Chain Rule for the Quantum Relative Entropy

Kun Fang
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom

Omar Fawzi
Univ Lyon, ENS Lyon, UCBL, CNRS, LIP, F-69342 LYON Cedex 07, France

Renato Renner and David Sutter
Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland

(Received 24 November 2019; accepted 10 February 2020; published 10 March 2020)

The chain rule for the classical relative entropy ensures that the relative entropy between probability distributions on multipartite systems can be decomposed into a sum of relative entropies of suitably chosen conditional distributions on the individual systems. Here, we prove a chain rule inequality for the quantum relative entropy. The new chain rule allows us to solve an open problem in the context of asymptotic quantum channel discrimination: surprisingly, adaptive protocols cannot improve the error rate for asymmetric channel discrimination compared to nonadaptive strategies.

DOI: 10.1103/PhysRevLett.124.100501

Introduction.—The von Neumann entropy $H(A)$ of a quantum system A is a fundamental measure of uncertainty. For example, it characterizes the optimal rates for basic information-theoretic tasks such as compression or entanglement manipulation [1] and it can be used to quantify entanglement and topological order in condensed matter systems [2–4]. The conditional von Neumann entropy $H(A_1|A_2)$ is defined via the relation

$$H(A_1A_2) = H(A_1) + H(A_2|A_1). \tag{1}$$

Iterative use of this defining relation yields expressions such as

$$H(A^n|B) = \sum_{i=1}^{n} H(A_i|A^{i-1}B), \tag{2}$$

as visualized in Fig. 1. This relation, called chain rule, allows us to view the entropy of a large composite system as a sum of entropies of its subsystems. The chain rule property is also crucial in the definition of entanglement measures such as the squashed entanglement [5].

In this Letter we propose a chain rule for the relative entropy defined as

$$D(\rho\|\sigma) \coloneqq \begin{cases} \mathrm{tr}\rho(\log\rho - \log\sigma) & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ +\infty & \text{otherwise,} \end{cases}$$

for any states ρ and σ (where the latter does not need to be normalized). The relative entropy is a more general entropic quantity than the von Neumann entropy. It contains the latter and other information measures, such as the mutual information, as a special case. It can be seen as a dissimilarity measure between quantum states and is used to define various important quantities such as the relative entropy of entanglement [6]. The relative entropy characterizes the error exponent for asymmetric hypothesis testing [7] or quantifies the amount of resource in a resource theory [8,9].

So far no chain rule for the quantum relative entropy has been proven. This is in sharp contrast with the classical case where a chain rule is known [[10] Theorem 2. 5. 3] for the relative entropy (also called the Kullback-Leibler divergence). For a pair of discrete random variables (X, Y) with alphabet $\mathcal{X} \times \mathcal{Y}$, we have

$$\begin{aligned} &D(P_{XY}\|Q_{XY}) \\ &\quad = D(P_X\|Q_X) + \sum_{x\in\mathcal{X}} P_X(x) D(P_{Y|X=x}\|Q_{Y|X=x}), \end{aligned}$$

where P_{XY} and Q_{XY} are joint probability distributions, but Q_{XY} does not need to be normalized. No quantum analog of

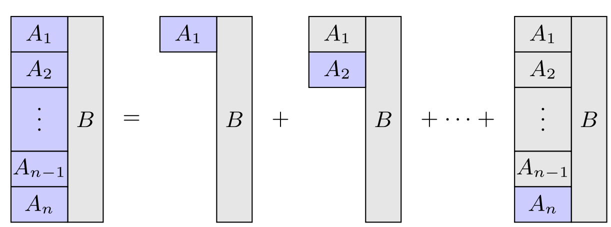

FIG. 1. A graphical visualization of the chain rule for the von Neumann entropy given in (2). The entropy is taken with respect to the blue systems conditioned on the gray systems.

such a chain rule is known, even if we relax the equality with the following inequalities

$$\begin{aligned} D(P_X\|Q_X) &+ \min_{x\in\mathcal{X}} D(P_{Y|X=x}\|Q_{Y|X=x}) \\ &\le D(P_{XY}\|Q_{XY}) \\ &\le D(P_X\|Q_X) + \max_{x\in\mathcal{X}} D(P_{Y|X=x}\|Q_{Y|X=x}). \end{aligned} \quad (3)$$

In this Letter, we prove a quantum version of the upper bound (3) (see Theorem 2). We also show that it is tight in the sense that there exist nontrivial scenarios where the chain rule is an equality (see Corollary 3).

To model the quantum setting, the conditional distributions are replaced by trace-preserving completely positive maps $\mathcal{E}$ and $\mathcal{F}$ from A to B and the initial states are density operators ρ_{RA} and σ_{RA}, where R denotes a reference system. To express the very last term in (3) in the quantum mechanical case we use a quantity called nonstabilized channel relative entropy, which is defined by

$$\bar{D}(\mathcal{E}\|\mathcal{F}) \coloneqq \max_{\phi_A\in \mathrm{S}(A)} D[\mathcal{E}(\phi_A)\|\mathcal{F}(\phi_A)],$$

where $\mathrm{S}(A)$ denotes the set of density operators on A. Its *stabilized* counterpart [11] is defined by $D(\mathcal{E}\|\mathcal{F}) \coloneqq \bar{D}(\mathcal{I}_A\otimes\mathcal{E}\|\mathcal{I}_A\otimes\mathcal{F})$, where $\mathcal{I}_A$ denotes the identity map on A. In the following we will omit identity maps if they are clear from the context. Motivated by the classical case (3), it is natural to ask whether the following chain rule

$$D[\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})] \overset{?}{\le} D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F}) \quad (4)$$

is correct [12].

Limitations on a chain rule.—It turns out that (4) does not hold in general because the channel relative entropy is not additive under the tensor product as shown next.

Proposition 1. There exist trace-preserving completely positive maps $\mathcal{E}$, $\mathcal{F}$ such that

$$D(\mathcal{E}\otimes\mathcal{E}\|\mathcal{F}\otimes\mathcal{F}) > 2D(\mathcal{E}\|\mathcal{F}). \quad (5)$$

This implies that there exist density operators ρ_{RA}, σ_{RA} for some finite-dimensional system R such that

$$D[\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})] > D(\rho_{RA}\|\sigma_{RA}) + D(\mathcal{E}\|\mathcal{F}). \quad (6)$$

Before proving the assertion we introduce another quantity called amortized channel relative entropy [13] defined by

$$\begin{aligned} D^A(\mathcal{E}\|\mathcal{F}) \coloneqq \sup_{\phi_{RA},\omega_{RA}\in \mathrm{S}(R\otimes A)} &\{D[\mathcal{E}(\phi_{RA})\|\mathcal{F}(\omega_{RA})] \\ &- D(\phi_{RA}\|\omega_{RA})\}, \end{aligned} \quad (7)$$

where R is a reference system whose dimension is not constrained [14].

Proof of proposition 1]—We start by proving that (5) implies (6). It is known [[15] Theorem 3 and 6] that

$$D(\mathcal{E}\|\mathcal{F}) \le D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \le D^A(\mathcal{E}\|\mathcal{F}),$$

where $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \coloneqq \lim_{n\to\infty}(1/n)D(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n})$. The statement (5) implies that the first inequality can be strict. By definition of the amortized channel relative entropy (7) this directly implies (6).

It thus remains to prove (5). To do so we construct an example of two trace-preserving completely positive maps $\mathcal{E}$ and $\mathcal{F}$ on qubits that satisfy (5). Consider the generalized amplitude damping channel $\mathcal{A}_{\gamma,\beta}(\rho) = \sum_{i=1}^4 A_i\rho A_i^\dagger$ for $\gamma,\beta\in[0,1]$ with the Kraus operators $A_1 = \sqrt{1-\beta}(|0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|)$, $A_2 = \sqrt{\gamma(1-\beta)}|0\rangle\langle 1|$, $A_3 = \sqrt{\beta}(\sqrt{1-\gamma}|0\rangle\langle 0| + |1\rangle\langle 1|)$, and $A_4 = \sqrt{\gamma\beta}|1\rangle\langle 0|$. For the two channels $\mathcal{E} = \mathcal{A}_{0.3,0}$ and $\mathcal{F} = \mathcal{A}_{0.5,0.9}$ their corresponding Choi matrices are given with respect to the computational basis by

$$J^{\mathcal{E}}_{RB} = \begin{pmatrix} 1 & 0 & 0 & \sqrt{0.7} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ \sqrt{0.7} & 0 & 0 & 0.7 \end{pmatrix}$$

$$J^{\mathcal{F}}_{RB} = \begin{pmatrix} 0.55 & 0 & 0 & \sqrt{0.5} \\ 0 & 0.45 & 0 & 0 \\ 0 & 0 & 0.05 & 0 \\ \sqrt{0.5} & 0 & 0 & 0.95 \end{pmatrix}.$$

Due to the joint-convexity of the relative entropy, the maximization of the channel relative entropy can be taken over all pure states. For an arbitrary density matrix $\rho\in \mathrm{S}(A)$ let $|\phi\rangle_{RA} = (\sqrt{\rho_R^{\intercal}}\otimes id_A)|\Omega\rangle_{RA}$ be its purification where $|\Omega\rangle_{RA} = \sum_i |i\rangle_R|i\rangle_A$ and where R is isomorphic to A and thus

$$\begin{aligned} \mathcal{E}(\phi_{RA}) &= (\mathcal{I}\otimes\mathcal{E})\left[\left(\sqrt{\rho_R^{\intercal}}\otimes id_A\right)|\Omega\rangle\langle\Omega|_{RA}\left(\sqrt{\rho_R^{\intercal}}\otimes id_A\right)\right] \\ &= \sqrt{\rho_R^{\intercal}}J^{\mathcal{E}}_{RB}\sqrt{\rho_R^{\intercal}}. \end{aligned}$$

Hence we find

$$\begin{aligned} D(\mathcal{E}\|\mathcal{F}) &= \max_{\rho_R\in \mathrm{S}(R)} D\left(\sqrt{\rho_R^{\intercal}}J^{\mathcal{E}}_{RB}\sqrt{\rho_R^{\intercal}}\middle\|\sqrt{\rho_R^{\intercal}}J^{\mathcal{F}}_{RB}\sqrt{\rho_R^{\intercal}}\right), \\ &= \max_{\rho_R=\mathrm{diag}(p,1-p)} D(\sqrt{\rho_R}J^{\mathcal{E}}_{RB}\sqrt{\rho_R}\|\sqrt{\rho_R}J^{\mathcal{F}}_{RB}\sqrt{\rho_R}). \end{aligned}$$

The final step follows since both $\mathcal{E}$ and $\mathcal{F}$ are covariant with respect to the Pauli-Z operator. Thus it suffices to perform the maximization over input states with respect to the

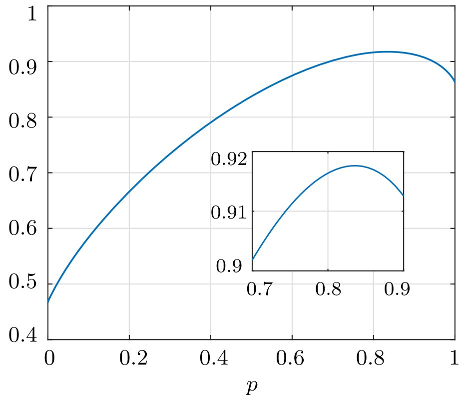

FIG. 2. Plot of the value $D(\sqrt{\rho_R}J^{\mathcal{E}}_{RB}\sqrt{\rho_R}\|\sqrt{\rho_R}J^{\mathcal{F}}_{RB}\sqrt{\rho_R})$ with respect to the input state $\rho_R = \mathrm{diag}(p, 1-p)$. The subfigure is an enlargement of the large plot. It is evident that $D(\mathcal{E}\|\mathcal{F})$ cannot be larger than 0.92.

one-parameter family of states $\rho_R = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ (see, e.g., [[16] Proposition II. 4.]). Using the FMINBND function in MATLAB, we find $D(\mathcal{E}\|\mathcal{F}) = 0.9176$ for an optimizer $\rho_R = \mathrm{diag}(0.8355, 1-0.8355)$. This can also be seen by plotting the value of the relative entropy over the interval $p \in [0,1]$ as shown in Fig. 2.

On the other hand, if we choose the input state $\rho_{R_1R_2} = \mathrm{diag}(0.8, 0, 0, 0.2)$ we have

$$\begin{aligned} &D(\mathcal{E}\otimes\mathcal{E}\|\mathcal{F}\otimes\mathcal{F}) \\ &\quad \geq D(\sqrt{\rho_{R_1R_2}}(J^{\mathcal{E}}_{RB})^{\otimes 2}\sqrt{\rho_{R_1R_2}}\|\sqrt{\rho_{R_1R_2}}(J^{\mathcal{F}}_{RB})^{\otimes 2}\sqrt{\rho_{R_1R_2}}) \\ &\quad = 1.9362 > 2\times 0.92 > 2D(\mathcal{E}\|\mathcal{F}), \end{aligned}$$

showing that the stabilized channel relative entropy is not additive under the tensor product. ■

More generally, as shown in Fig. 3, we can plot the difference $D(\mathcal{A}^{\otimes 2}_{\gamma_1,0}\|\mathcal{A}^{\otimes 2}_{\gamma_2,0.9}) - 2D(\mathcal{A}_{\gamma_1,0}\|\mathcal{A}_{\gamma_2,0.9})$ for a wide range of γ_1, γ_2. Exploiting the symmetry that $\mathcal{A}_{\gamma,\beta}$ is

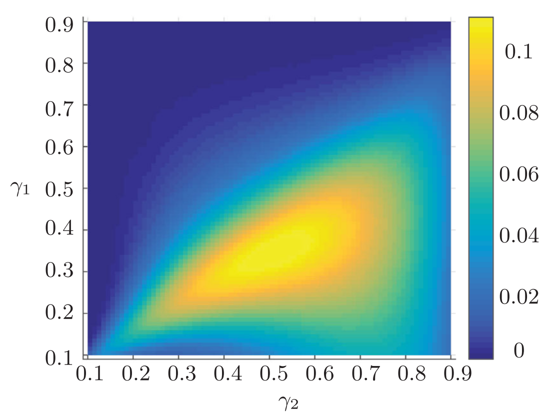

FIG. 3. A heat map of the value $D(\mathcal{A}^{\otimes 2}_{\gamma_1,0}\|\mathcal{A}^{\otimes 2}_{\gamma_2,0.9}) - 2D(\mathcal{A}_{\gamma_1,0}\|\mathcal{A}_{\gamma_2,0.9})$ where $\gamma_1, \gamma_2 \in [0.1, 0.9]$. This shows that the nonadditivity behavior of the stabilized channel relative entropy under tensor products occurs for many channels.

covariant with respect to the Pauli-Z operator and the tensor product channel $\mathcal{A}^{\otimes 2}_{\gamma,\beta}$ is also covariant under permutation, we can restrict the computation of $D(\mathcal{A}^{\otimes 2}_{\gamma_1,0}\|\mathcal{A}^{\otimes 2}_{\gamma_2,0.9})$ to a two-parameter state $\rho_{R_1R_2} = \mathrm{diag}(p_1, p_2, p_2, 1-p_1-2p_2)$ (see, e.g., [[16] Proposition II. 4.]) and we utilize the function QUANTUM_REL_ENTR from CVXQUAD [17]. We observe that the relative entropy is not additive for a wide range of parameters.

Proposition 1 justifies the definition of a (nonstabilized) regularized channel relative entropy as $\bar{D}^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \coloneqq \lim_{n\to\infty}(1/n)\bar{D}(\mathcal{E}^{\otimes n}\|\mathcal{F}^{\otimes n})$ and similarly for the stabilized quantity. This contrasts with the relative entropy for states that is additive under the tensor product.

Chain rule.—The main result of this Letter ensures that (4) becomes valid if we replace the channel relative entropy term with its regularized version. More precisely, the inequality is correct whenever $D_{\max}(\mathcal{E}\|\mathcal{F}) \coloneqq \max_{\phi_{RA}\in \mathrm{S}(R\otimes A)} D_{\max}[\mathcal{E}(\phi_{RA})\|\mathcal{F}(\phi_{RA})]$ is finite, where $D_{\max}(\rho\|\sigma) \coloneqq \inf\{\lambda\in\mathbb{R} : \rho \leq 2^{\lambda}\sigma\}$ is the max-relative entropy [18,19] and R is isomorphic to A.

Theorem 2 (Chain rule for relative entropy). Let ρ_A, σ_A be density operators and $\mathcal{E}$, $\mathcal{F}$ be trace-preserving completely positive maps such that $D_{\max}(\mathcal{E}\|\mathcal{F}) < \infty$. Then

$$D[\mathcal{E}(\rho_A)\|\mathcal{F}(\sigma_A)] \leq D(\rho_A\|\sigma_A) + \bar{D}^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}). \tag{8}$$

In addition, in case $A = A_1\otimes A_2$ and $\mathcal{F}$ is such that its output does not depend on the input on A_1, this inequality can be strengthened to

$$D[\mathcal{E}(\rho_{A_1A_2})\|\mathcal{F}(\sigma_{A_1A_2})] \leq D(\rho_{A_2}\|\sigma_{A_2}) + \bar{D}^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}). \tag{9}$$

Normalization properties of the relative entropy ensure that the chain rule remains valid if σ is not normalized and $\mathcal{F}$ is not trace preserving.

Proof sketch.—The full proof can be found in the Supplemental Material [20], which includes references [21–25]. Instead we sketch the proof idea. We start with the observation that the chain rule follows for the max-relative entropy from the triangle inequality [[26] Lemma 2. 1]—a property that does not hold for the relative entropy. To see this suppose $\mathcal{F} = \mathcal{R}\circ\mathcal{G}$ for some channels $\mathcal{R}$ and $\mathcal{G}$. Then the triangle inequality together with the data-processing inequality for the max-relative entropy [27] give

$$\begin{aligned} &D_{\max}[\mathcal{E}(\rho)\|\mathcal{F}(\sigma)] \\ &\quad \leq D_{\max}[\mathcal{E}(\rho)\|\mathcal{F}(\rho)] + D_{\max}[\mathcal{F}(\rho)\|\mathcal{F}(\sigma)], \\ &\quad = D_{\max}[\mathcal{E}(\rho)\|\mathcal{F}(\rho)] + D_{\max}[\mathcal{R}\circ\mathcal{G}(\rho)\|\mathcal{R}\circ\mathcal{G}(\sigma)], \\ &\quad \leq D_{\max}[\mathcal{G}(\rho)\|\mathcal{G}(\sigma)] + D_{\max}[\mathcal{E}(\rho)\|\mathcal{F}(\rho)]. \end{aligned} \tag{10}$$

In case $A = A_1\otimes A_2$ and $\mathcal{G}$ is the partial trace over A_1 this has the form of (9). Based on that insight we prove a variant of (10), where all three terms are replaced by smooth-max relative entropies together with an additive error term that depends on the smoothing parameter. Finally we use the asymptotic equipartition property [27,28], which ensures that for n-fold product states the smooth-max relative entropy converges to the relative entropy as $n \to \infty$, to obtain (9) from which (8) follows as the special case where A_1 is trivial.

In case $\mathcal{E} = \mathcal{F}$, inequality (8) simplifies to the data-processing inequality, i.e., $D[\mathcal{E}(\rho)\|\mathcal{E}(\sigma)] \le D(\rho\|\sigma)$. In this sense (8) may be viewed as a generalized data-processing inequality where not necessarily the same channel is applied to both arguments, which then is compensated with the regularized channel relative entropy term.

Note that, in the chain rule (1) the term $H(A_2|A_1)$ still depends on the state of A_1. However, if one instead considers the implication for any fixed $\rho_{A_1A_2}$

$$H(A_1A_2)_\rho \ge H(A_1)_\rho + \min_{\nu_{A_2|A_1}=\rho_{A_2|A_1}} H(A_2|A_1)_\nu, \qquad (11)$$

where the minimization is over all ν whose conditional state is identical to the conditional state of ρ, i.e., $\rho_{A_2|A_1} \coloneqq \rho_{A_1}^{-1/2}\rho_{A_1A_2}\rho_{A_1}^{-1/2}$, the second term becomes independent of the marginal state of A_1. This is particularly desirable for an iterative version analogous to (2), as the terms in the sum then only depend on the correlations between a subsystem A_i and the rest, but not on the state of the rest. One can now see that (11) indeed can be retrieved from our relative entropy chain rule (9) by expressing the conditional von Neumann entropy in terms of the relative entropy, i.e., $H(A_2|A_1) = -D(\rho_{A_1A_2}\|\rho_{A_1} \otimes id_{A_2})$ (see the Supplemental Material [20] for details).

An important corollary to Theorem 2 is when the maps are of the form $\mathcal{I} \otimes \mathcal{E}$ and $\mathcal{I} \otimes \mathcal{F}$. The right-hand side then simplifies to the more common (stabilized) relative entropy between channels.

Corollary 3. Let $\mathcal{E}$ be a trace-preserving completely positive map and $\mathcal{F}$ be a completely positive map. Then

$$D^A(\mathcal{E}\|\mathcal{F}) = D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}). \qquad (12)$$

Proof.—It is known [[15] Theorem 3 and 6] that $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \le D^A(\mathcal{E}\|\mathcal{F})$. Theorem 2 applied to channels $\mathcal{I}_R \otimes \mathcal{E}$ and $\mathcal{I}_R \otimes \mathcal{F}$ shows that

$$D[\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})] - D(\rho_{RA}\|\sigma_{RA}) \le \bar{D}^{\mathrm{reg}}(\mathcal{I}_R \otimes \mathcal{E}\|\mathcal{I}_R \otimes \mathcal{F}).$$

To conclude, it suffices to observe that for any system R we have $\bar{D}^{\mathrm{reg}}(\mathcal{I}_R \otimes \mathcal{E}\|\mathcal{I}_R \otimes \mathcal{F}) \le D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F})$. □

Corollary 3 shows that for any trace-preserving completely positive map $\mathcal{E}$ and any completely positive map $\mathcal{F}$ there exist states ρ_{RA} and σ_{RA} such that the chain rule holds with equality.

The regularization of relative entropy term in Theorem 2 is necessary in full generality, as shown by Proposition 1. However, for channels with a specific structure their stabilized channel relative entropy is additive under the tensor product which implies that $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) = D(\mathcal{E}\|\mathcal{F})$. Examples of such channels are (i) classical-quantum channels [[13] Lemma 25], (ii) covariant channels with respect to the unitary group [[16] Corollary II. 5], (iii) $\mathcal{E}$ arbitrary and $\mathcal{F}$ a replacer channel (i.e., $\mathcal{F}(X) = \omega \mathrm{tr} X$ for $\omega \in \mathrm{S}(B)$) [11,13], and (iv) environment-seizable channels [15].

We can single letterize the chain rule from Theorem 2 by replacing the regularized channel relative entropy term with the Belavkin-Staszewski channel relative entropy defined by $\hat{D}(\mathcal{E}\|\mathcal{F}) \coloneqq \max_{\phi_{RA}\in \mathrm{S}(R\otimes A)} \mathrm{tr}\mathcal{E}(\phi)\log[\mathcal{E}(\phi)^{1/2}\mathcal{F}(\phi)^{-1}\times \mathcal{E}(\phi)^{1/2}]$, where R is isomorphic to A. We note that the logarithmic trace inequality [29,30] (see also [[31] Theorem 4. 6]) ensures that $D(\mathcal{E}\|\mathcal{F}) \le \hat{D}(\mathcal{E}\|\mathcal{F})$. Furthermore, the Belavkin-Staszewski channel relative entropy is additive under tensor products [[32] Lemma 6]. Another benefit from this relaxation is the fact that $\hat{D}(\mathcal{A}\|\mathcal{B})$ has an explicit form and is thus efficiently computable [[32] Lemma 5].

Asymptotic quantum channel discrimination.—A fundamental task in quantum information theory is to distinguish between two quantum channels $\mathcal{E}, \mathcal{F}$. For this problem, one usually differentiates between two different classes of strategies:

Nonadaptive strategies (also called parallel strategies).—Here we are given "black-box" access to n uses of a channel $\mathcal{G}$, which is either $\mathcal{E}$ or $\mathcal{F}$, that can be used in parallel before performing a measurement. More precisely, for an arbitrary state $\rho_{A^nR} \in \mathrm{S}(A_1 \otimes, \ldots, \otimes A_n \otimes R)$ with a reference system R we create the state $\sigma_{B^nR} = \mathcal{G}^{\otimes n}(\rho_{A^nR})$ and perform a measurement on σ_{B^nR}. Based on the measurement outcome we try to guess if $\mathcal{G} = \mathcal{E}$ or $\mathcal{G} = \mathcal{F}$. The protocol is depicted in Fig. 4. It has been shown recently [[15] Theorem 3] that in the asymmetric regime where we fix the type-I error to be bounded by ε, the asymptotic optimal rate of the type-II error exponent is given by $D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F})$, when ε goes to 0. A type-I error is the rejection of a true null hypothesis while a type-II error is the nonrejection of a false null hypothesis.

Adaptive strategies (also called sequential strategies).—Here we are also given "black-box" access to n uses of a channel $\mathcal{G}$ which is either $\mathcal{E}$ or $\mathcal{F}$. However unlike in the

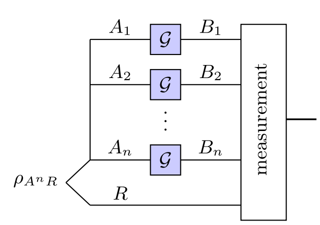

FIG. 4. General protocol for nonadaptive strategies for the task of channel discrimination. The channel $\mathcal{G}$ is either $\mathcal{E}$ or $\mathcal{F}$ and the task is to distinguish between these two cases.

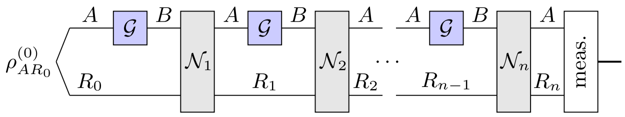

FIG. 5. General protocol for adaptive strategies for the task of channel discrimination.

nonadaptive scenario, after each use of a channel we are allowed to perform an adaptive trace-preserving completely positive map $\mathcal{N}_k$ before we perform a measurement at the end. More precisely, for an arbitrary state $\rho^{(0)}_{AR_0} \in \mathrm{S}(A \otimes R_0)$ we create $\rho^{(k)}_{AR_k} = (\mathcal{N}_k \circ \mathcal{G})(\rho^{(k-1)}_{AR_{k-1}})$ for $k = 1, \ldots, n$. Finally we perform a measurement on $\rho^{(n)}_{AR_n}$ and based on the outcome try to guess if $\mathcal{G} = \mathcal{E}$ or $\mathcal{G} = \mathcal{F}$. The strategy is depicted in Fig. 5. The asymptotically optimal rate of the type-II error exponent for this strategy is given by $D^A(\mathcal{E}\|\mathcal{F})$ [[15] Theorem 6].

Because a nonadaptive strategy can be viewed as a particular instance of an adaptive strategy [33], it follows that adaptive strategies are clearly as powerful as non-adaptive ones, which in technical terms means

$$D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) \leq D^A(\mathcal{E}\|\mathcal{F}).$$

It has been an open question if adaptive strategies can be more powerful for the task of asymptotic quantum channel discrimination [13,15,34,35]. For some special classes of channels, such as classical and classical-quantum channels it has been shown that adaptive protocols cannot improve the error rate for asymmetric channel discrimination [13,34]. Corollary 3 now proves that this is the case for all quantum channels because

$$D^{\mathrm{reg}}(\mathcal{E}\|\mathcal{F}) = D^A(\mathcal{E}\|\mathcal{F}).$$

We note that this is surprising for various reasons. In the symmetric Chernoff setting [36–38] adaptive protocols offer an advantage over nonadaptive ones. Furthermore, in the nonasymptotic setting adaptive protocols also outperform nonadaptive strategies [36,37,39]. Apart from its fundamental importance the channel discrimination problem features immediate applications in various areas ranging from quantum metrology [40] to the study of resource theories [41].

We thank Mario Berta for discussions about asymptotic quantum channel discrimination. We also thank Mark Wilde for presenting the question if adaptive strategies can outperform nonadaptive strategies for the task of asymmetric channel discrimination as an open problem in the Banff workshop on "Algebraic and Statistical ways into Quantum Resource Theories" [35]. K. F. is supported by the University of Cambridge Isaac Newton Trust Early Career Grant No. RG74916. O. F. acknowledges support from the French National Research Agency via Project No. ANR-18-CE47-0011 (ACOM). D. S. and R. R. acknowledge support from the Swiss National Science Foundation via the NCCR QSIT as well as Project No. 200020_165843 and from the Air Force Office of Scientific Research Grant No. FA9550-19-1-0202.

[1] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, England, 2013).
[2] M. B. Hastings, J. Stat. Mech. (2007) P08024.
[3] F. G. Brandão and M. Horodecki, Nat. Phys. **9**, 721 (2013).
[4] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).
[5] M. Christandl and A. Winter, J. Math. Phys. (N.Y.) **45**, 829 (2004).
[6] V. Vedral, Rev. Mod. Phys. **74**, 197 (2002).
[7] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).
[8] M. Berta and C. Majenz, Phys. Rev. Lett. **121**, 190503 (2018).
[9] A. Anshu, M.-H. Hsieh, and R. Jain, Phys. Rev. Lett. **121**, 190504 (2018).
[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley Interscience, New York, 2006).
[11] T. Cooney, M. Mosonyi, and M. M. Wilde, Commun. Math. Phys. **344**, 797 (2016).
[12] One may also ask whether a converse of Eq. (4), i.e., an inequality of the form $D[\mathcal{E}(\rho_{RA})\|\mathcal{F}(\sigma_{RA})] \geq D(\rho_{RA}\|\sigma_{RA}) + f(\mathcal{E},\mathcal{F})$, for some appropriately chosen f, could hold. It follows however from the data processing inequality that $f(\mathcal{E},\mathcal{E})$ cannot be positive, and hence cannot be a relative entropy.
[13] M. Berta, C. Hirche, E. Kaur, and M. M. Wilde, arXiv: 1808.01498.
[14] We note that this is in contrast to the stabilized channel relative entropy, where the reference system R can be assumed to be isomorphic to the input system A. It is unclear if this assumption can be made for the amortized channel relative entropy.
[15] X. Wang and M. M. Wilde, arXiv:1907.06306.
[16] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, Phys. Rev. A **97**, 012332 (2018).
[17] H. Fawzi, J. Saunderson, and P. A. Parrilo, Found. Comput. Math. **19**, 259 (2019).
[18] R. Renner, Ph.D. thesis, ETH Zurich, 2005, arXiv:quant-ph/0512258.
[19] N. Datta, IEEE Trans. Inf. Theory **55**, 2816 (2009).
[20] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.100501 for a complete proof of the main result, i.e., Theorem 2. Furthermore, we present more details on the connections between the chain rule for the relative and the conditional entropies.
[21] R. Bhatia, *Matrix Analysis* (Springer, New York, 1997).
[22] O. Fawzi and R. Renner, Commun. Math. Phys. **340**, 575 (2015).
[23] M. Tomamichel, Ph.D. thesis, ETH Zurich, 2012, arXiv: 1203.2142.
[24] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Trans. Inf. Theory **56**, 4674 (2010).

[25] M. Tomamichel and M. Hayashi, IEEE Trans. Inf. Theory **59**, 7693 (2013).

[26] D. Sutter and R. Renner, Ann. Henri Poincaré **19**, 3007 (2018).

[27] M. Tomamichel, in *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics Vol. 5 (Springer International Publishing, 2015).

[28] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Trans. Inf. Theory **55**, 5840 (2009).

[29] F. Hiai and D. Petz, Linear Algebra Appl. **181**, 153 (1993).

[30] T. Ando and F. Hiai, Linear Algebra Appl. **197–198**, 113 (1994).

[31] D. Sutter, in *Approximate Quantum Markov Chains*, SpringerBriefs in Mathematical Physics Vol. 28 (Springer International Publishing, 2018).

[32] K. Fang and H. Fawzi, arXiv:1909.05758.

[33] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180501 (2008).

[34] M. Hayashi, IEEE Trans. Inf. Theory **55**, 3807 (2009).

[35] M. M. Wilde, Open problems session at the Banff workshop about algebraic and statistical ways into quantum resource theories, https://www.birs.ca/workshops/2019/19w5120/files/19w5120-OpenProblems-20190723.mp4 (2019).

[36] R. Duan, Y. Feng, and M. Ying, Phys. Rev. Lett. **103**, 210501 (2009).

[37] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Phys. Rev. A **81**, 032339 (2010).

[38] R. Duan, C. Guo, C. Li, and Y. Li, in *2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2016), pp. 2259–2263.

[39] D. Puzzuoli and J. Watrous, Ann. Henri Poincaré **18**, 1153 (2017).

[40] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, npj Quantum Inf. **5**, 50 (2019).

[41] R. Takagi and B. Regula, Phys. Rev. X **9**, 031053 (2019).