rho \to 0} D_{\alpha}^{(\text{new})} (\rho \| \sigma) \) don’t exist, and hence we don’t define the values of \( D_{\alpha}^{(\text{old})} (0 \| \sigma) \) and \( D_{\alpha}^{(\text{new})} (0 \| \sigma) \). To see the latter, one can consider \( \rho_n := \frac{1}{n} |0\rangle \langle 0| + \frac{1}{n^2} |1\rangle \langle 1| \), and \( \sigma := |1\rangle \langle 1| \), where \( |0\rangle \langle 0| \) and \( |1\rangle \langle 1| \) are orthogonal rank 1 projections. It is easy to see that for \( \alpha < 1 \), \( \lim_{n \to +\infty} D_{\alpha}^{(\text{old})} (\rho_n \| \sigma) = \lim_{n \to +\infty} D_{\alpha}^{(\text{new})} (\rho_n \| \sigma) = \lim_{n \to +\infty} \frac{1}{1-\alpha} \log \frac{n^{1-\beta \alpha}}{1+n^{1-\beta \alpha}} \) depends on the value of \( \beta \). A similar example can be used for \( \alpha > 1 \).

Remark 3.2. Note that the definition of \( D_{\alpha}^{(\text{old})} \) makes sense also for \( \alpha = 0 \), and we get \( D_0 (\rho \| \sigma) = -\log \text{Tr} \rho^0 \sigma \). It is easy to see that if \( \text{supp } \rho \subseteq \text{supp } \sigma \) then

\[
D_{\alpha}^{(\text{old})} (\rho \| \sigma) := \lim_{\alpha \to +\infty} D_{\alpha}^{(\text{old})} (\rho \| \sigma) = \max \{ r/s : \text{Tr} P_\rho (\{r\}) P_\sigma (\{s\}) > 0 \},
\]

where \( P_\rho (\{r\}) \) and \( P_\sigma (\{s\}) \) are the spectral projections of \( \rho \) and \( \sigma \) corresponding to \( r \) and \( s \), respectively. If \( \text{supp } \rho \not\subseteq \text{supp } \sigma \) then obviously \( D_{\alpha}^{(\text{old})} (\rho \| \sigma) = +\infty \). In the case of \( D_{\alpha}^{(\text{new})} \), it was shown in \[10\] that

\[
D_{\alpha}^{(\text{new})} (\rho \| \sigma) := \lim_{\alpha \to +\infty} D_{\alpha}^{(\text{new})} (\rho \| \sigma) = D_{\text{max}} (\rho \| \sigma) := \log \inf \{ \gamma : \rho \leq \gamma \sigma \},
\]

where \( D_{\text{max}} \) is the max-relative entropy \[16, 48\]. The limit \( D_0^{(\text{new})} (\rho \| \sigma) := \lim_{\alpha \to 0} D_{\alpha}^{(\text{new})} (\rho \| \sigma) \) is in general different from \( D_0^{(\text{old})} (\rho \| \sigma) \); see, e.g. \[17, 17\].

According to the Araki-Lieb-Thirring inequality \[8, 33\], for any positive semidefinite operators \( A, B \),

\[
\text{Tr} A^\alpha B^n A^\alpha \leq \text{Tr} (ABA)^\alpha
\]

(4)
for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. A converse to the Araki-Lieb-Thirring inequality was given in [6], where it was shown that

$$
\text{Tr}(ABA)^\alpha \leq \left(\|B\|^\alpha \text{Tr} A^{2\alpha}\right)^{1-\alpha} \left(\text{Tr} A^{\alpha} B^{\alpha} A^{\alpha}\right)^\alpha
$$

for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. Applying (4) and (5) to $A := \rho^{\frac{1}{2}}$ and $B := \sigma^{\frac{1}{2} - \alpha}$, we get

$$
\text{Tr} \rho^{\alpha} \sigma^{1-\alpha} \leq \text{Tr} \left(\rho^{\frac{1}{2}} \sigma^{\frac{1}{2} - \alpha} \rho^{\frac{1}{2}}\right)^\alpha \leq \|\sigma\|^{(1-\alpha)^2} \left(\text{Tr} \rho^{\alpha}\right)^{1-\alpha} \left(\text{Tr} \rho^{\alpha} \sigma^{1-\alpha}\right)^\alpha
$$

for $\alpha \in (0,1)$, and the inequalities hold in the converse direction for $\alpha > 1$. This immediately yields the following:

**Lemma 3.3.** For any $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$ and $\alpha \in [0, +\infty) \setminus \{1\}$,

$$
D_{\alpha}^{(\text{old})} (\rho\|\sigma) \geq D_{\alpha}^{(\text{new})} (\rho\|\sigma) \geq \alpha D_{\alpha}^{(\text{old})} (\rho\|\sigma) + \log \text{Tr} \rho - \log \text{Tr} \rho^{\alpha} + (\alpha - 1) \log \|\sigma\|.
$$

**Remark 3.4.** The first inequality in (7) has already been noted in [56] for $\alpha > 1$.

It is straightforward to verify that $D_{\alpha}^{(\text{old})}$ yields Umegaki’s relative entropy in the limit $\alpha \to 1$; i.e., for any $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$,

$$
D_1 (\rho\|\sigma) := \lim_{\alpha \to 1} D_{\alpha}^{(\text{old})} (\rho\|\sigma) = \begin{cases} \frac{1}{\text{Tr} \rho} \text{Tr} (\log \rho - \log \sigma), & \text{supp} \rho \subseteq \text{supp} \sigma, \\ +\infty, & \text{otherwise.} \end{cases}
$$

This, together with lemma 3.3, yields immediately the following:

**Corollary 3.5.** For any two non-zero positive semidefinite operator $\rho, \sigma$,

$$
\lim_{\alpha \to 1} D_{\alpha}^{(\text{new})} (\rho\|\sigma) = D_1 (\rho\|\sigma).
$$

Taking into account (8)–(9) and Remark 3.2 we finally have the definitions of $D_{\alpha}^{(\text{old})}$ and $D_{\alpha}^{(\text{new})}$ for every parameter value $\alpha \in [0, +\infty]$.

**Remark 3.6.** The limit relation (9) has been shown in [40], and in [56] for $\alpha \searrow 1$, by explicitly computing the derivative of $\alpha \mapsto \log \text{Tr} \left(\rho^{\frac{1}{2}} \sigma^{\frac{1}{2} - \alpha} \rho^{\frac{1}{2}}\right)^\alpha$ at $\alpha = 1$.

It is easy to see (by computing its second derivative) that $\psi^{(\text{old})}(\alpha) := \log \text{Tr} \rho^{\alpha} \sigma^{1-\alpha}$ is a convex function of $\alpha$, which yields immediately that $D_{\alpha}^{(\text{old})} (\rho\|\sigma)$ is a monotonic increasing function of $\alpha$ for any fixed $\rho$ and $\sigma$. The following Proposition, due to [53] and [54], complements this monotonicity property around $\alpha = 1$, and in the same time gives a quantitative version of (8):

**Proposition 3.7.** Let $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$ be such that $\text{supp} \rho \subseteq \text{supp} \sigma$, let $\eta := 1 + \text{Tr} \rho^{3/2} \sigma^{1/2} - \text{Tr} \rho^{1/2} \sigma^{1/2}$, let $c > 0$, and $\delta := \min \left\{\frac{1}{2}, \frac{c}{2 \log \eta}\right\}$. Then

$$
D_1 (\rho\|\sigma) \geq D_{\alpha}^{(\text{old})} (\rho\|\sigma) \geq D_1 (\rho\|\sigma) - 4(1 - \alpha)(\log \eta)^2 \cosh c, \quad 1 - \delta < \alpha < 1,
$$

$$
D_1 (\rho\|\sigma) \leq D_{\alpha}^{(\text{old})} (\rho\|\sigma) \leq D_1 (\rho\|\sigma) - 4(1 - \alpha)(\log \eta)^2 \cosh c, \quad 1 < \alpha < 1 + \delta.
$$
The new Rényi divergences $D_\alpha^{(\text{new})}(\rho\|\sigma)$ are also monotonic increasing in $\alpha$, as was shown in in Theorem 6 of [40] (see also [39] for a different proof for the case $\alpha > 1$). Combining Proposition 3.7 with lemma 3.3 we obtain the following:

**Corollary 3.8.** In the setting of Proposition 3.7 we have

$$D_1(\rho\|\sigma) \geq D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_1(\rho\|\sigma) - 4\alpha(1-\alpha)(\log \eta)^2 \cosh c$$

$$+ \log \tr \rho - \log \tr \rho^\alpha + (1-\alpha) \log \|\sigma\|^{-1}, \quad 1 - \delta < \alpha < 1,$$

$$D_1(\rho\|\sigma) \leq D_\alpha^{(\text{new})}(\rho\|\sigma) \leq D_1(\rho\|\sigma) - 4(1-\alpha)(\log \eta)^2 \cosh c,$$

$$1 < \alpha < 1 + \delta.$$

**Remark 3.9.** The inequalities in the second line above have already appeared in [56].

Finally, we consider Lemma 3.3 in some special cases. Note that the monotonicity of the Rényi divergences in $\alpha$ yields that the Rényi entropies

$$S_\alpha(\rho) := -D_\alpha^{(\text{old})}(\rho\|I) = -D_\alpha^{(\text{new})}(\rho\|I) = \frac{1}{1-\alpha} \log \tr \rho^\alpha - \frac{1}{1-\alpha} \log \tr \rho$$

are monotonic decreasing in $\alpha$ for any fixed $\rho$, and hence,

$$\tr \rho^\alpha \leq (\tr \rho^0)^{(1-\alpha)} (\tr \rho)^\alpha$$

(10)

for every $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$.

Assume that $\alpha \in (0,1)$. Plugging (10) into (6), we get that for any $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$,

$$\tr \rho^\alpha \sigma^{1-\alpha} \leq \tr \left( \rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{2}} \rho^{\frac{1}{2}} \right)^\alpha \leq \|\sigma\|^{(1-\alpha)^2} (\tr \rho^0)^{(1-\alpha)^2} (\tr \rho)^{(1-\alpha)(1-\alpha)} (\tr \rho^0)^{(1-\alpha)(1-\alpha)}$$

for every $\alpha \in (0,1)$. This in turn yields that for every $\alpha \in (0,1)$,

$$D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\text{old})}(\rho\|\sigma) + (1-\alpha) \left( \log \tr \rho - \log \tr \rho^0 - \log \|\sigma\| \right).$$

In particular, if $\|\sigma\| \leq 1$ then

$$D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\text{old})}(\rho\|\sigma) + (1-\alpha) \left( \log \tr \rho - \log \tr \rho^0 \right).$$

(12)

Assume now that $\alpha > 1$. Then $\tr (\rho/\|\rho\|)^\alpha \leq \tr (\rho/\|\rho\|)$, and plugging it into (7) yields

$$D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\text{old})}(\rho\|\sigma) + (\alpha - 1) \left( \log \|\sigma\| - \log \|\rho\| \right).$$

In particular, if $\|\rho\| \leq 1$ then $\tr \sigma \leq \|\sigma\| \tr \sigma^0$ yields

$$D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\text{old})}(\rho\|\sigma) + (\alpha - 1) \left( \log \tr \sigma - \log \tr \sigma^0 \right).$$

(13)

**Corollary 3.10.** Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be density operators. For every $\alpha \in [0, +\infty)$,

$$D_\alpha^{(\text{old})}(\rho\|\sigma) \geq D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\text{old})}(\rho\|\sigma) - |\alpha - 1| \log(\dim \mathcal{H}).$$

**Proof.** Immediate from Lemma 3.3 (12) and (13).

(13)

Corollary 3.10 together with Proposition 3.7 yield the following version of Corollary 3.8 when $\rho$ and $\sigma$ are states:

**Corollary 3.11.** Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be density operators. With the notations of Proposition 3.7 we have

$$D_1(\rho\|\sigma) \geq D_\alpha^{(\text{new})}(\rho\|\sigma) \geq \alpha D_1(\rho\|\sigma) - (1-\alpha) \left[ 4\alpha(\log \eta)^2 \cosh c + \log(\dim \mathcal{H}) \right].$$

for every $1 - \delta < \alpha < 1$. 

6
3.2 Convexity properties

The general concavity result in [26, Theorem 2.1] implies as a special case that the quantity
\[ Q^{(\text{new})}_\alpha(\rho\|\sigma) := \text{Tr}\left(\sigma^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1-\alpha}{\alpha}}\right)^\alpha = \text{Tr}\left(\rho^\frac{1}{\alpha}\sigma^{\frac{1-\alpha}{\alpha}}\rho^{\frac{1}{\alpha}}\right)^\alpha \]  
(14)
is jointly concave for \( \alpha \in [1/2, 1) \). (See also [18] for a different proof of this). In [10, 56], joint convexity of \( Q^{(\text{new})}_\alpha \) was shown for \( \alpha \in [1, 2] \), which was later extended in [18], using a different proof method, to all \( \alpha > 1 \). That is, if \( \rho_i, \sigma_i \in \mathcal{B}(\mathcal{H})_+ \), \( i = 1, \ldots, r \), and \( \gamma_1, \ldots, \gamma_r \) is a probability distribution on \([r] := \{1, \ldots, r\}\), then
\[
Q^{(\text{new})}_\alpha \left( \sum_i \gamma_i \rho_i \right) \left( \sum_i \gamma_i \sigma_i \right) \geq \sum_i \gamma_i Q^{(\text{new})}_\alpha(\rho_i\|\sigma_i), \quad \frac{1}{2} \leq \alpha < 1, \tag{15}
\]
\[
Q^{(\text{new})}_\alpha \left( \sum_i \gamma_i \rho_i \right) \left( \sum_i \gamma_i \sigma_i \right) \leq \sum_i \gamma_i Q^{(\text{new})}_\alpha(\rho_i\|\sigma_i), \quad 1 < \alpha. \tag{16}
\]
(For the second inequality one also has to assume that \( \text{supp} \rho_i \subseteq \text{supp} \sigma_i \) for all \( i \).) This yields immediately that the Rényi divergences \( D^{(\text{new})}_\alpha \) are jointly quasi-convex for \( \alpha > 1 \) (see [56] for \( \alpha \in (1, 2] \)), and jointly convex for \( \alpha \in [1/2, 1) \) when restricted to \( \{\rho \in \mathcal{B}(\mathcal{H})_+ : \text{Tr} \rho = t\} \times \mathcal{B}(\mathcal{H})_+ \) for any fixed \( t > 0 \) [18].

Our goal here is to complement these inequalities to some extent. The following lemma is a special case of the famous Rotfel’d inequality (see, e.g., Section 4.5 in [25]). Below we provide an elementary proof for \( \alpha \in [0, 2] \).

**Lemma 3.12.** The function \( A \mapsto \text{Tr} A^\alpha \) is subadditive on positive semidefinite operators for every \( \alpha \in [0, 1] \), and superadditive for \( \alpha \geq 1 \). That is, if \( A, B \in \mathcal{B}(\mathcal{H})_+ \) then
\[
\text{Tr}(A + B)^\alpha \leq \text{Tr} A^\alpha + \text{Tr} B^\alpha, \quad \alpha \in [0, 1], \tag{17}
\]
\[
\text{Tr}(A + B)^\alpha \geq \text{Tr} A^\alpha + \text{Tr} B^\alpha, \quad 1 \leq \alpha. \tag{18}
\]

**Proof.** We only prove the case \( \alpha \in [0, 2] \). Assume first that \( A \) and \( B \) are invertible and let \( \alpha \in (0, 1) \). Then
\[
\text{Tr}(A + B)^\alpha - \text{Tr} A^\alpha = \int_0^1 \frac{d}{dt} \text{Tr}(A + tB)^\alpha dt = \int_0^1 \alpha \text{Tr} B(A + tB)^{\alpha-1} dt
\]
\[
\leq \int_0^1 \alpha \text{Tr} B(tB)^{\alpha-1} dt = \text{Tr} B^\alpha \int_0^1 \alpha t^{\alpha-1} dt = \text{Tr} B^\alpha,
\]
where in the first line we used the identity \( (d/dt) \text{Tr} f(A + tB) = \text{Tr} B f'(A + tB) \), and the inequality follows from the fact that \( x \mapsto x^{\alpha-1} \) is operator monotone decreasing on \((0, +\infty)\) for \( \alpha \in (0, 1) \). This proves (17) for invertible \( A \) and \( B \), and the general case follows by continuity. The proof for the case \( \alpha \in (1, 2] \) goes the same way, using the fact that \( x \mapsto x^{\alpha-1} \) is operator monotone increasing on \((0, +\infty)\) for \( \alpha \in (1, 2] \). The case \( \alpha = 1 \) is trivial, and the case \( \alpha = 0 \) follows by taking the limit \( \alpha \to 0 \) in (17). \(\square\)
Proposition 3.13. Let $\sigma, \rho_1, \ldots, \rho_r \in \mathcal{B}(\mathcal{H})_+$, and $\gamma_1, \ldots, \gamma_r$ be a probability distribution on $[r]$. We have

$$\sum_{i} \gamma_i Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma) \leq Q^{(\text{new})}_{\alpha}\left(\sum_{i} \gamma_i \rho_i \| \sigma\right) \leq \sum_{i} \gamma_i^\alpha Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma), \quad 0 < \alpha < 1, \quad (19)$$

$$\sum_{i} \gamma_i Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma) \geq Q^{(\text{new})}_{\alpha}\left(\sum_{i} \gamma_i \rho_i \| \sigma\right) \geq \sum_{i} \gamma_i^\alpha Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma), \quad 1 < \alpha. \quad (20)$$

Moreover, the second inequalities in (19) and (20) are valid for arbitrary non-negative $\gamma_1, \ldots, \gamma_r$ with $\gamma_1 + \ldots + \gamma_r > 0$.

Proof. By lemma 3.12 we have

$$\operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\left(\sum_{i=1}^{r} \gamma_i \rho_i\right)^{\frac{1-\alpha}{2\alpha}}\sigma^{\frac{1-\alpha}{2\alpha}}\right) \leq \sum_{i} \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\gamma_i \rho_i \sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha} = \sum_{i} \gamma_i^\alpha \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho_i \sigma^{\frac{1-\alpha}{2\alpha}}\right)^{\alpha}$$

for $\alpha \in (0, 1)$, and the inequality is reversed for $\alpha > 1$, which proves the second inequalities in (19) and (20).

The first inequalities follow the same way, by noting that $A \mapsto \operatorname{Tr} A^\alpha$ is concave for $\alpha \in (0, 1)$ and convex for $\alpha > 1$. \qed

Remark 3.14. Note that the first inequality in (20) follows from the joint convexity of $Q^{(\text{new})}_{\alpha}$, and the first inequality in (19) can be obtained from the joint concavity of $Q^{(\text{new})}_{\alpha}$ for $1/2 \leq \alpha < 1$; however, not for the range $0 < \alpha < 1/2$, where joint concavity fails.

Corollary 3.15. Let $\sigma, \rho_1, \ldots, \rho_r \in \mathcal{B}(\mathcal{H})_+$, and $\gamma_1, \ldots, \gamma_r$ be a probability distribution on $[r]$. For every $\alpha \in [0, +\infty]$,

$$\min_{i} D^{(\text{new})}_{\alpha}(\rho_i \| \sigma) + \log \min_{i} \gamma_i \leq D^{(\text{new})}_{\alpha}\left(\sum_{i=1}^{r} \gamma_i \rho_i \| \sigma\right) \leq \max_{i} D^{(\text{new})}_{\alpha}(\rho_i \| \sigma).$$

Proof. We prove the inequalities for $\alpha \in (1, +\infty)$; the proof for $\alpha \in (0, 1)$ goes exactly the same way, and the cases $\alpha = 0, 1, +\infty$ follow by taking the corresponding limit in $\alpha$. By the first inequality in (20), we have

$$D^{(\text{new})}_{\alpha}\left(\sum_{i=1}^{r} \gamma_i \rho_i \| \sigma\right) = \frac{1}{\alpha - 1} \log \frac{Q^{(\text{new})}_{\alpha}\left(\sum_{i=1}^{r} \gamma_i \rho_i \| \sigma\right)}{\sum_{i=1}^{r} \gamma_i \operatorname{Tr} \rho_i} \leq \frac{1}{\alpha - 1} \log \frac{\sum_{i=1}^{r} \gamma_i Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma)}{\sum_{i=1}^{r} \gamma_i \operatorname{Tr} \rho_i}$$

$$\leq \frac{1}{\alpha - 1} \log \min_{i} Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma),$$

proving the second inequality of the assertion. The second inequality in (20) yields

$$D^{(\text{new})}_{\alpha}\left(\sum_{i=1}^{r} \gamma_i \rho_i \| \sigma\right) = \frac{1}{\alpha - 1} \log \frac{Q^{(\text{new})}_{\alpha}\left(\sum_{i=1}^{r} \gamma_i \rho_i \| \sigma\right)}{\sum_{i=1}^{r} \gamma_i \operatorname{Tr} \rho_i} \geq \frac{1}{\alpha - 1} \log \frac{\sum_{i=1}^{r} \gamma_i^\alpha Q^{(\text{new})}_{\alpha}(\rho_i \| \sigma)}{\sum_{i=1}^{r} \gamma_i^\alpha \operatorname{Tr} \rho_i}.$$
We have
\[ \gamma_i^\alpha Q_\alpha^{(\text{new})}(\rho_i∥\sigma) \geq (\gamma_i \Tr \rho_i)\gamma_i^{α-1} \min_j \gamma_j^\alpha Q_\alpha^{(\text{new})}(\rho_j∥\sigma) \geq \gamma_i \Tr \rho_i \left( \min_j \gamma_j^{α-1} \right) \min_j Q_\alpha^{(\text{new})}(\rho_j∥\sigma) \],
and summing over \( i \) yields that
\[ \frac{1}{α - 1} \log \sum_i γ_i^α Q_α^{(\text{new})}(\rho_i∥\sigma) \geq \frac{1}{α - 1} \log \min_j Q_α^{(\text{new})}(\rho_j∥\sigma) + \log \min_j γ_j, \]
as required.

**Remark 3.16.** Note that the inequalities in (15) and (16) express joint concavity/convexity, whereas in the complements given in Proposition 3.13 and Corollary 3.15 we only took a convex combination in the first variable and not in the second. It is easy to see that this restriction is in fact necessary. Indeed, let \( ρ_1 := σ_2 := |x⟩⟨x| \) and \( ρ_2 := σ_1 := |y⟩⟨y| \), where \( x \) and \( y \) are orthogonal unit vectors in some Hilbert space. If we choose \( γ_1 = γ_2 = 1/2 \) then \( \sum_i γ_i ρ_i = \sum_i γ_i σ_i \), and hence
\[ D_α^{(\text{new})} \left( \sum_{i=1}^r γ_i ρ_i || \sum_{i=1}^r γ_i σ_i \right) = 0, \]
while \( D_α^{(\text{new})}(ρ_1∥σ_1) = D_α^{(\text{new})}(ρ_2∥σ_2) = +∞, \)
and hence no inequality of the form \( D_α^{(\text{new})} \left( \sum_{i=1}^r γ_i ρ_i || \sum_{i=1}^r γ_i σ_i \right) ≥ c_1 \min_i D_α^{(\text{new})}(ρ_i∥σ_i) - c_2 \) can hold for any positive constants \( c_1 \) and \( c_2 \).

The quantity
\[ Q_α^{(\text{old})}(ρ∥σ) := \Tr ρ^α σ^{1-α} \]
is jointly concave for \( α ∈ (0, 1) \) according to Lieb’s concavity theorem [32], and jointly convex for \( α ∈ (1, 2] \) according to Ando’s convexity theorem [1]; see also [47] for a different proof of both. That is, if \( ρ_i, σ_i ∈ B(H)_+, i = 1, \ldots, r, \) and \( γ_1, \ldots, γ_r \) is a probability distribution on \( [r] := \{1, \ldots, r\} \), then
\[ Q_α^{(\text{old})} \left( \sum_i γ_i ρ_i || \sum_i γ_i σ_i \right) ≥ \sum_i γ_i Q_α^{(\text{old})}(ρ_i∥σ_i), \ 0 ≤ α < 1, \]
\[ Q_α^{(\text{old})} \left( \sum_i γ_i ρ_i || \sum_i γ_i σ_i \right) ≤ \sum_i γ_i Q_α^{(\text{old})}(ρ_i∥σ_i), \ 1 ≤ α ≤ 2. \]
(For the second inequality, one has to assume that \( \supp ρ_i ⊆ \supp σ \) for all \( i \).) Note the difference in the ranges of joint convexity/concavity as compared to (15) and (16).

This yields immediately that \( D_α^{(\text{old})} \) is jointly convex for \( α ∈ (0, 1) \) when restricted to \( \{ρ ∈ B(H)_+ : \Tr ρ = t\} × B(H)_+ \) for any fixed \( t > 0 \), and it is jointly quasi-convex for \( α ∈ (1, 2] \). Moreover, it is convex in its second argument for \( α ∈ (1, 2] \), according to Theorem II.1 in [33]; see also Proposition 1.1 in [2]. It is not clear whether a subadditivity argument can be used to complement the above concavity/convexity properties. However, one can use the bounds for \( Q_α^{(\text{new})} \) and \( D_α^{(\text{new})} \) together with lemma 3.3 to obtain the following:
Proposition 3.17. Let $\sigma, \rho_1, \ldots, \rho_r \in B(H)_+$, and $\gamma_1, \ldots, \gamma_r$ be a probability distribution on $[r]$. We have
\begin{equation}
Q_\alpha^{(\text{old})} \left( \sum_i \gamma_i \rho_i \parallel \sigma \right) \leq \sum_i \gamma_i \alpha Q_\alpha^{(\text{old})} (\rho_i \parallel \sigma)^\alpha \parallel \sigma \parallel^{(1-\alpha)^2} (\text{Tr} \, \rho_i^\alpha)^{1-\alpha} \tag{21}
\end{equation}
for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. As a consequence,
\begin{equation}
D_\alpha^{(\text{old})} \left( \sum_i \gamma_i \rho_i \parallel \sigma \right) \geq \alpha \min_i D_\alpha^{(\text{old})} (\rho_i \parallel \sigma) + (\alpha - 1) \log \parallel \sigma \parallel + \log \min_i \left\{ \frac{\text{Tr} \, \rho_i}{\text{Tr} \, \rho_i^\alpha} \right\} \tag{22}
\end{equation}
for all $\alpha \in (0, +\infty) \setminus \{1\}$.

Proof. The inequality in (21) is immediate from (6) and Proposition 3.13. The same argument as in the proof of Corollary 3.15 yields (22). \qed

Remark 3.18. For $\alpha \in (0,1)$, we can use (10) to further bound the RHS of (22) from below and get
\begin{equation}
D_\alpha^{(\text{old})} \left( \sum_i \gamma_i \rho_i \parallel \sigma \right) \geq \alpha \min_i D_\alpha^{(\text{old})} (\rho_i \parallel \sigma) + (\alpha - 1) \log \parallel \sigma \parallel + \log \min_i \left\{ \gamma_i (\text{Tr} \, \rho_i)^{1-\alpha} (\text{Tr} \, \rho_i^0)^{\alpha-1} \right\} \tag{23}
\end{equation}

3.3 Rényi capacities

By a channel $W$ we mean a map $W: \mathcal{X} \rightarrow S(H)$, where $\mathcal{X}$ is some input alphabet (which can be an arbitrary non-empty set) and $H$ is a finite-dimensional Hilbert space. We recover the usual notion of a quantum channel when $\mathcal{X} = S(K)$ for some Hilbert space $K$, and $W$ is a completely positive trace-preserving linear map.

For an input alphabet $\mathcal{X}$, let $\{\delta_x\}_{x \in \mathcal{X}}$ be a set of rank-1 orthogonal projections in some Hilbert space $H_\mathcal{X}$, and for every channel $W: \mathcal{X} \rightarrow S(H)$ define
\begin{equation}
\hat{W}: x \mapsto \delta_x \otimes W_x,
\end{equation}
which is a channel from $\mathcal{X}$ to $S(H_\mathcal{X} \otimes H)$. Let $\mathcal{P}_f(\mathcal{X})$ denote the set of finitely supported probability measures on $\mathcal{X}$. The channels $W$ and $\hat{W}$ can naturally be extended to convex maps $W: \mathcal{P}_f(\mathcal{X}) \rightarrow S(H)$ and $\hat{W}: \mathcal{P}_f(\mathcal{X}) \rightarrow S(H_\mathcal{X} \otimes H)$, as
\begin{equation}
W(p) := \sum_{x \in \mathcal{X}} p(x) W(x), \quad \hat{W}(p) := \sum_{x \in \mathcal{X}} p(x) \hat{W}(p) = \sum_{x \in \mathcal{X}} p(x) \delta_x \otimes W(x).
\end{equation}

Note that $\hat{W}(p)$ is a classical-quantum state, and the marginals of $\hat{W}(p)$ are given by
\begin{equation}
\text{Tr}_H \hat{W}(p) = \hat{p} := \sum_x p(x) \delta_x \quad \text{and} \quad \text{Tr}_{H_\mathcal{X}} \hat{W}(p) = W(p).
\end{equation}
Let $D$ be a function on pairs of positive semidefinite operators. For a channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$, we define its corresponding $D$-capacity as

$$\hat{\chi}_D(W) := \sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi_D(W, p),$$

where

$$\chi_D(W, p) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D\left(\hat{W}(p)\|\hat{\sigma} \otimes \sigma\right), \quad p \in \mathcal{P}_f(\mathcal{X}).$$

For the cases $D = D^{(\text{old})}_\alpha$ and $D = D^{(\text{new})}_\alpha$, we use the shorthand notations $\chi^{(\text{old})}_\alpha(W, p)$, $\hat{\chi}^{(\text{old})}_\alpha(W)$ and $\chi^{(\text{new})}_\alpha(W, p)$, $\hat{\chi}^{(\text{new})}_\alpha(W)$, respectively. Note that these quantities generalize the Holevo quantity

$$\chi(W, p) := \chi^{(\text{old})}_\alpha(W, p) = \chi^{(\text{new})}_\alpha(W, p) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_1\left(\hat{W}(p)\|\hat{\sigma} \otimes \sigma\right)$$

$$= D_1\left(\hat{W}(p)\|\hat{\sigma} \otimes W(p)\right)$$

(24)

and the Holevo capacity

$$\hat{\chi}(W) := \sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi(W, p),$$

(25)

and hence we refer to them as generalized Holevo quantities for a general $D$, and generalized $\alpha$-Holevo quantities for the $\alpha$-divergences.

As it was pointed out in [31, 52],

$$D^{(\text{old})}_\alpha\left(\hat{W}(p)\|\hat{\sigma} \otimes \sigma\right) = \frac{\alpha}{\alpha - 1} \log \text{Tr} \omega(p) + D^{(\text{old})}_\alpha\left(\hat{\omega}(W, p)\|\sigma\right)$$

(26)

for any state $\sigma$, where

$$\hat{\omega}(W, p) := \omega(W, p) / \text{Tr} \omega(W, p), \quad \omega(W, p) := \left(\sum_x p(x)W(x)\right)^{\frac{1}{\alpha}}.$$

(27)

Since $D^{(\text{old})}_\alpha$ is non-negative on pairs of density operators, we get

$$\chi^{(\text{old})}_\alpha(W, p) = \frac{\alpha}{\alpha - 1} \log \text{Tr} \omega(p) = \frac{\alpha}{\alpha - 1} \log \text{Tr} \left(\sum_x p(x)W(x)^\alpha\right)^{\frac{1}{\alpha}}.$$

(28)

However, no such explicit formula is known for $\chi^{(\text{new})}_\alpha(W, p)$.

Note that $\max\{\text{Tr} \hat{W}(p)^0, \text{Tr}(\hat{\sigma} \otimes \sigma)^0\} \leq |\text{supp} p| \dim \mathcal{H}$, where $|\text{supp} p|$ denotes the cardinality of the support of $p$, and Lemma 3.3 with (12) and (13) yields that

$$\chi^{(\text{old})}_\alpha(W, p) \geq \chi^{(\text{new})}_\alpha(W, p) \geq \alpha \chi^{(\text{new})}_\alpha(W, p) - |\alpha - 1| \log (|\text{supp} p| \dim \mathcal{H})$$

(29)

for every $\alpha \in (0, +\infty)$. A more careful application of (12) and (13) yields the following improved bound:
Lemma 3.19. Let $W : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$ be a channel, and $\alpha \in (0, +\infty)$. For any $p \in \mathcal{P}_f(\mathcal{X})$ and any $\sigma \in \mathcal{S}(\mathcal{H})$, we have

$$D^{(\text{new})}_\alpha \left( \hat{W}(p) \| \hat{p} \otimes \sigma \right) \geq \alpha D^{(\text{old})}_\alpha \left( \hat{W}(p) \| \hat{p} \otimes \sigma \right) - |\alpha - 1| \log(\dim \mathcal{H}),$$

and hence,

$$\chi^{(\text{old})}_\alpha(W, p) \geq \chi^{(\text{new})}_\alpha(W, p) \geq \alpha \chi^{(\text{old})}_\alpha(W, p) - |\alpha - 1| \log(\dim \mathcal{H}).$$

Proof. Assume that $\alpha > 1$. By Corollary 3.10 we have

$$\text{Tr} \left( W(x) \frac{1}{2} \sigma^{-\frac{1}{\alpha} - \frac{1}{\alpha}} W(x) \frac{1}{2} \right) \geq (\dim \mathcal{H})^{-(\alpha - 1)^2} \left( \text{Tr} W(x)^\alpha \sigma^{1 - \alpha} \right)^\alpha$$

for every $x \in \mathcal{X}$, and hence,

$$D^{(\text{new})}_\alpha \left( \hat{W}(p) \| \hat{p} \otimes \sigma \right) = \frac{1}{\alpha - 1} \log \sum_x p(x) \text{Tr} \left( W(x) \frac{1}{2} \sigma^{-\frac{1}{\alpha} - \frac{1}{\alpha}} W(x) \frac{1}{2} \right)$$

$$\geq - (\alpha - 1) \log(\dim \mathcal{H}) + \frac{1}{\alpha - 1} \log \sum_x p(x) \left( \text{Tr} W(x)^\alpha \sigma^{1 - \alpha} \right)^\alpha$$

$$\geq - (\alpha - 1) \log(\dim \mathcal{H}) + \frac{1}{\alpha - 1} \log \left( \sum_x p(x) \text{Tr} W(x)^\alpha \sigma^{1 - \alpha} \right)^\alpha$$

$$= - (\alpha - 1) \log(\dim \mathcal{H}) + \alpha D^{(\text{old})}_\alpha \left( \hat{W}(p) \| \hat{p} \otimes \sigma \right),$$

where the second inequality is due to the convexity of $x \mapsto x^\alpha$. The proof for $\alpha \in (0, 1)$ goes exactly the same way.

Monotonicity of the Rényi divergences in $\alpha$ yields that the corresponding quantities $\chi^{(\text{old})}_\alpha(W, p)$ and $\chi^{(\text{new})}_\alpha(W, p)$ are also monotonic increasing in $\alpha$. A simple minimax argument shows (see, e.g. [38, Lemma B.3]) that

$$\lim_{\alpha \to 1} \chi^{(\text{old})}_\alpha(W, p) = \chi(W, p),$$

(30)

where $\chi(W, p)$ is the Holevo quantity. This, together with lemma 3.10 yields that also

$$\lim_{\alpha \to 1} \chi^{(\text{new})}_\alpha(W, p) = \chi(W, p).$$

Moreover, it was shown in [38, Proposition B.5] that if $\text{ran} W := \{W(x) : x \in \mathcal{X}\}$ is compact then

$$\lim_{\alpha \to 1} \hat{\chi}^{(\text{old})}_\alpha(W) = \hat{\chi}(W).$$

Applying lemma 3.19 to this, we obtain

$$\lim_{\alpha \to 1} \hat{\chi}^{(\text{new})}_\alpha(W) = \hat{\chi}(W).$$

(31)

Remark 3.20. Carathéodory’s theorem and the explicit formula (28) imply that in the definition $\hat{\chi}^{(\text{old})}_\alpha(W) := \sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi^{(\text{old})}_\alpha(W, p)$ it is enough to consider probability distributions with $|\text{supp} p| \leq (\dim \mathcal{H})^2 + 1$. However, this is not known for $\hat{\chi}^{(\text{new})}_\alpha(W)$, and hence (29) is insufficient to derive (31).
Finally, we point out a connection between $\alpha$-capacities and a special case of a famous convexity result by Carlen and Lieb \cite{12,13}. For any finite-dimensional Hilbert space $\mathcal{H}$ and $A_1, \ldots, A_n \in B(\mathcal{H})_+$, define

$$\Phi_{\alpha,q}(A_1, \ldots, A_n) := \left( \text{Tr} \left( \left( \sum_{i=1}^{n} A_i^\alpha \right)^{q/\alpha} \right) \right)^{1/q}, \quad \alpha \geq 0, \ q > 0.$$ 

Theorem 1.1 in \cite{13} says that for any finite-dimensional Hilbert space $\mathcal{H}$, $\Phi_{\alpha,q}$ is concave on $(B(\mathcal{H})_+)^n$ for $0 \leq \alpha \leq q \leq 1$, and convex for all $1 \leq \alpha \leq 2$ and $q \geq 1$. Below we give an elementary proof of the following weaker statement: $\Phi_{\alpha,1}$ is concave for $\alpha \in (0, 1)$ and convex for $\alpha \in (1, 2]$.

For a set $\mathcal{X}$, a finitely supported non-negative function $p : \mathcal{X} \to \mathbb{R}_+$, and a finite-dimensional Hilbert space $\mathcal{H}$, let $\hat{\Phi}_{p,\mathcal{H},\alpha} : (B(\mathcal{H})_+)^\mathcal{X} \to \mathbb{R}_+$ be defined as

$$\hat{\Phi}_{p,\mathcal{H},\alpha}(W) := \left( \text{Tr} \left( \sum_{x \in \mathcal{X}} p(x) W(x)^\alpha \right) \right)^{1/\alpha}, \quad W \in (B(\mathcal{H})_+)^\mathcal{X}.$$ 

The following Proposition is equivalent to our assertion:

**Proposition 3.21.** For any $\mathcal{X}$, $p$ and $\mathcal{H}$, $\hat{\Phi}_{p,\mathcal{H},\alpha}$ is concave on $(B(\mathcal{H})_+)^\mathcal{X}$ for $\alpha \in (0, 1)$ and convex for $\alpha \in (1, 2]$.

**Proof.** Exactly the same way as in \cite{26}--\cite{28}, we can see that

$$\frac{\alpha}{\alpha - 1} \log \text{Tr} \left( \sum_{x} p(x) W(x)^\alpha \right)^{1/\alpha} = \min_{\sigma \in S(\mathcal{H})} D_{\alpha,\text{old}}^{(\text{old})} \left( \hat{W}(p) \parallel \hat{\rho} \otimes \sigma \right). \quad (32)$$

Assume for the rest that $\alpha \in (1, 2]$; the proof for the case $\alpha \in (0, 1)$ goes exactly the same way. Let $r \in \mathbb{N}$, $W_1, \ldots, W_r \in (B(\mathcal{H})_+)^\mathcal{X}$, and $\gamma_1, \ldots, \gamma_r$ be a probability distribution. Then

$$\hat{\Phi}_{p,\mathcal{H},\alpha} \left( \sum_{i} \gamma_i W_i \right) = \min_{\sigma \in S(\mathcal{H})} Q_{\alpha,\text{old}}^{(\text{old})} \left( \sum_{i} \gamma_i \hat{W}(p) \parallel \hat{\rho} \otimes \sigma \right)$$

$$= \min_{\sigma_1, \ldots, \sigma_r \in S(\mathcal{H})} Q_{\alpha,\text{old}}^{(\text{old})} \left( \sum_{i} \gamma_i \hat{W}(p) \parallel \hat{\rho} \otimes \sum_{i} \gamma_i \sigma_i \right)$$

$$\leq \min_{\sigma_1, \ldots, \sigma_r \in S(\mathcal{H})} \sum_{i} \gamma_i Q_{\alpha,\text{old}}^{(\text{old})} \left( \hat{W}(p) \parallel \hat{\rho} \otimes \sigma_i \right)$$

$$= \sum_{i} \gamma_i \min_{\sigma_i} Q_{\alpha,\text{old}}^{(\text{old})} \left( \hat{W}(p) \parallel \hat{\rho} \otimes \sigma_i \right)$$

$$= \sum_{i} \gamma_i \hat{\Phi}_{p,\mathcal{H},\alpha}(W_i),$$

where the first and the last identities are due to \cite{32}, and the inequality follows from the joint convexity of $Q_{\alpha,\text{old}}^{(\text{old})}$ \cite{11,17}. (In the case $\alpha \in (0, 1)$, we have to use joint concavity \cite{32,17}.)
4 Applications to coding theorems

4.1 Preliminaries

For a self-adjoint operator $X$, we will use the notation $\{X > 0\}$ to denote the spectral projection of $X$ corresponding to the positive half-line $(0, +\infty)$. The spectral projections $\{X \geq 0\}$, $\{X < 0\}$ and $\{X \leq 0\}$ are defined similarly. The positive part $X_+$ and the negative part $X_-$ are defined as $X_+ := X\{X > 0\}$ and $X_- := X\{X < 0\}$, respectively, and the absolute value of $X$ is $|X| := X_+ + X_-$. The trace-norm of $X$ is $\|X\|_1 := \text{Tr}|X|$.

The following lemma is Theorem 1 from [4]; see also Proposition 1.1 in [28] for a simplified proof.

**Lemma 4.1.** Let $A, B$ be positive semidefinite operators on the same Hilbert space. For any $t \in [0, 1]$,

$$\text{Tr} A(I - \{A - B > 0\}) + \text{Tr} B\{A - B > 0\} = \frac{1}{2} \text{Tr}(A + B) - \frac{1}{2} \|A - B\|_1 \leq \text{Tr} A'B^{1-t}.$$

The next lemma is a reformulation of Lemma 2.6 in [34]. We include the proof for readers’ convenience.

**Lemma 4.2.** Let $(V, \|\cdot\|)$ be a finite-dimensional normed vector space, and let $D$ denote its real dimension. Let $\mathcal{N} \subset V$ be a subset. For every $\delta > 0$, there exists a finite subset $\mathcal{N}_{\delta} \subset \mathcal{N}$ such that

1. $|\mathcal{N}_{\delta}| \leq (1 + 2/\delta)^D$, and
2. for every $v \in \mathcal{N}$ there exists a $v_{\delta} \in \mathcal{N}_{\delta}$ such that $\|v - v_{\delta}\| < \delta$.

**Proof.** For every $\delta > 0$, let $\mathcal{N}_{\delta}$ be a maximal set in $\mathcal{N}$ such that $\|v - v'\| \geq \delta$ for every $v, v' \in \mathcal{N}_{\delta}$; then $\mathcal{N}_{\delta}$ clearly satisfies 2. On the other hand, the open $\|\cdot\|$-balls with radius $\delta/2$ around the elements of $\mathcal{N}_{\delta}$ are disjoint, and contained in the $\|\cdot\|$-ball with radius $1 + \delta/2$ and origin 0. Since the volume of balls scales with their radius on the power $D$, we obtain 1.

The *fidelity* of positive semidefinite operators $A$ and $B$ is defined as $F(A, B) := \text{Tr}(A^{1/2}BA^{1/2})^{1/2}$. The entanglement fidelity of a state $\rho$ and a completely positive trace-preserving map $\Phi$ is $F_e(\rho, \Phi) := F(|\psi_\rho\rangle\langle\psi_\rho|, (\text{id} \otimes \Phi)|\psi_\rho\rangle\langle\psi_\rho|)$, where $\psi_\rho$ is any purification of the state $\rho$; see Chapter 9 in [43] for details.

4.2 Quantum Stein’s lemma with composite null-hypothesis

Consider the asymptotic hypothesis testing problem with null-hypothesis $H_0 : \mathcal{N}_n \subset \mathcal{S} (\mathcal{H}_n)$ and alternative hypothesis $H_1 : \sigma_n \in \mathcal{S}(\mathcal{H}_n)$, $n \in \mathbb{N}$, where $\mathcal{H}_n$ is some finite-dimensional Hilbert space. Our goal is to decide between these two hypotheses based on the outcome of a binary POVM $(T_n(0), T_n(1))$ on $\mathcal{H}_n$, where 0 and 1 indicate the acceptance of $H_0$ and $H_1$, respectively. Since $T_n(1) = I - T_n(0)$, the POVM is uniquely determined by $T_n = T_n(0)$, and the only constraint on $T_n$ is that $0 \leq T_n \leq I_n$. We will
call such operators tests. Given a test \( T_n \), the probability of mistaking \( H_0 \) for \( H_1 \) (type I error) and the probability of mistaking \( H_1 \) for \( H_0 \) (type II error) are given by

\[
\alpha_n(T_n) := \sup_{\rho_n \in \mathcal{N}_n} \text{Tr} \rho_n (I - T_n), \quad \text{(type I),}
\quad \text{and} \quad \beta_n(T_n) := \text{Tr} \sigma_n T_n, \quad \text{(type II)}.
\]

**Definition 4.3.** We say that a rate \( R \geq 0 \) is achievable if there exists a sequence of tests \( T_n, n \in \mathbb{N} \), with

\[
\lim_{n \to +\infty} \alpha_n(T_n) = 0 \quad \text{and} \quad \limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(T_n) \leq -R.
\]

The largest achievable rate \( R(\mathcal{N}_n \| \sigma_n) \) is the direct rate of the hypothesis testing problem.

For what follows, we assume that \( \mathcal{H}_n = \mathcal{H}^{\otimes n}, n \in \mathbb{N} \), where \( \mathcal{H} = \mathcal{H}_1 \), and that the alternative hypothesis is i.i.d., i.e., \( \sigma_n = \sigma^{\otimes n}, n \in \mathbb{N} \), with \( \sigma = \sigma_1 \). We say that the null-hypothesis is composite i.i.d. if there exists a set \( \mathcal{N} \subset \mathcal{S}(\mathcal{H}) \) such that for all \( n \in \mathbb{N} \),\( \mathcal{N}_n = \mathcal{N}^{\otimes n} := \{ \rho^{\otimes n} : \rho \in \mathcal{N} \} \). The null-hypothesis is simple i.i.d. if \( \mathcal{N} \) consists of one single element, i.e., \( \mathcal{N} = \{ \rho \} \) for some \( \rho \in \mathcal{S}(\mathcal{H}) \). According to the quantum Stein’s lemma \[22\] \[10\], the direct rate in the simple i.i.d. case is given by \( D_1(\rho \| \sigma) \). The case of the general composite null-hypothesis was treated in \[9\] under the name of quantum Sanov theorem. There it was shown that there exists a sequence of tests \( \{ T_n \}_{n \in \mathbb{N}} \) such that \( \lim_{n \to +\infty} \text{Tr} \rho^{\otimes n} (I - T_n) = 0 \) for every \( \rho \in \mathcal{N} \), and \( \limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(T_n) \leq -D_1(\mathcal{N}\|\rho) \), where \( D_1(\mathcal{N}\|\rho) := \inf_{\rho \in \mathcal{N}} D_1(\rho \| \sigma) \). Note that this is somewhat weaker than \( D_1(\mathcal{N}\|\rho) \) being achievable in the sense of Definition 4.3. Achievability in this stronger sense has been shown very recently in \[44\], using the representation theory of the symmetric group and the method of types. The proof in both papers followed the approach in \[22\] of reducing the problem to a classical hypothesis testing problem by projecting all states onto the commutative algebra generated by \( \{ \sigma^{\otimes n} \}_{n \in \mathbb{N}} \).

Below we use a different proof technique to show that \( D_1(\mathcal{N}\|\rho) \) is achievable in the sense of Definition 4.3. Our proof is based solely on Audenaert’s trace inequality \[4\] and the subadditivity property of \( Q^{\text{(new)}} \), given in Proposition 3.13. We obtain explicit upper bounds on the error probabilities for any finite \( n \in \mathbb{N} \) for a sequence of Neyman-Pearson tests. Moreover, if a \( \delta \)-net can be explicitly constructed for \( \mathcal{N} \) for every \( \delta > 0 \) (this is trivially satisfied when \( \mathcal{N} \) is finite) then the tests can also be constructed explicitly. In \[9\], Stein’s lemma was stated with weak converse, while the results of \[44\] imply a strong converse. Here we use Nagaoka’s method to further strengthen the converse part by giving explicit bounds on the exponential rate with which the worst-case type I success probability goes to zero when the type II error decays with a rate larger than the optimal rate \( D_1(\mathcal{N}\|\rho) \).

Note that our proof technique doesn’t actually rely on the i.i.d. assumption, as we demonstrate in Theorem 4.9 where we give achievability bounds in the general correlated scenario. However, in the most general case we have to restrict to a finite null-hypothesis. We show examples in Remark 4.10 where the achievable rate of Theorem 4.9 can be expressed as the regularized relative entropy distance of the null-hypothesis and the alternative hypothesis, giving a direct generalization of the i.i.d. case. These
results complement those of [10], where it was shown that if $\Theta$ is a set of ergodic states on a spin chain, and $\Phi$ is a state on the spin chain such that for every $\Psi \in \Theta$, Stein’s lemma holds for the simple hypothesis testing problem $H_0 : \Psi, H_1 : \Phi$, then it also holds for the composite hypothesis testing problem $H_0 : \Theta, H_1 : \Phi$. This was also extended in [10] to the case where $\Theta$ consists of translation-invariant states, using ergodic decomposition.

Now let $\mathcal{N} \subset \mathcal{S}(\mathcal{H})$ be a non-empty set of states, and let $\sigma \in \mathcal{B}(\mathcal{H})_+$ be a positive semidefinite operator such that

$$\text{supp } \rho \subseteq \text{supp } \sigma, \quad \rho \in \mathcal{N}. \quad (33)$$

Note that in hypothesis testing $\sigma$ is usually assumed to be a state on $\mathcal{H}$; however, the proof for Stein’s lemma works the same way for a general positive semidefinite $\sigma$, and considering this more general case is actually useful e.g., for state compression. Let

$$\psi(t) := \sup_{\rho \in \mathcal{N}} \log Q_t^{(\text{new})}(\rho \| \sigma), \quad t > 0, \quad (34)$$

and for every $a \in \mathbb{R}$, let

$$\varphi(a) := \sup_{0 < t \leq 1} \{at - \psi(t)\}, \quad \hat{\varphi}(a) := \sup_{0 < t \leq 1} \{a(t - 1) - \psi(t)\} = \varphi(a) - a. \quad (35)$$

Note that $\varphi$ is the Legendre-Fenchel transform of $\psi$ on $(0, 1]$.

**Theorem 4.4.** For every $n \in \mathbb{N}$, let $\mathcal{N}(n) \subset \mathcal{N}$ be a finite subset, and let $\delta(\mathcal{N}(n)) := \sup_{\rho \in \mathcal{N}} \inf_{\rho' \in \mathcal{N}(n)} \| \rho - \rho' \|_1$. For every $a \in \mathbb{R}$, let $S_{n,a} := \left\{ e^{-na} \sum_{\rho \in \mathcal{N}(n)} \rho^{\otimes n} - \sigma^{\otimes n} > 0 \right\}$ be a Neyman-Pearson test. Then

$$\sup_{\rho \in \mathcal{N}} \text{Tr } \rho^{\otimes n}(I - S_{n,a}) \leq |\mathcal{N}(n)|e^{-n\hat{\varphi}(a)} + n\delta(\mathcal{N}(n)), \quad \text{and} \quad (36)$$

$$\text{Tr } \sigma^{\otimes n} S_{n,a} \leq |\mathcal{N}(n)|e^{-n\varphi(a)}. \quad (37)$$

In particular, let $\delta_n := e^{-nk}$ for some $\kappa > 0$, and $\mathcal{N}(n) := \mathcal{N}_{\delta_n} \subset \mathcal{N}$ as in lemma 4.2. Then

$$\limsup_{n \to +\infty} \frac{1}{n} \log \alpha_n(S_{n,a}) \leq - \min\{\kappa, \hat{\varphi}(a) - \kappa D(\mathcal{H})\}, \quad (38)$$

$$\limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(S_{n,a}) \leq - (\varphi(a) - \kappa D(\mathcal{H})). \quad (39)$$

**Proof.** For every $n \in \mathbb{N}$, let $\bar{\rho}_n := \sum_{\rho \in \mathcal{N}(n)} \rho^{\otimes n}$, $\sigma_n := \sigma^{\otimes n}$. Applying lemma 4.1 to $A := e^{-na}\bar{\rho}_n$ and $B := \sigma_n$ for some fixed $a \in \mathbb{R}$, we get

$$e_n(a) := e^{-na} \text{Tr } \bar{\rho}_n(I - S_{n,a}) + \text{Tr } \sigma_n S_{n,a} \leq e^{-nat} \text{Tr } \overline{\rho}_n \sigma_n^{1 - t} \quad (40)$$

for every $t \in [0, 1]$. This we can further upper bound as

$$\text{Tr } \overline{\rho}_n \sigma_n^{1 - t} \leq Q_t^{(\text{new})}(\bar{\rho}_n \| \sigma_n) \leq \sum_{\rho \in \mathcal{N}(n)} Q_t^{(\text{new})}(\rho^{\otimes n} \| \sigma^{\otimes n}) \leq |\mathcal{N}(n)| \sup_{\rho \in \mathcal{N}} Q_t^{(\text{new})}(\rho^{\otimes n} \| \sigma^{\otimes n})$$

$$= |\mathcal{N}(n)| \sup_{\rho \in \mathcal{N}} \left( Q_t^{(\text{new})}(\rho \| \sigma) \right)^n = |\mathcal{N}(n)|e^{n\psi(t)}, \quad (41)$$
where the first inequality is due to lemma 3.3, the second inequality is due to 10, the third inequality is obvious, the succeeding identity follows from the definition 14, and the last identity is due to the definition of \( \psi \). Since 10 holds for every \( t \in (0, 1) \), together with 11 it yields \( \varepsilon_n(a) \leq |N(n)|e^{-n\varphi(a)} \). Hence we have \( \sup a \in N[n,a] \leq \varepsilon_n(a) \leq |N(n)|e^{-n\varphi(a)} \), proving 30. Similarly, \( \sup a \in N[\bar{\rho}n] \leq e^{n\varphi(a)} \) yields

\[
\sup_\rho \in N(n) \text{ Tr } \rho^{\otimes n}(I - S_{n,a}) \leq \text{ Tr } \rho_{\bar{n}}(I - S_{n,a}) \leq e^{n\varphi(a)} |N(n)|e^{-n\varphi(a)} = |N(n)|e^{-n\varphi(a)}.
\]

The submultiplicativity of the trace-norm on tensor products yields that \( \sup_\rho \in N \) Tr \( \rho^{\otimes n}(I - S_{n,a}) \leq \sup_\rho \in N[a] \text{ Tr } \rho^{\otimes n}(I - S_{n,a}) + n\delta(N(n)) \). Combined with 42, this yields 37.

The inequalities in 38–39 are obvious from the choice of \( \delta_n \).

**lemma 4.5.** We have \( \varphi(a) \geq a \), and for every \( a < D_1(N||\sigma) \), we have \( \varphi(a) > 0 \).

**Proof.** Note that for any \( t \in (0, 1) \), \( a(t - 1) - \psi(t) = (t - 1)[a - \inf_{\rho \in N} D^\text{new}_t(\rho||\sigma)] \).

Moreover, by the assumption in 33, \( \rho \mapsto D^\text{new}_t(\rho||\sigma) \) is continuous on \( N \), and hence, \( \inf_{\rho \in N} D^\text{new}_t(\rho||\sigma) = \min_{\rho \in N} D^\text{new}_t(\rho||\sigma) \) for every \( t \in (0, 1) \). Note that \( N \) is compact, and for every \( \rho \in N \), \( t \mapsto D^\text{new}_t(\rho||\sigma) \) is monotone increasing, due to 10, Theorem 6. Applying now the minimax theorem from 38, Corollary A.2, we get \( \sup_{t \in (0, 1)} \inf_{\rho \in N} D^\text{new}_t(\rho||\sigma) = \min_{\rho \in N} \sup_{t \in (0, 1)} D^\text{new}_t(\rho||\sigma) \) for every \( t \in (0, 1) \). Thus, for any \( a < D_1(N||\sigma) \), there exists a \( t_a \in (0, 1) \) such that \( a - \inf_{\rho \in N} D^\text{new}_{t_a}(\rho||\sigma) < 0 \), and hence \( 0 < (t_a - 1)[a - \inf_{\rho \in N} D^\text{new}_{t_a}(\rho||\sigma)] \leq \varphi(a) \). Finally, note that assumption 33 yields that \( \psi(1) = 0 \), and hence \( \varphi(a) \geq a - \psi(1) = a \).

**Theorem 4.6.** The direct rate is lower bounded by \( D_1(N||\sigma) \), i.e.,

\[
R(\{N^{(\otimes n)}\}_{n \in N}|\{\sigma^{\otimes n}\}_{n \in N}) \geq D_1(N||\sigma).
\]

**Proof.** The proposition is trivial when \( D_1(N||\sigma) = 0 \), and hence for the rest we assume \( D_1(N||\sigma) > 0 \). By lemma 4.5, for every \( 0 < a < D_1(N||\sigma) \) we can find \( 0 < \kappa < \varphi(a)/D(H) \), so that 38–39 hold. Since we can take \( \kappa \) arbitrarily small, and \( a \) arbitrarily close to \( D_1(N||\sigma) \), we see that any rate below \( \sup_{0 < a < D_1(N||\sigma)} \varphi(a) \) is achievable. By lemma 4.5, \( \sup_{0 < a < D_1(N||\sigma)} \varphi(a) \geq \sup_{0 < a < D_1(N||\sigma)} a = D_1(N||\sigma) \), proving the assertion.

The strong converse for the simple i.i.d. case 46 yields immediately the strong converse for the composite i.i.d. case. We include a proof for completeness.

**Theorem 4.7.** If \( \limsup_{n \to +\infty} \frac{1}{n} \log \text{ Tr } \sigma^{\otimes n} T_n \leq -r \) for some sequence of tests \( T_n \), \( n \in N \), then

\[
\limsup_{n \to +\infty} \frac{1}{n} \log \inf_\rho \in N \text{ Tr } \rho^{\otimes n} T_n \leq \inf_{t > 1} \frac{t - 1}{t} \left[ -r + \inf_\rho \in N D^\text{new}_t(\rho||\sigma) \right].
\]

If \( r > D_1(N||\sigma) \) then the RHS of 44 is strictly negative, and hence the worst-case success probability \( \inf_\rho \in N \text{ Tr } \rho^{\otimes n} T_n \) goes to zero exponentially fast. As a consequence, 33 holds as an equality.
Proof. Following [41] (see also [39]), we can use the monotonicity of the Rényi divergences under measurements for \( \alpha > 1 \) [18 39 40 56] to obtain that for any sequence of tests \( T_n, n \in \mathbb{N} \), any \( \rho \in \mathcal{N} \), and any \( t > 1 \),
\[
Q_i^{\text{(new)}}(\rho \otimes^n \| \sigma \otimes^n) \geq Q_i^{\text{(new)}}\left( \{ \text{Tr} \rho \otimes^n T_n, \text{Tr} \rho \otimes^n (I_n - T_n) \} \| \{ \text{Tr} \sigma \otimes^n T_n, \text{Tr} \sigma \otimes^n (I_n - T_n) \} \right) \\
\quad \geq (\text{Tr} \rho \otimes^n T_n)^t (\text{Tr} \sigma \otimes^n T_n)^{1-t},
\]
which yields
\[
\frac{1}{n} \log \text{Tr} \rho \otimes^n T_n \leq \frac{t-1}{t} \left[ \frac{1}{n} \log \text{Tr} \sigma \otimes^n T_n + D_t^{\text{(new)}}(\rho \| \sigma) \right].
\]
Taking first the infimum in \( \rho \in \mathcal{N} \), and then the limsup in \( n \), we obtain (44).

Since \( \inf_{t>1} \inf_{\rho \in \mathcal{N}} D_t^{\text{(new)}}(\rho \| \sigma) = \inf_{\rho \in \mathcal{N}} \inf_{t>1} D_t^{\text{(new)}}(\rho \| \sigma) = D_1(\mathcal{N} \| \sigma) \), we see that if \( r > D_1(\mathcal{N} \| \sigma) \) then there exists a \( t > 1 \) such that \(-r + \inf_{t>1} \inf_{\rho \in \mathcal{N}} D_t^{\text{(new)}}(\rho \| \sigma) < 0 \), and hence the RHS of (14) is strictly negative. The rest of the statements follow immediately. \( \square \)

Remark 4.8. Theorem [4.6] shows the existence of a sequence of tests such that the type II error probability decays exponentially fast with rate \( D_1(\mathcal{N} \| \sigma) \), while the type I error probability goes to zero. Note that for this statement, it is enough to choose \( \delta_n \) polynomially decaying; e.g. \( \delta_n := 1/n^2 \) does the job, and we get an improved exponent for the type II error, \( \limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(S_{n,a}) \leq -\varphi(a) \).

Theorem [4.4] yields more detailed information in the sense that it shows that for any rate \( r \) below the optimal rate \( D_1(\mathcal{N} \| \sigma) \), there exists a sequence of tests along which the type II error decays with the given rate \( r \), while the type I error also decays exponentially fast; moreover, [38–39] provide a lower bound on the rate of the type I error. Note that if \( \mathcal{N} \) is finite then the approximation process can be omitted, and we obtain the bounds
\[
\limsup_{n \to +\infty} \frac{1}{n} \log \alpha_n(S_{n,a}) \leq -\hat{\varphi}(a), \quad \limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(S_{n,a}) \leq -\varphi(a).
\]
These bounds are not optimal; indeed, in the simple i.i.d. case the quantum Hoeffding bound theorem [5 21 23 41] shows that the above inequalities become equalities with \( \varphi \) and \( \hat{\varphi} \) replaced by \( \varphi^{(\text{old})}(a) := \sup_{0 < t \leq 1} \{ \alpha t - \log Q_i^{(\text{old})}(\rho \| \sigma) \} \), \( \hat{\varphi}^{(\text{old})}(a) := \varphi^{(\text{old})}(a) - a \), and if \( \rho \) and \( \sigma \) don’t commute then \( \varphi^{(\text{old})}(a) > \varphi(a) \) and \( \hat{\varphi}^{(\text{old})}(a) > \hat{\varphi}(a) \) for any \( 0 < a < D_1(\rho \| \sigma) \), due to the Araki-Lieb-Thirring inequality [3 33]. On the other hand, the RHS of (14) is known to give the exact strong converse rate in the simple i.i.d. case [39].

The above arguments can also be used to obtain bounds on the direct rate in the case of states with arbitrary correlations. In this case, however, it may not be possible to find a suitable approximation procedure, and hence we restrict our attention to the case of finite null-hypothesis. Thus, for every \( n \in \mathbb{N} \), our alternative hypothesis \( H_1 \) is given by some state \( \sigma_n \in \mathcal{S}(\mathcal{H}_n) \), where \( \mathcal{H}_n \) is some finite-dimensional Hilbert space, and the null-hypothesis \( H_0 \) is given by \( \mathcal{N}_n = \{ \rho_{n,1}, \ldots, \rho_{n,r} \} \subset \mathcal{S}(\mathcal{H}_n) \), where \( r \in \mathbb{N} \) is some fixed number. We assume that \( \text{supp} \rho_{i,n} \subseteq \text{supp} \sigma_n \) for every \( i \) and \( n \).

18
Theorem 4.9. In the above setting, we have

\[
\limsup_{n \to +\infty} \frac{1}{n} \log \alpha_n(S_{n,a}) \leq - \sup_{0 < t < 1} \left\{ a(t - 1) - \max_{1 \leq i \leq r} \limsup_{n \to +\infty} \frac{1}{n} \log Q^{(\text{new})}_t(\rho_{i,n} || \sigma_n) \right\}, \tag{45}
\]

\[
\limsup_{n \to +\infty} \frac{1}{n} \log \beta_n(S_{n,a}) \leq - \sup_{0 < t < 1} \left\{ at - \max_{1 \leq i \leq r} \limsup_{n \to +\infty} \frac{1}{n} \log Q^{(\text{new})}_t(\rho_{i,n} || \sigma_n) \right\} \leq -a, \tag{46}
\]

where \( S_{n,a} := \{ e^{-na} \sum_{i} \rho_{i,n} - \sigma_n > 0 \} \). As a consequence, the direct rate is lower bounded as

\[
R(\{ N_n \}_{n \in \mathbb{N}} \| \{ \sigma_n \}_{n \in \mathbb{N}}) \geq \sup_{0 < t < 1} \min \liminf_{n \to +\infty} \frac{1}{n} D^{(\text{new})}_t(\rho_{i,n} || \sigma_n). \tag{47}
\]

If \( \limsup_{n \to +\infty} \frac{1}{n} \log \dim \mathcal{H}_n < +\infty \) then we also have

\[
R(\{ N_n \}_{n \in \mathbb{N}} \| \{ \sigma_n \}_{n \in \mathbb{N}}) \geq \min_i \partial^- \psi^{(\text{old})}_i(1), \tag{48}
\]

where \( \partial^- \) stands for the left derivative, and \( \psi^{(\text{old})}_i(t) := \limsup_{n \to +\infty} \frac{1}{n} \log Q^{(\text{old})}_t(\rho_{i,n} || \sigma_n) \).

**Proof.** The same argument as in Theorem 4.4 yields (45) and (46), from which (47) follows immediately. Assume now that \( L := \limsup_{n \to +\infty} \frac{1}{n} \log \dim \mathcal{H}_n < +\infty \). By Corollary 3.10 we have

\[
\limsup_{n \to +\infty} \frac{1}{n} \log Q^{(\text{new})}_t(\rho_{i,n} || \sigma_n) \leq t \psi^{(\text{old})}_i(t) + (t - 1)^2 L. \tag{49}
\]

Note that \( \psi^{(\text{old})}_i(t) \) is the pointwise \( \limsup \) of convex functions, and hence it is convex, too. Moreover, the support condition \( \sup \rho_{i,n} \subseteq \sup \sigma_n \) implies \( \psi^{(\text{old})}_i(1) = 0 \). Hence, we have \( \lim_{t \to 1^-} \frac{1}{t - 1} \psi^{(\text{old})}_i(t) = \partial^- \psi^{(\text{old})}_i(1) \). Combining this with (45) and (49), we see that \( \limsup_{n \to +\infty} \frac{1}{n} \log \alpha_n(S_{n,a}) < 0 \) for all \( a < \min_i \partial^- \psi^{(\text{old})}_i(1) \). Taking into account (46), we get (48). \( \Box \)

**Remark 4.10.** Note that under suitable regularity, we have \( \partial^- \psi^{(\text{old})}_i(1) = \lim_{n \to +\infty} \frac{1}{n} D_1(\rho_{i,n} || \sigma_n) \), and hence

\[
R(\{ N_n \}_{n \in \mathbb{N}} \| \{ \sigma_n \}_{n \in \mathbb{N}}) \geq \min_i \lim_{n \to +\infty} \frac{1}{n} D_1(\rho_{i,n} || \sigma_n). \tag{50}
\]

This is clearly satisfied in the i.i.d. case, and we recover (43). There are also various important special cases of correlated states where the above holds. This is the case, for instance, if all the \( \rho_{i,n} \) and \( \sigma_n \) are \( n \)-site restrictions of gauge-invariant quasi-free states on a fermionic or bosonic chain (the latter type of states are also called Gaussian states). In this case \( \lim_{n \to +\infty} \frac{1}{n} D_1(\rho_{i,n} || \sigma_n) \) can be expressed by an explicit formula in terms of the symbols of the states; see [35] [36] for details. Another class of states where the above holds is when \( \rho_{i,n} \) and \( \sigma_n \) are group-invariant restrictions of i.i.d. states on a spin chain [23]. In this case one can use the same approximation procedure as in the i.i.d. case, and hence (50) holds for \( \mathcal{N}_n := \{ \rho_{i,n} : i \in \mathcal{I} \} \), where \( \mathcal{I} \) is an arbitrary (not necessarily finite) index set.
Finally, we show that the above considerations for the composite null-hypothesis yield the direct rate also for the averaged i.i.d. case. In this setting we have a probability measure $\mu$ on the Borel sets of $\mathcal{S}(\mathcal{H})$ such that $\bar{\rho}_n := \int_{\mathcal{S}(\mathcal{H})} \rho^{\otimes n} \, d\mu$ is well-defined for every $n \in \mathbb{N}$. The null-hypothesis is given by the sequence $\mathcal{N}_n = \{\bar{\rho}_n\}$, $n \in \mathbb{N}$. Note that in this case the null-hypotheses is simple, i.e., $\mathcal{N}_n$ consists of one single element, but it is not i.i.d. Let

$$D^* := \sup \left\{ \inf_{\rho \in \mathcal{S}(\mathcal{H}) \setminus H} D_1 (\rho \| \sigma) : H \subset \mathcal{S}(\mathcal{H}) \text{ Borel set with } \mu(H) = 0 \right\},$$

which is essentially the relative entropy distance of $\text{supp} \rho$ from $\sigma$, modulo subsets of zero measure. Assume that $D^* > 0$, since otherwise (51) holds trivially. For every $0 < a < D^*$, there exists a subset $\mathcal{N}(a)$ such that $a < D_1 (\mathcal{N}(a) \| \sigma) \leq D^*$ and $\mu(\mathcal{S}(\mathcal{H}) \setminus \mathcal{N}(a)) = 0$. Applying Theorem 4.4 to the composite i.i.d. problem with null-hypothesis $\mathcal{N}(a)$, we get the existence of a sequence of tests $T_n$, $n \in \mathbb{N}$, such that

$$\limsup_{n \to +\infty} \frac{1}{n} \log \text{Tr} \sigma^{\otimes n} T_n \leq -a,$$

$$\limsup_{n \to +\infty} \frac{1}{n} \log \text{Tr} \bar{\rho}_n (I - T_n) \leq \limsup_{n \to +\infty} \frac{1}{n} \log \sup_{\rho \in \mathcal{N}(a)} \text{Tr} \rho^{\otimes n} (I - T_n) < 0.$$

Hence, the direct rate for the averaged i.i.d. problem is lower bounded by $D^*$, i.e.,

$$R(\{\bar{\rho}_n\}_{n \in \mathbb{N}} \| \{\sigma^{\otimes n}\}_{n \in \mathbb{N}}) \geq D^*. \quad (51)$$

4.3 Universal state compression

Consider a sequence of Hilbert spaces $\mathcal{H}_n$, $n \in \mathbb{N}$, and for each $n$, let $\mathcal{N}_n \subset \mathcal{S}(\mathcal{H}_n)$ be a set of states. An asymptotic compression scheme is a sequence $(\mathcal{C}_n, \mathcal{D}_n)$, $n \in \mathbb{N}$, where $\mathcal{C}_n : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{K}_n)$ is the compression map, and $\mathcal{D}_n : \mathcal{B}(\mathcal{K}_n) \to \mathcal{B}(\mathcal{H}^{\otimes n})$ is the decompression. We use two different measures for the fidelity of $(\mathcal{C}_n, \mathcal{D}_n)$, defined as

$$F(\mathcal{C}_n, \mathcal{D}_n) := \inf_{\rho \in \mathcal{N}_n} F(\rho, (\mathcal{D}_n \circ \mathcal{C}_n) \rho), \quad \hat{F}(\mathcal{C}_n, \mathcal{D}_n) := \inf_{\rho \in \mathcal{N}_n} F(\rho, (\mathcal{D}_n \circ \mathcal{C}_n) \rho),$$

where $F$ stands for the fidelity, and $F_e$ for the the entanglement fidelity (see Section 1.1). Due to the monotonicity of the fidelity under partial trace, we have $F(\mathcal{C}_n, \mathcal{D}_n) \leq \hat{F}(\mathcal{C}_n, \mathcal{D}_n)$. Let $[\mathcal{C}_n(\mathcal{N}_n)]$ be the projection onto the subspace generated by the supports of $\mathcal{C}_n(\rho_n)$, $\rho_n \in \mathcal{N}_n$. We say that a compression rate $R$ is achievable if there exists an asymptotic compression scheme $(\mathcal{C}_n, \mathcal{D}_n)$, $n \in \mathbb{N}$, such that

$$\lim_{n \to +\infty} F(\mathcal{C}_n, \mathcal{D}_n) = 1 \quad \text{and} \quad \limsup_{n \to +\infty} \frac{1}{n} \log \text{Tr} [\mathcal{C}_n(\mathcal{N}_n)] \leq R.$$

The smallest achievable compression rate is the optimal compression rate $R(\{\mathcal{N}_n\}_{n \in \mathbb{N}})$. We say that the compression problem is i.i.d. if $\mathcal{H}_n = \mathcal{H}^{\otimes n}$ and $\mathcal{N}_n = \mathcal{N}^{\otimes n} := \{\rho^{\otimes n} : \rho \in \mathcal{N}\}$ for every $n \in \mathbb{N}$, where $\mathcal{H} = \mathcal{H}_1$, and $\mathcal{N} \subset \mathcal{S}(\mathcal{H})$. It was shown in [30] (see also [29]) that in the simple i.i.d. case, projecting the state onto its entropy-typical subspace
yields the entropy as an achievable coding rate, which is also optimal. In Section 10.3 of [20], Neyman-Pearson type projections were used instead of the typical projections, and exponential bounds were obtained for the error probability for suboptimal coding rates. An extension of the typical projection technique was used in [30] to obtain universal state compression, i.e., it was shown that for any \( s > 0 \) there exists a coding scheme of rate \( s \) that is asymptotically error-free for any state of entropy less than \( s \). Theorem 4.11 below shows that the use of Neyman-Pearson projections can also be extended to obtain universal state compression. Since Theorem 4.11 is essentially a special case of Theorems 4.4 and 4.7 with the choice \( \sigma := I \), we omit the proof. The only part that doesn’t follow immediately from Theorems 4.4 and 4.7 is relating the fidelity to the success probability of the generalized state discrimination problem; this, however, is standard and we refer the interested reader to Section 12.2.2 in [43].

Let \( \psi(t) \), \( \varphi(a) \) and \( \hat{\varphi}(a) \) be defined as in (34)–(35), with \( \sigma := I \). Note that in this case \( Q_i^{\text{new}}(\rho||\sigma) = Q_i^{\text{old}}(\rho||\sigma) = \text{Tr}(\rho) \).

Theorem 4.11. In the i.i.d. case, for every \( \kappa > 0 \), \( a \in \mathbb{R} \), and \( n \in \mathbb{N} \), let \( \delta_n := e^{-n\kappa} \), let \( \mathcal{N}_{\delta_n} \subset \mathcal{N}_n \) be a subset as in lemma 4.2, and let \( S_{n,a} := \left\{ e^{-na} \sum_{\rho \in \mathcal{N}_{\delta_n}} \rho^\otimes n - I_n > 0 \right\} \).

Define
\[
C_n(.) := S_{n,a}(.)S_{n,a} + |\psi_n\rangle\langle\psi_n| \text{Tr}(.) (I - S_{n,a}), \quad D_n := \text{id},
\]
where \( \psi_n \) is an arbitrary unit vector in the range of \( S_{n,a} \). For every \( a \in \mathbb{R} \) and \( \kappa > 0 \), we have
\[
\limsup_{n \to +\infty} \frac{1}{n} \log (1 - F(C_n, D_n)) \leq -\min \{ \kappa, \hat{\varphi}(a) - \kappa D(\mathcal{H}) \}, \tag{52}
\]
\[
\limsup_{n \to +\infty} \frac{1}{n} \log \text{Tr} [C_n(\mathcal{N}_n)] \leq -\varphi(a) + \kappa D(\mathcal{H}). \tag{53}
\]

On the other hand, for any coding scheme \((C_n, D_n), n \in \mathbb{N}\), we have
\[
\limsup_{n \to +\infty} \frac{1}{n} \log \hat{F}(C_n, D_n) \leq \inf_{t>1} \frac{t - 1}{t} \left[ \limsup_{n \to +\infty} \frac{1}{n} \log \text{Tr} [C_n(\mathcal{N}_n)] - \sup_{\rho \in \mathcal{N}} S_t(\rho) \right].
\]
where \( S_t(\rho) := \frac{1}{1-t} \log \text{Tr} \rho^t \) is the Rényi entropy of \( \rho \) with parameter \( t \).

Corollary 4.12. The optimal compression rate is equal to the maximum entropy, i.e.,
\[
R(\{\mathcal{N}_{n \in \mathbb{N}}\}) = \sup_{\rho \in \mathcal{N}} S(\rho).
\]

Remark 4.13. We recover the result of [30] by choosing \( \mathcal{N} := \{ \rho \in \mathcal{S}(\mathcal{H}) : S(\rho) \leq s \} \).

Remark 4.14. Theorem 4.11 and Corollary 4.12 can be extended to correlated states and averaged states the same way as the analogous results for state discrimination in Section 4.2. Since these extensions are trivial, we omit the details.
Remark 4.15. The simple i.i.d. state compression problem can also be formulated in an ensemble setting, which is in closer resemblance with the usual formulation of classical source coding. In that formulation, a discrete i.i.d. quantum information source is specified by a finite set \( \{\rho_x\}_{x \in \mathcal{X}} \subset \mathcal{S}(\mathcal{H}) \) of states and a probability distribution \( p \) on \( \mathcal{X} \). Invoking the source \( n \) times, we obtain a state \( \rho_{\underline{x}} := \rho_{x_1} \otimes \ldots \otimes \rho_{x_n} \), with probability \( p_{\underline{x}} := p(x_1) \ldots p(x_n) \). The fidelity of a compression-decompression pair \( (\mathcal{C}_n, \mathcal{D}_n) \) is then defined as \( F(\mathcal{C}_n, \mathcal{D}_n) := \sum_{\underline{x} \in \mathcal{X}} p(\underline{x}) F_e(\rho_{\underline{x}}, \mathcal{D}_n \circ \mathcal{C}_n) \). In the classical case the signals \( \rho_x \) can be identified with a system of orthogonal rank 1 projections, \( \mathcal{C}_n \) and \( \mathcal{D}_n \) are classical stochastic maps, and \( F(\mathcal{C}_n, \mathcal{D}_n) \) as defined above gives back the usual expression for the success probability. It follows from standard properties of the fidelity that the optimal compression rate, under the constraint that \( F(\mathcal{C}_n, \mathcal{D}_n) \) goes to 1 asymptotically, only depends on the average state \( \rho(p) := \sum_{x} p(x)\rho_x \), and is equal to \( S(\rho(p)) \). Theorem 4.11 and Corollary 4.12 thus also provide the optimal compression rate and exponential bounds on the error and success probabilities in the ensemble formulation, for multiple quantum sources.

4.4 Classical capacity of compound channels

Recall that by a channel \( \mathcal{W} \) we mean a map \( \mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H}) \), where \( \mathcal{X} \) is some input alphabet (which can be an arbitrary non-empty set) and \( \mathcal{H} \) is a finite-dimensional Hilbert space. For a channel \( \mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H}) \), we define its \( n \)-th i.i.d. extension \( \mathcal{W}^{\otimes n} \) as the channel \( \mathcal{W}^{\otimes n} : \mathcal{X}^n \rightarrow \mathcal{S}(\mathcal{H}^{\otimes n}) \), defined as

\[
\mathcal{W}^{\otimes n}(x_1, \ldots, x_n) := \mathcal{W}(x_1) \otimes \ldots \otimes \mathcal{W}(x_n), \quad x_1, \ldots, x_n \in \mathcal{X}.
\]

(54)

It is obvious from the explicit formula (28) for \( \chi^{(old)}_\alpha \) that

\[
\chi^{(old)}_\alpha(\mathcal{W}^{\otimes n}, \mathcal{P}^{\otimes n}) = n \chi^{(old)}_\alpha(\mathcal{W}, \mathcal{P}), \quad n \in \mathbb{N},
\]

(55)

where \( \mathcal{P}^{\otimes n} \in \mathcal{P}_f(\mathcal{X}^n) \) is the \( n \)-th i.i.d. extension of \( \mathcal{P} \), defined as \( \mathcal{P}^{\otimes n}(x_1, \ldots, x_n) := p(x_1) \cdot \ldots \cdot p(x_n) \), \( x_1, \ldots, x_n \in \mathcal{X} \). It is not known whether the same additivity property holds for \( \chi^{(new)}_\alpha \).

Remark 4.16. Note that in our definition of a channel, we didn’t make any assumption on the cardinality of the input alphabet \( \mathcal{X} \), nor did we require any further mathematical properties from \( \mathcal{W} \), apart from being a function to \( \mathcal{S}(\mathcal{H}) \). The usual notion of a quantum channel is a special case of this definition, where \( \mathcal{X} \) is the state space of some Hilbert space and \( \mathcal{W} \) is a completely positive trace-preserving convex map. In this case, however, our definition of the i.i.d. extensions are more restrictive than the usual definition of the tensor powers of a quantum channel. Indeed, our definition corresponds to the notion of quantum channels with product state encoding. Hence, our definition of the classical capacity below corresponds to the classical capacity of quantum channels with product state encoding.

Let \( \mathcal{W}_i : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H}), i \in \mathcal{I} \), be a set of channels with the same input alphabet \( \mathcal{X} \) and the same output Hilbert space \( \mathcal{H} \), where \( \mathcal{I} \) is any index set. A code \( \mathcal{C} = (\mathcal{C}_e, \mathcal{C}_d) \) for \( \{\mathcal{W}_i\}_{i \in \mathcal{I}} \) consists of an encoding \( \mathcal{C}_e : \{1, \ldots, M\} \rightarrow \mathcal{X} \) and a decoding \( \mathcal{C}_d : \{1, \ldots, M\} \rightarrow \mathcal{X} \).
$\mathcal{B}(\mathcal{H})_+$, where $\{\mathcal{C}_d(1), \ldots, \mathcal{C}_d(M)\}$ is a POVM on $\mathcal{H}$, and $M \in \mathbb{N}$ is the size of the code, which we will denote by $|\mathcal{C}|$. The worst-case average error probability of a code $\mathcal{C}$ is

\[
p_e(\{W_i\}_{i \in \mathcal{I}}, \mathcal{C}) := \sup_{i \in \mathcal{I}} \frac{1}{|\mathcal{C}|} \sum_{k=1}^{|\mathcal{C}|} \text{Tr} W_i(\mathcal{C}_e(k))(I - \mathcal{C}_d(k)).
\]

Consider now a sequence $\mathcal{W} := \{\mathcal{W}_n\}_{n \in \mathbb{N}}$, where each $\mathcal{W}_n$ is a set of channels with input alphabet $\mathcal{X}^m$ and output space $\mathcal{H}^{\otimes n}$. The classical capacity $C(\mathcal{W})$ of $\mathcal{W}$ is the largest number $R$ such that there exists a sequence of codes $C^{(n)} = (C_e^{(n)}, C_d^{(n)})$ with

\[
\lim_{n \to +\infty} p_e(\mathcal{W}_n, C_n) = 0 \quad \text{and} \quad \liminf_{n \to +\infty} \frac{1}{n} \log |C_n| \geq R.
\]

We say that $\mathcal{W}$ is simple i.i.d. if $\mathcal{W}_n$ consists of a single element $W^{\otimes n}$ for every $n \in \mathbb{N}$ with some fixed channel $W$. In this case we denote the capacity by $C(W)$. The Holevo-Schumacher-Westmoreland theorem \cite{27, 51} tells that in this case

\[
C(W) \geq \hat{\chi}(W) = \sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi(W, p),
\]

where $\chi(W, p)$ is the Holevo quantity \cite{24}, and $\hat{\chi}(W)$ is the Holevo capacity \cite{25} of the channel. It is easy to see that \cite{50} actually holds as an equality, i.e., no sequence of codes with a rate above $\sup_{p \in \mathcal{P}_f(\mathcal{X})} \chi(W, p)$ can have an asymptotic error equal to zero; this is called the weak converse to the channel coding theorem, while the strong converse theorem \cite{15, 57} says that such sequences of codes always have an asymptotic error equal to 1.

Here we will consider two generalizations of the simple i.i.d. case: In the compound i.i.d. case $\mathcal{W}_n = \{W_i^{\otimes n}\}_{i \in \mathcal{I}}$ for some fixed channels $W_i : \mathcal{X} \to \mathcal{S}(\mathcal{H})$. In the averaged i.i.d. case $\mathcal{W}_n$ consists of the single element $\mathcal{W}_n := \sum_{i \in \mathcal{I}} \gamma_i W_i^{\otimes n}$, where $\mathcal{I}$ is finite, and $\gamma$ is a probability distribution on $\mathcal{I}$. The capacity of finite averaged channels has been shown to be equal to $\sup_{p \in \mathcal{P}_f(\mathcal{X})} \min_i \chi(W_i, p)$ in \cite{15}, and the same formula for the capacity of a finite compound channel follows from it in a straightforward way.

The protocol used in \cite{15} to show the achievability was to use a certain fraction of the communication rounds to guess which channel the parties are actually using, and then code for that channel in the remaining rounds. These results were generalized to arbitrary index sets $\mathcal{I}$ in \cite{11}, using a different approach. The starting point in \cite{11} was the following random coding theorem from \cite{10} (for the exact form below, see \cite{37}).

**Theorem 4.17.** Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a channel. For any $M \in \mathbb{N}$, and any $p \in \mathcal{P}_f(\mathcal{X})$, there exists a code $\mathcal{C}$ such that $|\mathcal{C}| = M$ and

\[
p_e(W, \mathcal{C}) \leq \kappa(c, \alpha) M^{1-\alpha} \text{Tr} \hat{W}(p)^{\alpha} (\hat{p} \otimes W(p))^{1-\alpha}
\]

for every $\alpha \in (0, 1)$ and every $c > 0$, where $\kappa(c, \alpha) := (1 + c)^{\alpha}(2 + c + 1/c)^{1-\alpha}$.

Applying the general properties of the Rényi divergences, established in Section \ref{sect:renyi} together with the single-shot coding theorem of Theorem 4.17, we get a very simple
proof of the achievability part of the coding theorems in [15] and [11]. Since our primary interest is the applicability of the inequalities of Section 3, we only consider the achievability part and not the converse. The key step of our approach is the following extension of Theorem 4.17 to multiple channels.

**Theorem 4.18.** Let \( W_i : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H}), i \in \mathcal{I}, \) be a set of channels, where \( \mathcal{I} \) is a finite index set. For every \( R \geq 0, \) every \( n \in \mathbb{N} \) and every \( p \in \mathcal{P}_f(\mathcal{X}), \) there exists a code \( C_n, n \in \mathbb{N}, \) such that for every \( \alpha \in (0, 1), \)

\[
|C_n| \geq \exp(nR), \quad \text{and} \quad p_e(\{W_i^{\otimes n}\}_{i \in \mathcal{I}}, C_n) \leq 8|\mathcal{I}|^2 \exp \left( n(\alpha - 1) \left( \alpha \min_i \chi^{(\text{old})}_\alpha(W_i, p) - R - (\alpha - 1) \log \dim(\mathcal{H}) \right) \right). \tag{57}
\]

**Proof.** Let \( M_n := \lfloor \exp(nR) \rfloor, n \in \mathbb{N} \) and \( \gamma_i := 1/|\mathcal{I}|, i \in \mathcal{I}. \) Applying Theorem 4.17 to \( \tilde{W}_n = \sum_{i \in \mathcal{I}} \gamma_i W_i^{\otimes n}, M_n \) and \( p^{\otimes n} \), we get the existence of a code \( C_n \) with \( |C_n| = M_n \) and

\[
p_e(\tilde{W}_n, C_n) \leq 8M_n^{1-\alpha} Q^{(\text{old})}_\alpha \left( \sum_{i \in \mathcal{I}} \gamma_i \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \tag{58}
\]

for every \( \alpha \in (0, 1). \) Here we chose \( c = 1, \) and used the upper bound \( \kappa(1, \alpha) \leq 8. \) We can further upper bound the RHS above as

\[
Q^{(\text{old})}_\alpha \left( \sum_{i \in \mathcal{I}} \gamma_i \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \leq Q^{(\text{new})}_\alpha \left( \sum_{i \in \mathcal{I}} \gamma_i \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \tag{59}
\]

\[
\leq \sum_{i \in \mathcal{I}} \gamma_i^\alpha Q^{(\text{new})}_\alpha \left( \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \tag{60}
\]

\[
\leq \sum_{i \in \mathcal{I}} \gamma_i^\alpha \sup_{\sigma \in \mathcal{S}(H^{\otimes n})} Q^{(\text{new})}_\alpha \left( \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \tag{61}
\]

\[
\leq \sum_{i \in \mathcal{I}} \gamma_i^\alpha \sup_{\sigma \in \mathcal{S}(H^{\otimes n})} Q^{(\text{old})}_\alpha \left( \tilde{W}_i^{\otimes n}(p^{\otimes n}) \right) \tag{62}
\]

\[
= \sum_{i \in \mathcal{I}} \gamma_i^\alpha \exp \left( \alpha(\alpha - 1) \chi^{(\text{old})}_\alpha(W_i^{\otimes n}, p^{\otimes n}) \right) \tag{63}
\]

\[
= \sum_{i \in \mathcal{I}} \gamma_i^\alpha \exp \left( n\alpha(\alpha - 1) \chi^{(\text{old})}_\alpha(W_i, p) \right) \tag{64}
\]

\[
\leq |\mathcal{I}| \exp \left( n\alpha(\alpha - 1) \min_{i \in \mathcal{I}} \chi^{(\text{old})}_\alpha(W_i, p) \right) \tag{65}
\]

where (59) is due to (7), (60) is due to (19), (61) is trivial, (62) follows from Corollary 3.10 and (64) is due to (55). Note that

\[
p_e(\tilde{W}_n, C_n) = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} p_e(W_i^{\otimes n}, C_n) \geq \frac{1}{|\mathcal{I}|} \sup_{i \in \mathcal{I}} p_e(W_i^{\otimes n}, C_n). \tag{66}
\]

Combining (58), (64), and (66), we get (57). \( \square \)
Corollary 4.19. Let $W_i : X \to S(H)$, $i \in I := \{1, \ldots, r\}$, be a set of channels, and let $\gamma_1, \ldots, \gamma_r$ be a probability distribution on $I$ with strictly positive weights. Then

$$C \left( \left\{ \sum_i \gamma_i W_i^{\otimes n} \right\}_{n \in \mathbb{N}} \right) = C \left( \left\{ W_i^{\otimes n} \right\}_{n \in \mathbb{N}} \right) \geq \sup_{p \in \mathcal{P}_f(X)} \min_i \chi(W_i, p). \quad (67)$$

Proof. Let $R < \min_i \chi(W_i, p)$, and for every $n \in \mathbb{N}$, let $C_n$ be a code as in Theorem 4.18. Then $\lim_{n \to \infty} \frac{1}{n} \log |C_n| \geq R$, and

$$\limsup_{n \to \infty} \frac{1}{n} \log p_e \left( \left\{ W_i^{\otimes n} \right\}_{i \in I}, C_n \right) \leq (\alpha - 1) \left( \alpha \min_i \chi^{(\text{old})}(W_i, p) - R - (\alpha - 1) \log \dim(H) \right).$$

Note that

$$\lim_{\alpha \to 1} \left( \alpha \min_i \chi^{(\text{old})}(W_i, p) - R - (\alpha - 1) \dim(H) \right) = \chi(W, p) - R$$

due to (50), and hence there exists an $\alpha_0 \in (0, 1)$ such that the upper bound in (57) goes to zero exponentially fast for every $\alpha \in (\alpha_0, 1)$. This proves the inequality in (68), and the equality of the two capacities is trivial.

When the channels are completely positive trace-preserving affine maps on the state space of a Hilbert space, the above results can be extended to the case of infinitely many channels by a simple approximation argument. It is easy to see that the same argument doesn’t work when the channels can be arbitrary maps on an input alphabet. Note that the classical capacity considered in the theorem below is the product-state capacity.

Theorem 4.20. Let $H_{\text{in}}$ and $H$ be finite-dimensional Hilbert spaces, and $W_i : S(H_{\text{in}}) \to S(H)$, $i \in I$, be completely positive trace-preserving affine maps, where $I$ is an arbitrary index set. Then

$$C \left( \left\{ W_i^{\otimes n} \right\}_{n \in \mathbb{N}} \right) \geq \sup_{p \in \mathcal{P}_f(X)} \inf_i \chi(W_i, p). \quad (68)$$

Proof. We assume that $\sup_{p \in \mathcal{P}_f(X)} \inf_i \chi(W_i, p) > 0$, since otherwise the assertion is trivial. Let $V$ be the vector space of linear maps from $\mathcal{B}(H_{\text{in}})$ to $\mathcal{B}(H)$, equipped with the norm $\| \Phi \| := \sup \{ \| \Phi(X) \|_1 : \| X \|_1 \leq 1 \}$, and let $D$ denote the real dimension of $V$. Let $\kappa > 0$, and for every $n \in \mathbb{N}$, let $I(n)$ be a finite index set such that $|I(n)| \leq (1 + 2e^{\kappa n})^D$ and $\delta_n := \sup_{i \in I(n)} \inf_{j \in I(n)} \| W_i - W_j \| \leq e^{-\kappa n}$. The existence of such index sets is guaranteed by Lemma 4.2.

Let $p \in \mathcal{P}_f(S(H_{\text{in}}))$ be such that $\inf_i \chi(W_i, p) > 0$, and for every $n \in \mathbb{N}$, let $C_n$ be a code as in Theorem 4.18 with $\mathcal{I}(n)$ in place of $\mathcal{I}$. It is easy to see that

$$p_e \left( \left\{ W_i^{\otimes n} \right\}_{i \in \mathcal{I}(n)}, C_n \right) \geq p_e \left( \left\{ W_i^{\otimes n} \right\}_{i \in I}, C_n \right) - n \delta_n,$$

and hence we have

$$p_e \left( \left\{ W_i^{\otimes n} \right\}_{i \in I}, C_n \right) \leq 8|\mathcal{I}(n)|^2 \exp \left[ n(\alpha - 1) \left( \alpha \inf_{i \in I} \chi^{(\text{old})}(W_i, p) - R - (\alpha - 1) \log \dim(H) \right) \right] + ne^{-\kappa n}.$$
Let $0 < R < \inf_{i \in I} \chi_{\alpha}^{\text{old}}(W_i, p)$. By the same argument as in the proof of Corollary 4.19 there exists an $\alpha \in (0, 1)$ such that $\varphi := \alpha \inf_{i \in I} \chi_{\alpha}^{\text{old}}(W_i, p) - R - (\alpha - 1) \log \dim(\mathcal{H}) > 0$. Choosing then $\kappa$ such that $2 \kappa D/(1 - \alpha) < \varphi$, we see that the error probability goes to zero exponentially fast, while the rate is at least $R$. This shows that $C (\{W_i^{\otimes n} : i \in I\}_{n \in \mathbb{N}}) \geq \inf_i \chi(W_i, p)$, and taking the supremum over $p$ yields the assertion.

Acknowledgment

The author is grateful to Professor Fumio Hiai and Nilanjana Datta for discussions. This research was supported by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Programme. The author also acknowledges support by the European Research Council (Advanced Grant “IRQUAT”). Part of this work was done when the author was a Marie Curie research fellow at the School of Mathematics, University of Bristol.

References

[1] T. Ando: Concavity of certain maps and positive definite matrices and applications to Hadamard products; Linear Algebra Appl. 26, 203–241 (1979)

[2] T. Ando, F. Hiai: Operator log-convex functions and operator means Math. Annalen

[3] H. Araki: On an inequality of Lieb and Thirring; Letters in Mathematical Physics; Volume 19, Issue 2, pp. 167–170, (1990)

[4] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete: Discriminating states: the quantum Chernoff bound; Phys. Rev. Lett. 98 160501, (2007)

[5] K.M.R. Audenaert, M. Nussbaum, A. Szkoła, F. Verstraete: Asymptotic error rates in quantum hypothesis testing; Commun. Math. Phys. 279, 251–283 (2008)

[6] K.M.R. Audenaert: On the Araki-Lieb-Thirring inequality; Int. J. of Information and Systems Sciences 4, pp. 78–83, (2008)

[7] Koenraad M.R. Audenaert, Nilanjana Datta: $\alpha$-z-relative Renyi entropies, arXiv:1310.7178, (2013)

[8] Salman Beigi: Quantum Rényi divergence satisfies data processing inequality; J. Math. Phys., 54, 122202 (2013)

[9] Igor Bjelakovic, Jean-Dominique Deuschel, Tyll Krüger, Ruedi Seiler, Rainer Siegmund-Schultze, Arleta Szkoła: A quantum version of Sanov’s theorem; Communications in Mathematical Physics Vol. 260, Issue 3, pp. 659–671, (2005)
[10] Igor Bjelakovic, Jean-Dominique Deuschel, Tyll Krüger, Ruedi Seiler, Rainer Siegmund-Schultze, Arleta Szkoła: *Typical support and Sanov large deviations of correlated states*; Communications in Mathematical Physics Vol. 279, pp. 559–584, (2008)

[11] I. Bjelakovic, H. Boche: *Classical capacities of compound and averaged quantum channels*; IEEE Trans. Inform. Theory 55, 3360–3374, (2009)

[12] E.A. Carlen, E.H. Lieb: *A Minkowski type trace inequality and strong subadditivity of entropy*; Amer. Math. Soc. Transl. Ser. 2 textbf189, 59–68, (1999)

[13] Eric A. Carlen, Elliot H. Lieb: *A Minkowski type inequality and strong subadditivity of quantum entropy II: convexity and concavity*; Lett. Math. Phys. 83, pp. 107–126, (2008)

[14] I. Csiszár: *Generalized cutoff rates and Rényi’s information measures*; IEEE Trans. Inf. Theory 41, 26–34, (1995)

[15] N. Datta, T.C. Dorlas: *The Coding Theorem for a Class of Quantum Channels with Long-Term Memory*; Journal of Physics A: Mathematical and Theoretical, vol. 40, p. 8147, (2007)

[16] N. Datta: *Min- and Max-Relative Entropies and a New Entanglement Monotone*; IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2816–2826, (2009).

[17] Nilanjana Datta and Felix Leditzky: *A limit of the quantum Rényi divergence*; J. Phys. A: Math. Theor. 47, 045304, (2014)

[18] Rupert L. Frank and Elliott H. Lieb: *Monotonicity of a relative Rényi entropy*; arXiv:1306.5358, (2013).

[19] M. Hayashi, H. Nagaoka: *General Formulas for Capacity of Classical-Quantum Channels*; IEEE Trans. Inf. Theory 49, (2003)

[20] M. Hayashi: *Quantum Information Theory: An Introduction*; Springer, (2006).

[21] M. Hayashi: *Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding*; Phys. Rev. A 76, 062301, (2007)

[22] F. Hiai, D. Petz: *The proper formula for relative entropy and its asymptotics in quantum probability*; Comm. Math. Phys. 143, 99–114 (1991).

[23] F. Hiai, M. Mosonyi, T. Ogawa: *Error exponents in hypothesis testing for correlated states on a spin chain*; J. Math. Phys. 49, 032112, (2008)

[24] F. Hiai, M. Mosonyi, M. Hayashi, *Quantum hypothesis testing with group symmetry*, J. Math. Phys. 50, 103304 (2009)

[25] F. Hiai: *Matrix Analysis: Matrix Monotone Functions, Matrix Means, and Majorization* (GSIS selected lectures); Interdisciplinary Information Sciences 16, 139–248 (2010)
[26] F. Hiai: Concavity of certain matrix trace and norm functions; Linear Algebra and Appl. 439, 1568–1589, (2013)

[27] A.S. Holevo: The capacity of the quantum channel with general signal states; IEEE Transactions on Information Theory, vol. 44, no. 1, pp. 269-273, (1998)

[28] V. Jaksic, Y. Ogata, C.-A. Pillet, R. Seiringer: Quantum hypothesis testing and non-equilibrium statistical mechanics; Rev. Math. Phys. 24, no. 6, 1230002, (2012)

[29] R. Jozsa, B. Schumacher: A new proof of the quantum noiseless coding theorem; Journal of Modern Optics Volume 41, Issue 12, (1994)

[30] R. Jozsa, M. Horodecki, P. Horodecki, R. Horodecki: Universal quantum information compression; Phys. Rev. Lett. 81, 1714–1717, (1998)

[31] R. König and S. Wehner: A strong converse for classical channel coding using entangled inputs; Physical Review Letters, vol. 103, no. 7, 070504, (2009)

[32] E.H. Lieb: Convex trace functions and the Wigner-Yanase-Dyson conjecture; Adv. Math. 11, 267–288 (1973)

[33] E.H. Lieb, W. Thirring: Studies in mathematical physics; pp. 269–297. Princeton University Press, Princeton, (1976)

[34] Vitali D. Milman, Gideon Schechtman: Asymptotic Theory of Finite Dimensional Normed Spaces; Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, (1986)

[35] M. Mosonyi, F. Hiai, T. Ogawa, M. Fannes, Asymptotic distinguishability measures for shift-invariant quasi-free states of fermionic lattice systems, J. Math. Phys. 49, 072104, (2008)

[36] M. Mosonyi, Hypothesis testing for Gaussian states on bosonic lattices, J. Math. Phys. 50, 032104, (2009)

[37] M. Mosonyi, N. Datta: Generalized relative entropies and the capacity of classical-quantum channels; J. Math. Phys. 50, 072104 (2009)

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

[39] M. Mosonyi, T. Ogawa: Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies; arXiv:1308.3228, (2013)

[40] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, Marco Tomamichel: On quantum Rényi entropies: a new definition and some properties; J. Math. Phys. 54, 122203, (2013)

[41] H. Nagaoka: Strong converse theorems in quantum information theory; in the book “Asymptotic Theory of Quantum Statistical Inference” edited by M. Hayashi, World Scientific, (2005)
[42] H. Nagaoka: *The converse part of the theorem for quantum Hoeffding bound*; quant-ph/0611289

[43] M.A. Nielsen, I.L. Chuang: *Quantum Information and Quantum Computation*; Cambridge University Press, Cambridge, UK, (2000)

[44] J. Nötzel: *Hypothesis testing on invariant subspaces of the symmetric group, part I - quantum Sanov’s theorem and arbitrarily varying sources*; arXiv:1310.5553, (2013)

[45] T. Ogawa, H. Nagaoka: *Strong converse to the quantum channel coding theorem*; IEEE Transactions on Information Theory, vol. 45, no. 7, pp. 2486-2489, (1999)

[46] T. Ogawa, H. Nagaoka: *Strong converse and Stein’s lemma in quantum hypothesis testing*; IEEE Trans. Inform. Theory 47, 2428–2433 (2000).

[47] D. Petz: *Quasi-entropies for finite quantum systems*; Rep. Math. Phys. 23, 57–65, (1986)

[48] R. Renner: *Security of Quantum Key Distribution*, PhD dissertation, Swiss Federal Institute of Technology Zurich, Diss. ETH No. 16242, (2005).

[49] A. Rényi: *On measures of entropy and information*; Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, pp. 547–561, Univ. California Press, Berkeley, California, (1961)

[50] B. Schumacher: *Quantum coding*; Phys. Rev. A 51, 2738, (1995)

[51] B. Schumacher, M. Westmoreland: *Sending classical information via noisy quantum channels*; Physical Review A, vol. 56, no. 1, pp. 131-138, (1997)

[52] R. Sibson: *Information radius*; Z. Wahrscheinlichkeitsthe. Verw. Gebiete 14, 149–161, (1969)

[53] M. Tomamichel, R. Colbeck, R. Renner: *A fully quantum asymptotic equipartition property*; IEEE Trans. Inform. Theory 55, 5840–5847, (2009)

[54] M. Tomamichel: *A framework for non-asymptotic quantum information theory*; PhD thesis, ETH Zürich, (2006)

[55] H. Umegaki: *Conditional expectation in an operator algebra*; Kodai Math. Sem. Rep. 14, 59–85, (1962)

[56] Mark M. Wilde, Andreas Winter, Dong Yang: *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels*; arXiv:1306.1586, (2013)

[57] A. Winter: *Coding theorem and strong converse for quantum channels*; IEEE Transactions on Information Theory, vol. 45, no. 7, pp. 2481–2485, (1999)