Chain rules for quantum channels

Mario Berta  
Department of Computing, Imperial College London, UK  
AWS Center for Quantum Computing, Pasadena, USA  
California Institute of Technology, Pasadena, USA

Marco Tomamichel  
Center for Quantum Technologies &  
Department of Electrical and Computer Engineering  
National University of Singapore, Singapore

Abstract—Divergence chain rules for channels relate the divergence of a pair of channel inputs to the divergence of the corresponding channel outputs. An important special case of such a rule is the data-processing inequality, which tells us that if the same channel is applied to both inputs then the divergence cannot increase. Based on direct matrix analysis methods, we derive several Rényi divergence chain rules for channels in the quantum setting. Our results simplify and in some cases generalise previous derivations in the literature.

Index Terms—Quantum information measures, quantum Rényi divergences, quantum entropy inequalities, quantum statistics, quantum channel discrimination.

I. INTRODUCTION

A family of fundamental information measures is given by the Rényi divergences [34], defined for probability distributions $P, Q$ over finite alphabets $X$ and $\alpha \in (1, \infty)$ as

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \log Q_\alpha(P\|Q) \quad \text{with}$$

$$Q_\alpha(P\|Q) = \sum_{x \in X} P(x) \left(\frac{P(x)}{Q(x)}\right)^{\alpha-1},$$ (1)

if $P$ is absolutely continuous with respect to $Q$ and as $+\infty$ otherwise. This definition extends via the respective limits to $\alpha \to 1$ (the Kulback-Leibler divergence) and $\alpha \to \infty$ (the max-divergence). They are respectively given as

$$D(P\|Q) = D_1(P\|Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad \text{and}$$

$$D_\infty(P\|Q) = \max_x P(x) \log \frac{P(x)}{Q(x)}.$$ (2)

For $\alpha \in (0, 1)$ we rewrite the sum as $\sum_{x \in X} P(x)^\alpha Q(x)^{1-\alpha}$, indicating that absolutely continuity of $P$ with respect to $Q$ is not necessary to keep $D_\alpha$ finite for $\alpha < 1$.

For distributions over bipartite alphabets $X \times Y$ the Rényi divergences enjoy the chain rule property

$$D_\alpha(P_{XY}\|Q_{XY}) = D_\alpha(P_X\|Q_X) + \max_{x \in X} D_\alpha(P_{Y|X=x}\|Q_{Y|X=x}),$$ (4)

displaying how to split up Rényi divergences with respect to multipartite constituents. Eq. (4) is important in statistics because it can be seen as the crucial entropy inequality to quantify the asymptotic error exponents of adaptive channel discrimination [7], [22], [33], [35], [42].

In this report, we give various quantum generalizations of Eq. (4), where probability distributions and conditional probability distributions are replaced with positive operators and positive maps, respectively. Some of these quantum chain rules have been derived previously in the context of adaptive quantum channel discrimination and related tasks [10], [14], [15], [17]. The proofs of these earlier results are based on involved techniques from either complex interpolation theory [10], smooth entropies [15], or convex optimization theory [17]. In contrast, our derivations are based on direct matrix analysis methods.

The remainder of this report is structured as follows. First, we introduce in Section II our notation and state some mathematical facts from quantum information theory. We then state our main result in Section III, after which we sketch applications to quantum channel discrimination in Section IV. We conclude with some open question in Section V.

II. NOTATION AND METHODS

We consider quantum systems described by finite-dimensional inner product spaces labelled by $\mathcal{H}$. The set $S(\mathcal{H})$ of quantum states is given by positive semi-definite trace one operators on $\mathcal{H}$. For $\rho, \sigma \in S(\mathcal{H})$, we denote by $\rho \ll \sigma$ that the support of $\rho$ is contained in the support of $\sigma$, and inverses of operators are understood as generalized inverses on the support of the operator. Quantum channels are completely positive and trace preserving maps from $S(\mathcal{H})$ to $S(\mathcal{H'})$. Bipartite quantum systems are described by tensor product spaces $\mathcal{H}_A \otimes \mathcal{H}_R$ with correspondingly labelled quantum states $\rho_{AR} \in S(\mathcal{H}_A \otimes \mathcal{H}_R)$.

An important tool to lift entropy inequalities from the classical to the quantum setting is given by the asymptotic spectral pinching method (see [21]). Namely, the spectral pinching map $P_\sigma(\cdot)$ with respect to $\sigma \in S(\mathcal{H})$ is given as

$$P_\sigma : \rho \mapsto \sum_{\lambda} P_\lambda \rho P_\lambda,$$ (5)

where $\sigma = \sum_{\lambda} \lambda P_\lambda$ with eigenvalues $\lambda \in \text{spec}(\sigma)$ in the spectrum of $\sigma$ and corresponding mutually orthogonal eigenspectra $P_\lambda$. An alternative representation of the pinching map is

$$P_\sigma(\cdot) = \int \mu(dt) \sigma^{it}(\cdot)\sigma^{-it}.$$ (6)
for some probability measure $\mu$ on $\mathbb{R}$ [37] Lemma 2.1. We have that $P_{\sigma}(\rho)$ commutes with $\sigma$, $\text{Tr}[P_{\sigma}(\rho)|\sigma] = \text{Tr}[\rho|\sigma]$, and the pinching inequality [21]

$$P_{\sigma}(\rho) \geq |\text{spec}(\sigma)|^{-1} \rho,$$

(7)

where here and henceforth $\geq$ denotes the Löwner order, and $|\text{spec}(\sigma)|$ the number of distinct eigenvalues $\sigma$. Eq. (7) becomes asymptotically powerful as $|\text{spec}(\sigma^{\otimes n})| = O(\text{poly}(n))$ by a type-counting argument [12].

The generalization of the Rényi divergences from Eq. (1) to the quantum setting is not unique. Rather, one asks for any quantum Rényi divergence

$$D_{\alpha}(\rho\|\sigma) = \frac{1}{\alpha-1} \log Q_{\alpha}(\rho\|\sigma) \quad \text{on } \rho, \sigma \in \mathcal{S}(\mathcal{H})$$

(8)

to satisfy the following properties:

1) whenever $\rho$ and $\sigma$ commute, $Q_{\alpha}(\rho\|\sigma)$ simplifies to the corresponding classical functions in Eq. (4) and, moreover, for classical quantum states $\rho = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$ and $\sigma = \sum_x q_x |x\rangle\langle x| \otimes \sigma_x$, we have

$$Q_{\alpha}(\rho\|\sigma) = \sum_x p_x^\alpha q_x^{1-\alpha} Q_{\alpha}(\rho_x\|\sigma_x);$$

(9)

2) one has the data processing inequality

$$D_{\alpha}(\rho\|\sigma) \geq D_{\alpha}(\mathcal{E}(\rho),\mathcal{E}(\sigma))$$

(10)

for quantum channels $\mathcal{E} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}')$. The smallest such quantum divergences for $\alpha > 0$ are the measured Rényi divergences [31, 41, 13, 23]

$$D_{\alpha}^M(\rho\|\sigma) = \sup_{\mathcal{M}} D_{\alpha}(\mathcal{M}(\rho),\mathcal{M}(\sigma))$$

(11)

for the rank-one projective measurement maps $\mathcal{M}(\cdot) = \sum_x \langle x| \cdot |x\rangle\langle x|/\mathcal{H}$ with the supremum over orthonormal bases $\{|x\rangle\}_{x}$ of $\mathcal{H}$. On the other hand, for $\alpha \in (0,2]$ the largest such quantum divergences are the geometric Rényi divergences [14, 29]

$$\tilde{D}_{\alpha}(\rho\|\sigma) = \frac{1}{\alpha-1} \log \tilde{Q}_{\alpha}(\rho\|\sigma) \quad \text{with}$$

$$\tilde{Q}_{\alpha}(\rho\|\sigma) = \text{Tr} \left[\sigma^{1/2} \left(\sigma^{-1/2} \rho \sigma^{-1/2}\right)^{\alpha} \sigma^{1/2}\right] \quad \text{for } \rho \ll \sigma$$

(12)

and the respective limit for $\alpha \to 1$ as [2, 23]

$$\tilde{D}_1(\rho\|\sigma) = \text{Tr} \left[\rho \log \left(\rho^{1/2} \sigma^{-1/2} \rho^{1/2}\right)\right].$$

(13)

The formal definition from Eq. (12) can be extended to $\alpha > 0$, but only for $\alpha \in (0,2]$ the geometric Rényi divergences have the following operational characterisation.

Lemma 1 (Matsumoto [28, 29]). Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and $\alpha \in (0,2]$. Then, we have

$$\tilde{D}_\alpha(\rho\|\sigma) = \inf_{P\in \mathcal{M}(\mathcal{P})} D_\alpha(P\|Q),$$

(14)

where the infimum is over probability distributions $P, Q$ over finite alphabets $X$, and positive trace preserving maps with $\Gamma(P) = \rho$ and $\Gamma(Q) = \sigma$.

In fact, for the finite decomposition $\sigma^{-1/2} \rho \sigma^{-1/2} = \sum_{x\in X} \lambda_x \Pi_x$ the infimum is achieved as (see [39, Sec. 4.2.3])

$$\Gamma(\cdot) = \sum_{x\in X} \frac{\langle x| \cdot |x\rangle}{Q(x)} \sigma^{1/2} \Pi_x \sigma^{1/2},$$

(15)

$$P(x) = \lambda_x Q(x), \quad Q(x) = \text{Tr}[\sigma \Pi_x].$$

(16)

Additionally, for the axiomatic extension of the classical Rényi divergences from Eq. (1), the smallest quantum divergences that are also

3) additive on product states:

for $\alpha \in [1/2, \infty)$, the sandwiched Rényi divergences [31, 43]

$$\tilde{D}_{\alpha}(\rho\|\sigma) = \frac{1}{\alpha-1} \log \tilde{Q}_{\alpha}(\rho\|\sigma) \quad \text{with}$$

$$\tilde{Q}_{\alpha}(\rho\|\sigma) = \text{Tr} \left[\left(\frac{\rho}{\sigma}\right)^{\alpha} \frac{\rho}{\sigma}\right] \quad \text{for } \rho \ll \sigma$$

(17)

and the respective limits for $\alpha \to 1$ and $\alpha \to \infty$ as

$$\tilde{D}_1(\rho\|\sigma) = \text{Tr} [\rho (\log \rho - \log \sigma)] \quad \text{and}$$

$$\tilde{D}_\infty(\rho\|\sigma) = \inf \{ \lambda \in \mathbb{R} | \rho \leq 2^\lambda \sigma \},$$

(19)

respectively. We note that $\tilde{D}_1(\rho\|\sigma) \neq \tilde{D}_1(\sigma|\rho)$ unless $\rho$ and $\sigma$ commute. Moreover, the formal definition from Eq. (17) can be extended to $\alpha > 0$, but as showcased in [4, Theorem 7] only for $\alpha \in [1/2, \infty)$ one has the data-processing inequality from Eq. (10).

We record the following noteworthy property.

Lemma 2. Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, $\mathcal{E}, \mathcal{F}$ be positive maps from $\mathcal{B}(\mathcal{H})$ to $\mathcal{B}(\mathcal{H}')$, and $\alpha \in (0,\infty)$. Then, we have

$$\tilde{Q}_{\alpha}(\mathcal{E}(\rho\|\sigma)) \leq |\text{spec}(\sigma)|^{\alpha} \tilde{Q}_{\alpha}(\mathcal{E}(\mathcal{P}_\alpha(\rho))\|\mathcal{E}(\sigma)).$$

(20)

Proof. We follow [39, Eq. (4.52)]. From the pinching inequality in Eq. (7) together with the positivity of $\mathcal{E}$ we get

$$\mathcal{E}(\rho) \preceq |\text{spec}(\sigma)|^\alpha \mathcal{E}(\mathcal{P}_\alpha(\rho)),$$

(21)

Multiplying this from the left and the right with the positive operator $\mathcal{F}(\sigma) = \mathcal{E}(\mathcal{P}_\alpha(\rho))$ leads to

$$\mathcal{F}(\sigma) \preceq |\text{spec}(\sigma)|^\alpha \mathcal{E}(\mathcal{P}_\alpha(\rho)) \mathcal{F}(\sigma),$$

(22)

The claim then follows by the monotonicity of the trace of the monotone function $t \mapsto t^\alpha$.

Finally, other definitions of interest include the Petz Rényi divergences [32], the $\alpha$-z-divergences [1], and divergences based on convex optimization methods [9, 17]. We do not directly employ these further definitions in our work, but note that [17, Corollary 5.2] proves quantum extensions of the

---

1Including general measurement maps leads to the same quantities [4, Theorem 4].

2Divergences with properties 1)–3) are also termed relative entropies [19], also see [25] for more details.
chain rule from Eq. \(4\) in terms of divergences based on convex optimization methods.

All of aforementioned quantum Rényi divergences are extended to quantum channels \(\mathcal{E}, \mathcal{F}\) as \([11], [27]\)

\[
\mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F}) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} \mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\rho))
\]

or alternatively as the (generally larger) stabilized version \([7], [22]\)

\[
\mathcal{D}_{\text{stab}}^{\infty}(\mathcal{E}\|\mathcal{F}) = \sup_{\rho \in \mathcal{S}(\mathcal{H})} \mathcal{D}_\alpha((\mathcal{E}_A \otimes \mathcal{I}_R)(\rho_{AR})\|((\mathcal{F}_A \otimes \mathcal{I}_R)(\rho_{AR})),
\]

where \(\mathcal{I}_R\) denotes the identity channel on \(\mathcal{S}(\mathcal{H}_R)\), and the supremum is over all \(\rho_{AR} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_R)\) and finite dimensional \(\mathcal{H}_R\)\(^3\). In the general quantum case, the divergences \(\mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F})\) and \(\mathcal{D}_{\text{stab}}^{\infty}(\mathcal{E}\|\mathcal{F})\) are not additive on product channels (see, e.g., \([15], [20]\)), and thus one defines the corresponding regularized versions

\[
\mathcal{D}_\alpha^{\infty}(\mathcal{E}\|\mathcal{F}) = \lim_{n \to \infty} \frac{1}{n} \mathcal{D}_\alpha(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n})
\]

\[
\mathcal{D}_{\text{stab}}^{\infty}(\mathcal{E}\|\mathcal{F}) = \lim_{n \to \infty} \frac{1}{n} \mathcal{D}_{\text{stab}}^{\infty}(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n})
\]

Note that these limits are well-defined whenever \(\mathcal{D}_\alpha^{\infty}(\mathcal{E}\|\mathcal{F})\) is finite. To see this consider the sequences \(f_n = \frac{1}{n} \mathcal{D}_\alpha(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n})\) where \(n \in \mathbb{N}\). We first note that \(f_n\) is bounded from above by \(\mathcal{D}_\alpha^{\infty}(\mathcal{E}\|\mathcal{F})\) for all \(n\) since \(\mathcal{D}_\alpha(\rho\|\sigma) \leq \mathcal{D}_\alpha^{\infty}(\rho\|\sigma)\) for any pair of states (see, e.g., \([39]\)) and, furthermore, \(\mathcal{D}_\alpha^{\infty}(\mathcal{E}\|\mathcal{F})\) is additive for tensor product channels. We further get \(f_{n+1} \geq \frac{n}{n+1} f_n\) by choosing product inputs and hence the function is monotonic asymptotically and there can only be a single accumulation point.

### III. MAIN RESULTS

Our main result is the following Rényi divergences based simple meta chain rule for quantum channels.

**Proposition 3.** Let \(\rho, \sigma \in \mathcal{S}(\mathcal{H}), \mathcal{E}, \mathcal{F}\) be positive maps from \(\mathcal{B}(\mathcal{H})\) to \(\mathcal{B}(\mathcal{H'})\), and \(\alpha \in (0, \infty]\). Then, we have

\[
\mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\sigma)) \leq \mathcal{D}_\alpha(\rho\|\sigma) + \mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F}),
\]

where \(\mathcal{D}_\alpha\) denotes any Rényi divergence of order \(\alpha\).

**Proof.** Using Matsumoto’s construction (Lemma [1]), we write

\[
\mathcal{E}(\rho) = \sum_x \hat{p}_x \mathcal{E}(\Gamma(|x\rangle\langle x|)) \quad \text{and} \quad \mathcal{F}(\sigma) = \sum_x \hat{q}_x \mathcal{F}(\Gamma(|x\rangle\langle x|))
\]

We then use the data-processing inequality to get

\[
\mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\sigma)) \leq \mathcal{D}_\alpha \left( \sum_x \hat{p}_x |x\rangle\langle x| \otimes \mathcal{E}(\rho^x) \middle| \sum_x \hat{q}_x |x\rangle\langle x| \otimes \mathcal{F}(\rho^x) \right)
\]

\[
= \frac{1}{\alpha - 1} \log \left( \sum_x \hat{p}_x \hat{q}_x^{1 - \alpha} \mathcal{D}_\alpha(\mathcal{E}(\rho^x)\|\mathcal{F}(\rho^x)) \right)
\]

\[
\leq \mathcal{D}_\alpha(\rho\|\sigma) + \max_x \mathcal{D}_\alpha(\rho^x\|\mathcal{F}(\rho^x)),
\]

which implies the statement we set out to prove. \(\Box\)

The corresponding geometric chain rules for \(\alpha \in (0, 2]\) as

\[
\mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\sigma)) \leq \mathcal{D}_\alpha(\rho\|\sigma) + \mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F}),
\]

generalize the stabilized geometric chain rules from \([14]\) Lemma 7 that were proven with different techniques for completely positive maps \(\mathcal{E}_A, \mathcal{F}_A\) and \(\alpha \in (1, 2]\) as

\[
\mathcal{D}_\alpha((\mathcal{E}_A \otimes \mathcal{I}_R)(\rho_{AR})\|((\mathcal{F}_A \otimes \mathcal{I}_R)(\sigma_{AR})) \leq \mathcal{D}_\alpha(\rho_{AR}\|\sigma_{AR}) + \mathcal{D}_\alpha^{\text{stab}}(\mathcal{E}_A\|\mathcal{F}_A),
\]

as well as their extension to \(\alpha \in (0, 1]\) \([25]\) Proposition 45. We note that the stabilized version of the channel divergence was required in \([34]\), whereas our Proposition 3 does not require stabilization.

To obtain chain rules without featuring geometric Rényi divergences, we can apply pre-processing maps in terms of rank-one projective measurements \(\Lambda(\cdot)\). Namely, for \(\rho, \sigma \in \mathcal{S}(\mathcal{H})\) we have that the pre-processed inputs \(\Lambda(\rho), \Lambda(\sigma)\) commute and thus we find the chain rule

\[
\mathcal{D}_\alpha(\mathcal{E}(\Lambda(\rho))\|\mathcal{F}(\Lambda(\sigma))) \leq \mathcal{D}_\alpha,\Lambda(\rho\|\sigma) + \mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F}).
\]

We can then again eliminate the pre-processing map for the sandwiched Rényi divergences with \(\alpha \in (1, \infty]\) by choosing the overall rank-one projective measurement map \(\Lambda(\cdot) = \mathcal{M}_\sigma(P_{\sigma}(\cdot))\) with the pinching map \(P_{\sigma}(\cdot)\) and the projective rank-one measurement map \(\mathcal{M}_\alpha(\cdot)\) in the joint eigenbasis of \(\sigma\) and \(P_{\sigma}(\cdot)\)—then leaving those states invariant. Applying Lemma [2] for \(\alpha \in (1, \infty]\) immediately gives

\[
\mathcal{D}_\alpha(\mathcal{E}(\Lambda(\rho))\|\mathcal{F}(\Lambda(\sigma))) \geq \mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\sigma)) - \frac{\alpha}{\alpha - 1} \log |\text{spec}(\sigma)|
\]

and thus we find the following sandwiched chain rule.

**Corollary 4.** Let \(\rho, \sigma \in \mathcal{S}(\mathcal{H}), \mathcal{E}, \mathcal{F}\) be positive maps from \(\mathcal{B}(\mathcal{H})\) to \(\mathcal{B}(\mathcal{H'})\), and \(\alpha \in (1, \infty]\). Then, we have

\[
\mathcal{D}_\alpha(\mathcal{E}(\rho)\|\mathcal{F}(\sigma)) \leq \mathcal{D}_\alpha,\mathcal{M}(\rho\|\sigma) + \mathcal{D}_\alpha(\mathcal{E}\|\mathcal{F}) + \frac{\alpha}{\alpha - 1} \log |\text{spec}(\sigma)|.
\]
As a special case, applying above corollary to unital maps $\mathcal{F}$, the identity matrix $\sigma = 1$ and $\rho \in S(\mathcal{H})$, leads for $\alpha \in [1, \infty)$ to

$$H_\alpha(\mathcal{E}(\rho)) - H_\alpha(\rho) \geq \inf_{\mathcal{F} : \text{unital}} \tilde{D}_\alpha(\mathcal{E} || \mathcal{F}) ,$$

featuring the quantum Rényi entropies $H_\alpha(\rho) = \frac{1}{\alpha} \log \text{Tr} [\rho^\alpha]$ and in particular the von Neumann entropy $H_1(\rho) = -\text{Tr} [\rho \log \rho]$ in the limit $\alpha \to 1$. This generalizes the well-known monotonicity property $H_\alpha(\mathcal{E}(\rho)) - H_\alpha(\rho) \geq 0$ for unital quantum channels $\mathcal{E}$ to the Eq. (38) for generic positive maps.

Going back to the general case, but restricting our attention to tensor-stable positive maps $\mathcal{E}$ and $\mathcal{F}$, such that $\mathcal{E} \otimes n$ and $\mathcal{F} \otimes n$ are positive maps. Applying Corollary 4 to $\rho \otimes n$, $\sigma \otimes n$, $\mathcal{E} \otimes n$ and $\mathcal{F} \otimes n$, and using that $|\text{spec}(\sigma \otimes n)| \leq \text{poly}(n)$, we arrive at the corresponding regularized sandwiched chain rule.

**Corollary 5.** Let $\rho, \sigma \in S(\mathcal{H}), \mathcal{E}, \mathcal{F}$ be tensor-stable positive maps from $B(\mathcal{H})$ to $B(\mathcal{H})$, and $\alpha \in (1, \infty]$. Then, we have

$$\tilde{D}_\alpha(\mathcal{E}(\rho) || \mathcal{F}(\sigma)) \leq \tilde{D}_\alpha(\rho || \sigma) + \tilde{D}_\infty(\mathcal{E} || \mathcal{F}) .$$

(39)

This can be compared to [10 Theorem III.1] of which a special case states for the same range $\alpha \in (1, \infty]$ as in Corollary 5—but for positive trace preserving maps instead of tensor-stable maps—that

$$\tilde{D}_\alpha(\mathcal{E}(\rho) || \mathcal{F}(\sigma)) \leq \tilde{D}_\alpha(\rho || \sigma) + \tilde{D}_\infty(\mathcal{E} || \mathcal{F}) .$$

(40)

Applied to quantum channels, this is weaker than Corollary 5 as the sandwiched Rényi divergences are monotone in $\alpha \in [1/2, \infty]$ and one has $\tilde{D}_\infty(\mathcal{E} || \mathcal{F}) = \tilde{D}_\infty(\mathcal{E} || \mathcal{F})$ [42 Proposition 10]. Corollary 5 also generalizes the stabilized sandwiched chain rules from [17 Corollary 5.2] that hold for completely positive maps $\mathcal{E}_A, \mathcal{F}_A, |R| = |A|$, and $\alpha \in (1, \infty]$ as

$$\tilde{D}_\alpha((\mathcal{E}_A \otimes I_R)(\rho_{AR}) || (\mathcal{F}_A \otimes I_R)(\sigma_{AR})) \leq \tilde{D}_\alpha(\rho_{AR} || \sigma_{AR}) + \tilde{D}_\alpha((\mathcal{E}_A || \mathcal{F}_A))$$

(41)

and were proven with different techniques.

Note that for $\alpha \in (0, 1]$ it is unfortunately unclear how to eliminate the pre-processing map and directly regularizing Eq. (35) for general $\alpha \in [0, \infty]$ just leads to

$$\limsup_{n \to \infty} \sup_{\Lambda_n} \frac{1}{n} \tilde{D}_\alpha(\mathcal{E} \otimes n (\Lambda^n(\rho \otimes n)) || \mathcal{F} \otimes n (\Lambda^n(\sigma \otimes n))) \leq \tilde{D}_\alpha(\rho || \sigma) + \tilde{D}_\infty(\mathcal{E} || \mathcal{F}) ,$$

(42)

where the supremum is over all rank-one projective measurement maps $\Lambda^n(\cdot)$ acting on $n$ copies. It is then only for $\alpha \in (1, \infty]$ that the left-hand side of Eq. (42) is further lower bounded by $\tilde{D}_\alpha(\mathcal{E}(\rho) || \mathcal{F}(\sigma))$ (Lemma 2), whereas this

remains unclear for $\alpha \in (0, 1]$. Notwithstanding, at least for $\alpha = 1$ it is known that [15 Theorem 3.5]

$$\tilde{D}_1(\mathcal{E}(\rho) || \mathcal{F}(\sigma)) \leq \tilde{D}_1(\rho || \sigma) + \tilde{D}_\infty(\mathcal{E} || \mathcal{F}) .$$

(43)

However, this does not directly follow from our Rényi divergences chain rules, as the order of taking the limits in $n \to \infty$ and $\alpha \to 1$ might ultimately matter when regularizing Eq. (37). We briefly get back to this question in Section V and present some conjectures.

**IV. APPLICATIONS**

Rényi divergences chain rules such as Eq. (4) and its quantum extensions allow to characterize the asymptotic error exponents of adaptive channel discrimination [7], [14], [15], [17], [22], [33], [35], [42]. Namely, the crucial quantities to study are the so-called amortized channel divergences [7], [10], [42]

$$D_\alpha^a(\mathcal{E} || \mathcal{F}) = D_\alpha(\mathcal{E} || \mathcal{F}) .$$

(45)

This then means that adaptive or correlated input strategies are of no asymptotic advantage compared to product strategies and one immediately obtains a full asymptotic characterization of channel discrimination (see aforementioned references for details). In the general quantum case, one only has the stabilized and regularized chain rules as in Corollary 5 and for $\alpha \in (1, \infty]$ one then derives [15, 17]

$$\tilde{D}_\alpha^a(\mathcal{E} || \mathcal{F}) = \tilde{D}^\text{stab, \infty}_\alpha(\mathcal{E} || \mathcal{F}) ,$$

(46)

which includes $\alpha = 1$ thanks to [15 Theorem 3.5]. Consequently, this shows that adaptive strategies are of no asymptotic advantage over entangled input strategies—with the regularization reflecting that the latter are still needed compared to product input strategies [15]. Nonetheless, for a full asymptotic characterization of channel discrimination technical question concerning the $\alpha \to 1$ limit remain [17] (cf. the discussion in Section V). Alternatively, one might resort to single-letter converse bounds as provided by the geometric chain rules Eqs. (33)–(34) giving that

$$\tilde{D}_\alpha^a(\mathcal{E} || \mathcal{F}) = \tilde{D}^\text{stab}_\alpha(\mathcal{E} || \mathcal{F}) .$$

(47)

We refer to [14] for details. One advantage, of our novel geometric chain rule in Eq. (33) is that it does not require

4The definition for $D_\alpha(\rho || \sigma)$ and its properties straightforwardly apply to general positive operators $\sigma$ on $\mathcal{H}$ as well.

5In this context, Lemma 2 is not useful for $\alpha \in (0, 1]$ as one would have to divide the logarithm on both sides of Eq. (20) with the pre-factor $\alpha - 1$, which becomes negative in this range.

6No bound on the dimension $|R|$ is known.
a stabilization system and hence one might explore channel discrimination tasks without additional quantum memory as done in the classical case \cite{22} (or bounded quantum memory of fixed dimension).

V. OUTLOOK

We hope that our matrix analysis based techniques will be more broadly useful to derive entropy inequalities for quantum channels. Many fundamental questions remain with respect to the goal of giving for $\alpha > 0$ the most general Rényi divergence chain rules for quantum channels. For example, our main results from Section III immediately raise the question if

$$\limsup_{n \to \infty} \frac{1}{n} \tilde{D}_1 \left( \mathcal{E}^{\otimes n} \left( \mathcal{A}^n \left( \rho^{\otimes n} \right) \right) \| \mathcal{F}^{\otimes n} \left( \mathcal{A}^n \left( \sigma^{\otimes n} \right) \right) \right) \geq \tilde{D}_1 \left( \mathcal{E}(\rho) \| \mathcal{F}(\sigma) \right),$$

where the supremum is over all rank-one projective measurement maps $\Lambda^n(\cdot)$ acting on $n$ copies. Choosing the standard pinching based measurement map $\Lambda^n(\cdot) = \mathcal{M}_{\sigma^{\otimes n}}(P_{\sigma^{\otimes n}}(\cdot))$ in the form of Eq. \eqref{eq:pinch} only leads to

$$\frac{1}{n} \sup_{\Lambda^n} \tilde{D}_1 \left( \mathcal{E}^{\otimes n} \left( \mathcal{A}^n \left( \rho^{\otimes n} \right) \right) \| \mathcal{F}^{\otimes n} \left( \mathcal{A}^n \left( \sigma^{\otimes n} \right) \right) \right) \geq \frac{1}{n} \tilde{D}_1 \left( \mathcal{E}^{\otimes n} \left( \int \mu(dt) \left( \sigma^{it} \rho \sigma^{-it} \right)^{\otimes n} \right) \| \left( \mathcal{F}(\sigma) \right)^{\otimes n} \right) \geq \min_t \tilde{D}_1 \left( \mathcal{E} \left( \sigma^{it} \rho \sigma^{-it} \right) \| \mathcal{F}(\sigma) \right) - \frac{1}{n} \log \text{poly}(n),$$

where we used in the last inequality that by Carathéodory’s theorem the integral $\int \mu(dt) (\cdot)^{\otimes n}$ is discretized with most poly$(n)$ terms [5] Theorem D.5 and can then by [5] Lemma 2.5] be taken outside of the divergence at the price of the asymptotically small $\log \text{poly}(n)$ fudge term. The appearance of unitary operators like $\sigma^\mu$ is reminiscent of previous work on quantum entropy inequalities \cite{8, 10, 12, 16, 18, 19}, but here we would need to eliminate them.

As the $\alpha \in (1, \infty]$ analogue of Eq. \eqref{eq:chain_rule} holds by Corollary 4 we might also ask more generally what happens for the whole parameter range $\alpha \in (0, 1]$. This is strongly connected with the question if $\tilde{D}_1^{\alpha,\infty} (\mathcal{E}^{\otimes n} \| \mathcal{F}^{\otimes n})$ is continuous for $\alpha \to 1$ \cite{17}. Any insight on this question would have widespread applications in quantum information theory \cite{16} — starting from the fundamental statistical task of channel discrimination.

Note added: After submitting our paper, the work \cite{30} on generalised entropy accumulation appeared. A crucial proof step is their complementary Rényi divergence chain rule \cite{30} Theorem 3.1] for quantum channels and it would be interesting to explore the interplay with our purely matrix analysis based techniques.

Acknowledgements: This work was completed prior to MB joining the AWS Center for Quantum Computing. This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore and the Ministry of Education, Singapore under the Research Centres of Excellence programme. MT is also supported in part by NUS startup grants (R-263-000-E32-133 and R-263-000-E32-731). We thank David Sutter for discussions.

APPENDIX

We mention the following, alternative proof of the sandwiched chain rule (Corollary 4). This directly lifts the classical chain rule from Eq. \eqref{eq:chain_rule} to the quantum case.

Proof of Corollary 4 Let $\{ |x \rangle \}_{x}$ be a joint eigenbasis for both $\sigma$ and $\mathcal{P}_\sigma(\rho)$. We write

$$\mathcal{E}(\mathcal{P}_\sigma(\rho)) = \sum_{x} |x \rangle \langle x | \rho |x \rangle \langle x |$$

and by the data-processing for the partial trace we find

$$\tilde{Q}_\alpha \left( \mathcal{E}(\mathcal{P}_\sigma(\rho)) \| \mathcal{F}(\sigma) \right) \leq \sum_{x} p_x |x \rangle \langle x | \sigma^x \| \mathcal{F}(\sigma) \| \mathcal{F}(\sigma) \| \mathcal{F}(\sigma)$$

where in the last inequality we used that $\rho^x$ and $\sigma^x$ are channel outputs for the same input. Combining this with \eqref{eq:chain_rule} and taking the logarithm as well as multiplying by $\frac{1}{\alpha - 1}$, yields the desired result.

REFERENCES

[1] Koenraad M. R. Audenaert and Nilanjana Datta. $\alpha$-z-relative Rényi entropies. Journal of Mathematical Physics, 56:022202, 2015.

[2] V. P. Belavkin and P. Staszewski. C*-algebraic generalization of relative entropy and entropy. Communications in Mathematical Physics, 115(5):050501, 2015.

[3] M. Berta, O. Fawzi, and M. Tomamichel. Exploiting variational formulæ for quantum relative entropy. Letters in Mathematical Physics, 107(12):2239–2265, 2017.

[4] Mario Berta, Fernando G. S. L. Brandão, and Christoph Hirche. On composite quantum hypothesis testing. Communications in Mathematical Physics, 385(1):53–77, 2021.

[5] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. Communications in Mathematical Physics, 306(3):579–615, 2011.

[6] Mario Berta, Christoph Hirche, Eneet Kaur, and Mark M. Wilde. Stein’s lemma for classical-quantum channels. In IEEE International Symposium on Information Theory (ISIT), pages 2844–2848, 2016.

[7] Mario Berta, O. Fawzi, and M. Tomamichel. On variational expressions for quantum relative entropies. Letters in Mathematical Physics, 107(12):2239–2265, 2017.

[8] Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan Oppenheim, and Sergii Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. Physical Review Letters, 115(5):050501, 2015.

[9] Peter Brown, Hanza Fawzi, and Omar Fawzi. Device-independent lower bounds on the conditional von Neumann entropy. 2021.

[10] Matthias Christandl and Alexander Müller-Hermes. Relative entropy bounds on quantum, private and repeater capacities. Communications in Mathematical Physics, 353(2):821–852, 2017.
