An information-theoretic treatment of quantum dichotomies

Francesco Buscemi[1], David Sutter[2], and Marco Tomamichel[3]

[1]Graduate School of Informatics, Nagoya University, Nagoya, Japan

[2]Institute for Theoretical Physics, ETH Zurich, Switzerland

[3]Centre for Quantum Software and Information and School of Computer Science,
University of Technology Sydney, Sydney

Given two pairs of quantum states, we want to decide if there exists a quantum channel that transforms one pair into the other. The theory of quantum statistical comparison and quantum relative majorization provides necessary and sufficient conditions for such a transformation to exist, but such conditions are typically difficult to check in practice. Here, by building upon work by Keiji Matsumoto, we relax the problem by allowing for small errors in one of the transformations. In this way, a simple sufficient condition can be formulated in terms of one-shot relative entropies of the two pairs. In the asymptotic setting where we consider sequences of state pairs, under some mild convergence conditions, this implies that the quantum relative entropy is the only relevant quantity deciding when a pairwise state transformation is possible. More precisely, if the relative entropy of the initial state pair is strictly larger compared to the relative entropy of the target state pair, then a transformation with exponentially vanishing error is possible. On the other hand, if the relative entropy of the target state is strictly larger, then any such transformation will have an error converging exponentially to one. As an immediate consequence, we show that the rate at which pairs of states can be transformed into each other is given by the ratio of their relative entropies. We discuss applications to the resource theories of athermality and coherence, where our results imply an exponential strong converse for general state interconversion.

1 Introduction

Various pre- and partial orders have been the subject of extensive study both in mathematical statistics [24, 5, 32, 55, 2, 54, 21] and in information theory [49, 29, 20]. An example of paramount importance is that provided by the majorization preorder [34]: a probability distribution $\vec{p}_1$ is said to majorize another distribution $\vec{p}_2$, in formula $\vec{p}_1 \succeq \vec{p}_2$, whenever there exists a bistochastic[1] transformation T such that $T\vec{p}_1 = \vec{p}_2$. The majorization preorder is particularly relevant and useful because of a famous result by Hardy, Littlewood, and Pólya, according to which the relation $\vec{p}_1 \succeq \vec{p}_2$ can be expressed in terms of a finite set of inequalities [24] of the form $f_i(\vec{p}_1) \geq f_i(\vec{p}_2)$, intuitively capturing the

---

Marco Tomamichel: marco.tomamichel@uts.edu.au

[1]A transformation is said to be bistochastic if it transforms probability distributions to probability distributions, while keeping the uniform distribution fixed.

idea that $\vec{p}_1$ is "less uniform" than $\vec{p}_2$. Such inequalities can be conveniently visualized by plotting the Lorenz curve of $\vec{p}_1$ versus that of $\vec{p}_2$ [34].

As it involves the comparison of two probability distributions relative to a third one (i.e., the uniform distribution), the majorization preorder is naturally generalized by considering two pairs of probability distributions, that is, two *dichotomies* $(\vec{p}_1, \vec{q}_1)$ and $(\vec{p}_2, \vec{q}_2)$, where now $\vec{q}_1$ and $\vec{q}_2$ are arbitrary distributions. One then writes $(\vec{p}_1, \vec{q}_1) \succeq (\vec{p}_2, \vec{q}_2)$ whenever there exists a stochastic transformation simultaneously mapping $\vec{p}_1$ to $\vec{p}_2$ and $\vec{q}_1$ to $\vec{q}_2$. As a consequence of Blackwell's equivalence theorem [5], also the more general case of dichotomies is completely characterized by a finite set of simple inequalities, which directly reduce to those of Hardy, Littlewood and Polya if $\vec{q}_1$ and $\vec{q}_2$ are both taken to be uniform. Also in this more general scenario, a *relative* Lorenz curve can be associated to each dichotomy, and the preorder $\succeq$ visualized accordingly [47].

In relation to quantum information sciences, while the preorder of majorization has found early applications in entanglement theory [41], the notion of *relative* majorization have started being employed only more recently, especially due to its applications in quantum thermodynamics [26, 6, 7, 47, 10, 23] and generalized resource theories [16]. In the quantum setting, the objects of comparison are *quantum dichotomies*, namely, pairs of density matrices.

Let us consider two arbitrary finite-dimensional quantum dichotomies (ρ_1, σ_1) and (ρ_2, σ_2). In complete analogy with the classical case *à la* Blackwell, the relative majorization preorder $\succeq$ can be extended to the quantum setting by writing $(\rho_1, \sigma_1) \succeq (\rho_2, \sigma_2)$ whenever there exists a completely positive trace-preserving map $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \rho_2$ and $\mathcal{E}(\sigma_1) = \sigma_2$ simultaneously. However, in the quantum case[2], there is no known simple set of inequalities, analogously to the comparison of two relative Lorenz curves, able to completely capture the relative majorization preorder [44, 27, 38, 10]: statistical conditions can be derived [8, 36, 28] but they typically involve an infinite number of inequalities, thus becoming much more cumbersome to verify.

In order to overcome such problems, in this paper we build upon an information-theoretic approach for the comparison of quantum dichotomies first considered by Matsumoto in [37]. This involves the relaxation of the order $\succeq$ to allow for errors in the transformation, and the consideration of an asymptotic regime, in which an increasing number of identical copies of one dichotomy get transformed into something that resembles, up to an arbitrarily high level of accuracy, many copies of the other dichotomy. More precisely, while we allow for (small) errors in the transformation $\mathcal{E}(\rho_1) \approx \rho_2$, the condition $\mathcal{E}(\sigma_1) = \sigma_2$ must always be satisfied exactly, as if it were a sort of "conservation rule." We compute that the optimal rate at which such transformation can happen is given by the ratio between the quantum relative entropies $D(\rho_1 \| \sigma_1)$ and $D(\rho_2 \| \sigma_2)$. Our result hence shows that quantum dichotomies, while enjoying a very rich structure, are asymptotically characterized by a single number, that is, their relative entropy.

The remainder of the paper is structured as follows. After introducing the relevant one-shot divergences and their properties in Section 2, we establish our main technical results in Section 3. First, in Section 3.1 we derive sufficient conditions in terms of one-shot divergences for exact pairwise state transformations. In Section 3.2 we relax this to allow for an error on one of the states, and again find sufficient conditions in terms of smoothed one-shot entropies. This then allows us to derive our main results in Section 3.3, Theorems 3.4 and 3.6, which together show that the relative entropy fully characterizes when pairwise transformations are possible asymptotically. Section 4 then takes an information-

---

[2]With the notable exception of the qubit case [1, 12] and pure states [11, 50].

theoretic approach to the problem by studying the maximal rate at which independent copies of states can be transformed into each other in Theorem 4.1. We finally discuss applications of our results to the resource theories of athermality and coherence in Section 5 and end with a conjecture characterizing the second-order asymptotic behavior of resource transformations in Section 6.

Note added. Concurrent work [56, **?**] also derives some of our results in the setting of resource theories and thermodynamics. We explore these connections in Section 5 for the convenience of the reader.

2 Preliminaries

2.1 Notation

Let $S(\mathbb{C}^d)$ denote the set of *quantum states* on a d-dimensional Hilbert space $\mathbb{C}^d$. A *quantum channel* $\mathcal{E} : S(\mathbb{C}^d) \to S(\mathbb{C}^{d'})$ is a linear map that is *completely positive* and *trace-preserving* (CPTP). We denote by $\leq$ the Löwner partial order, i.e., for two Hermitian matrices X and Y the relation $X \geq Y$ means that $X - Y$ is positive semi-definite, and the relation $X \gg Y$ means that the support of Y is contained in the support of X. Throughout this paper we denote by log the logarithm to base 2.

We denote the *trace distance* between two states ρ and σ by $T(\rho, \sigma) := \frac{1}{2} \|\rho - \sigma\|_1$, where $\| \cdot \|_1$ denotes the Schatten 1-norm. The *fidelity* is given as $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$. We will also use the *sine* or *purified distance*, $P(\rho, \sigma) := \sqrt{1 - F(\rho, \sigma)}$. Both P and T satisfy the triangle inequality and are non-increasing when a CPTP map is applied to both states. They are related by the Fuchs-van de Graaf inequalities stating that $T(\rho, \sigma) \leq P(\rho, \sigma) \leq \sqrt{2T(\rho, \sigma) - T(\rho, \sigma)^2}$. When the choice of metric is arbitrary, we use Δ to stand for either T or P.

2.2 Some divergences and their properties

In this work we use several different non-commutative divergences. We will introduce here only the measures and properties that are needed for this work — an interested reader may consult [51] for a more comprehensive discussion with references to all the original papers. For $\rho, \sigma \in S(\mathbb{C}^d)$ the relative entropy is given by $D(\rho\|\sigma) := \operatorname{tr}\rho(\log\rho - \log\sigma)$ if $\sigma \gg \rho$, and $+\infty$ otherwise. To simplify the exposition in the following we assume throughout that the states σ always have full support and are thus invertible, avoiding such infinities.

For $\alpha \in (0, 1) \cup (1, 2]$ the *Petz quantum Rényi divergence* [43] is defined as

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

In the limit $\alpha \to 0$ the divergence converges to the *min-relative entropy* [48, 22], i.e.,

$$D_{\min}(\rho\|\sigma) := -\log \operatorname{tr} \sigma \Pi_\rho = \lim_{\alpha \to 0} \bar{D}_\alpha(\rho\|\sigma), \tag{2}$$

where Π_ρ is the projector onto the support of ρ.

Another non-commutative family of Rényi divergences is the *sandwiched quantum Rényi divergence* [40, 57], for $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty)$ defined as

$$\tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{tr}\big(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\big)^\alpha. \tag{3}$$

For $\alpha = \frac{1}{2}$ the sandwiched quantum Rényi divergence becomes $\tilde{D}_{1/2}(\rho\|\sigma) = -\log F(\rho,\sigma)$. In the limit $\alpha \to \infty$ the sandwiched quantum Rényi divergence converges to the so-called *max-divergence* [48, 22], i.e.,

$$D_{\max}(\rho\|\sigma) := \inf\left\{\lambda \in \mathbb{R} : \rho \leq 2^\lambda \sigma\right\}. \tag{4}$$

Both non-commutative families of Rényi divergences introduced above satisfy many desirable properties: they are monotonically increasing in α, they satisfy the data-processing inequality (i.e. they are non-increasing when the same CPTP map is applied to both states), and in the limit $\alpha \to 1$ they both converge to the relative entropy.

Smooth variants of the max and min-divergence will be useful to treat problems with finite errors. The *ε-smooth max-divergence* is defined as

$$D_{\max}^{\varepsilon,\Delta}(\rho\|\sigma) := \inf_{\tilde{\rho} \in \mathcal{B}_\varepsilon^\Delta(\rho)} D_{\max}(\tilde{\rho}\|\sigma), \tag{5}$$

where $\mathcal{B}_\varepsilon^\Delta(\rho) := \{\tilde{\rho} \in \mathrm{S}(\mathbb{C}^d) : \Delta(\rho,\tilde{\rho}) \leq \varepsilon\}$ for $\varepsilon \in (0,1)$. We will use this definition for both trace distance and purified distance, but note that in contrast to some other works the optimization here only goes over close quantum states, not sub-normalized states. As the smooth variant of $D_{\min}$ we introduce the so-called *hypothesis testing divergence*.[3] For any $\varepsilon \in (0,1)$ it is defined as

$$D_h^\varepsilon(\rho\|\sigma) := -\log\inf\left\{\operatorname{tr}\sigma Q : 0 \leq Q \leq \mathbb{1} \ \wedge \ \operatorname{tr}\rho Q \geq 1 - \varepsilon\right\}. \tag{6}$$

In the limit $\varepsilon \to 0$ we recover $D_{\min}(\rho\|\sigma)$. Finally, we note that both of these divergences satisfy the data-processing inequality, namely

$$D_{\max}^{\varepsilon,\Delta}(\rho\|\sigma) \geq D_{\max}^{\varepsilon,\Delta}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \qquad \text{and} \qquad D_h^\varepsilon(\rho\|\sigma) \geq D_h^\varepsilon(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \tag{7}$$

for any CPTP map $\mathcal{E}$.

The smooth max-divergence and the hypothesis testing divergence are closely related, as shown in [3, Theorem 4].

Proposition 2.1. *Let $\rho,\sigma \in \mathrm{S}(\mathbb{C}^d)$ and $\varepsilon \in (0,1)$ and $\nu \in (0, 1-\varepsilon)$. It holds that*

$$D_h^{1-\varepsilon}(\rho\|\sigma) \geq D_{\max}^{\sqrt{\varepsilon},P}(\rho\|\sigma) - \log\frac{1}{1-\varepsilon} \geq D_h^{1-\varepsilon-\nu}(\rho\|\sigma) - \log\frac{4}{\nu^2}. \tag{8}$$

We next recall two inequalities that relate the hypothesis testing relative entropy and the max-relative entropy to the Rényi divergences.

Proposition 2.2. *Let $\varepsilon \in (0,1)$ and let $\rho,\sigma \in \mathrm{S}(\mathbb{C}^d)$. Then,*

$$D_h^\varepsilon(\rho\|\sigma) \geq \bar{D}_\alpha(\rho\|\sigma) - \frac{\alpha}{1-\alpha}\log\frac{1}{\varepsilon} \qquad\qquad \text{for } \alpha \in [0,1), \quad \text{and} \tag{9}$$

$$D_{\max}^{\varepsilon,\Delta}(\rho\|\sigma) \leq \tilde{D}_\alpha(\rho\|\sigma) + \frac{1}{\alpha-1}\log\frac{1}{\varepsilon^2} + \log\frac{1}{1-\varepsilon^2} \qquad \text{for } \alpha \in (1,\infty]. \tag{10}$$

We can interchange $\bar{D}_\alpha$ and $\tilde{D}_\alpha$ in the above inequalities.

---

[3] In some earlier works (see, e.g., [22]) an optimization over a ball of ε-close states is used instead to define a smooth variant of $D_{\min}$, in analogy with $D_{\max}^\varepsilon$ defined above. However, the quantity we use here, first introduced in [9] as a generalization of $D_{\min}$, is more natural for our problem here.

Note that we can interchange the Rényi relative entropies in both inequalities since $\bar{D}_\alpha(\rho\|\sigma) \geq \tilde{D}_\alpha(\rho\|\sigma)$ for all states due to the Araki-Lieb-Thirring inequality. Inequality (9) follows immediately from [4, Proposition 3.2]. Inequality (10) is shown in [3, Theorem 3], tightening earlier bounds that were established as part of the fully quantum asymptotic equipartition property (QAEP) [52]. The QAEP states that, for all $\varepsilon \in (0,1)$, the regularized smooth entropies converge to the relative entropy

$$\lim_{n\to\infty} \frac{1}{n} D_{\max}^{\varepsilon,\Delta}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) = D(\rho\|\sigma) .\tag{11}$$

The analogous statement for D_h^ε is an immediate consequence from quantum Stein's lemma and its converse [25, 42].

3 Conditions for pairwise state transformation

In this section we derive sufficient conditions for the existence of a channel that transforms (ρ_1, σ_1) to (ρ_2, σ_2), where the first state is transformed either exactly or approximately, and the second state always has to be transformed exactly.

3.1 Conditions for exact state transformation

We start by considering the case of exact transformations. We note that conditions for an exact transformation derived in this section are very restrictive. If we allow for approximate transformations as done in Section 3.2 we will find conditions that are considerably easier to fulfill.

For what follows, we can restrict ourselves to a very special class of transformations, namely, *test-and-prepare channels* of the form

$$\mathcal{E}(\rho) = \gamma_1 \operatorname{tr} E\rho + \gamma_2 \operatorname{tr}(\mathbb{1} - E)\rho ,\tag{12}$$

where the γ_i's are density matrices and $0 \leq E \leq \mathbb{1}$. Hence test-and-prepare channels constitute a subset of measure-and-prepare channels, in which the measurement is a simple binary test. Ref. [10] provides a complete characterization of this case. The following lemma can be obtained as a consequence of the results in [10], but we provide an independent proof here for the sake of the reader (see also Ref. [39]).

Lemma 3.1. *Let $\rho_1, \sigma_1 \in \mathrm{S}(\mathbb{C}^2)$ be commuting qubit quantum states and let $\rho_2, \sigma_2 \in \mathrm{S}(\mathbb{C}^d)$. The following two conditions are equivalent:*

(i) there exists a CPTP map $\mathcal{E} : \mathrm{S}(\mathbb{C}^2) \to \mathrm{S}(\mathbb{C}^d)$ such that $\mathcal{E}(\rho_1) = \rho_2$ and $\mathcal{E}(\sigma_1) = \sigma_2$;

(ii) $D_{\max}(\rho_1\|\sigma_1) \geq D_{\max}(\rho_2\|\sigma_2)$ and $D_{\max}(\sigma_1\|\rho_1) \geq D_{\max}(\sigma_2\|\rho_2)$.

Proof. Since the implication (i) $\implies$ (ii) is just the data-processing inequality, we only need to prove the reverse implication (ii) $\implies$ (i).

By assumption, ρ_1 and σ_1 commute. Hence, we can see them as classical binary probability distributions, namely, $\rho_1 \leftrightarrow \vec{p}_1 = (p, 1-p)$ and $\sigma_1 \leftrightarrow \vec{q}_1 = (q, 1-q)$. Moreover, we can assume that $\frac{p}{q} \geq \frac{1-p}{1-q}$; otherwise, the first step is to map $(p, 1-p)$ and $(q, 1-q)$ into $(1-p, p)$ and $(1-q, q)$, respectively. Notice that this is equivalent to saying that, without loss of generality, we can always assume $p(1-q) \geq q(1-p)$, that is, $p \geq q$.

Let us now define $M := \frac{p}{q}$ and $m := \frac{1-p}{1-q}$. Notice that $\log M = D_{\max}(\rho_1\|\sigma_1)$ and $\log m = -D_{\max}(\sigma_1\|\rho_1)$. Problems with such definitions occur when $q = 0$ or $q = 1$. If

$q = 1$, since (as noticed above) we can assume that $p \geq q$, then also $p = 1$; that is, $\rho_1 = \sigma_1$ and $D_{\max}(\rho_1 \| \sigma_1) = -D_{\max}(\sigma_1 \| \rho_1) = 0$ (or, equivalently, $M = m = 1$). By Condition (ii), this implies that also $D_{\max}(\rho_2 \| \sigma_2) = 0$, namely, $\rho_2 = \sigma_2$, and there is nothing to prove. The case $q = 0$ will be addressed separately at the end of the proof.

For the moment, we thus assume that $\infty > M > 1 > m \geq 0$, that is, $p > q > 0$. Condition (ii) then guarantees that $M\sigma_2 - \rho_2 \geq 0$ and $\rho_2 - m\sigma_2 \geq 0$. We thus define a linear map $\mathcal{E} : \mathrm{S}(\mathbb{C}^2) \to \mathrm{S}(\mathbb{C}^d)$ as follows:

$$\mathcal{E}(\cdot) := \langle 0 | \cdot | 0 \rangle \frac{\rho_2 - m\sigma_2}{1 - m} + \langle 1 | \cdot | 1 \rangle \frac{M\sigma_2 - \rho_2}{M - 1} \; . \tag{13}$$

By construction, $\mathcal{E}$ is clearly CPTP. On the other hand, by direct inspection, $\mathcal{E}(\rho_1) = \rho_2$ and $\mathcal{E}(\sigma_1) = \sigma_2$.

In order to conclude the proof, we are left to address the case in which $q = 0$. In such a case, we consider a variant of the map in (13), that is,

$$\mathcal{F}(\cdot) := \langle 0 | \cdot | 0 \rangle \frac{\rho_2 - m\sigma_2}{1 - m} + \langle 1 | \cdot | 1 \rangle \sigma_2 \; .$$

Again, by direct inspection it is easy to check that $\mathcal{F}$ is CPTP and that $\mathcal{E}(\rho_1) = \rho_2$ and $\mathcal{E}(\sigma_1) = \sigma_2$, as claimed. $\qquad \square$

From the above, we obtain a sufficient condition for the existence of a transformation, more precisely a test-and-prepare channel, for arbitrary pairs of states.

Corollary 3.2. *Let $\rho_1, \sigma_1 \in \mathrm{S}(\mathbb{C}^{d_1})$ and $\rho_2, \sigma_2 \in \mathrm{S}(\mathbb{C}^{d_2})$. Then, if either*

$$D_{\min}(\rho_1 \| \sigma_1) \geq D_{\max}(\rho_2 \| \sigma_2) \qquad or \qquad D_{\min}(\sigma_1 \| \rho_1) \geq D_{\max}(\sigma_2 \| \rho_2) \, , \tag{14}$$

then there exists a test-and-prepare channel $\mathcal{E}$ such that $\mathcal{E}(\rho_1) = \rho_2$ and $\mathcal{E}(\sigma_1) = \sigma_2$.

Proof. It suffices to show the statement under the first condition; the second then follows by symmetry. Let us consider the measurement channel

$$\mathcal{M}(\cdot) := \mathrm{tr}[\cdot \, \Pi] | 0 \rangle \langle 0 | + \mathrm{tr}[\cdot \, (\mathbb{1} - \Pi)] | 1 \rangle \langle 1 | \, ,$$

where Π is the projector onto the support of ρ_1. Then, the binary classical probability distributions obtained from the pair (ρ_1, σ_1) are $\mathcal{M}(\rho_1) \leftrightarrow \vec{p}_1 = (p, 1-p) = (1, 0)$ and $\mathcal{M}(\sigma_1) \leftrightarrow \vec{q}_1 = (q, 1-q) = (\mathrm{tr}[\Pi \, \sigma_1], \mathrm{tr}[(\mathbb{1} - \Pi) \, \sigma_1])$.

We only need to show that, if Eq. (14) holds, then $\vec{p}_1$ and $\vec{q}_1$ satisfy condition (ii) in Lemma 3.1. That is indeed the case, since, on the one hand,

$$\frac{p}{q} = \frac{1}{\mathrm{tr}[\Pi \, \sigma_1]} = 2^{D_{\min}(\rho_1 \| \sigma_1)} \, ,$$

that is, $D_{\max}(\vec{p}_1 \| \vec{q}_1) = D_{\min}(\rho_1 \| \sigma_1)$, so that Eq. (14) guarantees that $D_{\max}(\vec{p}_1 \| \vec{q}_1) \geq D_{\max}(\rho_2 \| \sigma_2)$. Notice that, in the above equation, the case $q = 0$ can be excluded, as it is equivalent to ρ_1 and σ_1 being perfectly distinguishable, and thus exactly transformable into any other pair of states.

On the other hand, because

$$2^{-D_{\max}(\vec{q}_1 \| \vec{p}_1)} = \frac{1 - p}{1 - q} = 0 \leq 2^{-D_{\max}(\sigma_2 \| \rho_2)} \, .$$

the remaining condition, $D_{\max}(\vec{q}_1 \| \vec{p}_1) \geq D_{\max}(\sigma_2 \| \rho_2)$, is guaranteed by definition. In the above equation, the case $q = 1$ can be excluded, as it is equivalent to ρ_1 and σ_1 having the same support, that is, $D_{\min}(\rho_1 \| \sigma_1) = D_{\min}(\sigma_1 \| \rho_1) = 0$, which, by virtue of assumption (14), implies that $\rho_2 = \sigma_2$, so that there is nothing to prove. $\qquad \square$

We note that condition (14) is very strong in general. To see this, it is enough to simply consider two states with coinciding support, which is what happens generically (in the generic case, states are invertible with probability one). Then, since in such a case $D_{\min}(\rho_1\|\sigma_1) = D_{\min}(\sigma_1\|\rho_1) = 0$, the only dichotomies satisfying Eq. (14) are the trivial ones, i.e., those with $\rho_2 = \sigma_2$, as a consequence of the fact that max-relative entropy vanishes if and only if the two arguments coincide.

However, when errors are allowed in condition (14), in what follows we show that a much more flexible framework can be constructed, which is in fact sufficient to derive strong asymptotic results.

3.2 Sufficient condition for approximate state transformation

In this section we are interested in approximate state transformation, i.e., a transformation from (ρ_1, σ_1) to (ρ_2, σ_2) where we allow for a (small) error in the transformation $\rho_1 \to \rho_2$, while the transformation $\sigma_1 \to \sigma_2$ is required to be exact.

Lemma 3.3. *Let $\varepsilon_1, \varepsilon_2 \in (0,1)$, $\rho_1, \sigma_1 \in \mathrm{S}(\mathbb{C}^{d_1})$ and $\rho_2, \sigma_2 \in \mathrm{S}(\mathbb{C}^{d_2})$. If either*

$$D_h^{\varepsilon_1}(\rho_1\|\sigma_1) \geq D_{\max}^{\varepsilon_2,T}(\rho_2\|\sigma_2) \qquad or \qquad D_h^{\varepsilon_1}(\rho_1\|\sigma_1) \geq D_{\max}^{\varepsilon_2,P}(\rho_2\|\sigma_2)\,, \tag{15}$$

then there exists a test-and-prepare quantum channel $\mathcal{E} : \mathrm{S}(\mathbb{C}^{d_1}) \to \mathrm{S}(\mathbb{C}^{d_2})$ satisfying $\mathcal{E}(\sigma_1) = \sigma_2$ and $T(\mathcal{E}(\rho_1), \rho_2) \leq \varepsilon_1 + \varepsilon_2$ or $P(\mathcal{E}(\rho_1), \rho_2) \leq \sqrt{\varepsilon_1} + \varepsilon_2$, respectively.

Proof. Let Q^* denote the optimizer in (6) for $D_h^{\varepsilon_1}(\rho_1\|\sigma_1)$ that satisfies

$$2^{-D_h^{\varepsilon_1}(\rho_1\|\sigma_1)} = \mathrm{tr}\,\sigma_1 Q^* \qquad \text{and} \qquad \mathrm{tr}\sigma_1 Q^* \leq \mathrm{tr}\,\rho_1 Q^* = 1 - \varepsilon_1 < 1\,. \tag{16}$$

By definition of the smooth max-relative entropy, there furthermore exists a state $\tilde{\rho}_2 \in \mathcal{B}_{\varepsilon_2}^{\Delta}(\rho_2)$ for $\Delta \in \{T, P\}$, such that

$$D_{\max}^{\varepsilon_2,\Delta}(\rho_2\|\sigma_2) = D_{\max}(\tilde{\rho}_2\|\sigma_2)\,. \tag{17}$$

Consider now the mapping

$$\mathcal{E} : X \mapsto \mathcal{E}(X) = \tilde{\rho}_2 \mathrm{tr}(XQ^*) + \mathrm{tr}\big(X(\mathbb{1} - Q^*)\big)\frac{\sigma_2 - \tilde{\rho}_2\mathrm{tr}(\sigma_1 Q^*)}{\mathrm{tr}(\sigma_1(\mathbb{1} - Q^*))} \tag{18}$$

$$= \left(\mathrm{tr}X - \frac{\mathrm{tr}\big(X(\mathbb{1} - Q^*)\big)}{\mathrm{tr}(\sigma_1(\mathbb{1} - Q^*))}\right)\tilde{\rho}_2 + \left(\frac{\mathrm{tr}\big(X(\mathbb{1} - Q^*)\big)}{\mathrm{tr}(\sigma_1(\mathbb{1} - Q^*))}\right)\sigma_2\,. \tag{19}$$

We start by proving that $\mathcal{E}$ is a quantum channel, i.e., a trace-preserving completely positive map. To see that $\mathcal{E}$ is trace-preserving is straightforward since σ_2 and $\tilde{\rho}_2$ are density operators. We note that because $0 \leq Q^* \leq \mathbb{1}$ it suffices to show that $\sigma_2 \geq \tilde{\rho}_2\mathrm{tr}(\sigma_1 Q^*)$ in order prove that $\mathcal{E}$ is completely positive. By definition of the smooth max-relative entropy and by using (15) and (16) we find

$$2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)} = 2^{-D_{\max}^{\varepsilon_2,\Delta}(\rho_2\|\sigma_2)} \geq 2^{-D_h^{\varepsilon_1}(\rho_1\|\sigma_1)} = \mathrm{tr}\sigma_1 Q^*\,. \tag{20}$$

By definition of the max-relative entropy we thus have

$$\sigma_2 \geq \tilde{\rho}_2\, 2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)} \geq \tilde{\rho}_2\,\mathrm{tr}(\sigma_1 Q^*)\,. \tag{21}$$

We have seen that $\mathcal{E}$ is indeed a quantum channel. It thus remains to show that $\mathcal{E}(\sigma_1) = \sigma_2$ and $\Delta(\mathcal{E}(\rho_1), \rho_2) \leq \varepsilon_1 + \varepsilon_2$. The first property is straightforward to verify.

The second property requires some more work. Define $\bar{\rho}_2 := \mathcal{E}(\rho_1)$. The triangle inequality immediately yields

$$\Delta(\rho_2, \bar{\rho}_2) \leq \varepsilon_2 + \Delta(\tilde{\rho}_2, \bar{\rho}_2), \tag{22}$$

and it thus remains to bound the second term.

Let us first consider the case where $\Delta(\rho, \tau) = T(\rho, \tau) = \frac{1}{2}\|\rho - \tau\|_1$ is the trace distance. Substituting the expression in (19), we find

$$\|\tilde{\rho}_2 - \bar{\rho}_2\|_1 = \frac{1 - \mathrm{tr}\rho_1 Q^*}{1 - \mathrm{tr}\sigma_1 Q^*}\|\tilde{\rho}_2 - \sigma_2\|_1 \leq \varepsilon_1 \frac{\|\tilde{\rho}_2 - \sigma_2\|_1}{1 - 2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)}}, \tag{23}$$

where we used (16) and (20) in the final step. Using (21) we find for Π_+ denoting the projector on to the positive support of $\rho_2 - \sigma_2$

$$\|\tilde{\rho}_2 - \sigma_2\|_1 = 2\,\mathrm{tr}\,\Pi_+(\tilde{\rho}_2 - \sigma_2) \leq 2\Big(1 - 2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)}\Big)\mathrm{tr}\,\Pi_+\tilde{\rho}_2 \tag{24}$$

$$\leq 2\Big(1 - 2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)}\Big). \tag{25}$$

Combining this with (22) finally gives $\Delta\big(\mathcal{E}(\rho_1), \rho_2\big) \leq \varepsilon_1 + \varepsilon_2$, concluding the proof for the trace distance.

Let us now consider the case where $\Delta(\rho, \tau) = P(\rho, \tau) = \sqrt{1 - F(\rho, \tau)}$ is the purified distance. First, we recall that $\mathrm{tr}\rho_1 Q^* \geq \mathrm{tr}\sigma_1 Q^*$ by definition of Q^* as the optimizer for the hypothesis testing divergence. We then use the concavity of fidelity (see, e.g., [51, Lemma 3.4]) and the expression in (19) to bound

$$F(\tilde{\rho}_2, \bar{\rho}_2) \geq \left(1 - \frac{1 - \mathrm{tr}\rho_1 Q^*}{1 - \mathrm{tr}\sigma_1 Q^*}\right) + \left(\frac{1 - \mathrm{tr}\rho_1 Q^*}{1 - \mathrm{tr}\sigma_1 Q^*}\right) F(\tilde{\rho}_2, \sigma_2) \tag{26}$$

$$= 1 - \frac{1 - \mathrm{tr}\rho_1 Q^*}{1 - \mathrm{tr}\sigma_1 Q^*}\left(1 - F(\tilde{\rho}_2, \sigma_2)\right) \tag{27}$$

$$\geq 1 - \varepsilon_1 \frac{1 - F(\tilde{\rho}_2, \sigma_2)}{1 - 2^{-D_{\max}(\tilde{\rho}_2\|\sigma_2)}}. \tag{28}$$

Using the monotonicity of sandwiched Rényi relative entropy we find $D_{\max}(\tilde{\rho}_2\|\sigma_2) \geq -\log F(\tilde{\rho}_2, \sigma_2)$ and $F(\tilde{\rho}_2, \bar{\rho}_2) \geq 1 - \varepsilon_1$, concluding that $\Delta\big(\mathcal{E}(\rho_1), \rho_2\big) \leq \sqrt{\varepsilon_1} + \varepsilon_2$. $\qquad\square$

3.3 Conditions for asymptotic state transformation

In the following we will consider an asymptotic setting given by four sequences of states $\vec{\rho}_1 = \{\rho_1^n\}_n$, $\vec{\sigma}_1 = \{\sigma_1^n\}_n$, $\vec{\rho}_2 = \{\rho_2^n\}_n$, and $\vec{\sigma}_2 = \{\sigma_2^n\}_n$ for $n \in \mathbb{N}$. We assume no specific structure for these states and the underlying Hilbert spaces, i.e. in general we have $\rho_1^n, \sigma_1^n \in \mathrm{S}(\mathbb{C}^{d_1^n})$ and $\rho_2^n, \sigma_2^n \in \mathrm{S}(\mathbb{C}^{d_2^n})$ for arbitrary dimensions $\{d_1^n\}_n$ and $\{d_2^n\}_n$. The only requirement that we impose is that, for $i \in \{1, 2\}$, the limits

$$\tilde{D}_\alpha(\vec{\rho}_i\|\vec{\sigma}_i) := \lim_{n\to\infty} \frac{1}{n}\tilde{D}_\alpha(\rho_i^n\|\sigma_i^n) \tag{29}$$

exist and are continuous in α at $\alpha = 1$. To simplify notation we write $D(\vec{\rho}_i\|\vec{\sigma}_i) := \tilde{D}_1(\vec{\rho}_i\|\vec{\sigma}_i)$.

A simple sequence satisfying this condition is given by independent and identically distributed (iid) states, i.e. the sequences determined by $\rho_i^n = \rho_i^{\otimes n}$ and $\sigma_i^n = \sigma_i^{\otimes n}$ for $i \in \{1, 2\}$. In this case the expressions simplify due to the additivity of the Rényi divergence for product states and we have $\tilde{D}_\alpha(\vec{\rho}_i\|\vec{\sigma}_i) = \tilde{D}_\alpha(\rho_i\|\sigma_i)$ as well as $D(\vec{\rho}_i\|\vec{\sigma}_i) = D(\rho_i\|\sigma_i)$.

A slightly more involved example is constructed by taking tensor products of states, each randomly chosen from some finite sets. The resulting sequences of product states satisfy $\tilde{D}_\alpha(\vec{\rho}_i \| \vec{\sigma}_i) = \mathbb{E}[\tilde{D}_\alpha(\rho_i \| \sigma_i)]$ by the law of large numbers, where the expectation is taken over the joint distribution of ρ_1, σ_1, ρ_2 and σ_2.

In the following we show that sufficient and necessary conditions for asymptotic state transformations are determined by $\lambda_1 := D(\vec{\rho}_1 \| \vec{\sigma}_1)$ and $\lambda_2 := D(\vec{\rho}_2 \| \vec{\sigma}_2)$.

- If $\lambda_1 > \lambda_2$ we show the existence of a sequence of channels that transform (ρ_1^n, σ_1^n) to (ρ_2^n, σ_2^n) where the transformation $\rho_1^n \to \rho_2^n$ has an error that is vanishing exponentially as $n \to \infty$ and the transformation $\sigma_1^n \to \sigma_2^n$ is exact.

- On the other hand, if $\lambda_1 < \lambda_2$ we show that any transformation for which $\sigma_1^n \to \sigma_2^n$ is exact leads to an error exponentially approaching one as $n \to \infty$ in the transformation $\rho_1^n \to \rho_2^n$.

We note that the case where $\lambda_1 = \lambda_2$ is left as an open question. For example, in case of four states $\rho_1, \rho_2, \sigma_1,$ and σ_2 such that $D(\rho_1 \| \sigma_1) = D(\rho_2 \| \sigma_2)$ it is unknown if there exists a sequence of channels that take $\sigma_1^{\otimes n}$ to $\sigma_2^{\otimes n}$ for each n and that take $\rho_1^{\otimes n}$ to $\rho_2^{\otimes n}$ up to asymptotically vanishing error.

Our main technical results are formally presented in the next two theorems.

Theorem 3.4 (Achievability with exponentially small error)**.** *Let $\vec{\rho}_1$, $\vec{\sigma}_1$, $\vec{\rho}_2$ and $\vec{\sigma}_2$ for $n \in \mathbb{N}$ be sequences satisfying the condition in (29) and furthermore*

$$D(\vec{\rho}_1 \| \vec{\sigma}_1) > D(\vec{\rho}_2 \| \vec{\sigma}_2). \tag{30}$$

Then there exists $\gamma > 0$, $n_0 \in \mathbb{N}$, and sequence $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ of (test-and-prepare) quantum channels such that

$$\mathcal{E}_n(\sigma_1^n) = \sigma_2^n \quad \forall n \in \mathbb{N} \qquad \text{and} \qquad \Delta\big(\mathcal{E}_n(\rho_1^n), \rho_2^n\big) \leq 2^{-\gamma n} \quad \forall n \geq n_0. \tag{31}$$

Remark 3.5. The above theorem strengthens [37, Theorem 2.7] in that we are also able to show the exponential decay of the error in the transformation. Furthermore, we extend the result [37, Theorem 2.7] to hold for states that do not necessarily satisfy an iid assumption.

Proof. We show the statement for trace distance, and the statement for purified distance then follows by the Fuchs-van de Graaf inequality.

By the assumption of the theorem and the continuity guaranteed by the assumption at the beginning of the section, there exists a $\delta > 0$ and $\kappa > 0$ such that

$$\tilde{D}_{1-\delta}(\vec{\rho}_1 \| \vec{\sigma}_1) \geq \tilde{D}_{1+\delta}(\vec{\rho}_2 \| \vec{\sigma}_2) + \kappa. \tag{32}$$

Hence, by their definition as limits, there exists a $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$\tilde{D}_{1-\delta}(\rho_1^n \| \sigma_1^n) \geq \tilde{D}_{1+\delta}(\rho_2^n \| \sigma_2^n) + \frac{n\kappa}{2}. \tag{33}$$

Let us now set $\varepsilon_n = \frac{1}{2} 2^{-\gamma n}$ for some $\gamma > 0$ to be determined later. From Lemma 3.3 we learn that the maps $\mathcal{E}_n$ with the desired properties exist if

$$D_h^{\varepsilon_n}(\rho_1^n \| \sigma_1^n) \geq D_{\max}^{\varepsilon_n, T}(\rho_2^n \| \sigma_2^n). \tag{34}$$

Indeed, Proposition 2.2 together with (33) imply

$$D_h^{\varepsilon_n}(\rho_1^n \| \sigma_1^n) - D_{\max}^{\varepsilon_n, T}(\rho_2^n \| \sigma_2^n) \geq \frac{n\kappa}{2} - \frac{1-\delta}{\delta} \log \frac{1}{\varepsilon_n} - \frac{1}{\delta} \log \frac{1}{\varepsilon_n^2} - \log \frac{1}{1 - \varepsilon_n^2} \tag{35}$$

$$\geq \frac{n\kappa}{2} - \frac{3}{\delta} \log \frac{1}{\varepsilon_n} \tag{36}$$

$$\geq \frac{n\kappa}{2} - \frac{4\gamma n}{\delta}, \tag{37}$$

where to reach (36) we used that $1 - \varepsilon_n^2 \geq \varepsilon_n$ since $\varepsilon_n \leq \frac{1}{2}$, and in the last step we assumed $\gamma n \geq 3$. We conclude that the choice $\gamma = \frac{\kappa\delta}{8}$ ensures that (34) holds. $\qquad \square$

Theorem 3.6 (Exponential strong converse). *Let $\vec{\rho}_1$, $\vec{\sigma}_1$, $\vec{\rho}_2$ and $\vec{\sigma}_2$ for $n \in \mathbb{N}$ be sequences satisfying the condition in* (29) *and furthermore*

$$D(\vec{\rho}_1 \| \vec{\sigma}_1) < D(\vec{\rho}_2 \| \vec{\sigma}_2). \tag{38}$$

Then there exists $\gamma > 0$ such that for all sequences of quantum channels $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ that satisfy $\mathcal{E}_n(\sigma_1^n) = \sigma_2^n$ there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have

$$\Delta(\mathcal{E}_n(\rho_1^n), \rho_2^n) \geq 1 - 2^{-\gamma n}. \tag{39}$$

Proof. We show the statement for purified distance, and the statement for trace distance then follows by the Fuchs-van de Graaf inequality.

We again start by observing that the assumption of the theorem and the continuity guaranteed by the assumption at the beginning of the section imply the existence of a $\delta > 0$ and $\kappa > 0$ such that $\tilde{D}_{1+\delta}(\vec{\rho}_1 \| \vec{\sigma}_1) \leq \tilde{D}_{1-\delta}(\vec{\rho}_2 \| \vec{\sigma}_2) - \kappa$, and thus, there exists a $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have

$$\tilde{D}_{1+\delta}(\rho_1^n \| \sigma_1^n) - \tilde{D}_{1-\delta}(\rho_2^n \| \sigma_2^n) \leq -\frac{n\kappa}{2}. \tag{40}$$

It thus suffices to prove that (40) implies the desired property for all sequences of quantum channels. In the following we prove the contrapositive. Suppose that for all $\gamma > 0$, there exists a family of channels $\{\mathcal{E}_n\}_{n \in \mathbb{N}}$ such that for some $n \geq n_0$ we have both $\mathcal{E}_n(\sigma_1^n) = \sigma_2^n$ and $P(\mathcal{E}_n(\rho_1^n), \rho_2^n) < 1 - 2^{-\gamma n}$. Let us then fix a γ, to be determined later, and set $\varepsilon_n = \frac{1}{2} 2^{-\gamma n}$. By Proposition 2.2 we have

$$\tilde{D}_{1+\delta}(\rho_1^n \| \sigma_1^n) \geq D_{\max}^{\varepsilon_n, P}(\rho_1^n \| \sigma_1^n) - \frac{1}{\delta} \log \frac{1}{\varepsilon_n^2} - \log \frac{1}{1 - \varepsilon_n^2} \tag{41}$$

$$\geq D_{\max}^{\varepsilon_n, P}(\mathcal{E}_n(\rho_1^n) \| \sigma_2^n) - \frac{1}{\delta} \log \frac{1}{\varepsilon_n^2} - \log \frac{1}{\varepsilon_n} \tag{42}$$

$$\geq D_{\max}^{1-\varepsilon_n, P}(\rho_2^n \| \sigma_2^n) - \frac{1}{\delta} \log \frac{1}{\varepsilon_n^2} - \log \frac{1}{\varepsilon_n} \tag{43}$$

where the penultimate step uses the data-processing inequality for the smooth max-divergence and the fact that $1 - \varepsilon_n^2 \geq \varepsilon_n$ since $\varepsilon_n \leq \frac{1}{2}$. The final step follows from the definition of the smooth max-relative entropy as an optimization over a ball of close states and the triangle inequality of the purified distance.

Instantiating Proposition 2.1 with $\varepsilon = (1 - \varepsilon_n)^2$ and $\nu = \varepsilon_n - \varepsilon_n^2$ further yields

$$D_{\max}^{1-\varepsilon_n, P}(\rho_2^n \| \sigma_2^n) \geq D_h^{\varepsilon_n}(\rho_2^n \| \sigma_2^n) + \log \frac{1}{2\varepsilon_n - \varepsilon_n^2} - 2 \log \frac{2}{\varepsilon_n - \varepsilon_n^2} \tag{44}$$

$$\geq D_h^{\varepsilon_n}(\rho_2^n \| \sigma_2^n) - 2 \log \frac{4}{\varepsilon_n} \tag{45}$$

$$\geq \tilde{D}_{1-\delta}(\rho_2^n \| \sigma_2^n) - \frac{1-\delta}{\delta} \log \frac{1}{\varepsilon_n} - 2 \log \frac{4}{\varepsilon_n}, \tag{46}$$

where for the second inequality we used that $\varepsilon_n^2 \leq \frac{1}{2}\varepsilon_n$ since $\varepsilon_n \leq \frac{1}{2}$ and the last inequality is a consequence of Proposition 2.2. Combining (43) and (46) we find

$$\tilde{D}_{1+\delta}(\rho_1^n \| \sigma_1^n) - \tilde{D}_{1-\delta}(\rho_2^n \| \sigma_2^n) \geq -\frac{3}{\delta}\log\frac{1}{\varepsilon_n} - 2\log\frac{4}{\varepsilon_n} \tag{47}$$

$$= -n\gamma\left(\frac{3}{\delta} + 2\right) - \frac{3}{\delta} - 6. \tag{48}$$

However, this contradicts (40) for sufficiently large n_0 (and thus n) as long as γ chosen small enough so that $\gamma\left(\frac{3}{\delta} + 2\right) < \frac{\kappa}{2}$. $\qquad\square$

It is worth noting that we made no attempts to characterize the exact error exponents and strong converse exponents here. This is because finding the exact error exponent for this problem is still open even in simple commutative cases where, for example, σ_1 and σ_2 are proportional to the identity and thus commute with ρ_1 and ρ_2.

4   Rates for pairwise state transformations

From an information theoretic perspective, a natural question to ask is at what rate we can transform between pairs of states asymptotically. The main result of this section has been suggested already in the work of Matsumoto [37], and is a consequence of Theorems 3.4 and 3.6 of the previous section. It has been obtained independently in concurrent work [56]. Hence, in this section we ask at what rate we can transform a pair of states (ρ_1, σ_1) into (ρ_2, σ_2) up to some asymptotically vanishing error in the first transformation. To make this precise, let us first say that a triplet (n, m, ε) for $m, n \in \mathbb{N}$ and $\varepsilon \in [0, 1]$ is achievable by a transformation if and only if there exists a channel $\mathcal{E}$ such that

$$\mathcal{E}(\sigma_1^{\otimes n}) = \sigma_2^{\otimes m} \quad \text{and} \quad \Delta(\mathcal{E}(\rho_1^{\otimes n}), \rho_2^{\otimes m}) \leq \varepsilon. \tag{49}$$

We can now define the maximal achievable pairwise state transformation rate with error ε on input block length n as

$$\hat{R}_{\rho_1,\sigma_1 \to \rho_2,\sigma_2}^{\varepsilon,\Delta}(n) := \max\left\{\frac{m}{n} \,:\, (n, m, \varepsilon) \text{ is achievable}\right\}. \tag{50}$$

Our goal is to understand the asymptotics of this quantity for $n \to \infty$ when ε is constant. We can determine the first order asymptotics of $\hat{R}_{\rho_1,\sigma_1 \to \rho_2,\sigma_2}^{\varepsilon,\Delta}(n)$, which turns out to be independent of the metric $\Delta \in \{T, P\}$ and ε.

Theorem 4.1. *Let $\rho_1, \sigma_1 \in \mathrm{S}(\mathbb{C}^{d_1})$ and $\rho_2, \sigma_2 \in \mathrm{S}(\mathbb{C}^{d_2})$ be two pairs of states. For all $\varepsilon \in (0, 1)$, the pairwise state transformation rate satisfies*

$$\lim_{n\to\infty} \hat{R}_{\rho_1,\sigma_1 \to \rho_2,\sigma_2}^{\varepsilon,\Delta}(n) = \frac{D(\rho_1 \| \sigma_1)}{D(\rho_2 \| \sigma_2)}. \tag{51}$$

Remark 4.2. Another viewpoint on this question can be taken by fixing a rate R and asking how the minimal achievable error ε behaves as a function of n. Strong qualitative statements for when we transform above and below the critical rate $D(\rho_1 \| \sigma_1)/D(\rho_2 \| \sigma_2)$ are immediate from Theorems 3.4 and 3.6. Namely, for transformations below the critical rate the error will drop exponentially as $n \to \infty$, and for transformations above the critical rate the error will approach one exponentially fast as $n \to \infty$.

Proof. First consider the case $D(\rho_2\|\sigma_2) = 0$. This implies that $\rho_2 = \sigma_2$ and we can employ a sequence of constant output channels, $\mathcal{E}_n(\cdot) = \sigma_2^{\otimes m}$, for any $m \in \mathbb{N}$, which implies $\hat{R}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}^{\varepsilon,\Delta}(n) = \infty$.

In case $D(\rho_2\|\sigma_2) > 0$ we first show that the rate (51) can be achieved. Let $R = \frac{p}{q} > 0$ with $p, q \in \mathbb{N}$ with $D(\rho_1\|\sigma_1) > RD(\rho_2\|\sigma_2)$. The additivity of the relative entropy for tensor product states yields

$$D(\rho_1^{\otimes q}\|\sigma_1^{\otimes q}) > D(\rho_2^{\otimes p}\|\sigma_2^{\otimes p}). \tag{52}$$

Hence, Theorem 3.4 implies the existence of a sequence of channels $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ such that

$$\mathcal{E}_n(\sigma_1^{\otimes nq}) = \sigma_2^{\otimes Rnq} \qquad \text{and} \qquad \Delta(\mathcal{E}_n(\rho_1^{\otimes nq}), \rho_2^{\otimes Rnq}) \le \varepsilon \quad \text{for large enough } n \quad. \tag{53}$$

We thus showed that (nq, Rnq, ε) is achievable. Moreover, by just throwing away $s \in \{1, 2, \ldots, q-1\}$ systems, we also know that $(nq + s, Rnq, \varepsilon)$ is achievable. Hence,

$$\hat{R}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}^{\varepsilon,\Delta}(nq + s) \ge R \cdot \frac{nq}{nq + s},$$

and, thus, $\liminf_{N\to\infty} \hat{R}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}^{\varepsilon,\Delta}(N) \ge R$. Hence the achievability statement follows because $\mathbb{Q}$ is dense in $\mathbb{R}$.

We next prove a strong converse, i.e., we show that if ε is bounded away from 1 then for large n we must have $D(\rho_1\|\sigma_1) \ge RD(\rho_2\|\sigma_2)$. Let $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ be a sequence of channels satisfying (49). The fully quantum asymptotic equipartition property [52, 51] states that for any $\delta \in (0, 1 - \varepsilon)$ we have

$$D(\rho_1\|\sigma_1) = \lim_{n\to\infty} \frac{1}{n} D_{\max}^{\delta,\Delta}(\rho_1^{\otimes n}\|\sigma_1^{\otimes n}) \tag{54}$$

$$\ge \lim_{n\to\infty} \frac{1}{n} D_{\max}^{\delta,\Delta}(\mathcal{E}_n(\rho_1^{\otimes n})\|\mathcal{E}_n(\sigma_1^{\otimes n})) \tag{55}$$

$$= \lim_{n\to\infty} \frac{1}{n} D_{\max}^{\delta,\Delta}(\mathcal{E}_n(\rho_1^{\otimes n})\|\sigma_2^{\otimes\lceil Rn\rceil}), \tag{56}$$

where the penultimate step follows from the data-processing inequality for the smooth max-relative entropy [22]. Because $\Delta(\mathcal{E}_n(\rho_1^{\otimes n}), \rho_2^{\otimes\lceil Rn\rceil}) \le \varepsilon$, we further find

$$\lim_{n\to\infty} \frac{1}{n} D_{\max}^{\delta,\Delta}(\mathcal{E}_n(\rho_1^{\otimes n})\|\sigma_2^{\otimes\lceil Rn\rceil}) \ge \lim_{n\to\infty} \frac{1}{n} D_{\max}^{\delta+\varepsilon,\Delta}(\rho_2^{\otimes\lceil Rn\rceil}\|\sigma_2^{\otimes\lceil Rn\rceil}) = RD(\rho_2\|\sigma_2), \tag{57}$$

where the final step again uses the asymptotic equipartition property. Combining all of this gives $D(\rho_1\|\sigma_1) \ge RD(\rho_2\|\sigma_2)$, concluding the proof. $\qquad\square$

5 Applications to resource theories

The above results have some immediate consequences in resource theories: in what follows we consider in particular the resource theory of athermality and the resource theory of coherence. In concurrent work, Wang and Wilde [56] also derived some of our results, interpreting them in terms of a resource theory of "asymmetric distinguishability". Our perspective is different insofar as we interpret Theorems 3.4, 3.6 and 4.1 as building blocks that have applications in different resource theories.

Let us first consider the resource theory of athermality under Gibbs-preserving maps. There, we are given a Hamiltonian E, an inverse temperature β and a Gibbs state $\gamma = \frac{1}{Z}e^{-\beta E}$, where Z is the normalization factor. One then asks whether there exists a quantum

channel that has the Gibbs state γ of a quantum system as a fixed point and transforms ρ_1 to ρ_2. Theorem 3.4 reveals that there exists a sequence of Gibbs-preserving maps from $\rho_1^{\otimes n}$ to $\rho_2^{\otimes n}$, with exponentially vanishing error as $n \to \infty$, if

$$D(\rho_1 \| \gamma) > D(\rho_2 \| \gamma) \qquad \text{or, equivalently,} \qquad F_H(\rho_1) > F_H(\rho_2) \,, \tag{58}$$

where we used that the Helmholtz free energy $F_H = U - TH$, where U is the internal energy, T is the temperature and H is the entropy, that is,

$$F_H(\rho) = \operatorname{tr}\rho E + \frac{1}{\beta}\operatorname{tr}\rho \log \rho = \frac{1}{\beta}\left(D(\rho\|\gamma) - \log Z\right) \,. \tag{59}$$

Furthermore, as shown in Theorem 3.6, no such sequence of Gibbs-preserving maps can exist if the inequality in (58) is strictly reversed. In fact, any sequence of Gibbs-preserving maps would incur an error approaching one exponentially fast as $n \to \infty$. In concurrent work by Sagawa *et al.* [?] similar results are discussed for the resource theory of athermality with thermal operations (which is more restrictive than the Gibbs-preserving maps discussed here) and for sequences of states beyond the iid case. Since thermal operations are Gibbs-preserving, our strong converse results also apply to their setting, showing that there is no asymptotic advantage in allowing Gibbs-preserving maps.

Another resource theory in which Theorem 4.1 above plays a role is the resource theory of coherence, in particular, the resource theory of coherence based on dephasing-covariant incoherent operations (DIO) [15, 35]. A DIO operation $\mathcal{E}$ is such that its action commutes with the completely dephasing channel diag, that is

$$\mathcal{E} \circ \mathsf{diag} = \mathsf{diag} \circ \mathcal{E} \,. \tag{60}$$

In this framework, the rate at which coherence can be distilled from an initial resource state ρ is defined, as usual, as the optimal rate at which the transformation

$$\rho^{\otimes n} \xrightarrow{\text{DIO}} |+\rangle\langle+|^{\otimes m} \,, \tag{61}$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ is one unit of coherence, can be achieved with asymptotically vanishing error. Such a rate is known to be equal to $D(\rho\|\mathsf{diag}(\rho))$ [45, 14]. Recently, a relaxation of the DIO paradigm has been proposed and motivated [46]: instead of requiring that the transformation $\mathcal{E}$ and the completely dephasing channel diag commute on all states, the commutation relation is enforced only on the initial resource state ρ. In other words, one considers the ρ-DIO condition

$$(\mathcal{E} \circ \mathsf{diag})(\rho) = (\mathsf{diag} \circ \mathcal{E})(\rho) \,. \tag{62}$$

The existence of a ρ-DIO channel transforming ρ into σ can then be easily reformulated as the existence of a channel $\mathcal{E}$ achieving the following mapping of quantum dichotomies:

$$(\rho, \mathsf{diag}(\rho)) \xrightarrow{\mathcal{E}} (\sigma, \mathsf{diag}(\sigma)) \,. \tag{63}$$

Once formulated in this form, our Theorem 4.1 implies that the rate at which coherence can be distilled from ρ by means of ρ-DIO operations is given by $D(\rho\|\mathsf{diag}(\rho))$. The asymptotic distillation rate under ρ-DIO has been independently computed in Ref. [46]. Beyond that, Theorem 3.6 establishes an exponential strong converse which implies that if we try to distill at a rate exceeding $D(\rho\|\mathsf{diag}(\rho))$ then the error will go to one exponentially fast. Interestingly, since the asymptotic distillation rate is the same for both DIO and ρ-DIO operations, and since ρ-DIO operations constitute a larger set than DIO operations, we have that the above mentioned exponential strong converse property holds for DIO state distillation too. Beyond that, our results also yield an exponential strong converse for general DIO transformations between arbitrary states ρ and σ for any rate exceeding the ratio of $D(\rho\|\mathsf{diag}(\rho))$ and $D(\sigma\|\mathsf{diag}(\sigma))$.

6 Discussion

Given the result of Theorem 4.1, we are left to wonder how quickly $\hat{R}^{\varepsilon,\Delta}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}(n)$ approaches the asymptotic limit, or, in other words: what are the higher order terms in the expansion of $\hat{R}^{\varepsilon,\Delta}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}(n)$ for large n? Such questions have recently attracted a lot of interest in both classical and quantum information theory. In the quantum setting, the first tight results in this direction were achieved for hypothesis testing [33, 53]:

$$D^\varepsilon_h(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD(\rho\|\sigma) + \sqrt{nV(\rho\|\sigma)}\Phi^{-1}(\varepsilon) + O(\log n), \tag{64}$$

where $V(\rho\|\sigma) := \operatorname{tr}\rho\left(\log\rho - \log\sigma\right)^2 - D(\rho\|\sigma)^2$ is the *relative entropy variance*, i.e. a quantum generalization of the variance of the log-likelihood ratio, and $\Phi(\cdot)$ is the cumulative normal distribution function. This is the constant error regime of quantum hypothesis testing. Similar second-order expansions can be derived for the case where ε is not constant but approaches 0 and 1 slower than exponentially, the so-called moderate deviation regime [17, 13].

Our goal is to establish similar asymptotic expansions for $\hat{R}^{\varepsilon,\Delta}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}(n)$. While this is ultimately beyond the scope of this work, we will justify the following conjecture.

Conjecture 6.1. *Let $\rho_1,\sigma_1 \in \mathrm{S}(\mathbb{C}^{d_1})$ and $\rho_2,\sigma_2 \in \mathrm{S}(\mathbb{C}^{d_2})$ such that*

$$\nu := \frac{D(\rho_1\|\sigma_1)}{V(\rho_1\|\sigma_1)} \cdot \frac{V(\rho_2\|\sigma_2)}{D(\rho_2\|\sigma_2)} \, . \tag{65}$$

is finite, i.e. $V(\rho_1\|\sigma_1) > 0$ and $D(\rho_2\|\sigma_2) > 0$. Then, for any $\varepsilon \in (0,1)$, we have

$$\lim_{n\to\infty}\sqrt{n}\left(\hat{R}^{\varepsilon,P}_{\rho_1,\sigma_1\to\rho_2,\sigma_2}(n) - \frac{D(\rho_1\|\sigma_1)}{D(\rho_2\|\sigma_2)}\right) = \frac{\sqrt{V(\rho_1\|\sigma_1)}}{D(\rho_2\|\sigma_2)}Z^{-1}_\nu(\varepsilon)\,, \tag{66}$$

where $Z_\nu(\cdot)$ is the cumulative Rayleigh-normal distribution function [31].

This form is of interest since it shows a resonance behavior when $\nu = 1$, where the contribution in the second-order term turns positive even for arbitrarily small ε. This means that there exist pairs of states that can be transformed into each other without loss due to finite size effects (up to second order). This effect has been observed both analytically and numerically in the commutative case [30], and its applications to fully quantum resource theories remain to be explored.

The limit expression in (66) was shown to hold for the case where σ_1 and σ_2 are both proportional to the identity in the work of Kumagai and Hayashi [31], and the above conjecture thus constitutes a natural fully quantum generalization of their result. Building on that and an embedding technique from quantum thermodynamics, the equality was also shown for general σ_1 and σ_2 as long as ρ_1 and ρ_2 commute with σ_1 and σ_2, respectively [18].[4] The same special case can also be solved in the moderate error regime by adapting the results in [19].

Finally, note that since we are now concerned with higher order contribution that are a function of the error threshold ε, the limit does in general depend on how exactly we measure the error. It is thus not obvious how an appropriate conjecture for the trace distance would look like, for example.

---

[4]To be more precise, the cited work [18] only shows this for the case where $\sigma_1 = \sigma_2$, but the required extension follows after a close inspection of the proof.

References

[1] P. Alberti and A. Uhlmann. *"A problem relating to positive linear maps on matrix algebras"*. Reports on Mathematical Physics **18**(2): 163–176 (1980).

[2] P. M. Alberti and A. Uhlmann. *Stochasticity and partial order.* volume 9 of *Mathematics and Its Applications*, Deutscher Verlag der Wissenschaften (1982).

[3] A. Anshu, M. Berta, R. Jain, and M. Tomamichel. *"A minimax approach to one-shot entropy inequalities"*. Preprint, arXiv: 1906.00333 (2019).

[4] K. M. R. Audenaert, M. Mosonyi, and F. Verstraete. *"Quantum State Discrimination Bounds for Finite Sample Size"*. Journal of Mathematical Physics **53**(12): 122205 (2012).

[5] D. Blackwell. *"Equivalent Comparisons of Experiments"*. The Annals of Mathematical Statistics **24**(2): 265–272 (1953).

[6] F. G. S. L. Brandao, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner. *"The Second Laws of Quantum Thermodynamics"*. Proceedings of the National Academy of Sciences USA **112**(11): 3275–3279 (2014).

[7] F. Buscemi. *"Fully quantum second-law–like statements from the theory of statistical comparisons"*. Preprint, arXiv: 1505.00535v1 .

[8] F. Buscemi. *"Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency"*. Communications in Mathematical Physics **310**(3): 625–647 (2012).

[9] F. Buscemi and N. Datta. *"The Quantum Capacity of Channels With Arbitrarily Correlated Noise"*. IEEE Transactions on Information Theory **56**(3): 1447–1460 (2010).

[10] F. Buscemi and G. Gour. *"Quantum relative Lorenz curves"*. Physical Review A **95**(1) (2017).

[11] A. Chefles. *"Deterministic Quantum State Transformations"*. Physics Letters A **270**(1-2): 14–19 (2000).

[12] A. Chefles, R. Jozsa, and A. Winter. *"On the Existence of Physical Transformations between Sets of Quantum States"*. International Journal of Quantum Information **02**(01): 11–21 (2004).

[13] H.-C. Cheng and M.-H. Hsieh. *"Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing"*. IEEE Transactions on Information Theory **64**(2): 1385–1403 (2018).

[14] E. Chitambar. *"Dephasing-covariant operations enable asymptotic reversibility of quantum resources"*. Physical Review A **97**(5) (2018).

[15] E. Chitambar and G. Gour. *"Critical Examination of Incoherent Operations and a Physically Consistent Resource Theory of Quantum Coherence"*. Physical Review Letters **117**(3) (2016).

[16] E. Chitambar and G. Gour. *"Quantum resource theories"*. Reviews of Modern Physics **91**(2) (2019).

[17] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel. *"Moderate Deviation Analysis for Classical Communication over Quantum Channels"*. Communications in Mathematical Physics **355**(3): 1283–1315 (2017).

[18] C. T. Chubb, M. Tomamichel, and K. Korzekwa. *"Beyond the thermodynamic limit: finite-size corrections to state interconversion rates"*. Quantum **2**: 108 (2018).

[19] C. T. Chubb, M. Tomamichel, and K. Korzekwa. *"Moderate deviation analysis of majorization-based resource interconversion"*. Physical Review A **99**(3): 032332 (2019).

[20] J. E. Cohen, J. H. B. Kempermann, and G. Zbaganu. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*. Birkhäuser (1998).

[21] G. Dahl. *"Matrix majorization"*. Linear Algebra and its Applications **288**: 53 – 73 (1999).

[22] N. Datta. *"Min- and Max- Relative Entropies and a New Entanglement Monotone"*. IEEE Transactions on Information Theory **55**(6): 2816–2826 (2009).

[23] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian. *"Quantum majorization and a complete set of entropic conditions for quantum thermodynamics"*. Nature Communications **9**(1) (2018).

[24] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press (1934).

[25] F. Hiai and D. Petz. *"The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability"*. Communications in Mathematical Physics **143**(1): 99–114 (1991).

[26] M. Horodecki and J. Oppenheim. *"Fundamental limitations for quantum and nanoscale thermodynamics"*. Nature Communications **4**(1) (2013).

[27] A. Jenčová. *"Comparison of Quantum Binary Experiments"*. Reports on Mathematical Physics **70**(2): 237–249 (2012).

[28] A. Jenčová. *"Comparison of quantum channels and statistical experiments"*. In *Proc. IEEE ISIT 2016*, pages 2249–2253, (2016).

[29] J. Körner and K. Marton. *"Comparison of two noisy channels"*. Colloquia Mathematica Societatis Janos Bolyai, Topics in Information Theory **16**: 411–424, (1977).

[30] K. Korzekwa, C. T. Chubb, and M. Tomamichel. *"Avoiding Irreversibility: Engineering Resonant Conversions of Quantum Resources"*. Physical Review Letters **122**(11): 110403 (2019).

[31] W. Kumagai and M. Hayashi. *"Second Order Asymptotics of Optimal Approximate Conversion for Probability Distributions and Entangled States and Its Application to LOCC Cloning"*. Preprint, arXiv: 1306.4166 (2013).

[32] L. Le Cam. *"Sufficiency and Approximate Sufficiency"*. The Annals of Mathematical Statistics **35**(4): 1419–1455 (1964).

[33] K. Li. *"Second-Order Asymptotics for Quantum Hypothesis Testing"*. Annals of Statistics **42**(1): 171–189 (2014).

[34] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer (2011).

Accepted in ⟨ quantum 2019-11-22, click title to verify. Published under CC-BY 4.0.

16

[35] I. Marvian and R. W. Spekkens. *"How to quantify coherence: Distinguishing speakable and unspeakable notions"*. Physical Review A **94**(5) (2016).

[36] K. Matsumoto. *"A quantum version of randomization criterion"*. Preprint, arXiv: 1012.2650 (2010).

[37] K. Matsumoto. *"Reverse Test and Characterization of Quantum Relative Entropy"*. Preprint, arXiv: 1010.1030 (2010).

[38] K. Matsumoto. *"An example of a quantum statistical model which cannot be mapped to a less informative one by any trace preserving positive map"*. Preprint, arXiv: 1409.5658 (2014).

[39] K. Matsumoto. *"On the condition of conversion of classical probability distribution families into quantum families"*. Preprint, arXiv: 1412.3680 (2014).

[40] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. *"On Quantum Rényi Entropies: A New Generalization and Some Properties"*. Journal of Mathematical Physics **54**(12): 122203 (2013).

[41] M. A. Nielsen. *"Conditions for a Class of Entanglement Transformations"*. Physical Review Letters **83**(2): 436–439 (1999).

[42] T. Ogawa and H. Nagaoka. *"Strong Converse and Stein's Lemma in Quantum Hypothesis Testing"*. IEEE Transactions on Information Theory **46**(7): 2428–2433 (2000).

[43] D. Petz. *"Quasi-entropies for Finite Quantum Systems"*. Reports on Mathematical Physics **23**(1): 57–65 (1986).

[44] D. Reeb, M. J. Kastoryano, and M. M. Wolf. *"Hilbert's projective metric in quantum information theory"*. Journal of Mathematical Physics **52**(8): 082201 (2011).

[45] B. Regula, K. Fang, X. Wang, and G. Adesso. *"One-Shot Coherence Distillation"*. Physical Review Letters **121**(1) (2018).

[46] B. Regula, V. Narasimhachar, F. Buscemi, and M. Gu. *"Coherence manipulation with dephasing-covariant operations"*. Preprint, arXiv: 1907.08606 (2019).

[47] J. M. Renes. *"Relative submajorization and its use in quantum resource theories"*. Journal of Mathematical Physics **57**(12): 122202 (2016).

[48] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, (2005). Available at arXiv: quant-ph/0512258.

[49] C. E. Shannon. *"A note on a partial ordering for communication channels"*. Information and control **1**(4): 390–397 (1958).

[50] V. Siddhu and R. B. Griffiths. *"Degradable Quantum Channels using Pure-State to Product-of-Pure-State Isometries"*. Physical Review A **94**(5): 052331 (2016).

[51] M. Tomamichel. *Quantum Information Processing with Finite Resources — Mathematical Foundations*. volume 5 of *SpringerBriefs in Mathematical Physics*, Springer International Publishing (2016).

[52] M. Tomamichel, R. Colbeck, and R. Renner. *"A Fully Quantum Asymptotic Equipartition Property"*. IEEE Transactions on Information Theory **55**(12): 5840–5847 (2009).

[53] M. Tomamichel and M. Hayashi. *"A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks"*. IEEE Transactions on Information Theory **59**(11): 7693–7710 (2013).

[54] E. Torgersen. *Comparison of statistical experiments.* volume 36 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press (1991).

[55] E. N. Torgersen. *"Comparison of experiments when the parameter space is finite".* Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **16**(3): 219–249 (1970).

[56] X. Wang and M. M. Wilde. *"Resource theory of asymmetric distinguishability".* Preprint, arXiv: 1905.11629 (2019).

[57] M. M. Wilde, A. Winter, and D. Yang. *"Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy".* Communications in Mathematical Physics **331**(2): 593–622 (2014).