Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions

Milán Mosonyi[1,2,*] and Tomohiro Ogawa[3,†]

[1] *MTA-BME Lendület Quantum Information Theory Research Group*
[2] *Mathematical Institute, Budapest University of Technology and Economics, Egry József u 1., Budapest, 1111 Hungary.*
[3] *Graduate School of Informatics and Engineering, University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan.*

Abstract

There are different inequivalent ways to define the Rényi capacity of a channel for a fixed input distribution P. In a 1995 paper [16] Csiszár has shown that for classical discrete memoryless channels there is a distinguished such quantity that has an operational interpretation as a generalized cutoff rate for constant composition channel coding. We show that the analogous notion of Rényi capacity, defined in terms of the sandwiched quantum Rényi divergences, has the same operational interpretation in the strong converse problem of classical-quantum channel coding. Denoting the constant composition strong converse exponent for a memoryless classical-quantum channel W with composition P and rate R as $sc(W,R,P)$, our main result is that

$$sc(W,R,P)=\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^*(W,P)\right], \tag{a.1}$$

where $\chi_\alpha^*(W,P)$ is the P-weighted sandwiched Rényi divergence radius of the image of the channel.

I. INTRODUCTION

A classical-quantum channel $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ models a device with a set of possible inputs $\mathcal{X}$, which, on an input $x\in\mathcal{X}$, outputs a quantum system with finite-dimensional Hilbert space $\mathcal{H}$ in state $W(x)$. For every such channel $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$, we define the lifted channel

$$\mathbb{W}:\mathcal{X}\to\mathcal{S}(\mathcal{H}_\mathcal{X}\otimes\mathcal{H}),\qquad \mathbb{W}(x):=|x\rangle\langle x|\otimes W(x).$$

Here, $\mathcal{H}_\mathcal{X}$ is an auxiliary Hilbert space, and $\{|x\rangle : x\in\mathcal{X}\}$ is an orthonormal basis in it. As a canonical choice, one can use $\mathcal{H}_\mathcal{X}=l^2(\mathcal{X})$, the L^2-space on $\mathcal{X}$ with respect to the counting measure, and choose $|x\rangle:=\mathbf{1}_{\{x\}}$ to be the characteristic function (indicator function) of the singleton $\{x\}$. Note that this is well-defined irrespectively of the cardinality of $\mathcal{X}$. The classical-quantum state $\mathbb{W}(P):=\sum_{x\in\mathcal{X}}P(x)\,|x\rangle\langle x|\otimes W(x)$ plays the role of the joint distribution of the input and the output of the channel for a fixed finitely supported input probability distribution $P\in\mathcal{P}_f(\mathcal{X})$.

Given a quantum divergence Δ, i.e., some sort of generalized distance of quantum states, there are various natural-looking but inequivalent ways to define the corresponding capacity of the channel for a fixed input probability distribution P. One possibility is a mutual information-type quantity

$$I_\Delta(W,P):=\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\Delta\left(\mathbb{W}(P)\|P\otimes\sigma\right), \tag{I.2}$$

where one measures the Δ-distance of the joint input-output state of the channel from the set of uncorrelated states, while the first marginal is kept fixed. This can be interpreted as a measure of the maximal amount of correlation that can be created between the input and the output of the channel with a fixed input distribution. The idea is that the more correlated the input and

---

* milan.mosonyi@gmail.com
† ogawa@is.uec.ac.jp

the output can be made, the more useful the channel is for information transmission. Another option is to use the P-weighted Δ-radius of the image of the channel, defined as

$$\chi_\Delta(W,P) := \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x)\,\Delta(W(x)\|\sigma). \tag{I.3}$$

This approach is geometrically motivated, and the idea behind is that the further away some states are in Δ-distance (weighted by the input distribution P), the more distinguishable they are, and the information transmission capacity of the channel is related to the number of far away states among the output states of the channel. In the case where W is a classical channel, and Δ is a Rényi divergence, these quantities were studied by Sibson [55] and Augustin [8], respectively; see [16], and the recent works [15, 52, 53] for more references on the history and applications of these quantities.

It was shown in [16] that for classical channels, both quantities yield the same channel capacity when $\Delta = D_\alpha$ is a Rényi α-divergence, i.e.,

$$C_\Delta(W) := \sup_{P\in\mathcal{P}_f(\mathcal{X})} \chi_\Delta(W,P) = \sup_{P\in\mathcal{P}_f(\mathcal{X})} I_\Delta(W,P) = \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sup_{x\in\mathcal{X}} \Delta(W(x)\|\sigma), \tag{I.4}$$

where the last quantity is the Δ-radius of the image of the channel. This was extended to the case of classical-quantum channels and a variety of quantum Rényi divergences in a series of work [36, 43, 44, 60]. Moreover, it was shown in [44] (extending the corresponding classical result of [16, 17, 21]) that the strong converse exponent $\mathrm{sc}(W,R)$ of a classical-quantum channel W for coding rate R can be expressed as

$$\mathrm{sc}(W,R) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[R - \chi_\alpha^*(W)\right],$$

where $\chi_\alpha^*(W) = \chi_{D_\alpha^*}(W)$ is the Rényi capacity of W corresponding to the sandwiched Rényi divergences D_α^* [45, 60], thus giving an operational interpretation to the sandwiched Rényi capacities with parameter $\alpha > 1$.

After this, it is natural to ask which of the two quantities presented in (I.2) and (I.3) has an operational interpretation, and for what divergence. Note that the standard channel coding problem does not yield an answer to this question, essentially due to (I.4), so to settle this problem, one needs to consider a refinement of the channel coding problem where the input distribution P appears on the operational side. This can be achieved by considering constant composition coding, where the codewords are required to have the same empirical distribution for each message, and these empirical distributions are required to converge to a fixed distribution P on $\mathcal{X}$ as the number of channel uses goes to infinity. It was shown in [16] that in this setting

$$\mathrm{sc}(W,R,P) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[R - \chi_\alpha(W,P)\right], \tag{I.5}$$

for any classical channel W and input distribution P, where $\mathrm{sc}(W,R,P)$ is the strong converse exponent for coding rate R, and $\chi_\alpha(W,P) = \chi_{D_\alpha}(W,P)$. This shows that, maybe somewhat surprisingly, it is not the perhaps more intuitive-looking concept of mutual information (I.2) but the geometric quantity (I.3) that correctly captures the information transmission capacity of a classical channel.

Our main result is an exact analogue of (I.5) for classical-quantum channels, as given in (a.1). Thus we establish that the operationally relevant notion of Rényi capacity with fixed input distribution P for a classical-quantum channel in the strong converse domain is the sandwiched Rényi divergence radius of the channel.

The structure of the paper is as follows. After collecting some technical preliminaries in Section II, we study the concepts of divergence radius and center for general divergences in Section III A and for Rényi divergences in Section III B. One of our main results is the additivity of the weighted Rényi divergence radius for classical-quantum channels, given in Section III D. We prove it using a representation of the minimizing state in (I.3) when $\Delta = D_{\alpha,z}$ is a quantum α-z Rényi divergence [6], as the fixed point of a certain map on the state space. Analogous results have been derived very recently by Nakiboğlu in [53] for classical channels, and by Cheng, Li and Hsieh in [15] for classical-quantum channels and the Petz-type Rényi divergences. Our results extend these with a different proof method, which in turn is closely related to the approach of Hayashi and Tomamichel for proving the additivity of the sandwiched Rényi mutual information [29].

In Section IV, we prove our main result, (a.1). The non-trivial part of this is the inequality LHS≤RHS, which we prove using a refinement of the arguments in [44]. First, in Proposition IV.4 we employ a suitable adaptation of the techniques of Dueck and Körner [21] and obtain the inequality in terms of the log-Euclidean Rényi divergence, which gives a suboptimal bound. Then in Proposition IV.5 we use the asymptotic pinching technique developed in [44] to arrive at an upper bound in terms of the regularized sandwiched Rényi divergence radii, and finally we use the previously established additivity property of these quantities to arrive at the desired bound. In the proof of Proposition IV.4 we need a constant composition version of the classical-quantum channel coding theorem. Such a result was established, for instance, by Hayashi in [27], and very recently by Cheng, Hanson, Datta and Hsieh in [14], with a different exponent, by refining another random coding argument by Hayashi [26]. We give a slightly modified proof in Appendix C. Further appendices contain various technical ingredients of the proofs, and in Appendix A we give a more detailed discussion of the concepts of divergence radius and mutual information for general divergences and α-z Rényi divergences, which may be of independent interest.

II. PRELIMINARIES

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the set of all linear operators on $\mathcal{H}$, and let $\mathcal{B}(\mathcal{H})_{\rm sa}$, $\mathcal{B}(\mathcal{H})_{+}$, and $\mathcal{B}(\mathcal{H})_{++}$ denote the set of self-adjoint, non-zero positive semi-definite (PSD), and positive definite operators, respectively. For an interval $J \subseteq \mathbb{R}$, let $\mathcal{B}(\mathcal{H})_{{\rm sa},J} := \{A \in \mathcal{B}(\mathcal{H})_{\rm sa} : \operatorname{spec}(A) \subseteq J\}$, i.e., the set of self-adjoint operators on $\mathcal{H}$ with all their eigenvalues in J. Let $\mathcal{S}(\mathcal{H}) := \{\varrho \in \mathcal{B}(\mathcal{H})_{+},\ \operatorname{Tr}\varrho = 1\}$ denote the set of *density operators*, or *states*, on $\mathcal{H}$.

For a self-adjoint operator A, let $P_a^A := \mathbf{1}_{\{a\}}(A)$ denote the spectral projection of A corresponding to the singleton $\{a\}$. The projection onto the support of A is $\sum_{a\neq 0} P_a^A$; in particular, if A is positive semi-definite, it is equal to $\lim_{\alpha\searrow 0} A^\alpha =: A^0$. In general, we follow the convention that real powers of a positive semi-definite operator A are taken only on its support, i.e., for any $x\in\mathbb{R}$, $A^x := \sum_{a>0} a^x P_a^A$.

Given a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})_{\rm sa}$, the *pinching* by A is the operator $\mathcal{F}_A : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, $\mathcal{F}_A(.) := \sum_a P_a^A (.) P_a^A$, i.e., the block-diagonalization with the eigen-projectors of A. By the pinching inequality [25],

$$X \le |\operatorname{spec}(A)| \mathcal{F}_A(X), \qquad X \in \mathcal{B}(\mathcal{H})_{+}.$$

Since $\mathcal{F}_A$ can be written as a convex combination of unitary conjugations,

$$f\left(\mathcal{F}_A(B)\right) \le \mathcal{F}_A(f(B)), \qquad \operatorname{Tr} g\left(\mathcal{F}_A(B)\right) \le \operatorname{Tr} g(B), \tag{II.6}$$

for any operator convex function f, and any convex function g on an interval J, and any $B \in \mathcal{B}(\mathcal{H})_{{\rm sa},J}$. The second inequality above is due to the following well-known fact:

Lemma II.1 *Let $J \subseteq \mathbb{R}$ be an interval and $f : J \to \mathbb{R}$ be a function.*

(i) If f is monotone increasing then $\operatorname{Tr} f(.)$ is monotone increasing on $\mathcal{B}(\mathcal{H})_{{\rm sa},J}$.

(ii) If f is convex then $\operatorname{Tr} f(.)$ is convex on $\mathcal{B}(\mathcal{H})_{{\rm sa},J}$.

For a differentiable function f defined on an interval $J \subseteq \mathbb{R}$, let $f^{[1]} : J \times J \to \mathbb{R}$ be its *first divided difference function*, defined as

$$f^{[1]}(a,b) := \begin{cases} \frac{f(a)-f(b)}{a-b}, & a \neq b, \\ f'(a), & a = b, \end{cases} \qquad a, b \in J.$$

The proof of the following can be found, e.g., in [10, Theorem V.3.3] or [30, Theorem 2.3.1]:

Lemma II.2 *If f is a continuously differentiable function on an open interval $J \subseteq \mathbb{R}$ then for any finite-dimensional Hilbert space $\mathcal{H}$, $A \mapsto f(A)$ is Fréchet differentiable on $\mathcal{B}(\mathcal{H})_{\mathrm{sa},J}$, and its Fréchet derivative $Df(A)$ at a point A is given by*

$$Df(A)(Y) = \sum_{a,b} f^{[1]}(a,b) P_a^A Y P_b^A, \qquad Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}.$$

It is straightforward to verify that in the setting of Lemma II.2, the function $A \mapsto \operatorname{Tr} f(A)$ is also Fréchet differentiable on $\mathcal{B}(\mathcal{H})_{\mathrm{sa},J}$, and its Fréchet derivative $D(\operatorname{Tr} \circ f)(A)$ at a point A is given by

$$D(\operatorname{Tr} \circ f)(A)(Y) = \operatorname{Tr} f'(A) Y, \qquad Y \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}, \tag{II.7}$$

where f' is the derivative of f as a real-valued function.

An operator $A \in \mathcal{B}(\mathcal{H}^{\otimes n})$ is *symmetric*, if $U_\pi A U_\pi^* = A$ for all permutations $\pi \in S_n$, where U_π is defined by $U_\pi x_1 \otimes \ldots \otimes x_n = x_{\pi^{-1}(1)} \otimes \ldots \otimes x_{\pi^{-1}(n)}$, $x_i \in \mathcal{H}$, $i \in [n]$. As it was shown in [27], for every finite-dimensional Hilbert space $\mathcal{H}$ and every $n \in \mathbb{N}$, there exists a *universal symmetric state* $\sigma_{u,n} \in \mathcal{S}(\mathcal{H}^{\otimes n})$ such that it is symmetric, it commutes with every symmetric state, and for every symmetric state $\omega \in \mathcal{S}(\mathcal{H}^{\otimes n})$,

$$\omega \leq v_{n,d} \sigma_{u,n},$$

where $v_{n,d}$ only depends on $d = \dim \mathcal{H}$ and n, and it is polynomial in n.

By a *generalized classical-quantum (gcq) channel* we mean a map $W : \mathcal{X} \to \mathcal{B}(\mathcal{H})_+$, where $\mathcal{X}$ is a non-empty set, and $\mathcal{H}$ is a finite-dimensional Hilbert space. It is a *classical-quantum (cq) channel* if $\operatorname{ran} W \subseteq \mathcal{S}(\mathcal{H})$, i.e., each output of the channel is a normalized quantum state. A (generalized) classical-quantum channel is *classical*, if $W(x)W(y) = W(y)W(x)$ for all $x, y \in \mathcal{X}$. We remark that we do not require any further structure of $\mathcal{X}$ or the map W, and in particular, $\mathcal{X}$ need not be finite. Given a finite number of gcq channels $W_i : \mathcal{X}_i \to \mathcal{B}(\mathcal{H}_i)_+$, their *product* is the gcq channel

$$\begin{aligned} W_1 \otimes \ldots \otimes W_n : & \mathcal{X}_1 \times \ldots \times \mathcal{X}_n \to \mathcal{B}(\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n)_+ \\ & (x_1, \ldots, x_n) \mapsto W_1(x_1) \otimes \ldots \otimes W_n(x_n). \end{aligned}$$

In particular, if all W_i are the same channel W then we use the notaion $W^{\otimes n} = W \otimes \ldots \otimes W$.

We say that a function $P : \mathcal{X} \to [0,1]$ is a *probability density function* on a set $\mathcal{X}$ if $\sum_{x \in \mathcal{X}} P(x) = 1$. The *support* of P is $\operatorname{supp} P := \{x \in \mathcal{X} : P(x) > 0\}$. We say that P is finitely supported if supp is a finite set, and we denote by $\mathcal{P}_f(\mathcal{X})$ the set of all finitely supported probability distributions. The *Shannon entropy* of a $P \in \mathcal{P}_f(\mathcal{X})$ is defined as

$$H(P) := -\sum_{x \in \mathcal{X}} P(x) \log P(x).$$

For a sequence $\underline{x} \in \mathcal{X}^n$, the *type* $P_{\underline{x}} \in \mathcal{P}_f(\mathcal{X})$ of $\underline{x}$ is the empirical distribution of $\underline{x}$, defined as

$$P_{\underline{x}} := \frac{1}{n} \sum_{i=1}^n \delta_{x_i} : \quad y \mapsto \frac{1}{n} \left| \{k : x_k = y\} \right|, \qquad y \in \mathcal{X},$$

where δ_x is the Dirac measure concentrated at x. We say that a probability distribution P on $\mathcal{X}$ is an *n-type* if there exists an $\underline{x} \in \mathcal{X}^n$ such that $P = P_{\underline{x}}$. We denote the set of n-types by $\mathcal{P}_n(\mathcal{X})$. For an n-type P, let $\mathcal{X}_P^n := \{\underline{x} \in \mathcal{X}^n : P_{\underline{x}} = P\}$ be the set of sequences with the same type P. A key property of types is that $\underline{x}, \underline{y} \in \mathcal{X}^n$ have the same type if and only if they are permutations of each other, and for any $\underline{x}, \underline{y}$ with $P_{\underline{x}} = P_{\underline{y}}$, we have

$$P_{\underline{x}}^{\otimes n}(\underline{y}) = e^{-nH(P_{\underline{x}})}. \tag{II.8}$$

By Lemma 2.3 in [18], for any $P \in \mathcal{P}_n(\mathcal{X})$,

$$(n+1)^{-|\operatorname{supp} P|} e^{nH(P)} \leq |\mathcal{X}_P^n| \leq e^{nH(P)}. \tag{II.9}$$

The following lemma is an extension of the minimax theorems due to Kneser [35] and Fan [22] to the case where f can take the value $+\infty$. For a proof, see [23, Theorem 5.2].

Lemma II.3 *Let X be a compact convex set in a topological vector space V and Y be a convex subset of a vector space W. Let $f : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be such that*

(i) $f(x,.)$ is concave on Y for each $x \in X$, and

(ii) $f(.,y)$ is convex and lower semi-continuous on X for each $y \in Y$.

Then

$$\inf_{x\in X} \sup_{y\in Y} f(x,y) = \sup_{y\in Y} \inf_{x\in X} f(x,y), \tag{II.10}$$

and the infima in (II.10) *can be replaced by minima.*

III. DIVERGENCE RADII

A. General divergences

By a *divergence* Δ we mean a family of maps $\Delta_{\mathcal{H}} : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to [-\infty, +\infty]$, defined for every finite-dimensional Hilbert space $\mathcal{H}$. We will normally not indicate the dependence on the Hilbert space, and simply use the notation Δ instead of $\Delta_{\mathcal{H}}$. We will only consider divergences that are invariant under isometries, i.e., for any $\varrho, \sigma \in \mathcal{B}(\mathcal{H})_+$ and $V : \mathcal{H} \to \mathcal{K}$ isometry, $\Delta(V\varrho V^* \| V\sigma V^*) = \Delta(\varrho\|\sigma)$. Note that this implies that Δ is invariant under extensions with pure states, i.e., $\Delta(\varrho \otimes |\psi\rangle\langle\psi| \,\|\sigma \otimes |\psi\rangle\langle\psi|) = \Delta(\varrho\|\sigma)$, where ψ is an arbitrary unit vector in some Hilbert space. Further properties will often be important. In particular, we say that a divergence Δ is

- *positive* if $\Delta(\varrho\|\sigma) \geq 0$ for all density operators ϱ, σ, and it is *strictly positive* if $\Delta(\varrho\|\sigma) = 0 \iff \varrho = \sigma$, again for density operators;

- *monotone under CPTP maps* if for any $\varrho, \sigma \in \mathcal{B}(\mathcal{H})_+$ and any CPTP (completely positive and trace-preserving) map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$,

$$\Delta(\Phi(\varrho)\|\Phi(\sigma)) \leq \Delta(\varrho\|\sigma);$$

- *jointly convex* if for all $\varrho_i, \sigma_i \in \mathcal{B}(\mathcal{H})$, $i \in [r]$, and probability distribution $(p_i)_{i=1}^r$,

$$\Delta\left(\sum_{i=1}^r p_i\varrho_i \middle\| \sum_{i=1}^r p_i\sigma_i\right) \leq \sum_{i=1}^r p_i\,\Delta(\varrho_i\|\sigma_i);$$

- *block additive* if for any $\varrho_1, \varrho_2, \sigma_1, \sigma_2$ such that $\varrho_1^0 \vee \sigma_1^0 \perp \varrho_2^0 \vee \sigma_2^0$, we have

$$\Delta(\varrho_1 + \varrho_2\|\sigma_1 + \sigma_2) = \Delta(\varrho_1\|\sigma_1) + \Delta(\varrho_2\|\sigma_2);$$

- *homogeneous* if

$$\Delta(\lambda\varrho\|\lambda\sigma) = \lambda\,\Delta(\varrho\|\sigma), \qquad \varrho, \sigma \in \mathcal{B}(\mathcal{H})_+, \quad \lambda \in (0, +\infty).$$

Typical examples for divergences with some or all of the above properties are the relative entropy and some Rényi divergences and related quantities; see Section III B.

Remark III.1 *It is well-known [49, 58] that a block additive and homogenous divergence is monotone under CPTP maps if and only if it is jointly convex. The "only if" direction follows by applying monotonicity to $\widehat{\varrho} := \sum_i p_i |i\rangle\langle i|_E \otimes \varrho_i$ and $\widehat{\sigma} := \sum_i p_i |i\rangle\langle i|_E \otimes \sigma_i$ under the partial trace over the E system, where $(|i\rangle)_{i=1}^r$ is an ONS in $\mathcal{H}_E$. The "if" direction follows by using a Stinespring dilation $\Phi(.) = \operatorname{Tr}_E V(.)V^*$ with an isometry $V : \mathcal{H} \to \mathcal{K} \otimes \mathcal{H}_E$, and writing the partial trace as a convex combination of unitary conjugations (e.g., by the discrete Weyl unitaries).*

Given a non-empty set of positive semi-definite operators $S \subseteq \mathcal{B}(\mathcal{H})_+$, its Δ-*radius* $R_\Delta(S)$ is defined as

$$R_\Delta(S) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\varrho \in \mathcal{S}} \Delta(\varrho \| \sigma). \tag{III.11}$$

If the above infimum is attained at some $\sigma \in \mathcal{S}(\mathcal{H})$ then σ is called a Δ-*center* of S. A variant of this notion is when, instead of minimizing the maximal Δ-distance, we minimize an averaged distance according to some finitely supported probability distribution $P \in \mathcal{P}_f(S)$. This yields the notion of the P-*weighted* Δ-*radius:*

$$R_{\Delta,P}(S) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{\varrho \in S} P(\varrho)\, \Delta(\varrho \| \sigma). \tag{III.12}$$

If the above infimum is attained at some $\sigma \in \mathcal{S}(\mathcal{H})$ then σ is called a P-*weighted* Δ-*center* for S.

Remark III.2 *For applications in channel coding, S will be the image of a classical-quantum channel, and hence a subset of the state space. In this case minimizing over density operators σ in* (III.11) *and* (III.12) *seems natural, while it is less obviously so when the elements of S are general positive semi-definite operators. We discuss this further in Appendix A.*

Remark III.3 *Note that for any finitely supported probability distribution P on $\mathcal{B}(\mathcal{H})_+$, and any* $\operatorname{supp} P \subseteq S \subseteq \mathcal{B}(\mathcal{H})_+$, *we have*

$$R_{\Delta,P}(S) = R_{\Delta,P}(\operatorname{supp} P) = R_{\Delta,P}(\mathcal{B}(\mathcal{H})_+).$$

That is, $R_{\Delta,P}(S)$ does not in fact depend on S, it is a function only of P. Hence, if no confusion arises, we may simply denote it as $R_{\Delta,P}$.

Remark III.4 *The concepts of the divergence radius and P-weighted divergence radius can be unified (to some extent) by the notion of the (P,β)-weighted Δ-radius, which we explain in Section A 1.*

We will mainly be interested in the above concepts when $\mathcal{S}$ is the image of a gcq channel $W : \mathcal{X} \to \mathcal{B}(\mathcal{H})_+$, in which case we will use the notation

$$\chi_\Delta(W,P) := R_{\Delta,P\circ W^{-1}}(\operatorname{ran} W) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sum_{x \in \mathcal{X}} P(x)\, \Delta(W(x) \| \sigma), \tag{III.13}$$

where $(P \circ W^{-1})(\varrho) := \sum_{x \in \mathcal{X} :\, W(x) = \varrho} P(x)$. Note that, as far as these quantities are concerned, the channel simply gives a parametrization of its image set, and the previously considered case can be recovered by parametrizing the set by itself, i.e., by taking the gcq channel $\mathcal{X} := \mathcal{S}$ and $W := \mathrm{id}_{\mathcal{S}}$. We will call (III.13) the P-*weighted* Δ-*radius* of the channel W, and any state achieving the infimum in its definition a P-*weighted* Δ-*center* for W. We define the Δ-*capacity* of the channel W as

$$\chi_\Delta(W) := \sup_{P \in \mathcal{P}_f(\mathcal{X})} \chi_\Delta(W,P).$$

In the relevant cases for information theory, the Δ-capacity coincides with the Δ-radius of the image of the channel, i.e.,

$$\chi_\Delta(W) = R_\Delta(\operatorname{ran} W) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{x \in \mathcal{X}} \Delta(W(x) \| \sigma);$$

see Proposition A.1.

We will mainly be interested in the above quantities when Δ is a quantum Rényi divergence. For some further properties of these quantities for general divergences, see Appendix A 1.

B. Quantum Rényi divergences

In this section we specialize to various notions of quantum Rényi divergences. For every pair of positive definite operators $\varrho, \sigma \in \mathcal{B}(\mathcal{H})_{++}$ and every $\alpha \in (0,+\infty) \setminus \{1\}$, $z \in (0,+\infty)$ let

$$Q_{\alpha,z}(\varrho\|\sigma) := \operatorname{Tr}\left(\varrho^{\frac{\alpha}{2z}}\sigma^{\frac{1-\alpha}{z}}\varrho^{\frac{\alpha}{2z}}\right)^{z}.$$

These quantities were first introduced in [34] and further studied in [6]. The cases

$$Q_{\alpha}(\varrho\|\sigma) := Q_{\alpha,1}(\varrho\|\sigma) = \operatorname{Tr}\varrho^{\alpha}\sigma^{1-\alpha}, \tag{III.14}$$

$$Q_{\alpha}^{*}(\varrho\|\sigma) := Q_{\alpha,\alpha}(\varrho\|\sigma) = \operatorname{Tr}\left(\varrho^{\frac{1}{2}}\sigma^{\frac{1-\alpha}{\alpha}}\varrho^{\frac{1}{2}}\right)^{\alpha}, \tag{III.15}$$

and

$$Q_{\alpha}^{\flat}(\varrho\|\sigma) := Q_{\alpha,+\infty}(\varrho\|\sigma) := \lim_{z\to+\infty} Q_{\alpha,z}(\varrho\|\sigma) = \operatorname{Tr} e^{\alpha\log\varrho+(1-\alpha)\log\sigma} \tag{III.16}$$

are of special significance. (The last identity in (III.16) is due to the Lie-Trotter formula.) Here and henceforth (t) stands for one of the three possible values $(t) = \{\ \}$, $(t) = *$ or $(t) = \flat$, where $\{\ \}$ denotes the empty string, i.e., $Q_{\alpha}^{(t)}$ with $(t) = \{\ \}$ is simply Q_{α}.

These quantities are extended to general, not necessarily invertible positive semi-definite operators $\varrho, \sigma \in \mathcal{B}(\mathcal{H})_{+}$ as

$$Q_{\alpha,z}(\varrho\|\sigma) := \lim_{\varepsilon\searrow 0} Q_{\alpha,z}(\varrho+\varepsilon I\|\sigma+\varepsilon I) \tag{III.17}$$

$$= \lim_{\varepsilon\searrow 0} Q_{\alpha,z}(\varrho\|\sigma+\varepsilon I) = \lim_{\varepsilon\searrow 0} Q_{\alpha,z}(\varrho\|(1-\varepsilon)\sigma+\varepsilon I/d) = s(\alpha)\sup_{\varepsilon>0}\overline{Q}_{\alpha,z}(\varrho\|\sigma+\varepsilon I), \tag{III.18}$$

for every $z \in (0,+\infty)$, where $d := \dim\mathcal{H}$,

$$s(\alpha) := \operatorname{sgn}(\alpha-1) = \begin{cases} -1, & \alpha<1, \\ 1, & \alpha>1 \end{cases}, \qquad \overline{Q}_{\alpha,z} := s(\alpha)Q_{\alpha,z},$$

and the identities are easy to verify. For $z = +\infty$, the extension is defined by (III.17); see [33, 44] for details.

Various further divergences can be defined from the above quantities. The *quantum α-z Rényi divergences* [6] are defined as

$$D_{\alpha,z}(\varrho\|\sigma) := \frac{1}{\alpha-1}\log\frac{Q_{\alpha,z}(\varrho\|\sigma)}{\operatorname{Tr}\varrho} \tag{III.19}$$

for any $\alpha \in (0,+\infty)\setminus\{1\}$ and $z \in (0,+\infty]$. It is easy to see that

$$\alpha>1, \quad \varrho^{0} \not\leq \sigma^{0} \implies Q_{\alpha,z} = D_{\alpha,z} = +\infty$$

for any z. Moreover, if $\alpha \mapsto z(\alpha)$ is continuously differentiable in a neighbourhood of 1, on which $z(\alpha) \neq 0$, or $z(\alpha) = +\infty$ for all α, then, according to [41, Theorem 1] and [44, Lemma 3.5],

$$D_{1}(\varrho\|\sigma) := \lim_{\alpha\to 1} D_{\alpha,z(\alpha)}(\varrho\|\sigma) = \frac{1}{\operatorname{Tr}\varrho}D(\varrho\|\sigma) =: D_{1,z}(\varrho\|\sigma), \qquad z \in (0,+\infty],$$

where $D(\varrho\|\sigma)$ is Umegaki's relative entropy [59], defined as

$$D(\varrho\|\sigma) := \operatorname{Tr}\varrho(\log\varrho - \log\sigma)$$

for positive definite operators, and extended as above for non-zero positive semidefinite operators. Of the Rényi divergences corresponding to the special Q_{α} quantities discussed above, D_{α} is usually called the *Petz-type* Rényi divergence, D_{α}^{*} the *sandwiched* Rényi divergence [45, 60], and

$D_\alpha^\flat$ the *log-Euclidean* Rényi divergence. For more on the above definitions and a more detailed reference to their literature, see, e.g., [44].

To discuss some important properties of the above quantities, let us introduce the following regions of the α-z plane:

$K_0 : 0 < \alpha < 1,\ z < \min\{\alpha, 1-\alpha\};$ $\quad K_1 : 0 < \alpha < 1,\ \alpha \le z \le 1-\alpha;$

$K_2 : 0 < \alpha < 1,\ \max\{\alpha, 1-\alpha\} \le z \le 1;$ $\quad K_3 : 0 < \alpha < 1,\ 1-\alpha \le z \le \alpha;$

$K_4 : 0 < \alpha < 1,\ 1 \le z;$ $\quad K_5 : 1 < \alpha,\ \alpha/2 \le z \le 1;$

$K_6 : 1 < \alpha,\ \max\{\alpha-1, 1\} \le z \le \alpha;$ $\quad K_7 : 1 < \alpha \le z;$

The (α, z) values for which $D_{\alpha,z}$ is monotone under CPTP maps have been completely characterized in [2, 9, 13, 24, 31, 62] (cf. also [6, Theorem 1]). This can be summarized as follows.

Lemma III.5 *$D_{\alpha,z}$ is monotone under CPTP maps $\Longleftrightarrow$ $\overline{Q}_{\alpha,z}$ is monotone under CPTP maps $\Longleftrightarrow$ $\overline{Q}_{\alpha,z}$ is jointly convex $\Longleftrightarrow$ $(\alpha, z) \in K_2 \cup K_4 \cup K_5 \cup K_6$.*

Corollary III.6 *$D_{\alpha,z}$ is jointly convex if $(\alpha, z) \in K_2 \cup K_4$.*

Proof Immediate from Lemma III.5, as the joint convexity of $\overline{Q}_{\alpha,z}$ implies the joint convexity of $D_{\alpha,z} = \frac{1}{\alpha-1}\log s(\alpha)\overline{Q}_{\alpha,z}$ whenever $\alpha \in (0,1)$. □

Recall that a function $f : C \to \mathbb{R}\cup\{+\infty\}$ on a convex set C is *quasi-convex* if $f((1-t)x+ty) \le \max\{f(x), f(y)\}$ for all $x, y \in C$ and $t \in [0,1]$.

Lemma III.7 *On top of the cases discussed in Lemma III.5 and Corollary III.6, $D_{\alpha,z}$ is convex in its second argument if $(\alpha, z) \in K_3 \cup K_6 \cup K_7$, and $\overline{Q}_{\alpha,z}$ is convex in its second argument if $(\alpha, z) \in K_3 \cup K_7$. Moreover, $D_{\alpha,z}$ is jointly quasi-convex if $(\alpha, z) \in K_5$.*

Proof The assertion about the quasi-convexity of $D_{\alpha,z}$ is immediate from the joint convexity of $\overline{Q}_{\alpha,z}$ when $(\alpha, z) \in K_5$.

Note that it is enough to prove convexity in the second argument for positive definite operators, due to (III.18).

Assume that $(\alpha, z) \in K_2 \cup K_3$, i.e., $0 < \alpha < 1$, $1-\alpha \le z \le 1$. Then $0 < \frac{1-\alpha}{z} \le 1$, and hence $\sigma \mapsto \sigma^{\frac{1-\alpha}{z}}$ is concave. Since $\mathcal{B}(\mathcal{H})_+ \ni A \mapsto \operatorname{Tr} A^z$ is both monotone and concave (see Lemma II.1), we get that $\sigma \mapsto Q_{\alpha,z}(\varrho\|\sigma)$ is concave, from which the convexity of both $\overline{Q}_{\alpha,z}$ and $D_{\alpha,z}$ in their second argument follows for $(\alpha, z) \in K_3$ (and also for K_2, although that is already covered by joint convexity).

Assume next that $(\alpha, z) \in K_6 \cup K_7$, i.e., $1 < \alpha$, and $\max\{1, \alpha-1\} \le z$. Then $-1 \le \frac{1-\alpha}{z} < 0$, and hence $f : t \mapsto t^{\frac{1-\alpha}{z}}$ is a non-negative operator monotone decreasing function on $(0, +\infty)$. Applying the duality of the Schatten p-norms to $p = z$, we have

$$D_{\alpha,z}(\varrho\|\sigma) = \sup_{\tau\in\mathcal{S}(\mathcal{H})} \frac{z}{\alpha-1}\log \operatorname{Tr} \varrho^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{z}} \varrho^{\frac{\alpha}{2z}} \tau^{1-\frac{1}{z}} = \sup_{\tau\in\mathcal{S}(\mathcal{H})} \frac{z}{\alpha-1}\log \omega_\tau\left(f(\sigma)\right),$$

where $\omega_\tau(.) := \operatorname{Tr} \varrho^{\frac{\alpha}{2z}}(.)\varrho^{\frac{\alpha}{2z}}\tau^{1-\frac{1}{z}}$ is a positive functional. By [3, Proposition 1.1], $D_{\alpha,z}(\varrho\|.)$ is the supremum of convex functions on $\mathcal{B}(\mathcal{H})_{++}$, and hence is itself convex. This immediately implies that $\overline{Q}_{\alpha,z}$ is convex in its second argument when $(\alpha, z) \in K_6 \cup K_7$ (of which the case K_6 also follows from joint convexity). □

Lemma III.8 *For any fixed $\varrho \in \mathcal{B}(\mathcal{H})_+$, the maps*

$$\sigma \mapsto \overline{Q}_{\alpha,z}(\varrho\|\sigma) \quad \text{and} \quad \sigma \mapsto D_{\alpha,z}(\varrho\|\sigma)$$

are lower semi-continuous on $\mathcal{B}(\mathcal{H})_+$ for any $\alpha \in (0, +\infty)\setminus\{1\}$ and $z \in (0, +\infty)$, and for $z = +\infty$ and $\alpha > 1$.

Proof The cases $\alpha \in (0, +\infty) \setminus \{1\}$ and $z \in (0, +\infty)$ are obvious from the last expression in (III.18), and the case $z = +\infty$ was discussed in [44, Lemma 3.27]. □

It is known that D_α, D_α^* and $D_\alpha^\flat$ are non-negative on pairs of states [44, 45, 49], but it seems that the non-negativity of general α-z Rényi divergences has not been analyzed in the literature so far. We show in Appendix A 4 that they are indeed non-negative for any pair of parameters (α, z).

C. The Rényi divergence center

Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a gcq channel. Specializing to $\Delta = D_{\alpha,z}$ in (III.13) yields the *P-weighted Rényi (α, z) radii* of the channel for a finitely supported input probability distribution $P \in \mathcal{P}_f(\mathcal{X})$,

$$\chi_{\alpha,z}(W,P) := \chi_{D_{\alpha,z}}(W,P) = \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x) D_{\alpha,z}(W(x)\|\sigma) = \min_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x) D_{\alpha,z}(W(x)\|\sigma). \tag{III.20}$$

The existence of the minimum is guaranteed by the lower semi-continuity stated in Lemma III.8. We will call any state σ achieving the minimum in (III.20) a *P-weighted $D_{\alpha,z}$ center* for W.

It is sometimes convenient that it is enough to consider the infimum above over invertible states, i.e., we have

$$\chi_{\alpha,z}(W,P) = \inf_{\sigma\in\mathcal{S}(\mathcal{H})_{++}} \sum_{x\in\mathcal{X}} P(x) D_{\alpha,z}(W(x)\|\sigma), \tag{III.21}$$

which is obvious from the second expression in (III.18). Moreover, any minimizer of (III.20) has the same support as the joint support of the channel states $\{W_x\}_{x\in\operatorname{supp} P}$, with projection

$$\bigvee_{x\in\operatorname{supp} P} W(x)^0 = W(P)^0,$$

at least for a certain range of (α, z) values, as we show below.

Lemma III.9 *Let σ be a P-weighted $D_{\alpha,z}$ center for W. If (α, z) is such that $D_{\alpha,z}$ is quasi-convex in its second argument then $\sigma^0 \le W(P)^0$.*

Proof Define $\mathcal{F}(X) := W(P)^0 X W(P)^0 + (I - W(P)^0) X (I - W(P)^0)$, $X \in \mathcal{B}(\mathcal{H})$, and let $\tilde\sigma := W(P)^0 \sigma W(P)^0 / \operatorname{Tr} W(P)^0 \sigma$. We will show that $D_{\alpha,z}(W(x)\|\tilde\sigma) \le D_{\alpha,z}(W(x)\|\sigma)$ for all $x \in \operatorname{supp} P$, which will yield the assertion. Note that we can assume without loss of generality that $W(P)^0\sigma \ne 0$, since otherwise $D_{\alpha,z}(W(x)\|\sigma) = +\infty$ for all $x \in \operatorname{supp} P$, and hence σ clearly cannot be a minimizer for (III.20).

According to the decomposition $\mathcal{H} = \operatorname{ran} W(P)^0 \oplus \operatorname{ran}(I - W(P)^0)$, define the block-diagonal unitary $U := \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}$, so that $\mathcal{F}(.) = \frac{1}{2}((.) + U(.)U^*)$. For every $x \in \mathcal{X}$,

$$\begin{aligned} D_{\alpha,z}(W(x)\|\mathcal{F}(\sigma)) &\le \max\{D_{\alpha,z}(W(x)\|\sigma), D_{\alpha,z}(W(x)\|U\sigma U^*)\} \\ &= \max\{D_{\alpha,z}(W(x)\|\sigma), D_{\alpha,z}(UW(x)U^*\|U\sigma U^*)\} = D_{\alpha,z}(W(x)\|\sigma), \end{aligned}$$

where the first inequality is due to quasi-convexity, and the first equality is due to the fact that $UW(x)U^* = W(x)$. On the other hand,

$$\begin{aligned} D_{\alpha,z}(W(x)\|\mathcal{F}(\sigma)) &= D_{\alpha,z}(W(x)\|(\operatorname{Tr} W(P)^0\sigma)\tilde\sigma) \\ &= D_{\alpha,z}\left(W(x)\|\tilde\sigma\right) - \log \operatorname{Tr} W(P)^0\sigma \ge D_{\alpha,z}\left(W(x)\|\tilde\sigma\right), \end{aligned}$$

where the inequality is strict unless $\sigma^0 \le W(P)^0$. □

For fixed W and P, we define

$$F(\sigma) := \sum_{x\in\mathcal{X}} P(x) D_{\alpha,z}(W(x)\|\sigma), \qquad \sigma \in \mathcal{B}(\mathcal{H})_+.$$

In the following, we may naturally interpret $W(x)$ as an operator acting on $\operatorname{ran} W(x)$ or on $\operatorname{ran} W(P)$.

Lemma III.10 *F is Fréchet-differentiable at every $\sigma \in \mathcal{B}(\mathcal{H})_{++}$, with Fréchet-derivative $DF(\sigma)$ given by*

$$\begin{aligned} DF(\sigma):\ Y \mapsto & \frac{z}{\alpha-1}\sum_{x\in\mathcal{X}} P(x)\frac{1}{Q_{\alpha,z}(W(x)\|\sigma)} \\ & \cdot \operatorname{Tr}\sum_{a,b} h^{[1]}_{\alpha,z}(a,b)P^\sigma_a W(x)^{\frac{\alpha}{2z}}\left(W(x)^{\frac{\alpha}{2z}}\sigma^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{2z}}\right)^{z-1}W(x)^{\frac{\alpha}{2z}}P^\sigma_b Y, \end{aligned} \tag{III.22}$$

where $h^{[1]}_{\alpha,z}$ is the first divided difference function of $h_{\alpha,z}(t) := t^{\frac{1-\alpha}{z}}$.

Proof We have $F = \sum_{x\in\mathcal{X}} P(x)(g_x \circ \iota_x \circ H_{\alpha,z})$, where $H_{\alpha,z} : \mathcal{B}(\mathcal{H})_+ \to \mathcal{B}(\mathcal{H})$, $H_{\alpha,z}(\sigma) := \sigma^{\frac{1-\alpha}{z}}$ is Fréchet differentiable at every $\sigma \in \mathcal{B}(\mathcal{H})_{++}$ with $DH_{\alpha,z}(\sigma) : Y \mapsto \sum_{a,b} h^{[1]}_{\alpha,z}(a,b)P^\sigma_a Y P^\sigma_b$, according to Lemma II.2. For a fixed x, $\iota_x : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\operatorname{ran}(W(x))$ is defined as $A \mapsto W(x)^{\frac{\alpha}{2z}}AW(x)^{\frac{\alpha}{2z}}$, and, as a linear map, it is Fréchet differentiable at every $A \in \mathcal{B}(\mathcal{H})$, with its derivative being equal to itself. Finally, $g_x : \mathcal{B}(\operatorname{ran} W(x)) \to \mathbb{R}$ is defined as $g_x(T) := \operatorname{Tr} T^z$, and it is Fréchet differentiable at every $T \in \mathcal{B}(\operatorname{ran} W(x))_{++}$, with Fréchet derivative $Dg_x(T) : Y \mapsto z\operatorname{Tr} T^{z-1}Y$, according to (II.7). If $\sigma \in \mathcal{B}(\operatorname{ran} W(P))_{++}$ then $H_{\alpha,z}(\sigma) \in \mathcal{B}(\operatorname{ran} W(P))_{++}$, and $\iota_x(H_{\alpha,z}(\sigma)) \in \mathcal{B}(\operatorname{ran} W(x))_{++}$. Hence, we can apply the chain rule for derivatives, and obtain (III.22). □

Lemma III.11 *Let σ be a P-weighted $D_{\alpha,z}$ center for W. If $\alpha \geq 1$ or $\alpha \in (0,1)$ and $1-\alpha < z < +\infty$ then $W(P)^0 \leq \sigma^0$.*

Proof When $\alpha > 1$ and $W(P)^0 \nleq \sigma^0$, there exists an $x \in \operatorname{supp} P$ with $W^0_x \nleq \sigma^0$ so that $D_{\alpha,z}(W(x)\|\sigma) = +\infty$. Hence, σ cannot be a minimizer for (III.20).

Assume for the rest that $\alpha \in (0,1)$, and σ is such that $W(P)^0 \nleq \sigma^0$; this is equivalent to the existence of an $x_0 \in \operatorname{supp} P$ such that $W_{x_0}P^\sigma_0 \neq 0$. Let us define the state $\omega := cP^\sigma_0$, with $c := 1/\operatorname{Tr} P^\sigma_0$. For every $t \in [0,1]$, let

$$\sigma_t := (1-t)\sigma + t\omega = \sum_{\lambda\in\operatorname{spec}(\sigma)\setminus\{0\}} (1-t)\lambda P^\sigma_\lambda + tcP^\sigma_0,$$

so that $\sigma_t \in \mathcal{B}(\mathcal{H})_{++}$ for every $t \in (0,1]$. Note that if $t < t_0 := \lambda_{\min}(\sigma)/(c+\lambda_{\min}(\sigma))$, where $\lambda_{\min}(\sigma)$ is the smallest non-zero eigenvalue of σ, then $P^{\sigma_t}_{ct} = P^\sigma_0$, and $P^{\sigma_t}_{(1-t)\lambda} = P^\sigma_\lambda$, $\lambda \in \operatorname{spec}(\sigma)\setminus\{0\}$.

By Lemma III.10, the derivative of $f(t) := F(\sigma_t)$ at any $t \in (0,1)$ is given by

$$\begin{aligned} f'(t) &= DF(\sigma_t)(\omega-\sigma) \\ &= \frac{z}{\alpha-1}\sum_{x\in\mathcal{X}} P(x)\frac{1}{Q_{\alpha,z}(W(x)\|\sigma_t)}\Bigg[h'_{\alpha,z}(ct)\,c\operatorname{Tr} A_{x,t}P^\sigma_0 \\ &\qquad\qquad - \sum_{\lambda\in\operatorname{spec}(\sigma)\setminus\{0\}} h'_{\alpha,z}((1-t)\lambda)\,\lambda\operatorname{Tr} A_{x,t}P^\sigma_\lambda\Bigg] \\ &= \sum_{x\in\mathcal{X}} P(x)\frac{1}{Q_{\alpha,z}(W(x)\|\sigma_t)}\Bigg[(1-t)^{\frac{1-\alpha}{z}-1}\sum_{\lambda\in\operatorname{spec}(\sigma)\setminus\{0\}} \lambda^{\frac{1-\alpha}{z}}\operatorname{Tr} A_{x,t}P^\sigma_\lambda \\ &\qquad\qquad - t^{\frac{1-\alpha}{z}-1}c^{\frac{1-\alpha}{z}}\operatorname{Tr} A_{x,t}P^\sigma_0\Bigg], \end{aligned}$$

where $A_{x,t} := W(x)^{\frac{\alpha}{2z}}\left(W(x)^{\frac{\alpha}{2z}}\sigma_t^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{2z}}\right)^{z-1}W(x)^{\frac{\alpha}{2z}}$.

Our aim will be to show that $\lim_{t\searrow 0} f'(t) = -\infty$. This implies that $f(t) < f(0)$ for small enough $t > 0$, contradicting the assumption that F has a global minimum at σ. Note that $\lim_{t\searrow 0} Q_{\alpha,z}(W(x)\|\sigma_t) = Q_{\alpha,z}(W(x)\|\sigma)$, which is strictly positive for every $x \in \operatorname{supp} P$. Indeed, the contrary would mean that $D_{\alpha,z}(W(x)\|\sigma) = +\infty$, contradicting again the assumption that F

has a global minimum at σ. Hence, the proof will be complete if we show that $t^{\frac{1-\alpha}{z}-1}\operatorname{Tr}A_{x_0,t}P_0^\sigma$ diverges to $+\infty$ while $\operatorname{Tr}A_{x,t}P_\lambda^\sigma$ is bounded as $t\searrow 0$ for any $x\in\operatorname{supp}P$ and $\lambda\in\operatorname{spec}(\sigma)\setminus\{0\}$.

Note that for any $t\in(0,t_0)$ and $z\geq 1$,

$$\begin{aligned}
tcI\leq\sigma_t\leq I \quad &\Longrightarrow\quad (tc)^{\frac{1-\alpha}{z}}I\leq\sigma_t^{\frac{1-\alpha}{z}}\leq I\\
&\Longrightarrow\quad (tc)^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{z}}\leq W(x)^{\frac{\alpha}{2z}}\sigma_t^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{2z}}\leq W(x)^{\frac{\alpha}{z}}\\
&\Longrightarrow\quad t^{\frac{1-\alpha}{z}}c_1W(x)^0\leq W(x)^{\frac{\alpha}{2z}}\sigma_t^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{2z}}\leq c_3W(x)^0\\
&\Longrightarrow\quad t^{\frac{1-\alpha}{z}(z-1)}c_2W(x)^0\leq\left[W(x)^{\frac{\alpha}{2z}}\sigma_t^{\frac{1-\alpha}{z}}W(x)^{\frac{\alpha}{2z}}\right]^{z-1}\leq c_4W(x)^0 &\text{(III.23)}\\
&\Longrightarrow\quad t^{\frac{1-\alpha}{z}(z-1)}c_2W(x)^{\frac{\alpha}{z}}\leq A_{x,t}\leq c_4W(x)^{\frac{\alpha}{z}}, &\text{(III.24)}
\end{aligned}$$

where $c_1:=c^{\frac{1-\alpha}{z}}\lambda_{\min}(W(x))^{\frac{\alpha}{z}}>0$, $c_2:=c_1^{z-1}>0$, $c_3:=\|W(x)\|^{\frac{\alpha}{z}}>0$, $c_4:=c_3^{z-1}>0$, and the inequalities in (III.23)–(III.24) hold in the opposite direction when $z\in(0,1)$. This immediately implies that

$$\begin{aligned}
t^{\frac{1-\alpha}{z}-1}\operatorname{Tr}A_{x_0,t}P_0^\sigma &\geq t^{\frac{1-\alpha}{z}-1+\frac{1-\alpha}{z}(z-1)}c_2\operatorname{Tr}W(x_0)^{\frac{\alpha}{z}}P_0^\sigma\xrightarrow[t\searrow 0]{}+\infty, && z\geq 1,\\
t^{\frac{1-\alpha}{z}-1}\operatorname{Tr}A_{x_0,t}P_0^\sigma &\geq t^{\frac{1-\alpha}{z}-1}c_4\operatorname{Tr}W(x_0)^{\frac{\alpha}{z}}P_0^\sigma\xrightarrow[t\searrow 0]{}+\infty, && z\in(0,1),
\end{aligned}$$

since $\operatorname{Tr}W(x_0)^{\frac{\alpha}{z}}P_0^\sigma>0$ by assumption, $\frac{1-\alpha}{z}-1+\frac{1-\alpha}{z}(z-1)=-\alpha<0$, and $\frac{1-\alpha}{z}-1<0$ iff $1-\alpha<z$ when $z\in(0,1)$.

Next, observe that

$$(1-t)P_\lambda^\sigma\leq\sigma_t\quad\Longrightarrow\quad(1-t)^{\frac{1-\alpha}{z}}P_\lambda^\sigma\leq\sigma_t^{\frac{1-\alpha}{z}}$$

where the inequality follows since, by assumption, $0<\frac{1-\alpha}{z}<1$, and $x\mapsto x^\gamma$ is operator monotone on $(0,+\infty)$ for $\gamma\in(0,1)$. Hence,

$$0\leq\operatorname{Tr}A_{x,t}P_\lambda^\sigma\leq(1-t)^{\frac{\alpha-1}{z}}\operatorname{Tr}A_{x,t}\sigma_t^{\frac{1-\alpha}{z}}=(1-t)^{\frac{\alpha-1}{z}}Q_{\alpha,z}(W(x)\|\sigma_t)\xrightarrow[t\searrow 0]{}Q_{\alpha,z}(W(x)\|\sigma),$$

which is finite. This finishes the proof. □

Remark III.12 *Note that the region of (α,z) values given in Lemma III.11 covers $z=1$ for all $\alpha\in(0,+\infty]$, i.e., all the Petz-type Rényi divergences, and $\{(\alpha,\alpha):\,\alpha\in(1/2,+\infty]\}$, i.e., the sandwiched Rényi divergences for every parameter α for which they are monotone under CPTP maps, except for $\alpha=1/2$. It is an open question whether the condition $z>1-\alpha$ in Lemma III.11 can be improved, or maybe completely removed.*

Remark III.13 *Note that the case $\alpha>1$ in Lemma III.11 is trivial, and this is the case that we actually need for the strong converse exponent of constant composition classical-quantum channel coding in Section IV; more precisely, we need the case $z=\alpha>1$.*

Let us define Γ_D to be the set of (α,z) values such that for any gcq channel W and any input probability distribution P, any P-weighted $D_{\alpha,z}$ center σ for W satisfies $\sigma^0=W(P)^0$. Then Corollary III.6 and Lemmas III.7, III.9 and III.11 yield

$$\Gamma_D\supseteq\{(\alpha,z):\,\alpha\in(0,1),\,1-\alpha<z+\infty\}\cup\{(\alpha,z):\,\alpha>1,\,z\geq\max\{\alpha/2,\alpha-1\}\}\,.$$

The following characterization of the weighted $D_{\alpha,z}$ centers will be crucial in proving the additivity of the weighted sandwiched Rényi divergence radius of a gcq channel.

Theorem III.14 *Assume that $(\alpha,z)\in\Gamma_D$ are such that $D_{\alpha,z}$ is convex in its second variable. Then σ is a P-weighted $D_{\alpha,z}$ center for W if and only if it is a fixed point of the map*

$$\Phi_{W,P,D_{\alpha,z}}(\sigma):=\sum_{x\in\mathcal{X}}P(x)\frac{1}{Q_{\alpha,z}(W(x)\|\sigma)}\left(\sigma^{\frac{1-\alpha}{2z}}W(x)^{\frac{\alpha}{z}}\sigma^{\frac{1-\alpha}{2z}}\right)^z \tag{III.25}$$

defined on $\mathcal{S}_{W,P}(\mathcal{H})_{++}:=\{\sigma\in\mathcal{S}(\mathcal{H})_+:\,\sigma^0=W(P)^0\}$.

Proof By the assumption that $(\alpha, z) \in \Gamma_D$, we may restrict the Hilbert space to be $\operatorname{ran} W(P)^0$, and assume that σ is invertible. Let $F(A) := \sum_{x\in\mathcal{X}} P(x) D_{\alpha,z}(W(x)\|A)$, $A \in \mathcal{B}(\mathcal{H})_{++}$. Due to the assumption that $D_{\alpha,z}$ is convex in its second variable, σ is a minimizer of F if and only if $DF(\sigma)(Y) = 0$ for all self-adjoint traceless Y. By Lemma III.10, this condition is equivalent to

$$\lambda I = \frac{z}{\alpha-1} \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q_{\alpha,z}(W(x)\|\sigma)} \sum_{a,b} h^{[1]}_{\alpha,z}(a,b) P^\sigma_a W(x)^{\frac{\alpha}{2z}} \left(W(x)^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{z}} W(x)^{\frac{\alpha}{2z}}\right)^{\alpha-1} W(x)^{\frac{\alpha}{2z}} P^\sigma_b$$

for some $\lambda \in \mathbb{R}$. Multiplying both sides by $\sigma^{1/2}$ from the left and the right, and taking the trace, we get $\lambda = -1$. Hence, the above is equivalent to (by multiplying both sides by $\sigma^{1/2}$ from the left and the right)

$$\begin{aligned} \sigma &= \frac{z}{1-\alpha} \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q_{\alpha,z}(W(x)\|\sigma)} \sum_{a,b} a^{1/2} b^{1/2} h^{[1]}_{\alpha,z}(a,b) P^\sigma_a W(x)^{\frac{\alpha}{2z}} \left(W(x)^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{z}} W(x)^{\frac{\alpha}{2z}}\right)^{z-1} W(x)^{\frac{\alpha}{2z}} P^\sigma_b \\ &= \sum_{a,b} P^\sigma_a \left(\frac{z}{1-\alpha} a^{1/2} b^{1/2} h^{[1]}_{\alpha,z}(a,b) \widehat{\Phi}_{W,P,\alpha,z}(\sigma)\right) P^\sigma_b, \end{aligned} \tag{III.26}$$

where

$$\widehat{\Phi}_{W,P,D_{\alpha,z}}(\sigma) := \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q^*_{\alpha,z}(W(x)\|\sigma)} W(x)^{\frac{\alpha}{2z}} \left(W(x)^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{z}} W(x)^{\frac{\alpha}{2z}}\right)^{z-1} W(x)^{\frac{\alpha}{2z}}.$$

Writing the operators in (III.26) in block form according to the spectral decomposition of σ, we see that (III.26) is equivalent to

$$\begin{aligned} \forall a,b: \quad \delta_{a,b} a^{\frac{\alpha-1}{z}+1} P^\sigma_a = P^\sigma_a \widehat{\Phi}_{W,P,\alpha,z}(\sigma) P^\sigma_b \quad &\iff \quad \sigma^{\frac{\alpha-1}{z}+1} = \widehat{\Phi}_{W,P,D_{\alpha,z}}(\sigma) \\ &\iff \quad \sigma = \sigma^{\frac{1-\alpha}{2z}} \widehat{\Phi}_{W,P,D_{\alpha,z}}(\sigma) \sigma^{\frac{1-\alpha}{2z}}. \end{aligned}$$

This can be rewritten as

$$\begin{aligned} \sigma &= \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q_{\alpha,z}(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2z}} W(x)^{\frac{\alpha}{2z}} \left(W(x)^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{z}} W(x)^{\frac{\alpha}{2z}}\right)^{z-1} W(x)^{\frac{\alpha}{2z}} \sigma^{\frac{1-\alpha}{2z}} \\ &= \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q_{\alpha,z}(W(x)\|\sigma)} \left(\sigma^{\frac{1-\alpha}{2z}} W(x)^{\frac{\alpha}{z}} \sigma^{\frac{1-\alpha}{2z}}\right)^z, \end{aligned}$$

where the last identity follows from $Xf(X^*X)X^* = (\mathrm{id}_{\mathbb{R}} f)(XX^*)$. □

Remark III.15 *The special case $z = 1$ yields the characterization of the P-weighted Petz-type Rényi divergence center as the fixed point of the map*

$$\Phi_{W,P,D_\alpha}(\sigma) := \sum_{x\in\mathcal{X}} P(x) \frac{1}{Q_\alpha(W(x)\|\sigma)} \sigma^{\frac{1-\alpha}{2}} W(x)^\alpha \sigma^{\frac{1-\alpha}{2}}, \qquad \sigma \in \mathcal{S}_{W,P}(\mathcal{H})_{++},$$

for any $\alpha \in (0, +\infty)\setminus\{1\}$. Note that in the classical case $D_{\alpha,z}$ is independent of z, i.e., $D_{\alpha,z} = D_\alpha$ for all $z > 0$, and the above characterization of the minimizer has been derived recently by Nakiboğlu in [53, Lemma 13], using very different methods. Following Nakiboğlu's approach, Cheng, Li and Hsieh has derived the above characterization for the Petz-type Rényi divergence center in [15, Proposition 4]. The advantage of Nakiboğlu's approach is that it also provides quantitative bounds of the deviation of $\sum_x P(x) D_\alpha(W(x)\|\sigma)$ from $\chi_{\alpha,z}(W,P)$ for an arbitrary state σ; however, it is not clear whether this approach can be extended to the case $z \neq 1$, in particular, for $z = \alpha$, which is the relevant case for the strong converse exponent of constant composition classical-quantum channel coding, as we will see in Section IV.

Remark III.16 *A similar approach as in the above proof of Theorem III.14 was used by Hayashi and Tomamichel in [29, Appendix C] to characterize the optimal state for the sandwiched Rényi mutual information as the fixed point of a non-linear map on the state space. We comment on this in more detail in Section A 2.*

Example III.17 *We say that a cq channel W is noiseless on* $\operatorname{supp} P$ *if* $W(x)W(y) = 0$ *for all* $x, y \in \operatorname{supp} P$, $x \neq y$, *i.e., the output states corresponding to inputs in* $\operatorname{supp} P$ *are perfectly distinguishable. A straightforward computation shows that if W is noiseless on* $\operatorname{supp} P$ *then* $\sigma := W(P) = \sum_x P(x)W(x)$ *satisfies the fixed point equation* (III.25) *for any pair* (α, z). *Hence, if* (α, z) *satisfies the conditions of Proposition III.14 then* $W(P)$ *is a minimizer for* (III.20), *and we have*

$$\chi_{\alpha,z}(W,P) = \sum_{x\in\mathcal{X}} P(x)D_{\alpha,z}(W(x)\|W(P)) = H(P) := -\sum_{x\in\mathcal{X}} P(x)\log P(x).$$

Thus, the Rényi (α, z) *radius of W is equal to the Shannon entropy of the input distribution, independently of the value of* (α, z).

Corollary III.18 *If* (α, z) *satisfies the conditions of Proposition III.14, and* $D_{\alpha,z}$ *is monotone under CPTP maps then*

$$\chi_{\alpha,z}(W,P) \le H(P) \tag{III.27}$$

for any cq channel W and input distribution P.

Proof We may assume without loss of generality that $\mathcal{X} = \operatorname{supp} P$. Let $\tilde{W}(x) := |e_x\rangle\langle e_x|$ for some orthonormal basis $(e_x)_{x\in\operatorname{supp} P}$ in a Hilbert space $\mathcal{K}$, and let $\Phi(.) := \sum_{x\in\operatorname{supp} P} W(x)\,\langle e_x|\,(.)\,|e_x\rangle$, which is a CPTP map from $\mathcal{B}(\mathcal{K})$ to $\mathcal{B}(\mathcal{H})$. We have $W = \Phi\circ\tilde{W}$, and the assertion follows from Example III.17. □

Remark III.19 *Our approach to prove* (III.27) *follows that of Csiszár [16]. A (much) simpler approach to prove the inequality* (III.27) *was given by Nakiboğlu [53, Lemma 13] (see also [15, Proposition 4] for an adaptation to various quantum Rényi divergences). Obviously,*

$$\chi_{\alpha,z}(W,P) \le \sum_{x\in\mathcal{X}} P(x)D_{\alpha,z}(W(x)\|W(P)). \tag{III.28}$$

Assume now that $D_{\alpha,z}$ *satisfies the monotonicity property* $\mathcal{B}(\mathcal{H})_+ \ni \sigma_1 \le \sigma_2 \Longrightarrow D_{\alpha,z}(\varrho\|\sigma_1) \ge D_{\alpha,z}(\varrho\|\sigma_2)$ *for any* $\varrho \in \mathcal{B}(\mathcal{H})_+$. *It is easy to see that this holds for every* (α, z) *with* $z \ge |\alpha-1|$. *In this case, we can lower bound* $W(P)$ *by* $P(x)W(x)$, *and hence* $D_{\alpha,z}(W(x)\|W(P)) \le D_{\alpha,z}(W(x)\|P(x)W(x)) = -\log P(x)$, *whence the RHS of* (III.28) *can be upper bounded by* $H(P)$.

D. Additivity of the weighted Rényi radius

Let $W^{(i)} : \mathcal{X}^{(i)} \to \mathcal{B}(\mathcal{H}^{(i)})_+$, $i = 1, 2$, be gcq channels, and $P^{(i)} \in \mathcal{P}_f(\mathcal{X}^{(i)})$ be input probability distributions. For any $\alpha \in (0, +\infty)$ and $z \in (0, +\infty]$,

$$\chi_{\alpha,z}\left(W^{(1)}\otimes W^{(2)}, P^{(1)}\otimes P^{(2)}\right) \tag{III.29}$$

$$\le \inf_{\sigma_i\in\mathcal{S}(\mathcal{H}_i)} \sum_{x_1\in\mathcal{X}^{(1)},\, x_2\in\mathcal{X}^{(2)}} P^{(1)}(x_1)P^{(2)}(x_2)D_{\alpha,z}\left(W^{(1)}(x_1)\otimes W^{(1)}(x_1)\|\sigma_1\otimes\sigma_2\right) \tag{III.30}$$

$$= \chi_{\alpha,z}\left(W^{(1)}, P^{(1)}\right) + \chi_{\alpha,z}\left(W^{(2)}, P^{(2)}\right), \tag{III.31}$$

by definition, i.e., $\chi_{\alpha,z}$ is subadditive. In particular, for fixed $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ and $P \in \mathcal{P}_f(\mathcal{X})$, the sequence $m \mapsto \chi_{\alpha,z}(W^{\otimes m}, P^{\otimes m})$ is subadditive, and hence

$$\lim_{m\to+\infty}\frac{1}{m}\chi_{\alpha,z}(W^{\otimes m}, P^{\otimes m}) = \inf_{m\in\mathbb{N}}\frac{1}{m}\chi_{\alpha,z}(W^{\otimes m}, P^{\otimes m}) \le \chi_{\alpha,z}(W,P).$$

In fact,

$$\frac{1}{m}\chi_{\alpha,z}(W^{\otimes m}, P^{\otimes m}) \le \chi_{\alpha,z}(W,P) \tag{III.32}$$

for all $m \in \mathbb{N}$.

As it turns out, we also have the stronger property of additivity, at least for (α, z) pairs for which the optimal σ can be characterized by the fixed point equation (III.25).

Theorem III.20 *(Additivity of the weighted Rényi radius) Let $W^{(1)}:\mathcal{X}^{(1)}\to\mathcal{S}(\mathcal{H}^{(1)})$ and $W^{(2)}:\mathcal{X}^{(2)}\to\mathcal{S}(\mathcal{H}^{(2)})$ be gcq channels, and $P^{(i)}\in\mathcal{P}_f(\mathcal{X}^{(i)})$, $i=1,2$, be input distributions. Assume, moreover, that α and z satisfy the conditions of Theorem III.14. Then*

$$\chi_{\alpha,z}\left(W^{(1)}\otimes W^{(2)},P^{(1)}\otimes P^{(2)}\right)=\chi_{\alpha,z}\left(W^{(1)},P^{(1)}\right)+\chi_{\alpha,z}\left(W^{(2)},P^{(2)}\right). \tag{III.33}$$

Proof Let σ_i be a minimizer of (III.20) for $(W^{(i)},P^{(i)})$. By Theorem III.14, this means that $\Phi_{W^{(i)},P^{(i)},D_{\alpha,z}}(\sigma_i)=\sigma_i$. It is easy to see that

$$\Phi_{W^{(1)}\otimes W^{(2)},P^{(1)}\otimes P^{(2)},D_{\alpha,z}}(\sigma_1\otimes\sigma_2)=\Phi_{W^{(1)},P^{(1)},D_{\alpha,z}}(\sigma_1)\otimes\Phi_{W^{(2)},P^{(2)},D_{\alpha,z}}(\sigma_2)=\sigma_1\otimes\sigma_2.$$

Hence, again by Proposition III.14, $\sigma_1\otimes\sigma_2$ is a minimizer of (III.20) for $(W^{(1)}\otimes W^{(2)},P^{(1)}\otimes P^{(2)})$. This proves the assertion. □

Corollary III.21 *For any gcq channel $W:\mathcal{X}\to\mathcal{B}(\mathcal{H})_+$, any $P\in\mathcal{P}_f(\mathcal{X})$, and any pair (α,z) satisfying the conditions in Theorem III.14, we have*

$$\chi_{\alpha,z}(W^{\otimes m},P^{\otimes m})=m\chi_{\alpha,z}(W,P),\qquad m\in\mathbb{N}.$$

We will need the following special case for the application to classical-quantum channel coding in the next section:

Corollary III.22 *For any gcq channel $W:\mathcal{X}\to\mathcal{B}(\mathcal{H})_+$, any $P\in\mathcal{P}_f(\mathcal{X})$, and any $\alpha\in(1/2,+\infty]$,*

$$\chi_\alpha^*(W^{\otimes m},P^{\otimes m})=m\chi_\alpha^*(W,P),\qquad m\in\mathbb{N}.$$

Remark III.23 *As far as we are aware, the idea of proving the additivity of an information quantity by characterizing some optimizer state as the fixed point of a non-linear operator on the state space appeared first in [29]. We comment on this in more detail in Appendix A 2.*

IV. STRONG CONVERSE EXPONENT WITH CONSTANT COMPOSITION

Let $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ be a classical-quantum channel. A *code* $\mathcal{C}_n$ for n uses of the channel is a pair $\mathcal{C}_n=(\mathcal{E}_n,\mathcal{D}_n)$, where $\mathcal{E}_n:[M_n]\to\mathcal{X}^n$, $\mathcal{D}_n:[M_n]\to\mathcal{B}(\mathcal{H}^{\otimes n})_+$, where $|\mathcal{C}_n|:=M_n\in\mathbb{N}$ is the size of the code, and $\mathcal{D}_n$ is a POVM, i.e., $\sum_{i=1}^{M_n}\mathcal{D}_n(i)=I$. The average success probability of a code $\mathcal{C}_n$ is

$$P_s(W^{\otimes n},\mathcal{C}_n):=\frac{1}{|\mathcal{C}_n|}\sum_{m=1}^{|\mathcal{C}_n|}\operatorname{Tr}W^{\otimes n}(\mathcal{E}_n(m))\mathcal{D}_n(m).$$

A sequence of codes $\mathcal{C}_n=(\mathcal{E}_n,\mathcal{D}_n)$, $n\in\mathbb{N}$, is called a sequence of *constant composition codes with asymptotic composition* $P\in\mathcal{P}_f(\mathcal{X})$ if there exists a sequence of types $P_n\in\mathcal{P}_n(\mathcal{X})$, $n\in\mathbb{N}$, such that $\lim_{n\to+\infty}\|P_n-P\|_1=0$, and $\mathcal{E}_n(k)\in\mathcal{X}^n_{P_n}$ for all $k\in\{1,\dots,|\mathcal{C}_n|\}$, $n\in\mathbb{N}$. (See Section II for the notation and basic facts concerning types.) For any rate $R\geq0$, the strong converse exponents of W with composition constraint P are defined as

$$\underline{\mathrm{sc}}(W,R,P):=\inf\left\{\liminf_{n\to+\infty}-\frac{1}{n}\log P_s(W^{\otimes n},\mathcal{C}_n):\,\liminf_{n\to+\infty}\frac{1}{n}\log|\mathcal{C}_n|\geq R\right\}, \tag{IV.34}$$

$$\overline{\mathrm{sc}}(W,R,P):=\inf\left\{\limsup_{n\to+\infty}-\frac{1}{n}\log P_s(W^{\otimes n},\mathcal{C}_n):\,\liminf_{n\to+\infty}\frac{1}{n}\log|\mathcal{C}_n|\geq R\right\}, \tag{IV.35}$$

where the infima are taken over code sequences of constant composition P.

Our main result is the following:

Theorem IV.1 *For any classical-quantum channel W, and finitely supported probability distribution P on the input of W, and any rate R,*

$$\underline{\mathrm{sc}}(W,R,P)=\overline{\mathrm{sc}}(W,R,P)=\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^*(W,P)\right]. \tag{IV.36}$$

We will prove the equality in (IV.36) as two separate inequalities in Propositions IV.3 and IV.5. Before starting with that, we point out the following complementary result by Dalai and Winter [19]:

Remark IV.2 *Similarly to the strong converse exponents, one can define the direct exponents as*

$$\underline{d}(W,R,P):=\sup\left\{\liminf_{n\to+\infty}-\frac{1}{n}\log(1-P_s(W^{\otimes n},\mathcal{C}_n))\,:\,\liminf_{n\to+\infty}\frac{1}{n}\log|\mathcal{C}_n|\ge R\right\},$$
$$\overline{d}(W,R,P):=\sup\left\{\limsup_{n\to+\infty}-\frac{1}{n}\log(1-P_s(W^{\otimes n},\mathcal{C}_n))\,:\,\liminf_{n\to+\infty}\frac{1}{n}\log|\mathcal{C}_n|\ge R\right\},$$

where the suprema are taken over code sequences of constant composition P. The following, so-called sphere packing bound has been shown in [19]:

$$\overline{d}(W,R,P)\le\sup_{0<\alpha<1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha(W,P)\right]. \tag{IV.37}$$

Note that the right-hand sides of (IV.36) *and* (IV.37) *are very similar to each other, except that the range of optimization is $\alpha>1$ in the former and $\alpha\in(0,1)$ in the latter, and the weighted Rényi radii corresponding to the sandwiched Rényi divergences appear in the former, and to the Petz-type Rényi divergences in the latter. Also, while* (IV.36) *holds for any $R>0$ (and is non-trivial for $R>\chi_1(W,P)$), it is known that* (IV.37) *holds as an equality only for high enough rates (and is non-trivial for $R<\chi_1(W,P)$) for classical channels, and it is a long-standing open problem if the same equality is true for classical-quantum channels.*

The following lower bound follows by a standard argument, due to Nagaoka [46], as was also observed, e.g., in [14]. For readers' convenience, we write out the details in Appendix B.

Proposition IV.3 *For any $R>0$,*

$$\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^*(W,P)\right]\le\underline{\mathrm{sc}}(W,R,P).$$

Our aim in the rest is to show that the second term is upper bounded by the rightmost term in (IV.36). We will follow the approach of [44], which in turn was inspired by [21]. We start with the following:

Proposition IV.4 *For any $R>0$,*

$$\overline{\mathrm{sc}}(W,R,P)\le\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^\flat(W,P)\right]. \tag{IV.38}$$

Proof We will show that

$$\overline{\mathrm{sc}}(W,R,P)\le\min\{F_1(W,R,P),F_2(W,R,P)\}, \tag{IV.39}$$

where

$$F_1(W,R,P):=\inf_{V:\,\chi(V,P)>R}D(V\|W|P),$$
$$F_2(W,R,P):=\inf_{V:\,\chi(V,P)\le R}\left[D(V\|W|P)+R-\chi(V,P)\right].$$

Here, the infima are over channels $V:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ satisfying the indicated properties. It was shown in [44, Theorem 5.12] that the RHS of (IV.39) is the same as the RHS of (IV.38).

We first show that $\overline{\mathrm{sc}}(W,R,P)\le F_1(W,R,P)$. To this end, let $r>F_1(W,R,P)$; then, by definition, there exists a channel $V:\mathcal{X}\to\mathcal{S}(\mathcal{H})$ such that

$$D(V\|W|P)<r \qquad \text{and} \qquad \chi(V,P)>R.$$

Due to $\chi(V,P)>R$, Corollary C.3 yields the existence of a sequence of constant composition codes $\mathcal{C}_n$ with composition P_n, $n\in\mathbb{N}$, such that $\operatorname{supp}P_n\subseteq\operatorname{supp}P$ for all n, $\lim_{n\to+\infty}\|P_n-P\|_1=0$, the rate is lower bounded as $\frac{1}{n}\log|\mathcal{C}_n|\ge R$, $n\in\mathbb{N}$, and $\lim_{n\to+\infty}P_s(V^{\otimes n},\mathcal{C}_n)=1$. Note that for any message k,

$$\operatorname{Tr}\left(V^{\otimes n}(\mathcal{E}_n(k))-e^{nr}W^{\otimes n}(\mathcal{E}_n(k))\right)_+\ge\operatorname{Tr}\left(V^{\otimes n}(\mathcal{E}_n(k))-e^{nr}W^{\otimes n}(\mathcal{E}_n(k))\right)\mathcal{D}_n(k),$$

and hence

$$\operatorname{Tr}W^{\otimes n}(\mathcal{E}_n(k))\mathcal{D}_n(k)\ge e^{-nr}\left[\operatorname{Tr}V^{\otimes n}(\mathcal{E}_n(k))\mathcal{D}_n(k)-\operatorname{Tr}\left(V^{\otimes n}(\mathcal{E}_n(k))-e^{nr}W^{\otimes n}(\mathcal{E}_n(k))\right)_+\right],$$

This in turn yields, by averaging over k, that

$$\begin{aligned}P_s(W^{\otimes n},\mathcal{C}_n)&\ge e^{-nr}\left[P_s(V^{\otimes n},\mathcal{C}_n)-\frac{1}{|\mathcal{C}_n|}\sum_{k=1}^{|\mathcal{C}_n|}\operatorname{Tr}\left(V^{\otimes n}(\mathcal{E}_n(k))-e^{nr}W^{\otimes n}(\mathcal{E}_n(k))\right)_+\right]\\&=e^{-nr}\left[P_s(V^{\otimes n},\mathcal{C}_n)-\operatorname{Tr}\left(V^{\otimes n}(\underline{x}^{(n)})-e^{nr}W^{\otimes n}(\underline{x}^{(n)})\right)_+\right],\end{aligned}$$

where $\underline{x}^{(n)}$ is any sequence in $\mathcal{X}^n$ with type P_n. Since $D(V\|W|P)<r$, Corollary D.2 yields that $\lim_{n\to+\infty}\operatorname{Tr}\left(V^{\otimes n}(\underline{x}^{(n)})-e^{nr}W^{\otimes n}(\underline{x}^{(n)})\right)_+=0$, and so finally

$$\liminf_{n\to+\infty}\frac{1}{n}\log P_s(W^{\otimes n},\mathcal{C}_n)\ge -r,\qquad\text{whence}\qquad \overline{\mathrm{sc}}(W,R,P)\le r.$$

Since this holds for any $r>F_1(W,R,P)$, we get $\overline{\mathrm{sc}}(W,R,P)\le F_1(W,R,P)$.

From this, one can prove that also $\overline{\mathrm{sc}}(W,R,P)\le F_2(W,R,P)$, the same way as it was done in [44, Lemma 5.11] (which in turn followed the proof in [21, Lemma 2]); one only has to make sure that the extension of the code can be done in a way that it remains constant composition with composition P, but that is easy to verify. □

From the above result, we can obtain the desired upper bound.

Proposition IV.5 *For any $R>0$,*

$$\overline{\mathrm{sc}}(W,R,P)\le\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^*(W,P)\right]. \tag{IV.40}$$

Proof We employ the asymptotic pinching technique from [44]. Let $W_m:\mathcal{X}^m\to\mathcal{S}(\mathcal{H}^{\otimes m})$ be defined as

$$W_m(\underline{x}):=\mathcal{F}_m\,W^{\otimes m}(\underline{x}),$$

where $\mathcal{F}_m$ is the pinching by the universal symmetric state $\sigma_{u,m}$, introduced in Section II. Employing Proposition IV.4 with $W\mapsto W_m$, $R\mapsto Rm$ and $P\mapsto P^{\otimes m}$, we get that for any $R>0$, there exists a sequence of codes $\mathcal{C}_k^{(m)}=(\mathcal{E}_k^{(m)},\mathcal{D}_k^{(m)})$ with constant composition $P_k^{(m)}\in\mathcal{P}_k(\mathcal{X}^n)$, $k\in\mathbb{N}$, such that $\frac{1}{k}\log|\mathcal{C}_k^{(m)}|\ge mR$ for all k, $\lim_{k\to+\infty}\left\|P_k^{(m)}-P^{\otimes m}\right\|_1=0$, and

$$\limsup_{k\to+\infty}-\frac{1}{k}\log P_s(W_m^{\otimes k},\mathcal{C}_k^{(m)})\le\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[mR-\chi_\alpha^\flat(W_m,P^{\otimes m})\right].$$

For every $k\in\mathbb{N}$, define $\mathcal{C}_{km}:=(\mathcal{E}_k^{(m)},\mathcal{F}_m\,\mathcal{D}_k^{(m)})$, which can be considered a code for $W^{\otimes km}$, with the natural identifications $(\mathcal{X}^m)^k\equiv\mathcal{X}^{km}$ and $(\mathcal{H}^{\otimes m})^{\otimes k}\equiv\mathcal{H}^{\otimes km}$. For a general $n\in\mathbb{N}$, choose $k \in \mathbb{N}$ such that $km \le n < (k+1)m$, and for every $i = 1, \dots, |\mathcal{C}_k^{(m)}|$, define $\mathcal{E}_n(i)$ to be $\mathcal{E}_{km}(i)$ concatenated with $n - km$ copies of some fixed $x_0 \in \operatorname{supp} P$, independent of i and n, and let $\mathcal{D}_n(i) := \mathcal{D}_{km}(i) \otimes I_{\mathcal{H}}^{\otimes(n-km)}$. Then it is easy to see that

$$\liminf_{n\to+\infty} \frac{1}{n} \log |\mathcal{C}_n| \ge R,$$

and

$$\limsup_{n\to+\infty} -\frac{1}{n} \log P_s(W^{\otimes n}, \mathcal{C}_n) \le \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left[R - \frac{1}{m}\chi_\alpha^\flat(W_m, P^{\otimes m})\right]. \tag{IV.41}$$

We need to show that the above sequence of codes is of constant composition P. Let $\mathbf{x} = (\underline{x}_1, \dots, \underline{x}_k) \in (\mathcal{X}^m)^k \equiv \mathcal{X}^{km}$ be a codeword for $\mathcal{C}_k^{(m)}$, and let $P_{\mathbf{x}}^{(m)}$ and $P_{\mathbf{x}}$ denote the corresponding types when $\mathbf{x}$ is considered as an element of $(\mathcal{X}^m)^k$ and of $\mathcal{X}^{km}$, respectively. For any $a \in \mathcal{X}$,

$$P_{\mathbf{x}}(a) = \frac{1}{km} \sum_{\underline{x}\in\mathcal{X}^m} \#\{i : \mathbf{x}_i = \underline{x}\} \cdot \#\{j : x_j = a\} = \sum_{\underline{x}\in\mathcal{X}^m} P_{\mathbf{x}}^{(m)}(\underline{x}) P_{\underline{x}}(a) = \sum_{\underline{x}\in\mathcal{X}^m} P_k^{(m)}(\underline{x}) P_{\underline{x}}(a)$$

only depends on $\mathbf{x}$ through its type $P_{\mathbf{x}}^{(m)} = P_k^{(m)}$, which is independent of $\mathbf{x}$. Thus, the type of $\mathcal{E}_{km}(i)$ is independent of i, i.e., $\mathcal{C}_{km}$ is a constant composition code for every $k \in \mathbb{N}$. For a general $n \in \mathbb{N}$ with $km \le n < (k+1)m$, we have

$$P_{\mathcal{E}_n(i)}(a) = \frac{km}{n} P_{\mathcal{E}_{km}(i)}(a) + \delta_{a,x_0} \frac{n-km}{n}, \qquad i \in \{1, \dots, |\mathcal{C}_n| = |\mathcal{C}_{km}|\},$$

and hence $\mathcal{C}_n$ is also of constant composition.

Next, we show that $\lim_{n\to+\infty} \|P_n - P\|_1 = 0$, where P_n is the type of $\mathcal{C}_n$. For $km \le n < (k+1)m$, we have

$$\sum_{a\in\mathcal{X}} |P_n(a) - P(a)| \le \sum_{a\in\mathcal{X}} |P_n(a) - P_{km}(a)| + \sum_{a\in\mathcal{X}} |P_{km}(a) - P(a)|,$$

and

$$\sum_{a\in\mathcal{X}} |P_n(a) - P_{km}(a)| = \sum_{a\in\mathcal{X}} \left(1 - \frac{km}{n}\right) P_{km}(a) + \left(1 - \frac{km}{n}\right) = 2\left(1 - \frac{km}{n}\right) \to 0$$

as $k \to +\infty$. For the second term, we get

$$\begin{aligned}
\sum_{a\in\mathcal{X}} |P_{km}(a) - P(a)| &= \sum_{a\in\mathcal{X}} \left| \sum_{\underline{x}\in\mathcal{X}^m} P_k^{(m)}(\underline{x}) P_{\underline{x}}(a) - \sum_{\underline{x}\in\mathcal{X}^m} P^{\otimes m}(\underline{x}) P_{\underline{x}}(a) \right| \\
&\le \sum_{\underline{x}\in\mathcal{X}^m} \left| P_k^{(m)}(\underline{x}) - P^{\otimes m}(\underline{x}) \right| \sum_{a\in\mathcal{X}} P_{\underline{x}}(a) = \left\| P_k^{(m)} - P^{\otimes m} \right\|_1,
\end{aligned}$$

where in the first identity we used Lemma E.1, and the last expression goes to 0 as $k \to +\infty$ by assumption.

Since we have established that the codes used in (IV.41) are of constant composition P, we get that for any $m \in \mathbb{N}$,

$$\overline{\mathrm{sc}}(W, R, P) \le \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left[R - \frac{1}{m}\chi_\alpha^\flat(W_m, P^{\otimes m})\right]. \tag{IV.42}$$

According to [44, Lemma 4.10], $\chi_\alpha^\flat(W_m, P^{\otimes m}) \ge \chi_\alpha^*(W^{\otimes m}, P^{\otimes m}) - 3\log v_{m,d}$, and hence

$$\begin{aligned}
\overline{\mathrm{sc}}(W, R, P) &\le \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left[R - \frac{1}{m}\chi_\alpha^*(W^{\otimes m}, P^{\otimes m})\right] + 3\frac{\log v_{m,d}}{m} && \text{(IV.43)} \\
&\le \sup_{\alpha>1} \frac{\alpha-1}{\alpha} \left[R - \inf_{m\in\mathbb{N}} \frac{1}{m}\chi_\alpha^*(W^{\otimes m}, P^{\otimes m})\right] + 3\frac{\log v_{m,d}}{m} && \text{(IV.44)}
\end{aligned}$$

for every $m \in \mathbb{N}$, from which

$$\overline{\mathrm{sc}}(W,R,P) \le \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[R - \inf_{m\in\mathbb{N}} \frac{1}{m}\chi_\alpha^*(W^{\otimes m},P^{\otimes m})\right]. \tag{IV.45}$$

Finally, Corollary III.22 yields the desired bound (IV.40). □

As it has been shown in [44], the strong converse exponent of a cq channel W with unconstrained coding is given by

$$\mathrm{sc}(W,R) := \underline{\mathrm{sc}}(W,R) = \overline{\mathrm{sc}}(W,R) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[R - \chi_\alpha^*(W)\right]$$

for any $R>0$, where $\chi_\alpha^*(W) = \sup_{P\in\mathcal{P}_f(\mathcal{X})} \chi_\alpha^*(W,P)$, and $\underline{\mathrm{sc}}(W,R)$ and $\overline{\mathrm{sc}}(W,R)$ are defined analogously to (IV.34)–(IV.35) by dropping the constant composition constraint. It is natural to ask whether this optimal value can be achieved, or at least arbitrarily well approximated, by constant composition codes, i.e., whether we have

$$\inf_{P\in\mathcal{P}_f(\mathcal{X})} sc(W,R,P) = sc(W,R). \tag{IV.46}$$

In view of Theorem IV.1, this is equivalent to whether

$$\inf_{P\in\mathcal{P}_f(\mathcal{X})} \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[R - \chi_\alpha^*(W,P)\right] = \sup_{\alpha>1} \inf_{P\in\mathcal{P}_f(\mathcal{X})} \frac{\alpha-1}{\alpha}\left[R - \chi_\alpha^*(W,P)\right]. \tag{IV.47}$$

By introducing the new variable $u := \frac{\alpha-1}{\alpha}$, and $f(P,u) := u\chi^*_{\frac{1}{1-u}}(W,P)$, (IV.47) can be rewritten as

$$\inf_{P\in\mathcal{P}_f(\mathcal{X})} \sup_{0<u<1} \{uR - f(P,u)\} = \sup_{0<u<1} \inf_{P\in\mathcal{P}_f(\mathcal{X})} \{uR - f(P,u)\}. \tag{IV.48}$$

One can extend f to $[0,1]$ by $f(0) := 0$ and $f(1) := \lim_{u\nearrow 1} f(u) = \chi^*_{+\infty}(W,P)$, where the latter is the P-weighted max-relative entropy [20, 51] radius of P. It is not difficult to see (similarly to [44, Lemma 5.13]) that (IV.48) is equivalent to

$$\inf_{P\in\mathcal{P}_f(\mathcal{X})} \sup_{0\le u\le 1} \{uR - f(P,u)\} = \sup_{0\le u\le 1} \inf_{P\in\mathcal{P}_f(\mathcal{X})} \{uR - f(P,u)\}. \tag{IV.49}$$

It is clear that f is a concave function of P, hence the above minimiax equality follows from Lemma II.3 if $f(P,.)$ is convex on $[0,1]$ for every $P\in\mathcal{P}_f(\mathcal{X})$. This turns out to be highly non-trivial, and has been proved very recently in [15] using interpolation techniques. The identity (IV.48) has also been stated in [15], although only as a formal identity, as the operational interpretation of $\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha^*(W,P)\right]$, (i.e., Theorem IV.1), and hence the equivalence of (IV.48) and (IV.46), had not yet been known then.

The convexity of $f(P,.)$ also plays an important role in establishing the weighted sandwiched Rényi divergence radii as generalized cutoff rates in the sense of Csiszár [16]. Following [16], for a fixed $\kappa>0$, we define the generalized κ-cutoff rate $C_\kappa(W,P)$ for a cq channel W and input distribution P as the smallest number R_0 satisfying

$$\underline{\mathrm{sc}}(W,R,P) \ge \kappa(R-R_0), \qquad R>0. \tag{IV.50}$$

The following extends the analogous result for classical channels in [16] to classical-quantum channels, and gives a direct operational interpretation of the weighted sandwiched Rényi divergence radius of a cq channel as a generalized cutoff rate.

Proposition IV.6 *For any $\kappa\in(0,1)$,*

$$C_\kappa(W,P) = \chi^*_{\frac{1}{1-\kappa}}(W,P),$$

or equivalently, for any $\alpha>1$,

$$\chi_\alpha^*(W,P) = C_{\frac{\alpha-1}{\alpha}}(W,P).$$

Proof By Theorem IV.1, we have

$$\underline{\mathrm{sc}}(R,W,P)=\sup_{0<u<1}\{uR-f(P,u)\}\ge \kappa R-f(P,\kappa)=\kappa\left(R-\frac{1}{\kappa}f(P,\kappa)\right),\qquad \kappa\in(0,1),$$

where the inequality is trivial. Since $f(P,.)$ is convex according to [15], its left and right derivatives at κ, $\partial^- f(P,.)(\kappa)$ and $\partial^+ f(P,.)(\kappa)$, exist. Obviously, for any $\partial^- f(P,.)(\kappa)\le R\le \partial^+ f(P,.)(\kappa)$,

$$\sup_{0<u<1}\{uR-f(P,u)\}=\kappa R-f(P,\kappa)=\kappa\left(R-\frac{1}{\kappa}f(P,\kappa)\right),$$

showing that $\frac{1}{\kappa}f(P,\kappa)=\chi^*_{\frac{1}{1-\kappa}}(W,P)$ is the minimal R_0 for which (IV.50) holds for all $R>0$. $\square$

Appendix A: Further properties of divergence radii

1. General divergences

Here we consider, among others, the connection between the divergence radius and the weighted divergence radius for a general divergence Δ. The following is a common generalization and simplification of several results of the same kind for various Rényi divergences [36, 43, 44, 60].

Proposition A.1 *Assume that Δ is convex and lower semi-continuous in its second argument. Then*

$$\sup_{P\in\mathcal{P}_f(S)} R_{\Delta,P}(S)=R_\Delta(S) \tag{A.1}$$

for any $S\subseteq\mathcal{B}(\mathcal{H})_+$.

Proof We have

$$\begin{aligned}R_\Delta(S)=\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{\varrho\in S}\Delta(\varrho\|\sigma)&=\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{P\in\mathcal{P}_f(S)}\sum_{\varrho\in S}P(\varrho)\,\Delta(\varrho\|\sigma)\\&=\sup_{P\in\mathcal{P}_f(S)}\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sum_{\varrho\in S}P(\varrho)\,\Delta(\varrho\|\sigma)=\sup_{P\in\mathcal{P}_f(S)}R_{\Delta,P}(S).\end{aligned}$$

The first equality above is by definition, and the second one is trivial. The third one follows from Lemma II.3 by noting that $\sum_{\varrho\in S}P(\varrho)\,\Delta(\varrho\|\sigma)$ is convex and lower semi-continuous in σ on the compact set $\mathcal{S}(\mathcal{H})$, and it is trivially concave (in fact, affine) on the convex set $\mathcal{P}_f(S)$. The last equality is again by definition. $\square$

When Δ is non-negative on $\operatorname{supp}P$ for some $P\in\mathcal{P}_f(\mathcal{B}(\mathcal{H}))$ in the sense that $\Delta(\varrho\|\sigma)\ge 0$ for all $\varrho\in\operatorname{supp}P$ and $\sigma\in\mathcal{S}(\mathcal{H})$, it is possible to define a continuous approximation between the Δ-radius and the P-weighted Δ-radius as follows. Define for every $\beta\in[1,+\infty]$ the (P,β)*-weighted divergence radius* of a set $S\subseteq\mathcal{B}(\mathcal{H})_+$ as

$$R_{\Delta,P,\beta}(S):=\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\|\Delta(.\|\sigma)\|_{P,\beta}:=\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\begin{cases}\left(\sum_{\varrho\in S}P(\varrho)\,\Delta(\varrho\|\sigma)^\beta\right)^{1/\beta}, & \beta\in[1,+\infty),\\ \sup_{\varrho\in\operatorname{supp}P}\Delta(\varrho\|\sigma), & \beta=+\infty.\end{cases} \tag{A.2}$$

Just like before, a $\sigma\in\mathcal{S}(\mathcal{H})$ is called a (P,β)*-weighted Δ-centre* if it attains the infimum in (A.2). Again, $R_{\Delta,P,\beta}(S)$ is in fact independent of S, and hence we will often drop it from the notation.

Note that $R_{\Delta,P,1}=R_{\Delta,P}$, and when S is finite and $\operatorname{supp}P=S$ then $R_{\Delta,P,+\infty}(S)=R_\Delta(S)$. In general, though, we need a further optimization to recover R_Δ from $R_{\Delta,P,+\infty}$. According to well-known properties of the β-norms,

$$R_{\Delta,P,\beta_1}\le R_{\Delta,P,\beta_2}\quad\text{when}\quad\beta_1\le\beta_2,\quad\text{and}\quad R_{\Delta,P,\beta}\nearrow R_{\Delta,P,+\infty}\quad\text{as}\quad\beta\nearrow+\infty \tag{A.3}$$

for any $P \in \mathcal{P}_f(\mathcal{B}(\mathcal{H})_+)$. Moreover, it is clear from the definitions that

$$\sup_{P \in \mathcal{P}_f(S)} R_{\Delta,P,\beta} \leq R_\Delta(S) \tag{A.4}$$

for any $S \subseteq \mathcal{B}(\mathcal{H})_+$ and $\beta \in [1, +\infty]$. Under the conditions of Proposition A.1, the above holds as an equality:

Proposition A.2 *Assume that Δ is non-negative on some $S \subseteq \mathcal{B}(\mathcal{H})_+$, and convex and lower semi-continuous in its second argument. Then*

$$\sup_{P \in \mathcal{P}_f(S)} R_{\Delta,P,\beta}(S) = R_\Delta(S) \tag{A.5}$$

for any $\beta \in [1, +\infty]$.

Proof Due to (A.4) and the monotonicity (A.3), it is enough to prove the assertion for $\beta = 1$, which has already been established in Proposition A.1. □

In the rest of the section we explore some properties of the divergence radius R_Δ. We will denote the set of Δ-centers of S by $C_\Delta(S)$.

Lemma A.3 *The Δ-radius satisfies the following simple properties.*

(i) The Δ-radius is a monotone function of S, i.e., if $S \subseteq S'$ then $R_\Delta(S) \leq R_\Delta(S')$.

(ii) If $S \subseteq S'$ and $R_\Delta(S) = R_\Delta(S')$ then $C_\Delta(S') \subseteq C_\Delta(S)$.

(iii) If Δ is quasi-convex in its first argument then $R_\Delta(S) = R_\Delta(\operatorname{conv}(S))$ and $C_\Delta(S) = C_\Delta(\operatorname{conv}(S))$.

(iv) If Δ is lower semi-continuous in its first argument then $R_\Delta(S) = R_\Delta(\overline{S})$ and $C_\Delta(S) = C_\Delta(\overline{S})$.

Proof The monotonicity in (i) is obvious.

Assume that the conditions of (ii) hold, and that σ is a Δ-centre for S' (if $C_\Delta(S) = \emptyset$ then the assertion holds trivially). Then

$$R_\Delta(S) \leq \sup_{\varrho \in S} \Delta(\varrho\|\sigma) \leq \sup_{\varrho \in S'} \Delta(\varrho\|\sigma) = R_\Delta(S') = R_\Delta(S),$$

from which $\sup_{\varrho \in S} \Delta(\varrho\|\sigma) = R_\Delta(S)$, i.e., σ is a Δ-centre of S.

As a consequence of (i) and (ii), in (iii) we only have to prove that $R_\Delta(S) \geq R_\Delta(\operatorname{conv}(S))$ and $C_\Delta(S) \subseteq C_\Delta(\operatorname{conv}(S))$, and analogously in (iv), with $\overline{S}$ in place of $\operatorname{conv}(S)$.

Assume that Δ is quasi-convex in its first argument. For any $\varrho' \in \operatorname{conv}(S)$, there exists a finitely supported probability distribution $P_{\varrho'} \in \mathcal{P}_f(S)$ such that $\varrho' = \sum_{\varrho \in S} P_{\varrho'}(\varrho)\varrho$, and hence $\Delta(\varrho'\|\sigma) \leq \max_{\varrho \in \operatorname{supp} P_{\varrho'}} \Delta(\varrho\|\sigma) \leq \sup_{\varrho \in S} \Delta(\varrho\|\sigma)$ for any $\sigma \in \mathcal{S}(\mathcal{H})$. Taking the supremum in $\varrho' \in \operatorname{conv}(S)$ and then the infimum in $\sigma \in \mathcal{S}(\mathcal{H})$ yields $R_\Delta(\operatorname{conv}(S)) \leq R_\Delta(S)$. If $\sigma \in C_\Delta(S)$ then

$$\begin{aligned} R_\Delta(\operatorname{conv}(S)) \leq \sup_{\varrho' \in \operatorname{conv}(S)} \Delta(\varrho'\|\sigma) \leq \sup_{\varrho' \in \operatorname{conv}(S)} \sup_{\varrho \in \operatorname{supp} P_{\varrho'}} \Delta(\varrho\|\sigma) = \sup_{\varrho \in S} \Delta(\varrho\|\sigma) = R_\Delta(S) \\ = R_\Delta(\operatorname{conv}(S)), \end{aligned}$$

from which $R_\Delta(\operatorname{conv}(S)) = \sup_{\varrho' \in \operatorname{conv}(S)} \Delta(\varrho'\|\sigma)$, i.e., $\sigma \in C_\Delta(\operatorname{conv}(S))$.

Assume now that Δ is l.s.c. in its first argument, and let $\varrho' \in \overline{S}$. Then there exists a sequence $(\varrho_n)_{n \in \mathbb{N}} \subseteq S$ converging to ϱ', and hence, by lower semi-continuity,

$$\Delta(\varrho'\|\sigma) \leq \liminf_{n \to +\infty} \Delta(\varrho_n\|\sigma) \leq \sup_{\varrho \in S} \Delta(\varrho\|\sigma), \qquad \sigma \in \mathcal{S}(\mathcal{H}). \tag{A.6}$$

Taking the supremum in $\varrho' \in \overline{S}$ and then the infimum in $\sigma \in \mathcal{S}(\mathcal{H})$ yields $R_\Delta(\overline{S}) \le R_\Delta(S)$. If $\sigma \in C_\Delta(S)$ then

$$R_\Delta(\overline{S}) \le \sup_{\varrho' \in \overline{S}} \Delta(\varrho' \| \sigma) \le \sup_{\varrho \in S} \Delta(\varrho \| \sigma) = R_\Delta(S) = R_\Delta(\overline{S}),$$

where the second inequality is due to (A.6). This yields that $R_\Delta(\overline{S}) = \sup_{\varrho' \in \overline{S}} \Delta(\varrho' \| \sigma)$, i.e., $\sigma \in C_\Delta(\overline{S})$. $\square$

Corollary A.4 *If Δ is quasi-convex and lower semi-continuous in its first argument then*

$$R_\Delta(S) = R_\Delta(\overline{\mathrm{conv}}(S)) \qquad \text{and} \qquad C_\Delta(S) = C_\Delta(\overline{\mathrm{conv}}(S)).$$

for any S.

As a consequence, when studying the divergence radius for a divergence with the above properties, we can often restrict our investigation to closed convex sets without loss of generality.

Proposition A.5 *Assume that Δ satisfies* (A.1) *for any $S \subseteq \mathcal{B}(\mathcal{H})_+$. Then R_Δ is continuous on monotone increasing nets of subsets of $\mathcal{B}(\mathcal{H})_+$, i.e., if $\mathcal{S} \subseteq P(\mathcal{B}(\mathcal{H})_+)$ (all the subsets of $\mathcal{B}(\mathcal{H})_+$) such that for all $S, S' \in \mathcal{S}$ there exists an $S'' \in \mathcal{S}$ with $S \cup S' \subseteq S''$ then*

$$R_\Delta(\cup \mathcal{S}) = \sup_{S \in \mathcal{S}} R_\Delta(S). \tag{A.7}$$

In particular, for any $S \subseteq \mathcal{B}(\mathcal{H})_+$,

$$R_\Delta(S) = \sup\{R_\Delta(S') : \, S' \subseteq S, \, S' \text{ finite}\}. \tag{A.8}$$

Proof It is clear from the monotonicity stated in Lemma A.3 that

$$R_\Delta(\cup \mathcal{S}) \ge \sup_{S \in \mathcal{S}} R_\Delta(S),$$

and hence we only have to prove the converse inequality. To this end, let

$$c < R_\Delta(\cup \mathcal{S}) = \sup_{P \in \mathcal{P}_f(\cup \mathcal{S})} R_{\Delta,P}(\cup \mathcal{S}),$$

where the equality holds by assumption. Then there exists a $P \in \mathcal{P}_f(\cup \mathcal{S})$ for which $c < R_{\Delta,P}(\cup \mathcal{S})$, and there exists an $S \in \mathcal{S}$ such that $\operatorname{supp} P \subseteq S$, and hence $R_{\Delta,P}(\cup \mathcal{S}) = R_{\Delta,P}(S)$. From this we get (A.7), and (A.8) follows immediately. $\square$

Remark A.6 *Theorem 3.5 in [48] states A.1 for the relative entropy. However, in their proof they use* (A.8) *without any explanation. In the proof above we assumed that A.1 holds, so the question arises whether the proof in [48] can be made complete in some other way.*

2. Generalized mutual information

For a general divergence Δ and a gcq channel $W : \mathcal{X} \to \mathcal{B}(\mathcal{H})_+$, we may define the Δ-mutual information between the classical input and the quantum output of the channel for a fixed input distribution $P \in \mathcal{P}_f(\mathcal{X})$ as

$$I_\Delta(W, P) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \Delta(\mathbb{W}(P) \| P \otimes \sigma). \tag{A.9}$$

The mutual information and the weighted divergence radius are different quantities in general (as we see below). However, they are equal if the divergence satisfies some simple properties:

Lemma A.7 *Assume that Δ is block additive and homogeneous. Then*

$$I_\Delta(W, P) = \chi_\Delta(W, P)$$

for any gcq channel W and input distribution P.

Proof We have

$$\begin{aligned}
I_\Delta(W,P) &= \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \Delta(\mathbb{W}(P)\|P\otimes\sigma)\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \Delta\left(\sum_{x\in\mathcal{X}} P(x)\,|x\rangle\langle x|\otimes W(x)\Big\|\sum_{x\in\mathcal{X}} P(x)\,|x\rangle\langle x|\otimes\sigma\right)\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} \Delta\left(P(x)\,|x\rangle\langle x|\otimes W(x)\|P(x)\,|x\rangle\langle x|\otimes\sigma\right)\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x)\,\Delta\left(|x\rangle\langle x|\otimes W(x)\|\,|x\rangle\langle x|\otimes\sigma\right)\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x)\,\Delta\left(W(x)\|\sigma\right)\\
&= \chi_\Delta(W,P),
\end{aligned}$$

where the first two equalities are by definition, the third follows by block additivity, the fourth by homogeneity, the fifth by isometric invariance (see the beginning of Section III A), and the last one is again by definition. □

In particular, all $\overline{Q}_{\alpha,z}$ are block additive and homogenous, and hence we have

Corollary A.8 *For any (α,z), any gcq channel W and input distribution $P\in\mathcal{P}_f(\mathcal{X})$, we have*

$$I_{\overline{Q}_{\alpha,z}}(W,P)=\chi_{\overline{Q}_{\alpha,z}}(W,P).$$

Note that the relative entropy $D=D_1$ is also block additive and homogeneous, and hence the corresponding mutual information and weighted D-radius coincide for any input distribution P:

$$\chi_1(W,P)=I_1(W,P). \tag{A.10}$$

Moreover, the relative entropy is even more special, as for any cq channel W, the P-weighted D-center coincides with the minimizer in (A.9), and can be explicitly given as $W(P)$. Indeed, for any state $\sigma\in\mathcal{S}(\mathcal{H})$, we have the simple identities (attributed to Donald)

$$\begin{aligned}
D\left(\mathbb{W}(P)\|P\otimes\sigma\right)=\sum_{x\in\mathcal{X}}P(x)D(W(x)\|\sigma) &= D(W(P)\|\sigma)+\sum_{x\in\mathcal{X}}P(x)D(W(x)\|W(P))\\
&= D(W(P)\|\sigma)+D\left(\mathbb{W}(P)\|P\otimes W(P)\right),
\end{aligned}$$

and the assertion follows from the strict positivity of the relative entropy on pairs of states. (Note that for this it is necessary that all the $W(x)$ are normalized on $\operatorname{supp}P$.) $\chi_1(W,P)=I_1(W,P)$ is called the *Holevo quantity* of the ensemble $\{W(x),P(x)\}_{x\in\operatorname{supp}P}$ in the quantum information theory literature.

It is easy to see that $D_{\alpha,z}$ is not block additive if $\alpha\neq 1$, and in general the $D_{\alpha,z}$ mutual information $I_{\alpha,z}(W,P):=I_{D_{\alpha,z}}(W,P)$ and the weighted $D_{\alpha,z}$ divergence radius $\chi_{\alpha,z}(W,P)$ are different quantities. However, they can be related by a simple inequality, as

$$\begin{aligned}
I_{\alpha,z}(W,P) &= \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\frac{1}{\alpha-1}\log Q_{\alpha,z}\left(\sum_{x\in\mathcal{X}}P(x)\,|x\rangle\langle x|\otimes W(x)\Big\|\sum_{x\in\mathcal{X}}P(x)\,|x\rangle\langle x|\otimes\sigma\right) && \text{(A.11)}\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}}P(x)Q_{\alpha,z}(W(x)\|\sigma) && \text{(A.12)}\\
&\le \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sum_{x\in\mathcal{X}}P(x)\frac{1}{\alpha-1}\log Q_{\alpha,z}(W(x)\|\sigma) \qquad (\text{if } \alpha\in(0,1))\\
&= \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sum_{x\in\mathcal{X}}P(x)D_{\alpha,z}(W(x)\|\sigma)\\
&= \chi_{\alpha,z}(W,P),
\end{aligned}$$

where the second equality follows as in the proof of Lemma A.7, and the inequality is due to the concavity of the logarithm. Obviously, the inequality holds in the opposite direction when $\alpha > 1$.

The above inequality is in general strict. As an example, consider a noiseless channel as in Example III.17 with a non-uniform input distribution P. Then we have

$$\begin{aligned} I_{\alpha,z}(W,P) &\le D_{\alpha,z}(\mathbb{W}(P)\|P\otimes W(P)) = \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} P(x)Q_{\alpha,z}(W(x)\|W(P)) \\ &= \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} P(x)P(x)^{1-\alpha} < \sum_{x\in\mathcal{X}} P(x)\frac{1}{\alpha-1}\log P(x)^{1-\alpha} \\ &= H(P) = \chi_{\alpha,z}(W,P), \end{aligned}$$

where the first inequality is by definition, the strict inequality follows from the strict concavity of the logarithm, and the last equality is due to Example III.17.

While the mutual information $I_{\alpha,z}$ for $D_{\alpha,z}$ differs from the weighted channel radius $\chi_{\alpha,z}$ for $D_{\alpha,z}$, it is a simple function of the weighted channel radius for $Q_{\alpha,z}$ (and thus, by Corollary A.8, also of the mutual information for $Q_{\alpha,z}$); indeed, moving the infimum over σ behind the logarithm in (A.11) and (A.12), respectively, yields

$$I_{\alpha,z}(W,P) = \frac{1}{\alpha-1}\log s(\alpha)I_{\overline{Q}_{\alpha,z}}(W,P) = \frac{1}{\alpha-1}\log s(\alpha)\chi_{\overline{Q}_{\alpha,z}}(W,P), \tag{A.13}$$

Moreover, the difference between $I_{\alpha,z}$ and $D_{\alpha,z}$ vanishes after optimizing over the input distribution:

Corollary A.9 *If (α,z) are such that $D_{\alpha,z}$ is convex in its second argument then*

$$\sup_{P\in\mathcal{P}_f(\mathcal{X})} \chi_{\alpha,z}(W,P) = R_{D_{\alpha,z}}(\operatorname{ran} W), \tag{A.14}$$

and if (α,z) are such that $\overline{Q}_{\alpha,z}$ is convex in its second argument and $\alpha\neq 1$ then

$$\sup_{P\in\mathcal{P}_f(\mathcal{X})} I_{\alpha,z}(W,P) = R_{D_{\alpha,z}}(\operatorname{ran} W) \tag{A.15}$$

for any gcq channel $W:\mathcal{X}\to\mathcal{S}(\mathcal{H})$.

Proof The identity in (A.14) is immediate from Proposition A.1. The identity in (A.15) follows as

$$\begin{aligned} \sup_{P\in\mathcal{P}_f(\mathcal{X})} I_{\alpha,z}(W,P) &= \frac{1}{\alpha-1}\log s(\alpha)\sup_{P\in\mathcal{P}_f(\mathcal{X})}\chi_{\overline{Q}_{\alpha,z}}(W,P) \\ &= \frac{1}{\alpha-1}\log s(\alpha)R_{\overline{Q}_{\alpha,z}}(\operatorname{ran} W) \\ &= \frac{1}{\alpha-1}\log s(\alpha)\inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{x\in\mathcal{X}}\overline{Q}_{\alpha,z}(W(x)\|\sigma) \\ &= \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{x\in\mathcal{X}}\frac{1}{\alpha-1}\log s(\alpha)\overline{Q}_{\alpha,z}(W(x)\|\sigma) \\ &= \inf_{\sigma\in\mathcal{S}(\mathcal{H})}\sup_{x\in\mathcal{X}} D_{\alpha,z}(W(x)\|\sigma) \\ &= R_{D_{\alpha,z}}(\operatorname{ran} W) \end{aligned}$$

where the first equality is due to (A.13), the second one is due to Proposition A.1, and the rest are obvious. □

Remark A.10 *The above proof method is due to Csiszár [16], and extends various prior results for different quantum Rényi divergences and ranges of parameters (α,z) in [36, 43, 44, 60]. The special case*

$$\sup_{P\in\mathcal{P}_f(\mathcal{X})}\chi_1(W,P) = R_D(\operatorname{ran} W) = \sup_{P\in\mathcal{P}_f(\mathcal{X})} I_1(W,P)$$

was proved by different methods in [48, 54, 57].

Remark A.11 *Mutual information quantifies the amount of correlation in a bipartite state by measuring its distance from the set of uncorrelated states. Following this general idea, one may give two alternative definitions of mutual information between the input and the output of a gcq channel W for a fixed input distribution P using a general divergence Δ: the Δ-"distance" of $\mathbb{W}(P)$ from the product of its marginals $P \otimes W(P)$:*

$$I_{\Delta}^{(2)}(W,P) := \Delta\left(\mathbb{W}(P)\|P \otimes W(P)\right),$$

or the Δ-"distance" of $\mathbb{W}(P)$ from the set of uncorrelated states:

$$I_{\Delta}^{(1)}(W,P) := \inf_{\omega\in\mathcal{S}(\mathcal{H}_{\mathcal{X}}),\,\sigma\in\mathcal{S}(\mathcal{H})} \Delta\left(\mathbb{W}(P)\|\omega\otimes\sigma\right).$$

Both of these options seem more natural than the curiously asymmetric optimization in (A.9)*, and one might wonder why it is nevertheless the capacity $\chi_{\alpha}^{*}(W) := \sup_{P\in\mathcal{P}_f(\mathcal{X})} I_{\alpha}^{*}(W,P)$ corresponding to the version* (A.9) *(for the sandwiched Rényi divergence) that seems to obtain an operational significance in channel coding, according to [16, 44].*

There seems to be at least two different resolutions of this question: First, as the more detailed analysis of constant composition channel coding shows (according to [16] and our Corollary IV.6), it is in fact not the mutual information, but the weighted divergence radius that obtains a natural operational interpretation in channel coding; the mutual information only enters the picture because its optimized version over all input distributions "happens to" coincide with the optimized version of the divergence radius. In this context it would be interesting to know if any of the trivial inequalities

$$\sup_{p\in\mathcal{P}_f(\mathcal{X})} I_{\Delta}^{(1)}(W,P) \le \sup_{p\in\mathcal{P}_f(\mathcal{X})} I_{\Delta}(W,P) \le \sup_{p\in\mathcal{P}_f(\mathcal{X})} I_{\Delta}^{(2)}(W,P)$$

holds as an equality for general W and P (at least for $\Delta = D_{\alpha}^{*}$ with $\alpha > 1$).

Second, according to (A.13)*, the mutual information for Rényi divergences is simply a function of another weighted divergence radius, corresponding to the $\overline{Q}$ quantities rather than the Rényi divergences, and the optimization over σ is simply the optimization over the candidates for the weighted divergences centers, which is very natural, and in this context the problem of asymmetry does not even makes sense.*

The same arguments as in Sections III C and III D yield the following statements, and hence we omit their proofs.

Lemma A.12 *Let σ be a P-weighted $\overline{Q}_{\alpha,z}$ center for W.*

(1) If (α,z) is such that $\overline{Q}_{\alpha,z}$ is quasi-convex in its second argument then $\sigma^0 \le W(P)^0$.

(2) If $\alpha > 1$ or $\alpha\in(0,1)$ and $1-\alpha < z < +\infty$ then $W(P)^0 \le \sigma^0$.

Let us define $\Gamma_{\overline{Q}}$ to be the set of (α,z) values such that for any gcq channel W and any input probability distribution P, any P-weighted $\overline{Q}_{\alpha,z}$ center σ for W satisfies $\sigma^0 = W(P)^0$. Then Lemmas III.5, III.7, and A.12 yield

$$\Gamma_{\overline{Q}} \supseteq \{(\alpha,z):\ \alpha\in(0,1),\, 1-\alpha < z + \infty\} \cup \{(\alpha,z):\ \alpha > 1,\, z \ge \max\{\alpha/2,\alpha\}\}.$$

Proposition A.13 *Assume that $(\alpha,z)\in\Gamma_{\overline{Q}}$ are such that $\overline{Q}_{\alpha,z}$ is convex in its second variable. Then σ is a P-weighted $\overline{Q}_{\alpha,z}$ center for W if and only if it is a fixed point of the map*

$$\Phi_{W,P,\overline{Q}_{\alpha,z}}(\sigma) := \frac{1}{\tau}\sum_{x\in\mathcal{X}} P(x)\left(\sigma^{\frac{1-\alpha}{2z}} W(x)^{\frac{\alpha}{z}} \sigma^{\frac{1-\alpha}{2z}}\right)^{z}, \tag{A.16}$$

$\sigma\in\mathcal{S}_{W,P}(\mathcal{H})_{++}$, where τ is the normalization factor

$$\tau := \sum_{x\in\mathcal{X}} P(x) Q_{\alpha,z}(W(x)\|\sigma).$$

Remark A.14 *Note that the fixed point equations in* (A.16) *and* (III.25) *look very similar, except that the normalization of sigma is obtained "globally" in the former, and for each x individually in the latter.*

Proposition A.15 *(Additivity of the $D_{\alpha,z}$ mutual information) Let $W^{(1)} : \mathcal{X}^{(1)} \to \mathcal{S}(\mathcal{H}^{(1)})$ and $W^{(2)} : \mathcal{X}^{(2)} \to \mathcal{S}(\mathcal{H}^{(2)})$ be gcq channels, and $P^{(i)} \in \mathcal{P}_f(\mathcal{X}^{(i)})$, $i = 1, 2$, be input distributions. Assume, moreover, that α and z satisfy the conditions of Proposition A.13. Then*

$$\chi_{\overline{Q}_{\alpha,z}}\left(W^{(1)} \otimes W^{(2)}, P^{(1)} \otimes P^{(2)}\right) = \chi_{\overline{Q}_{\alpha,z}}\left(W^{(1)}, P^{(1)}\right) \cdot \chi_{\overline{Q}_{\alpha,z}}\left(W^{(2)}, P^{(2)}\right), \tag{A.17}$$

$$I_{\alpha,z}\left(W^{(1)} \otimes W^{(2)}, P^{(1)} \otimes P^{(2)}\right) = I_{\alpha,z}\left(W^{(1)}, P^{(1)}\right) + I_{\alpha,z}\left(W^{(2)}, P^{(2)}\right). \tag{A.18}$$

Remark A.16 *In [29], a generalized notion of mutual information was studied, where the bipartite state need not be classical-quantum, and the first marginal of the second argument is fixed but not necessarily equal to the first marginal of the first argument. For sandwiched Rényi divergences a characterization of the optimal state in terms of a fixed point equation, as well as the additivity of the generalized mutual information was obtained. Our approach is essentially the same as that of [29], and Propositions A.13 and A.15 are special cases of the results of [29] when $z = \alpha$.*

Remark A.17 *As we have mentioned above, the relative entropy ($\alpha = 1$) is special in that for cq channels the minimizer for the mutual information can be explicitly determined (as $W(P)$), and hence also the mutal information can be given by an explicit formula. The only other known family of quantum Rényi divergences with these properties are the Petz-type Rényi divergences (corresponding to $z = 1$). Indeed, it is easy to verify that in this case we have the classical-quantum Sibson identity [36, Lemma 2.2] $D_{\alpha,1}(\mathbb{W}(P)\|P \otimes \sigma) = D_{\alpha,1}(\sigma_{\alpha,1}\|\sigma) + \frac{1}{\alpha-1}\log(\operatorname{Tr}\omega(\alpha))^{\alpha}$, where*

$$\omega(\alpha) := \left(\sum_x P(x)W(x)^{\alpha}\right)^{1/\alpha}, \qquad \textit{and} \qquad \sigma_{\alpha,1} := \omega(\alpha)/\operatorname{Tr}\omega(\alpha).$$

As a consequence, $\sigma_{\alpha,1}$ is the unique P-weighted $\overline{Q}_{\alpha,1}$ center, and

$$\chi_{\overline{Q}_{\alpha,1}}(W, P) = \sum_{x \in \mathcal{X}} P(x)\overline{Q}_{\alpha,1}(W(x)\|\sigma_{\alpha,1}) = s(\alpha)(\operatorname{Tr}\omega(\alpha))^{\alpha} = s(\alpha)\left(\operatorname{Tr}\left(\sum_x P(x)W(x)^{\alpha}\right)^{1/\alpha}\right)^{\alpha}.$$

3. PSD divergence center and radius

Note that while we defined the divergence radius and center for an arbitrary non-empty subset S of $\mathcal{B}(\mathcal{H})_+$, in the definition of the center we restricted to density operators, i.e., PSD operators with trace 1. This is a natural choice when the set S consists of density operators itself, and, even more importantly, it leads to operationally relevant information measures as our main result, Theorem IV.1 shows.

Moreover, restricting the set of possible divergence centers to density operators may be operationally motivated even when the elements of the set S are not normalized to have trace 1. Indeed, the optimal success probability of discriminating quantum states $\varrho_1, \ldots, \varrho_r$ with prior probabilities $p_1, \ldots, p_r$ can be expressed as

$$P_s^* = \exp\left(R_{D_\infty^*}(\{p_1\varrho_1, \ldots, p_r\varrho_r\})\right),$$

where $D_\infty^* := \lim_{\alpha\to+\infty} D_\alpha^* = \lim_{\alpha\to+\infty} D_{\alpha,\alpha}$ is the max-relative entropy [20, 45, 51]. This follows by simply rewriting the optimal success probability $P_s^* := \max\{\sum_{i=1}^r p_i \operatorname{Tr}\varrho_i M_i : (M_i)_{i=1}^r \text{ POVM}\}$ using the duality of linear programming, as was done in [61] (see also [37]).

In this section we consider the alternative approach where the divergence center is allowed to be a general PSD operator. This is largely motivated by recent investigations in matrix analysis regarding various concepts of multi-variate geometric matrix means, in particular, by the approach of [11]. We comment on this in more detail at the end of the section.

For a general divergence Δ, we define the *P-weighted PSD Δ-radius* as

$$\widetilde{R}_{\Delta,P}(S) := \inf_{\sigma\in\mathcal{B}(\mathcal{H})_+} \sum_{\varrho\in S} P(\varrho)\,\Delta(\varrho\|\sigma),$$

and we call any $\sigma \in \mathcal{B}(\mathcal{H})_+$ that attains the above infimum a *P-weighted PSD Δ-center*. For a gcq channel $W : \mathcal{X} \to \mathcal{B}(\mathcal{H})_+$ and $P \in \mathcal{P}_f(\mathcal{X})$ we define P-weighted PSD Δ-radius of W as before:

$$\widetilde{\chi}_\Delta(W,P) := \tilde{R}_{\Delta,P\circ W^{-1}}(\operatorname{ran} W).$$

Any PSD minimizer in the definition of $\tilde{R}_{\Delta,P\circ W^{-1}}(\operatorname{ran} W)$ will be called a P-weighted PSD Δ-center for the channel W.

It is easy to see that these quantities are meaningless for the Rényi divergences considered before, as we always have

$$\widetilde{R}_{D_{\alpha,z},P}(S) = -\infty, \qquad \widetilde{R}_{\overline{Q}_{\alpha,z},P}(S) = \begin{cases} -\infty, & \alpha\in(0,1),\\ 0, & \alpha>1,\end{cases}$$

due to the scaling laws

$$D_{\alpha,z}(\varrho\|\lambda\sigma) = D_{\alpha,z}(\varrho\|\sigma) - \log\lambda, \qquad \overline{Q}_{\alpha,z}(\varrho\|\lambda\sigma) = \lambda^{1-\alpha}\overline{Q}_{\alpha,z}(\varrho\|\sigma), \qquad \lambda\in(0,+\infty).$$

Hence, in order to make sense of the PSD divergence radius, it seems necessary to modify the notion of Rényi divergence for PSD operators.

We consider two such options, motivated by Proposition A.23 in Section A 4. One is a simple rescaling of $D_{\alpha,z}$, defined as

$$\begin{aligned}\widehat{D}_{\alpha,z}(\varrho\|\sigma) &:= \frac{1}{\alpha-1}\log\frac{Q_{\alpha,z}(\varrho\|\sigma)}{(\operatorname{Tr}\varrho)^\alpha(\operatorname{Tr}\sigma)^{1-\alpha}} = \frac{1}{\alpha-1}\log Q_{\alpha,z}(\varrho\|\sigma) - \frac{\alpha}{\alpha-1}\log\operatorname{Tr}\varrho + \log\operatorname{Tr}\sigma\\ &= D_{\alpha,z} - \log\operatorname{Tr}\varrho + \log\operatorname{Tr}\sigma = D_{\alpha,z}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right).\end{aligned}$$

The limit $\alpha\to1$ yields

$$\widehat{D}_1(\varrho\|\sigma) := \lim_{\alpha\to1}\widehat{D}_{\alpha,z(\alpha)}(\varrho\|\sigma) = D_1(\varrho\|\sigma) + \log\operatorname{Tr}\varrho - \log\operatorname{Tr}\sigma = D\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right)$$

for any function $\alpha\mapsto z(\alpha)$ that is continuously differentiable in a neighbourhood of 1, on which $z(\alpha)\neq0$ [41]. Obviously, $\widehat{D}_{\alpha,z}(\varrho\|\sigma) = D_{\alpha,z}(\varrho\|\sigma)$ for any pair of states ϱ,σ. Note that $\widehat{D}_{\alpha,z}$ is a *projective divergence*, i.e., $\widehat{D}_{\alpha,z}(\lambda\varrho\|\sigma) = \widehat{D}_{\alpha,z}(\varrho\|\lambda\sigma) = \widehat{D}_{\alpha,z}(\varrho\|\sigma)$ for any $\varrho,\sigma\in\mathcal{B}(\mathcal{H})_+$ and $\lambda\in(0,+\infty)$. As an immediate consequence, we have

$$\widetilde{R}_{\widehat{D}_{\alpha,z},P}(S) = R_{D_{\alpha,z},P}(S) - \sum_{\varrho} P(\varrho)\log\operatorname{Tr}\varrho,$$

i.e., we essentially recover the previously considered concept of the $D_{\alpha,z}$-radius, apart from an uninteresting term. Obviously, if σ is a PSD $\widehat{D}_{\alpha,z}$-center then so is $\lambda\sigma$ for all $\lambda\in(0,+\infty)$. Moreover, the normalized PSD $\widehat{D}_{\alpha,z}$-centers can be characterized by the fixed point equation in Theorem III.14 for the (α,z) pairs for which Theorem III.14 holds.

The other option we consider is a modification of the Tsallis quantum Rényi divergences, or Tsallis relative entropies, defined as

$$T_{\alpha,z}(\varrho\|\sigma) := \frac{1}{1-\alpha}\left(\alpha\operatorname{Tr}\varrho + (1-\alpha)\operatorname{Tr}\sigma - Q_{\alpha,z}(\varrho\|\sigma)\right).$$

We will call these quantities Tsallis (α,z)-divergences. The limit $\alpha\to1$ yields

$$T_1(\varrho\|\sigma) := \lim_{\alpha\to1} T_{\alpha,z(\alpha)}(\varrho\|\sigma) = D(\varrho\|\sigma) - \operatorname{Tr}\varrho + \operatorname{Tr}\sigma,$$

for any function $\alpha\mapsto z(\alpha)$ that is continuously differentiable in a neighbourhood of 1, on which $z(\alpha)\neq0$ [41].

Remark A.18 *The usual way to define the quantum Tsallis relative entropy is*

$$T'_\alpha(\varrho\|\sigma) := \frac{1}{1-\alpha}\left(\operatorname{Tr}\varrho - Q_{\alpha,1}(\varrho\|\sigma)\right).$$

Note that T'_α coincides with $T_{\alpha,1}$ on pairs of density operators but, unlike $T_{\alpha,1}$, T'_α is not positive on pairs of PSD operators for any (α, z). Moreover, the PSD divergence radius problem is trivial for these quantities, as

$$\widetilde{R}_{T'_{\alpha,z}}(S) = \begin{cases} -\infty, & \alpha \in (0,1), \\ \frac{1}{1-\alpha}\sum_\varrho P(\varrho)\varrho, & \alpha > 1. \end{cases}$$

We consider the PSD divergence center for $T_{\alpha,z}$ in the gcq channel formalism, for easier comparison with Theorem III.14 and Proposition A.13.

Proposition A.19 *(i) For any $\alpha \in (0,+\infty)\setminus\{1\}$ and $z \in (0,+\infty)$,*

$$\widetilde{\chi}_{T_{\alpha,z}}(W,P) = \frac{\alpha}{1-\alpha}\left[\operatorname{Tr} W(P) - \left(s(\alpha)\chi_{\overline{Q}_{\alpha,z}}(W,P)\right)^{1/\alpha}\right]. \tag{A.19}$$

(ii) If σ is a P-weighted PSD $T_{\alpha,z}$-center for W then $\operatorname{Tr}\sigma = \sum_x P(x)Q_{\alpha,z}(W(x)\|\sigma)$, and $\overline{\sigma} = \sigma/\operatorname{Tr}\sigma$ is a P-weighted $\overline{Q}_{\alpha,z}$-center for W.

(iii) If $\overline{\sigma} \in \mathcal{S}(\mathcal{H})$ is a P-weighted $\overline{Q}_{\alpha,z}$-center for W then

$$\sigma := \left(\sum_x P(x)Q_{\alpha,z}(W(x)\|\overline{\sigma})\right)^{1/\alpha}\overline{\sigma}$$

is a P-weighted PSD $T_{\alpha,z}$-center for W.

(iv) If (α, z) satisfy the conditions of Proposition A.13 then σ is a P-weighted PSD $T_{\alpha,z}$-center for W if and only if it is a solution of the fixed point equation

$$\sigma = \sum_{x\in\mathcal{X}} P(x)\left(\sigma^{\frac{1-\alpha}{2z}}W(x)^{\frac{\alpha}{z}}\sigma^{\frac{1-\alpha}{2z}}\right)^z. \tag{A.20}$$

Proof We have

$$\begin{aligned}
\widetilde{\chi}_{T_{\alpha,z}}(W,P) &= \inf_{\sigma\in\mathcal{B}(\mathcal{H})_+}\sum_x P(x)T_{\alpha,z}(W(x)\|\sigma) \\
&= \frac{\alpha}{1-\alpha}\operatorname{Tr} W(P) + \inf_{\sigma\in\mathcal{B}(\mathcal{H})_+}\left[\operatorname{Tr}\sigma - \frac{1}{1-\alpha}\sum_x P(x)Q_{\alpha,z}(W(x)\|\sigma)\right] \\
&= \frac{\alpha}{1-\alpha}\operatorname{Tr} W(P) + \inf_{\overline{\sigma}\in\mathcal{S}(\mathcal{H})}\inf_{\lambda>0}\left[\lambda - \frac{1}{1-\alpha}\lambda^{1-\alpha}\sum_x P(x)Q_{\alpha,z}(W(x)\|\overline{\sigma})\right].
\end{aligned}$$

Differentiating w.r.t. λ yields that the optimal λ is

$$\lambda = \left(\sum_x P(x)Q_{\alpha,z}(W(x)\|\overline{\sigma})\right)^{1/\alpha} = \sum_x P(x)Q_{\alpha,z}(W(x)\|\sigma), \tag{A.21}$$

with $\sigma := \lambda\overline{\sigma}$. Writing it back to the previous equation, we get

$$\widetilde{\chi}_{T_{\alpha,z}}(W,P) = \frac{\alpha}{1-\alpha}\operatorname{Tr} W(P) + \inf_{\overline{\sigma}\in\mathcal{S}(\mathcal{H})}\frac{\alpha}{\alpha-1}\left(\sum_x P(x)Q_{\alpha,z}(W(x)\|\overline{\sigma})\right)^{1/\alpha},$$

which is exactly (A.19). This proves (i), and (ii) and (iii) are clear from the above argument. Finally, (iv) is immediate from the above and Proposition A.13. □

Remark A.20 *Note that (i)–(iii) of Proposition A.19 hold true for any pair of divergences* T^q_α *and* Q^q_α *related as*

$$T^q_\alpha(\varrho\|\sigma) = \frac{1}{1-\alpha}\left(\alpha \operatorname{Tr}\varrho + (1-\alpha)\operatorname{Tr}\sigma - Q^q_\alpha(\varrho\|\sigma)\right),$$

provided that Q^q_α *is a non-negative divergence that satisfies the scaling property* $Q^q_\alpha(\varrho\|\lambda\sigma) = \lambda^{\alpha-1}Q^q_\alpha(\varrho\|\sigma)$*, and that there exists a* $\sigma \in \mathcal{B}(\mathcal{H})_+$ *such that* $Q^q_\alpha(W(x)\|\sigma) < +\infty$ *for all* $x \in \operatorname{supp} P$*.*

The above Proposition extends Theorem 8 in [12], where the fixed point characterization (A.20) was obtained in the case $z = \alpha = 1/2$, by a somewhat different proof than above. The existence of a solution of the fixed point equation (A.20) was studied in [1, Theorem 6.1] for the case $z = \alpha = 1/2$, and their proof extends without alteration for more general (α, z) pairs as below.

Proposition A.21 *Let* $\alpha \in (0,1)$*. If there exist positive numbers* λ, η*, such that* $W(x) \in [\lambda I, \eta I] := \{A \in \mathcal{B}(\mathcal{H})_+ : \lambda I \le A \le \eta I\}$ *for all* $x \in \operatorname{supp} P$ *then the fixed point equation* (A.20) *has a solution, which is also in* $[\lambda I, \eta I]$*.*

Proof It is easy to see that the map on the RHS of (A.20) maps the compact convex set $[\lambda I, \eta I]$ into itself, and hence, by Brouwer's fixed point theorem, it has a fixed point. □

Remark A.22 *The case of the Petz-type Tsallis divergences (*$z = 1$*) is again special in that the weighted PSD* $T_{\alpha,1}$ *center and radius can be given explicitly. Indeed, by Remark A.17 and Proposition A.19, the unique* $T_{\alpha,1}$ *center is given by*

$$\tilde{\sigma}_{\alpha,1} = \left(\sum_x P(x) Q_{\alpha,1}(W(x)\|\sigma_{\alpha,1})\right)^{1/\alpha} \frac{\omega(\alpha)}{\operatorname{Tr}\omega(\alpha)} = \omega(\alpha) = \left(\sum_x P(x) W(x)^\alpha\right)^{1/\alpha},$$

and

$$\widetilde{\chi}_{T_{\alpha,z}}(W,P) = \frac{\alpha}{\alpha-1}\left[\operatorname{Tr} W(P) - \operatorname{Tr}\left(\sum_x P(x)W(x)^\alpha\right)^{1/\alpha}\right].$$

Let us now explain how the above considerations are related to recent research in matrix analysis. First, we recall that if $\varrho_1, \dots \varrho_r$ are points in a metric space (M,d), and $(p_i)_{i=1}^r$ is a probability distribution then the *p-weighted Fréchet variance* is $\inf_{\sigma\in M}\sum_i p_i d^2(\varrho_i,\sigma)$, and if σ attains this infimum then it is called a *p-weighted Fréchet mean* of $\varrho_1, \dots \varrho_r$. These are analogous to our notions of weighted divergence radius and center. Note, however, that we only defined divergences on PSD operators, which is more restrictive than the setting of the Fréchet means, while it is also more general in the sense that a divergence does not need to be the square of a metric. (Although some properties of the relative entropy are reminiscent to those of a squared Euclidean distance.) In particular, if f is an injective continuous function on an interval $M \subseteq \mathbb{R}$, then $d(x,y) := |f(x) - f(y)|$ is a metric on M, and a straightforward computation shows that for any $\varrho_1, \dots, \varrho_r \in M$, and any probability distribution $(p_i)_{i=1}^r$, there is a unique Fréchet mean, which is exactly the generalized f-mean $f^{-1}\left(\sum_{i=1}^r p_i f(\varrho_i)\right)$, and the Fréchet variance is $\sum_i p_i f(\varrho_i)^2 - \left(\sum_i p_i f(\varrho_i)\right)^2$.

Of particular importance to us are the cases $M := (0,+\infty)$, $f(t) := t^\alpha$, which yields the *α-power mean* $\left(\sum_i p_i \varrho_i^\alpha\right)^{1/\alpha}$, and $f(t) := \log t$ on the same set, which yields the *geometric mean* $\prod_i \varrho_i^{p_i}$. These special cases are closely related to each other, as the geometric mean can be recovered from the α-power mean in the limit $\alpha \searrow 0$,

$$\lim_{\alpha\searrow 0}\left(\sum_i p_i \varrho_i^\alpha\right)^{1/\alpha} = \prod_i \varrho_i^{p_i},$$

while the α-power mean is the unique solution of the fixed point equation

$$\sigma = \sum_i p_i \mathcal{G}_\alpha(\varrho_i \| \sigma), \tag{A.22}$$

where $\mathcal{G}_\alpha(\varrho_i\|\sigma) := \varrho_i^\alpha \sigma_i^{1-\alpha}$ is the α-geometric mean of ϱ_i and σ.

There are various ways to extend the α-geometric mean to operators, and these are closely related to different definitions of Rényi divergences. Indeed, for every (α, z), the quantity

$$\mathcal{G}_{\alpha,z}(\varrho\|\sigma) := \left(\sigma^{\frac{1-\alpha}{2z}} \varrho^{\frac{\alpha}{z}} \sigma^{\frac{1-\alpha}{2z}}\right)^z$$

is well-defined if $\alpha \in (0,1)$, or $\alpha > 1$ and $\varrho^0 \le \sigma^0$, and it reduces to the α-geometric mean for scalars for every $z > 0$. This is related to the (α, z)-Rényi divergences via $Q_{\alpha,z}(\varrho\|\sigma) = \operatorname{Tr} \mathcal{G}_{\alpha,z}(\varrho\|\sigma)$. The fixed point equation (A.20) is an exact analogy of (A.22), and hence we may call the solution of (A.20) the P-weighted (α, z)-power mean of $(W(x))_{x\in\mathcal{X}}$, and denote it as $\mathcal{P}_{\alpha,z}(W, P)$, provided that it exists and is unique. (From here we switch to the gcq channel formalism for better comparison with the preceding part of the paper, but this is equivalent to considering subsets of PSD operators and probability distributions on them.) That is, the P-weighted (α, z)-power mean is nothing but the P-weighted $T_{\alpha,z}$-center for W (or, in other terminology, the $P \circ W^{-1}$-weighted $T_{\alpha,z}$-center of $\operatorname{ran} W$). In general, there is no explicit formula for it; the case $z = 1$, discussed in Remark A.22, is an exception, and the resulting formula is probably the most straightforward extension of the α-power mean from numbers to operators. Following the ideas of [40], a family of multivariate geometric means for operators may be defined as $\mathcal{G}(W, P) := \lim_{\alpha\searrow 0} \mathcal{G}_{\alpha,z(\alpha)}(W, P)$, where $z(\alpha)$ is some well-behaved funcion of α. It is an interesting question if this limit always exists, how it depends on the choice of $z(\alpha)$, and how it relates to other notions of multivariate geometric means.

Probably the most studied notion of α-geometric mean for a pair of operators is the *Kubo-Ando geometric mean* [38]

$$\mathcal{G}_\alpha^{\max}(\varrho\|\sigma) := \sigma^{1/2}\left(\sigma^{-1/2}\varrho\sigma^{-1/2}\right)^\alpha \sigma^{1/2},$$

introduced by Kubo and Ando for $\alpha \in [0, 1]$. This is also a special instance of a *maximal f-divergence* [32, 42] (with a minus sign for $\alpha \in (0, 1)$), and can be extended to $\alpha > 1$. It gives rise to the *maximal Rényi divergence* [56] $D_\alpha^{\max}$ and the maximal Tsallis divergence $T_\alpha^{\max}$ via

$$D_\alpha^{\max}(\varrho\|\sigma) := \frac{1}{\alpha-1}\log\frac{1}{\operatorname{Tr}\varrho} Q_\alpha^{\max}(\varrho\|\sigma),$$
$$T_\alpha^{\max}(\varrho\|\sigma) := \frac{1}{1-\alpha}\left(\alpha \operatorname{Tr}\varrho + (1-\alpha)\operatorname{Tr}\sigma - Q_\alpha^{\max}(\varrho\|\sigma)\right),$$

where $Q_\alpha^{\max}(\varrho\|\sigma) := \operatorname{Tr}\mathcal{G}_\alpha^{\max}(\varrho\|\sigma)$ for positive definite ϱ and σ, and it is extended to general PSD operators via the smoothing procedure in (III.17). (In fact, $T_\alpha^{\max}$ is also a maximal f-divergence, corresponding to the convex function $f(t) = \frac{1}{1-\alpha}(\alpha t + (1-\alpha) - t^\alpha)$, which is operator convex if and only if $\alpha \in [0, 2]$.) The positive version of $D_\alpha^{\max}$ can be defined again as $\widehat{D}_\alpha^{\max}(\varrho\|\sigma) := D_\alpha^{\max}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right)$. Lim and Pálfia [40] showed that when all $W(x)$ are positive definite, the fixed point equation

$$\sigma = \sum_x P(x)\mathcal{G}_\alpha^{\max}(W(x)\|\sigma)$$

has a unique solution, which we will call the *max α-power mean* and denote it as $\mathcal{P}_\alpha^{\max}(W, P)$. Moreover,

$$\lim_{\alpha\searrow 0}\mathcal{P}_\alpha^{\max}(W, P) = \mathcal{G}_K(W, P),$$

where the latter is the *Karcher mean*, which, in our terminology, is nothing else but the $P \circ W^{-1}$-weighted PSD $(D_\infty^*)^2$-center of $\operatorname{ran} W$, i.e.,

$$\mathcal{G}_K(W, P) = \operatorname{argmin}_{\sigma\in\mathcal{B}(\mathcal{H})_+}\sum_x P(x)(D_\infty^*(W(x)\|\sigma))^2.$$

By Remark A.20, (i)–(iii) of Proposition A.19 hold true for the pair $T_\alpha^{\max}$ and $Q_\alpha^{\max}$. However, it has been shown in [50] that (iv) of Proposition A.19 is no longer true in this case. More precisely, an example is shown in [50] for W and P for which the max 1/2 power mean does not coincide with the P-weighted $T_{1/2}^{\max}$-center for W.

4. Positive Rényi divergences

In this section we establish the non-negativity of $\widehat{D}_{\alpha,z}$ and $T_{\alpha,z}$ for all $\alpha \in (0,+\infty)\setminus\{1\}$ and $z \in (0,+\infty]$.

Proposition A.23 *For every $\varrho,\sigma \in \mathcal{B}(\mathcal{H})_+$, and every $z \in (0,+\infty]$, we have*

$$Q_{\alpha,z}(\varrho\|\sigma) \le (\operatorname{Tr}\varrho)^\alpha(\operatorname{Tr}\sigma)^{1-\alpha} \le \alpha\operatorname{Tr}\varrho + (1-\alpha)\operatorname{Tr}\sigma, \qquad \alpha\in(0,1), \tag{A.23}$$

$$Q_{\alpha,z}(\varrho\|\sigma) \ge (\operatorname{Tr}\varrho)^\alpha(\operatorname{Tr}\sigma)^{1-\alpha} \ge \alpha\operatorname{Tr}\varrho + (1-\alpha)\operatorname{Tr}\sigma, \qquad \alpha>1, \tag{A.24}$$

or equivivalently, $\widehat{D}_{\alpha,z}(\varrho\|\sigma)\ge 0$ and $T_{\alpha,z}(\varrho\|\sigma)\ge 0$ for all $\alpha\in(0,+\infty)\setminus\{1\}$ and $z\in(0,+\infty]$.

Proof Note that the bounds are independent of z; in particular, it is enough to prove them for all finite z, and the case $z=+\infty$ follows by taking the limit $z\to+\infty$. Note also that the second inequalities in (A.23)–(A.24) follow by the trivial identity $x^\alpha y^{1-\alpha} = y(x/y)^\alpha$, and lower bounding the convex function $t\mapsto s(\alpha)t^\alpha$ on $[0,+\infty)$ by its tangent line at 1.

To prove the first inequalities, we may assume w.l.o.g. that ϱ and σ are positive definite, due to (III.17). Assume first that ϱ and σ are both diagonal in the same orthonormal basis with diagonal elements $r_1,\dots,r_d$ and $s_1,\dots,s_d$, respectively, where $d:=\dim\mathcal{H}$. Concavity of $t\mapsto t^\alpha$ for $\alpha\in(0,1)$ yields

$$\frac{(\operatorname{Tr}\varrho)^\alpha}{(\operatorname{Tr}\sigma)^\alpha} = \left(\sum_{i=1}^d \frac{s_i}{\operatorname{Tr}\sigma}\cdot\frac{r_i}{s_i}\right)^\alpha \ge \sum_{i=1}^d \frac{s_i}{\operatorname{Tr}\sigma}\cdot\left(\frac{r_i}{s_i}\right)^\alpha = (\operatorname{Tr}\sigma)^{-1}\sum_{i=1}^d r_i^\alpha s_i^{1-\alpha}, \tag{A.25}$$

which is exactly the first inequality in (A.23). When $\alpha>1$, the inequality in (A.25) is in the opposite direction, which yields the first inequality in (A.24).

For the general case, we follow the approach at the beginning of Section 3 in [7]. Let $s_1(X)\ge \ldots \ge s_d(X)$ denote the decreasingly ordered singular values of an operator X. By the Gelfand-Naimark majorization theorem (see, e.g. [30, Theorem 4.3.4]), we have

$$\prod_{j=1}^k s_{i_j}\left(\sigma^{\frac{1-\alpha}{2z}}\right) s_{d+1-i_j}\left(\varrho^{\frac{\alpha}{2z}}\right) \le \prod_{i=1}^k s_i\left(\sigma^{\frac{1-\alpha}{2z}}\varrho^{\frac{\alpha}{2z}}\right) \le \prod_{i=1}^k s_i\left(\sigma^{\frac{1-\alpha}{2z}}\right) s_i\left(\varrho^{\frac{\alpha}{2z}}\right)$$

for every $1\le k\le d$ and $1\le i_1<\ldots<i_j\le d$. Taking it to the power $2z$ yields

$$\prod_{j=1}^k s_{i_j}(\sigma)^{1-\alpha} s_{d+1-i_j}(\varrho)^\alpha \le \prod_{i=1}^k s_i\left(\left(\varrho^{\frac{\alpha}{2z}}\sigma^{\frac{1-\alpha}{z}}\varrho^{\frac{\alpha}{2z}}\right)^z\right) \le \prod_{i=1}^k s_i(\sigma)^{1-\alpha} s_i(\varrho)^\alpha .$$

Using that weak log-majorization implies weak majorization (see, e.g. [30, Proposition 4.1.6]), we obtain

$$\sum_{j=1}^k s_{i_j}(\sigma)^{1-\alpha} s_{d+1-i_j}(\varrho)^\alpha \le \sum_{i=1}^k s_i\left(\left(\varrho^{\frac{\alpha}{2z}}\sigma^{\frac{1-\alpha}{z}}\varrho^{\frac{\alpha}{2z}}\right)^z\right) \le \sum_{i=1}^k s_i(\sigma)^{1-\alpha} s_i(\varrho)^\alpha$$

for every $1\le k\le d$. Taking now $k=d$, we see that the middle sum is equal to $Q_{\alpha,z}(\varrho\|\sigma)$, the rightmost sum can be upper bounded by $\left(\sum_{i=1}^d s_i(\varrho)\right)^\alpha\left(\sum_{i=1}^d s_i(\sigma)\right)^{1-\alpha} = (\operatorname{Tr}\varrho)^\alpha(\operatorname{Tr}\sigma)^{1-\alpha}$ when $\alpha\in(0,1)$, by the same argument as in (A.25), and similarly, the leftmost sum can be lower bounded by $(\operatorname{Tr}\varrho)^\alpha(\operatorname{Tr}\sigma)^{1-\alpha}$ when $\alpha>1$. □

Remark A.24 *The first inequalities in* (A.23)–(A.24) *for* $(\alpha, z) \in K_2 \cup K_4 \cup K_5 \cup K_7$ *are immediate from the monotonicity under taking the trace of both* ϱ *and* σ*, according to Lemma III.5.*

We can also establish strict positivity of $\widehat{D}_{\alpha,z}$ and $T_{\alpha,z}$, except for the region

$$K_0: \quad 0 < \alpha < 1,\, z < \min\{\alpha, 1-\alpha\}.$$

In the proof we also give an alternative proof for the first inequalities in (A.23)–(A.24) when $(\alpha, z) \notin K_0$.

Proposition A.25 *The second inequalities in* (A.23)–(A.24) *hold as equalities if and only if* $\operatorname{Tr}\varrho = \operatorname{Tr}\sigma$. *If* $(\alpha, z) \in ((0,+\infty) \times (0,+\infty]) \setminus K_0$ *then the second inequalities in* (A.23)–(A.24) *hold as equalities if and only if* $\varrho / \operatorname{Tr}\varrho = \sigma / \operatorname{Tr}\sigma$.

Proof The assertion about the second inequalities is trivial from the strict convexity of $t \mapsto s(\alpha)t^\alpha$ when $\alpha \in (0,+\infty) \setminus \{1\}$. Hence, for the rest we analyze the first inequalities.

It has been pointed out, e.g., in [41, Proposition 1], that the Araki-Lieb-Thirring inequality [4, 39] implies the monotonicity

$$Q_{\alpha,z_1}(\varrho\|\sigma) = \operatorname{Tr}\left(\varrho^{\frac{\alpha}{2z_1}} \sigma^{\frac{1-\alpha}{z_1}} \varrho^{\frac{\alpha}{2z_1}}\right)^{\frac{z_1}{z_2} z_2} \leq \operatorname{Tr}\left(\varrho^{\frac{\alpha}{2z_2}} \sigma^{\frac{1-\alpha}{z_2}} \varrho^{\frac{\alpha}{2z_2}}\right)^{z_2} = Q_{\alpha,z_2}(\varrho\|\sigma), \qquad z_2 \leq z_1. \tag{A.26}$$

Hence, for $\alpha > 1$ we have

$$D_{\alpha,z}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) = \widehat{D}_{\alpha,z}(\varrho\|\sigma) \geq \widehat{D}_{\alpha,+\infty}(\varrho\|\sigma) = D_{\alpha,+\infty}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) \geq 0,$$

with equality if and only if $\varrho / \operatorname{Tr}\varrho = \sigma / \operatorname{Tr}\sigma$, according to [44, Proposition 3.22]. This is exactly the second inequality in (A.24) with the equality condition. If $\alpha \in (0, 1/2]$ and $z \geq \alpha$ then

$$D_{\alpha,z}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) = \widehat{D}_{\alpha,z}(\varrho\|\sigma) \geq \widehat{D}_{\alpha,\alpha}(\varrho\|\sigma) = D_{\alpha,\alpha}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) \geq 0,$$

with equality if and only if $\varrho / \operatorname{Tr}\varrho = \sigma / \operatorname{Tr}\sigma$, according to [9, Theorem 5] (see also [44, Proposition 3.22]). If $\alpha \in [1/2, 1)$ and $z \geq 1 - \alpha$ then

$$\begin{aligned} D_{\alpha,z}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) = \widehat{D}_{\alpha,z}(\varrho\|\sigma) \geq \widehat{D}_{\alpha,1-\alpha}(\varrho\|\sigma) &= D_{\alpha,1-\alpha}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) \\ &= \frac{\alpha}{1-\alpha} D_{1-\alpha,1-\alpha}\left(\frac{\varrho}{\operatorname{Tr}\varrho}\middle\|\frac{\sigma}{\operatorname{Tr}\sigma}\right) \geq 0, \end{aligned}$$

with equality if and only if $\varrho / \operatorname{Tr}\varrho = \sigma / \operatorname{Tr}\sigma$, by the same argument as above. □

Corollary A.26 *Let* $\varrho, \sigma \in \mathcal{B}(\mathcal{H})_+$ *and* $\alpha \in (0,+\infty)$, $z \in (0,+\infty]$. *Then*

$$D_{\alpha,z}(\varrho\|\sigma) \geq \log \operatorname{Tr}\varrho - \log \operatorname{Tr}\sigma, \tag{A.27}$$

$$\widehat{D}_{\alpha,z}(\varrho\|\sigma) \geq 0, \tag{A.28}$$

$$T_{\alpha,z}(\varrho\|\sigma) \geq 0. \tag{A.29}$$

If, moreover, $(\alpha, z) \notin K_0$ *then equality holds in* (A.27) *or* (A.28) *if and only if* $\varrho = \lambda\sigma$ *for some* $\lambda \in (0,+\infty)$, *and equality holds in* (A.29) *if and only if* $\varrho = \sigma$.

Proof Immediate from Proposition A.25. □

5. Further properties of the Rényi divergence radii

Lemma A.27 *For any $\sigma \in \mathcal{S}(\mathcal{H})_{++}$ and any $\alpha \in (0,+\infty)\setminus\{1\}$, $z \in (0,+\infty)$, the map $x \mapsto D_{\alpha,z}(W(x)\|\sigma)$ is bounded on $\mathcal{X}$ for any qc channel W, and hence the map $P \mapsto \sum_{x\in\mathcal{X}} P(x)D_{\alpha,z}(W(x)\|\sigma)$ is continuous on $\mathcal{P}_f(\mathcal{X})$ in the variational norm.*

Proof We prove the case $\alpha > 1$; the case $\alpha \in (0,1)$ follows the same way with all inequalities reversed. If $\sigma \in \mathcal{S}(\mathcal{H})_{++}$ then $\sigma \geq \lambda_{\min}(\sigma)I$, and hence $\sigma^{\frac{1-\alpha}{z}} \leq \lambda_{\min}(\sigma)^{\frac{1-\alpha}{z}}I$. Using the monotonicity of $A \mapsto \operatorname{Tr} A^z$ on $\mathcal{B}(\mathcal{H})_+$ for $z > 0$, we get

$$\begin{aligned} D_{\alpha,z}(W(x)\|\sigma) &\leq \frac{1}{\alpha-1}\log\operatorname{Tr}\left(W(x)^{\frac{\alpha}{2z}}\lambda_{\min}(\sigma)^{\frac{1-\alpha}{z}}IW(x)^{\frac{\alpha}{2z}}\right)^z \\ &= -\log\lambda_{\min}(\sigma) + \frac{1}{\alpha-1}\log\operatorname{Tr} W(x)^\alpha \leq -\log\lambda_{\min}(\sigma), \end{aligned}$$

proving the boundedness, and the assertion on continuity is immediate from this. □

Corollary A.28 *The map $P \mapsto \chi^*_\alpha(W,P)$ is concave and upper semi-continuous on $\mathcal{P}_f(\mathcal{X})$ in the variational norm. In particular, if $P \in \mathcal{P}_f(\mathcal{X})$ and $P_n \in \mathcal{P}_f(\mathcal{X})$, $n \in \mathbb{N}$, are such that $\lim_{n\to+\infty}\|P_n - P\|_1 = 0$ then*

$$\limsup_{n\to+\infty}\chi^*_\alpha(W,P_n) \leq \chi^*_\alpha(W,P). \tag{A.30}$$

Proof A combination or (III.21) and Lemma A.27 shows that χ^*_α, as the infimum of continuous affine functions, is upper semi-continuous and concave. Upper semi-continuity imples (A.30). □

Appendix B: Proof of Proposition IV.3

Proof of Proposition IV.3: Define the classical-quantum states

$$R_n := \frac{1}{|\mathcal{C}_n|}\sum_{k=1}^{|\mathcal{C}_n|}|k\rangle\langle k|\otimes W^{\otimes n}(\mathcal{E}_n(k)), \qquad S_n := \frac{1}{|\mathcal{C}_n|}\sum_{k=1}^{|\mathcal{C}_n|}|k\rangle\langle k|\otimes\sigma^{\otimes n},$$

and the POVM element

$$T_n := \sum_{k=1}^{|\mathcal{C}_n|}|k\rangle\langle k|\otimes\mathcal{D}_n(k),$$

where $(|k\rangle\langle k|)_{k=1}^{|\mathcal{C}_n|}$ is a set of orthogonal rank 1 projections in some Hilbert space, and $\sigma \in \mathcal{S}(\mathcal{H})$ is an arbitrary state. Then we have

$$\begin{aligned} \operatorname{Tr} R_nT_n &= \frac{1}{|\mathcal{C}_n|}\sum_{k=1}^{|\mathcal{C}_n|}\operatorname{Tr} W^{\otimes n}(\mathcal{E}_n(k))\mathcal{D}_n(k) = P_s(W^{\otimes n},\mathcal{C}_n), \\ \operatorname{Tr} S_nT_n &= \frac{1}{|\mathcal{C}_n|}\sum_{k=1}^{|\mathcal{C}_n|}\operatorname{Tr}\sigma^{\otimes n}\mathcal{D}_n(k) = \frac{1}{|\mathcal{C}_n|}. \end{aligned}$$

For any $\sigma \in \mathcal{S}(\mathcal{H}^{\otimes n})$ such that $W^{\otimes n}(\mathcal{E}_n(k))^0 \le \sigma^0$ for all k, and for all $\alpha > 1$, we get

$$\begin{aligned}
P_s(W^{\otimes n}, \mathcal{C}_n)^\alpha \left(\frac{1}{|\mathcal{C}_n|}\right)^{1-\alpha} &= (\operatorname{Tr} R_n T_n)^\alpha (\operatorname{Tr} S_n T_n)^{1-\alpha} \le Q_\alpha^*(R_n \| S_n) \\
&= \frac{1}{|\mathcal{C}_n|} \sum_{k=1}^{|\mathcal{C}_n|} Q_\alpha^*\left(W^{\otimes n}(\mathcal{E}_n(k)) \| \sigma^{\otimes n}\right) \\
&= \frac{1}{|\mathcal{C}_n|} \sum_{k=1}^{|\mathcal{C}_n|} \prod_{x \in \mathcal{X}} Q_\alpha^*(W(x) \| \sigma)^{n P_n(x)} \\
&= \exp\left(n(\alpha - 1) \sum_{x \in \mathcal{X}} P_n(x) D_\alpha^*(W(x) \| \sigma)\right),
\end{aligned}$$

where the inequality is due to the monotonicity of the sandwiched Rényi divergence for $\alpha > 1$. Note that the inequality between the first and the last expressions above holds trivially when $W^{\otimes n}(\mathcal{E}_n(k))^0 \not\le \sigma^0$ for some k. From this we get

$$\frac{1}{n} \log P_s(W^{\otimes n}, \mathcal{C}_n) \le -\frac{\alpha - 1}{\alpha} \left[\frac{1}{n} \log |\mathcal{C}_n| - \sum_{x \in \mathcal{X}} P_n(x) D_\alpha^*(W(x) \| \sigma)\right] \tag{B.1}$$

for any $\sigma \in \mathcal{S}(\mathcal{H})$ and $\alpha > 1$. Hence, if $\frac{1}{n} \log |\mathcal{C}_n| \ge R$ for some n then

$$\frac{1}{n} \log P_s(W^{\otimes n}, \mathcal{C}_n) \le -\sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} \left[R - \chi_\alpha^*(W, P_n)\right]. \tag{B.2}$$

Assuming instead that $\liminf_{n \to +\infty} \frac{1}{n} \log |\mathcal{C}_n| \ge R$, taking first the limit $n \to +\infty$ in (B.1) for a fixed $\sigma \in \mathcal{S}(\mathcal{H})_{++}$, using Lemma A.27, then taking the infimum over σ, and finally the infimum over $\alpha > 1$, we get

$$\limsup_{n \to +\infty} \frac{1}{n} \log P_s(W^{\otimes n}, \mathcal{C}_n) \le -\sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} \left[R - \chi_\alpha^*(W, P)\right], \tag{B.3}$$

which is what we wanted to prove.

Remark B.1 *Note that* (B.2) *is valid for every n for which $\frac{1}{n} \log |\mathcal{C}_n| \ge R$, and hence it gives a stronger statement then the asymptotic version* (B.3)*.*

Remark B.2 *We can also arrive at* (B.3) *by taking* (B.2) *without the supremum over α, taking the limit $n \to +\infty$ and using Corollary A.28, and then taking the infimum over $\alpha > 1$.*

Appendix C: Random coding exponent with constant composition

Below we give a slightly different proof of the constant composition random coding bound first proved in [14]. We start with the following random coding bound, given in [26, 28, 47].

Lemma C.1 *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, $n \in \mathbb{N}$, $R > 0$, $M_n := \lceil e^{nR} \rceil$, and $Q_n \in \mathcal{P}_f(\mathcal{X}^n)$. For every $\mathbf{x} = (\underline{x}_1, \dots, \underline{x}_{M_n}) \in (\mathcal{X}^n)^{M_n}$, there exists a code $\mathcal{C}_{n,\mathbf{x}} = (\mathcal{E}_{n,\mathbf{x}}, \mathcal{D}_{n,\mathbf{x}})$ for $W^{\otimes n}$ such that $\mathcal{E}_{n,\mathbf{x}}(k) = \underline{x}_k$, $k \in [M_n]$, and*

$$\begin{aligned}
\mathbb{E}_{Q_n^{\otimes M_n}} P_e(W^{\otimes n}, \mathcal{C}_{n,\mathbf{x}}) &\le \sum_{\underline{x} \in \mathcal{X}^n} Q_n(\underline{x}) \operatorname{Tr} W^{\otimes n}(\underline{x}) \{W^{\otimes n}(\underline{x}) - e^{nR} W^{\otimes n}(Q_n) \le 0\} \\
&\quad + e^{nR} \sum_{\underline{x} \in \mathcal{X}^n} Q_n(\underline{x}) \operatorname{Tr} W^{\otimes n}(Q_n) \{W^{\otimes n}(\underline{x}) - e^{nR} W^{\otimes n}(Q_n) > 0\} \\
&\le e^{nR(1-\alpha)} \sum_{\underline{x} \in \mathcal{X}^n} Q_n(\underline{x}) \operatorname{Tr} W^{\otimes n}(\underline{x})^\alpha W^{\otimes n}(Q_n)^{1-\alpha}.
\end{aligned} \tag{C.1}$$

In particular, there exists an $\underline{x} \in \operatorname{supp} Q_n$ such that $P_e(W^{\otimes n}, \mathcal{C}_{n,\mathbf{x}})$ is upper bounded by the RHS of (C.1)*.*

From this, we can obtain the following:

Proposition C.2 *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, and let $R > 0$. For every $n \in \mathbb{N}$, and every type $P_n \in \mathcal{P}_n(\mathcal{X})$, there exists a code $\mathcal{C}_n$ of constant composition P_n with rate $\frac{1}{n}\log|\mathcal{C}_n| \geq R$ such that*

$$\frac{1}{n}\log P_e(W^{\otimes n},\mathcal{C}_n) \leq -\sup_{0\leq\alpha\leq 1}(\alpha-1)\left[R-\sum_{x\in\mathcal{X}}P_n(x)D_\alpha(W(x)\|W(P_n))+|\operatorname{supp}P_n|\frac{\log(n+1)}{n}\right]. \tag{C.2}$$

Proof Let $\mathcal{X}_n := \mathcal{X}^n_{P_n} \subseteq \mathcal{X}^n$ be the set of sequences with type P_n. Choosing $Q_n := \frac{1}{|\mathcal{X}_n|}\mathbf{1}_{\mathcal{X}_n}$ in Lemma C.1, we get the existence of codes $\mathcal{C}_n$ with constant composition P_n such that

$$P_e(W^{\otimes n},\mathcal{C}_n) \leq e^{nR(1-\alpha)}\operatorname{Tr}W^{\otimes n}(\underline{x})^\alpha W^{\otimes n}(Q_n)^{1-\alpha}$$

for any $\underline{x} \in \mathcal{X}_n$, where we used that $W^{\otimes n}(Q_n) = \frac{1}{|\mathcal{X}_n|}\sum_{\underline{y}\in\mathcal{X}_n}W^{\otimes n}(\underline{y})$ is permutation-invariant. Now we use the well-known facts that $|\mathcal{X}^n_{P_n}| \geq (n+1)^{-|\operatorname{supp}P_n|}e^{nH(P_n)}$, and that for any $\underline{y} \in \mathcal{X}^n_{P_n}$, $P_n^{\otimes n}(\underline{y}) = e^{-nH(P_n)}$, (see (II.8) and (II.9)), to obtain that

$$\begin{aligned}
W^{\otimes n}(Q_n) = \frac{1}{|\mathcal{X}_n|}\sum_{\underline{y}\in\mathcal{X}_n}W^{\otimes n}(\underline{y}) &\leq (n+1)^{|\operatorname{supp}P_n|}e^{-nH(P_n)}\sum_{\underline{y}\in\mathcal{X}_n}W^{\otimes n}(\underline{y})\\
&= (n+1)^{|\operatorname{supp}P_n|}\sum_{\underline{y}\in\mathcal{X}_n}P_n^{\otimes n}(\underline{y})W^{\otimes n}(\underline{y})\\
&\leq (n+1)^{|\operatorname{supp}P_n|}\sum_{\underline{y}\in\mathcal{X}^n}P_n^{\otimes n}(\underline{y})W^{\otimes n}(\underline{y})\\
&= (n+1)^{|\operatorname{supp}P_n|}W(P_n)^{\otimes n}.
\end{aligned}$$

Using that $t \mapsto t^{1-\alpha}$ is operator monotone on $\mathbb{R}_+$ for $\alpha \in [0,1]$, we get that

$$P_e(W^{\otimes n},\mathcal{C}_n) \leq (n+1)^{(1-\alpha)|\operatorname{supp}P_n|}e^{nR(1-\alpha)}\operatorname{Tr}W^{\otimes n}(\underline{x})^\alpha\left(W(P_n)^{\otimes n}\right)^{1-\alpha}.$$

From this (C.2) follows by simple algebra. □

Corollary C.3 *Let $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ be a classical-quantum channel, let $R > 0$, and $P \in \mathcal{P}_f(\mathcal{X})$. There exists a sequence of codes $(\mathcal{C}_n)_{n\in\mathbb{N}}$, where all $\mathcal{C}_n$ are of constant composition with some $P_n \in \mathcal{P}_n(\mathcal{X})$ and $\operatorname{supp}P_n \subseteq \operatorname{supp}P$, $\lim_{n\to+\infty}\|P_n - P\|_1 = 0$, and every $\mathcal{C}_n$ has rate $\frac{1}{n}\log|\mathcal{C}_n| \geq R$, such that*

$$\limsup_{n\to+\infty}\frac{1}{n}\log P_e(W^{\otimes n},\mathcal{C}_n) \leq -\sup_{0\leq\alpha\leq 1}(\alpha-1)\left[R-\sum_{x\in\mathcal{X}}P(x)D_\alpha(W(x)\|W(P))\right]. \tag{C.3}$$

If, moreover, $R < \sum_{x\in\mathcal{X}}P(x)D(W(x)\|W(P))$ then $P_e(W^{\otimes n},\mathcal{C}_n)$ goes to zero exponentially fast.

Proof For any given P we can find an approximating sequence $P_n \in \mathcal{P}_n(\mathcal{X})$ such that $\operatorname{supp}P_n \subseteq \operatorname{supp}P$ for all $n \in \mathbb{N}$, and $\lim_{n\to+\infty}\|P_n - P\|_1 = 0$. Applying Proposition C.2 to this sequence, the assertion follows. □

A variant of Corollary C.3 was given by Csiszár and Körner in [18, Theorem 10.2] for classical channels, where the RHS is

$$-\sup_{0<\alpha\leq 1}\frac{\alpha-1}{\alpha}\left[R-\chi_\alpha(W,P)\right].$$

Note that $\frac{\alpha-1}{\alpha}$ gives a strictly better prefactor, while $\chi_\alpha(W,P) \le \sum_{x\in\mathcal{X}} P(x) D_\alpha(W(x)\|W(P))$. However, the Csiszár-Körner bound is optimal for high enough rates, and hence it would be desirable to obtain an exact analogue of it for classical-quantum channels.

Constant composition exponents were obtained also for classical-quantum channels before; for instance, the following was stated in [27]: Let $\mathcal{X}$ be a finite set, $\mathcal{H}$ be a finite-dimensional Hilbert space, let P be a probability mass function on $\mathcal{X}$, and $R > 0$. Then there exists a sequence of codes $(\mathcal{C}_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty} \frac{1}{n} \log |\mathcal{C}_n| = R,$$

and for any classical-quantum channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$,

$$\liminf_{n\to+\infty} -\frac{1}{n} \log P_e(W^{\otimes n}, \mathcal{C}_n) \ge \sup_{0\le\alpha\le 1} \frac{\alpha-1}{2-\alpha} \left[R - I_\alpha(W,P)\right]. \tag{C.4}$$

While this is not sufficiently detailed in [27], the codes above can indeed be chosen to be of constant composition.

Note that the bound in (C.4) is not as strong as the classical universal random coding exponent given by Csiszár and Körner, as $0 < \alpha < 2 - \alpha$ for all $\alpha < 1$, and $\chi_\alpha(W,p) \ge I_\alpha(W,p)$, with the inequality being strict in general.

Appendix D: Evaluation of the information spectrum quantity

For every $n \in \mathbb{N}$, let $\mathcal{H}_n$ be a finite-dimensional Hilbert space, and let $\varrho_n, \sigma_n \in \mathcal{S}(\mathcal{H}_n)$. Let

$$\psi(\alpha) := \begin{cases} \limsup_{n\to+\infty} \frac{1}{n} \log \operatorname{Tr} \varrho_n^\alpha \sigma_n^{1-\alpha}, & \alpha \in (0,1], \\ \limsup_{n\to+\infty} \frac{1}{n} \log \operatorname{Tr} \left(\varrho_n^{1/2} \sigma_n^{\frac{1-\alpha}{\alpha}} \varrho_n^{1/2}\right)^\alpha, & \alpha \in (1,+\infty). \end{cases}$$

The following statement is a central observation in the information spectrum method, and its proof follows by standard arguments. We include a complete proof for the readers' convenience.

Lemma D.1 *In the above setting,*

$$\lim_{n\to+\infty} \operatorname{Tr}(\varrho_n - e^{nr}\sigma_n)_+ = 1 \quad \begin{cases} \text{for all } r \in \mathbb{R}, & \text{if } \psi(1) < 0, \\ \text{for } r < \partial^-\psi(1), & \text{if } \psi(1) = 0. \end{cases} \tag{D.1}$$

If $\varrho_n^0 \le \sigma_n^0$ for all large enough n, then $\psi(1) = 0$, and

$$\lim_{n\to+\infty} \operatorname{Tr}(\varrho_n - e^{nr}\sigma_n)_+ = 0 \quad \text{for all} \quad r > \partial^+\psi(1). \tag{D.2}$$

Proof Let $S_{n,r} := \{\varrho_n - e^{nr}\sigma_n > 0\}$ be the Neyman-Pearson test with parameter r. Then

$$0 \le \operatorname{Tr}(\varrho_n - e^{nr}\sigma_n)_+ = \operatorname{Tr} \varrho_n S_{n,r} - e^{nr} \operatorname{Tr} \sigma_n S_{n,r} = 1 - \alpha_n(S_{n,r}) - e^{nr}\beta_n(S_{n,r}),$$

with $\alpha_n(S_{n,r}) := \operatorname{Tr} \varrho_n(I - S_{n,r})$ and $\beta_n(S_{n,r}) := \operatorname{Tr} \sigma_n S_{n,r}$ being the type I and the type II error probabilities corresponding to the test $S_{n,r}$, respectively. In particular,

$$0 \le e^{nr} \operatorname{Tr} \sigma_n S_{n,r} \le \operatorname{Tr} \varrho_n S_{n,r}. \tag{D.3}$$

By Audenaert's inequality [5, Theorem 1],

$$e_n(r) := \alpha_n(S_{n,r}) + e^{nr}\beta_n(S_{n,r}) = \operatorname{Tr} \varrho_n(I - S_{n,r}) + e^{nr} \operatorname{Tr} \sigma_n S_{n,r} \le e^{nr(1-\alpha)} \operatorname{Tr} \varrho_n^\alpha \sigma_n^{1-\alpha}$$

for all $\alpha \in [0,1]$, and hence

$$\limsup_{n\to+\infty} \frac{1}{n} \log e_n(r) \le -\sup_{0\le\alpha\le 1} (\alpha-1)\{r - \psi(\alpha)/(\alpha-1)\}. \tag{D.4}$$

Since $\lim_{\alpha\nearrow 1}\psi(\alpha)/(\alpha-1)=\partial^-\psi(1)$ if $\psi(1)=0$, and $+\infty$ otherwise, we obtain that $e_n(r)\to 0$ as $n\to+\infty$ for any $r\in\mathbb{R}$ when $\psi(1)<0$ and for $r<\partial^-\psi(1)$ when $\psi(1)=0$. Using that $\max\{\alpha_n(S_{n,r}),e^{nr}\beta_n(S_{n,r})\}\le e_n(r)$, we see that in these cases, also $e^{nr}\beta_n(S_{n,r})\to 0$ and $1-\alpha_n(S_{n,r})\to 1$, and therefore $\operatorname{Tr}(\varrho_n-e^{nr}\sigma_n)_+\to 1$. This proves (D.1).

Next, we prove (D.2). By the monotonicity of the sandwiched Rényi divergences, we have, for every $0\le T_n\le I$, and every $\alpha>1$,

$$\operatorname{Tr}\left(\varrho_n^{1/2}\sigma_n^{\frac{1-\alpha}{\alpha}}\varrho_n^{1/2}\right)^{\alpha}\ge\left(\operatorname{Tr}\varrho_nT_n\right)^{\alpha}\left(\operatorname{Tr}\sigma_nT_n\right)^{1-\alpha},$$

and therefore

$$\limsup_{n\to+\infty}\frac{1}{n}\log\operatorname{Tr}\varrho_nT_n\le\inf_{\alpha>1}\frac{\alpha-1}{\alpha}\left\{\limsup_{n\to+\infty}\frac{1}{n}\log\operatorname{Tr}\sigma_nT_n+\frac{\psi(\alpha)}{\alpha-1}\right\}. \tag{D.5}$$

Now, let $T_n:=S_{n,r}$ with some r. Then, by (D.4), we have

$$\limsup_{n\to+\infty}\frac{1}{n}\log\operatorname{Tr}\sigma_nT_n\le-\sup_{0\le\alpha\le 1}\{\alpha r-\psi(\alpha)\}\le -r.$$

Hence, if $r>\partial^+\psi(1)$, then the RHS of (D.5) is negative, and thus $\operatorname{Tr}\varrho_nS_{n,r}\to 0$ as $n\to+\infty$. Using (D.3), we get that also $e^{nr}\operatorname{Tr}\sigma_nS_{n,r}\to 0$, and hence $\operatorname{Tr}(\varrho_n-e^{nr}\sigma_n)_+\to 0$, as required. □

Corollary D.2 *For every $n\in\mathbb{N}$, let P_n be an n-type and $\underline{x}^{(n)}\in\mathcal{X}^n$ be of type P_n. Assume that $(P_n)_{n\in\mathbb{N}}$ converges to some $P\in\mathcal{P}_f(\mathcal{X})$ and $\cup_{n\in\mathbb{N}}\operatorname{supp}P_n$ is finite. If $V(x)^0\le W(x)^0$ for all $x\in\operatorname{supp}P_n$ and all n large enough, then*

$$\lim_{n\to+\infty}\operatorname{Tr}\left(V^{\otimes n}(\underline{x}^{(n)})-e^{nr}W^{\otimes n}(\underline{x}^{(n)})\right)_+=\begin{cases}1, & r<\sum_{x\in\mathcal{X}}P(x)D(V(x)\|W(x)),\\ 0, & r>\sum_{x\in\mathcal{X}}P(x)D(V(x)\|W(x)).\end{cases}$$

Proof We use Lemma D.1 with $\varrho_n:=V^{\otimes n}(\underline{x}^{(n)})$, $\sigma_n:=W^{\otimes n}(\underline{x}^{(n)})$,. Then we have

$$\psi_n(\alpha):=\frac{1}{n}\log\operatorname{Tr}\varrho_n^{\alpha}\sigma_n^{1-\alpha}=\sum_{x\in\mathcal{X}}P_n(x)\log\operatorname{Tr}V(x)^{\alpha}W(x)^{1-\alpha},$$

for $\alpha\in(0,1]$, and

$$\psi_n(\alpha):=\frac{1}{n}\log\operatorname{Tr}\left(\varrho_n^{1/2}\sigma_n^{\frac{1-\alpha}{\alpha}}\varrho_n^{1/2}\right)^{\alpha}=\sum_{x\in\mathcal{X}}P_n(x)\log\operatorname{Tr}\left(V(x)^{1/2}W(x)^{\frac{1-\alpha}{\alpha}}V(x)^{1/2}\right)^{\alpha},$$

for $\alpha>1$. Hence, $\psi(\alpha)=\lim_{n\to+\infty}\psi_n(\alpha)$ exists as a limit, and

$$\psi(\alpha)=\begin{cases}\sum_{x\in\mathcal{X}}P(x)\log\operatorname{Tr}V(x)^{\alpha}W(x)^{1-\alpha}, & \alpha\in(0,1],\\ \sum_{x\in\mathcal{X}}P(x)\log\operatorname{Tr}\left(V(x)^{1/2}W(x)^{\frac{1-\alpha}{\alpha}}V(x)^{1/2}\right)^{\alpha}, & \alpha>1.\end{cases}$$

By assumption, $\psi(1)=0$, and

$$\partial^-\psi(1)=\sum_{x\in\mathcal{X}}P(x)\lim_{\alpha\nearrow 1}D_\alpha(V(x)\|W(x))=\sum_{x\in\mathcal{X}}P(x)D(V(x)\|W(x),$$
$$\partial^+\psi(1)=\sum_{x\in\mathcal{X}}P(x)\lim_{\alpha\searrow 1}D_\alpha^*(V(x)\|W(x))=\sum_{x\in\mathcal{X}}P(x)D(V(x)\|W(x).$$

Hence, the assertion follows immediately from Lemma D.1. □

Appendix E: A type lemma

The following simple lemma is probably well known; however, we are not aware of a reference, so we give a proof for readers' convenience.

Lemma E.1 *For any $P \in \mathcal{P}_f(\mathcal{X})$, and any $m \in \mathbb{N}$,*

$$\sum_{\underline{x} \in \mathcal{X}^m} P^{\otimes m}(\underline{x}) P_{\underline{x}} = P.$$

Proof Note that if $x_i \notin \operatorname{supp} P$ for some $\underline{x} \in \mathcal{X}^m$ and $i \in [m]$ then $P^{\otimes m}(\underline{x}) = 0$, and hence the LHS above is equal to $\sum_{\underline{x} \in (\operatorname{supp} P)^m} P^{\otimes m}(\underline{x}) P_{\underline{x}}$. For any $a \in \mathcal{X}$, we have

$$\begin{aligned}
\sum_{\underline{x} \in (\operatorname{supp} P)^m} P^{\otimes m}(\underline{x}) P_{\underline{x}}(a) &= \sum_{k=0}^{m} \frac{k}{m} \sum_{\substack{\underline{x} \in (\operatorname{supp} P)^m \\ \#\{i:\, x_i = a\} = k}} P^{\otimes m}(\underline{x}) \\
&= \sum_{k=1}^{m} \frac{k}{m} P(a)^k \binom{m}{k} \sum_{\substack{\underline{x} \in (\operatorname{supp} P)^{m-k} \\ \forall i:\, x_i \neq a}} P^{\otimes (m-k)}(\underline{x}) \\
&= \sum_{k=1}^{m} \binom{m-1}{k-1} P(a)^k P(\operatorname{supp} P \setminus \{a\})^{m-k} \\
&= P(a) \sum_{k=1}^{m} \binom{m-1}{k-1} P(a)^{k-1} (1 - P(a))^{m-k} \\
&= P(a).
\end{aligned}$$

□

ACKNOWLEDGMENTS

We are grateful to Fumio Hiai, Péter Vrana, Andreas Winter, Hao-Chung Heng, Min-Hsiu Hsieh, Marco Tomamichel, József Pitrik and Dániel Virosztek for discussions. This work was partially funded by the JSPS grant no. JP16K00012 (TO), by the National Research, Development and Innovation Office of Hungary via the research grants K124152, KH129601, the Quantum Technology National Excellence Program, Project Nr. 2017-1.2.1-NKP-2017-00001, and by a Bolyai János Fellowship of the Hungarian Academy of Sciences (MM).

---

[1] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2000.

[3] Tsuyoshi Ando and Fumio Hiai. Operator log-convex functions and operator means. *Mathematische Annalen*, 350:611–630, 2011.

[4] H. Araki. On an inequality of Lieb and Thirring. *Letters in Mathematical Physics*, 19:167–170, 1990.

[5] K.M.R. Audenaert, J. Calsamiglia, R. Mu noz Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Physical Review Letters*, 98:160501, 2007. arXiv:quant-ph/0610027.

[6] Koenraad M. R. Audenaert and Nilanjana Datta. α-z-relative Renyi entropies. *J. Math. Phys.*, 56:022202, 2015. arXiv:1310.7178.

[7] Koenraad M. R. Audenaert and Fumio Hiai. Reciprocal Lie-Trotter formula. *Linear and Multilinear Algebra*, 64(6):1220–1235, 2016.

[8] U. Augustin. *Noisy channels*. Habilitation thesis, Univ. of Erlangen-Numberg, 1978.

[9] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, December 2013. arXiv:1306.5920.

[10] Rajendra Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, 1997.

[11] Rajendra Bhatia, Stephane Gaubert, and Tanvi Jain. Matrix versions of the Hellinger distance. *Letters in Mathematical Physics*, 2019.

[12] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 2018.

[13] E. A. Carlen, R. L. Frank, and E. H. Lieb. Some operator and trace function convexity theorems. *Linear Algebra Appl.*, 490:174–185, 2016.

[14] Hao-Chung Cheng, Eric P. Hanson, Nilanjana Datta, and Min-Hsiu Hsieh. Duality between source coding with quantum side information and c-q channel coding. arXiv:1809.11143, 2018.

[15] Hao-Chung Cheng, Gao Li, and Min-Hsiu Hsieh. Properties of noncommutative Renyi and Augustin information. arXiv:1811.04218, 2018.

[16] Imre Csiszár. Generalized cutoff rates and Rényi's information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, January 1995.

[17] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless channels*. Akadémiai Kiadó, Budapest, 1981.

[18] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless channels, 2nd edition*. Cambridge University Press, 2011.

[19] Marco Dalai and Andreas Winter. Constant compositions in the sphere packing bound for classical-quantum channels. *IEEE Transactions on Information Theory*, 63(9):5603 – 5617, 2017. arXiv:1509.00715.

[20] Nilanjana Datta. Min- and max-relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, 2009.

[21] G. Dueck and J. Körner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Transactions on Information Theory*, 25:82–85, 1979.

[22] Ky Fan. Minimax theorems. *Proc Natl Acad Sci USA.*, 39(1):42–47, 1953.

[23] Bálint Farkas and Szilárd Révész. Potential theoretic approach to rendezvous numbers. *Monatshefte für Mathematik*, 148(4):309–331, 2006.

[24] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12):122201, December 2013. arXiv:1306.5358.

[25] Masahito Hayashi. Optimal sequence of POVM's in the sense of Stein's lemma in quantum hypothesis testing. *J. Phys. A: Math. Gen.*, 35:10759–10773, 2002.

[26] Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76(6):062301, December 2007. arXiv:quant-ph/0611013.

[27] Masahito Hayashi. Universal coding for classical-quantum channel. *Commun. Math. Phys.*, 289(3):1087–1098, May 2009.

[28] Masahito Hayashi and Hiroshi Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49(7):1753–1768, July 2003. arXiv:quant-ph/0206186.

[29] Masahito Hayashi and Marco Tomamichel. Correlation detection and an operational interpretation of the Rényi mutual information. *Journal of Mathematical Physics*, 57:102201, 2016.

[30] F. Hiai. Matrix analysis: Matrix monotone functions, matrix means, and majorization. *Interdisciplinary Information Sciences*, 16:139–248, 2010.

[31] F. Hiai. Concavity of certain matrix trace and norm functions. *Linear Algebra Appl.*, 439:1568–1589, 2013.

[32] F. Hiai and M. Mosonyi. Different quantum f-divergences and the reversibility of quantum operations. *Rev. Math. Phys.*, 29, 2017.

[33] Fumio Hiai and Dénes Petz. The Golden-Thompson trace inequality is complemented. *Linear Algebra Appl.*, 181:153–185, 1993.

[34] V. Jaksic, Y. Ogata, Y. Pautrat, and C.-A. Pillet. Entropic fluctuations in quantum statistical mechanics. an introduction. In *Quantum Theory from Small to Large Scales, August 2010*, volume 95 of *Lecture Notes of the Les Houches Summer School*. Oxford University Press, 2012.

[35] Hellmuth Kneser. Sur un téorème fondamental de la théorie des jeux. *C. R. Acad. Sci. Paris*, 234:2418–2420, 1952.

[36] Robert Koenig and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103(7):070504, August 2009. arXiv:0903.2838.

[37] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.

[38] F. Kubo and T. Ando. Means of positive linear operators. *Math. Ann.*, 246:205–224, 1980.

[39] E.H. Lieb and W. Thirring. *Studies in mathematical physics*. University Press, Princeton, 1976.

[40] Yongdo Lim and Miklós Pálfia. Matrix power means and the Karcher mean. *Journal of Functional Analysis*, 262(4):1498–1514, 2012.

[41] Mingyan Simon Lin and Marco Tomamichel. Investigating properties of a family of quantum renyi divergences. *Quantum Information Processing*, 14(4):1501–1512, 2015.

[42] K. Matsumoto. A new quantum version of f-divergence. In *Nagoya Winter Workshop 2015: Reality and Measurement in Algebraic Quantum Theory*, pages 229–273, 2018. arXiv:1311.4722.

[43] Milán Mosonyi and Fumio Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, April 2011.

[44] Milán Mosonyi and Tomohiro Ogawa. Strong converse exponent for classical-quantum channel coding. *Communications in Mathematical Physics*, 355(1):373–426, 2017.

[45] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv:1306.3142.

[46] Hiroshi Nagaoka. Strong converse theorems in quantum information theory. *Proceedings of ERATO Workshop on Quantum Information Science*, page 33, 2001. Also appeared in Asymptotic Theory of Quantum Statistical Inference, ed. M. Hayashi, World Scientific, 2005.

[47] Tomohiro Ogawa. An information-spectrum approach to hash properties for quantum states and relation to channel coding. In *The 28th Symposium on Information Theory and its Applications (SITA2015)*, November 2015.

[48] Masanori Ohya, Dénes Petz, and Noboru Watanabe. On capacities of quantum channels. *Probability and Mathematical Statistics*, 17:179–196, 1997.

[49] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports in Mathematical Physics*, 23:57–65, 1986.

[50] József Pitrik and Dániel Virosztek. Quantum Hellinger distances revisited. arXiv:1903.10455, 2019.

[51] Renato Renner. *Security of Quantum Key Distribution.* PhD thesis, Swiss Federal Institute of Technology Zurich, 2005. Diss. ETH No. 16242.

[52] Bariş Nakiboˇ glu. The Renyi capacity and center. arXiv:1608.02424, 2016.

[53] Bariş Nakiboˇ glu. The Augustin capacity and center. arXiv:1803.07937, 2018.

[54] Benjamin Schumacher and Michael D. Westmoreland. Optimal signal ensembles. *Physical Review A*, 63:022308, 2001.

[55] R. Sibson. Information radius. *Z. Wahrscheinlichkeitsth. Verw. Gebiete*, 14:149–161, 1969.

[56] M. Tomamichel. *Quantum Information Processing with Finite Resources*, volume 5 of *Mathematical Foundations, SpringerBriefs in Math. Phys.* Springer, 2016.

[57] Marco Tomamichel and Vincent Y.F. Tan. Second-order asymptotics for the classical capacity of image-additive quantum channels. *Communications in Mathematical Physics*, 338:103–137, 2015.

[58] A. Uhlmann. Endlich dimensionale dichtmatrizen, ii. *Wiss. Z. Karl-Marx-University Leipzig*, 22:139–177, 1973.

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

[60] Mark M. Wilde, Andreas Winter, and Dong 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, October 2014. arXiv:1306.1586.

[61] H. P. Yuen, R. S. Kennedy, and M. Lax. Optimum testing of multiple hypothesis in quantum detection theory. *IEEE Trans. Inf. Theory*, 21(2):125–134, 1975.

[62] Haonan Zhang. Carlen-Frank-Lieb conjecture and monotonicity of α-z Rényi relative entropy. arXiv:1811.01205, 2018.