Stein's Lemma for Classical-Quantum Channels

Mario Berta*, Christoph Hirche†, and Eneet Kaur‡, Mark M. Wilde‡§
*Department of Computing, Imperial College London, United Kingdom
†Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona
‡Hearne Institute for Theoretical Physics, Department of Physics and Astronomy,
Louisiana State University, Baton Rouge, Louisiana 70803, USA
§Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

***Abstract*—It is well known that for the discrimination of classical and quantum channels in the finite, non-asymptotic regime, adaptive strategies can give an advantage over non-adaptive strategies. However, Hayashi [IEEE Trans. Inf. Theory 55(8), 3807 (2009)] showed that in the asymptotic regime, the exponential error rate for the discrimination of classical channels is not improved in the adaptive setting. We show that, for the discrimination of classical-quantum channels, adaptive strategies do not lead to an asymptotic advantage. As our main result, this establishes Stein's lemma for classical-quantum channels. Our proofs are based on the concept of amortized distinguishability of channels, which we analyse using entropy inequalities.**

I. Introduction

A fundamental task in quantum statistics is to distinguish between two (or multiple) non-orthogonal quantum states. After considerable efforts, the resource trade-off is by now well understood in the information-theoretic limit of asymptotically many copies and quantified by quantum Stein's lemma [1], [2], the quantum Chernoff bound [3], [4], as well as refinements thereof [5], [6], [7]. As a natural extension of quantum state discrimination, we study the task of distinguishing between two quantum channels in the information-theoretic limit of asymptotically many repetitions. Whereas the mathematical properties of states and channels are strongly intertwined, channel discrimination is qualitatively different from state discrimination for a variety of reasons. Most importantly, when distinguishing between two quantum channels, one can employ adaptive protocols that make use of a quantum memory [8], as depicted in Figure 1. For the finite, non-asymptotic regime, such protocols are then also known to give an advantage over non-adaptive protocols [9, Sect. 5], the latter of which are restricted to picking a fixed input state and then executing standard state discrimination for the channel outputs [10]. In fact, the advantage of adaptive protocols in this regime already manifests itself for the discrimination of classical channels [9].

Somewhat surprisingly, Hayashi showed that this advantage disappears for classical channel discrimination in the information-theoretic limit of a large number of repetitions [11]. In particular, the optimal exponential error rate for the discrimination of classical channels in the sense of Stein and Chernoff is achieved by just picking a large number of copies of the best possible product-state input and then performing state discrimination for the product output states.

In the following, we extend some of the seminal classical results to the quantum setting by providing a framework for deriving upper bounds on the power of adaptive protocols for asymptotic quantum channel discrimination. In particular, in order to quantify the largest distinguishability that can be realized between two quantum channels, we introduce the concept of amortized channel divergence. This then allows to give converse bounds for adaptive channel discrimination protocols in the asymmetric hypothesis testing setting in the sense of Stein and Chernoff. Now, whenever the amortized channel divergences collapse to standard channel divergences, we immediately get single-letter converse bounds on the power of adaptive protocols for channel discrimination. Most importantly, we arrive at the characterization of the strong Stein's lemma for classical-quantum channels. Namely, as a full extension of the corresponding classical result [11, Cor. 1], we find that picking many copies of the best possible product-state input and then applying quantum Stein's lemma for the product output states is asymptotically optimal.

Intriguingly, we leave open the question of whether adaptive protocols improve the exponential error rate for quantum channel discrimination in the asymmetric Stein setting. We emphasise that this might already occur for entanglement breaking channels or even quantum-classical channels (measurements). This would also be consistent with the known advantage of adaptive protocols in the symmetric Chernoff setting [10], [9].

II. Notation

Quantum systems are denoted by A, B, R and have finite dimensions $|A|, |B|, |R|$, respectively. Quantum states on a system A are linear, positive semi-definite operators of trace

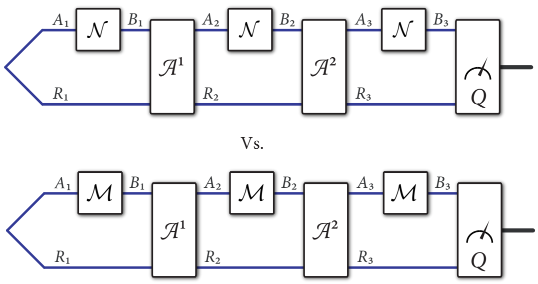

Figure 1. A protocol for channel discrimination when the channel $\mathcal{N}$ or $\mathcal{M}$ is called three times.

one and denoted by $\rho_A, \sigma_A, \tau_A \in \mathcal{S}(A)$. Quantum channels are completely positive and trace-preserving maps from the linear operators on A to the linear operators on B and denoted by $\mathcal{N}_{A\to B}$ or $\mathcal{M}_{A\to B}$. The relative entropy for quantum states ρ, σ is defined as [12]

$$D(\rho\|\sigma) := \begin{cases} \operatorname{tr}\big[\rho\,(\log\rho - \log\sigma)\,\big] & \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma) \\ \infty & \text{otherwise} \end{cases} \tag{1}$$

where in the above and throughout the paper all logarithms are evaluated using base two. For quantum states ρ, σ and $\alpha \in (0,1) \cup (1,\infty)$, the Petz-Rényi divergences [13] and the sandwiched Rényi divergences [14], [15] are defined as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log\operatorname{tr}\left[\rho^\alpha\sigma^{1-\alpha}\right] \tag{2}$$
$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log\operatorname{tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] \tag{3}$$

whenever either $\alpha \in (0,1)$ and ρ is not orthogonal to σ in Hilbert-Schmidt inner product or $\alpha > 1$ and $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$.[1] Otherwise we set $D_\alpha(\rho\|\sigma) := \infty =: \widetilde{D}_\alpha(\rho\|\sigma)$.

III. Asymptotic Channel Discrimination

The problem of quantum channel discrimination is made mathematically precise by the following hypothesis testing problem. Given two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ acting on an input system A and an output system B, a general adaptive strategy for discriminating them is as follows. We allow the preparation of an arbitrary input state $\rho_{A_1R_1} = \tau_{A_1R_1}$, where R_1 is an ancillary register. The ith use of a channel accepts the register A_i as input and produces the register B_i as output. After each invocation of the channel $\mathcal{N}_{A\to B}$ or $\mathcal{M}_{A\to B}$, an (adaptive) channel $\mathcal{A}^{(i)}_{B_iR_i\to A_{i+1}R_{i+1}}$ is applied to the registers B_i and R_i, yielding a quantum state $\rho_{A_{i+1}R_{i+1}}$ or $\tau_{A_{i+1}R_{i+1}}$ in registers $A_{i+1}R_{i+1}$, depending on whether the channel is equal to $\mathcal{N}_{A\to B}$ or $\mathcal{M}_{A\to B}$. That is,

$$\rho_{A_{i+1}R_{i+1}} := \mathcal{A}^{(i)}_{B_iR_i\to A_{i+1}R_{i+1}}(\rho_{B_iR_i}), \tag{4}$$
$$\rho_{B_iR_i} := \mathcal{N}_{A_i\to B_i}(\rho_{A_iR_i}), \tag{5}$$
$$\tau_{A_{i+1}R_{i+1}} := \mathcal{A}^{(i)}_{B_iR_i\to A_{i+1}R_{i+1}}(\tau_{B_iR_i}), \tag{6}$$
$$\tau_{B_iR_i} := \mathcal{M}_{A_i\to B_i}(\tau_{A_iR_i}), \tag{7}$$

for every $1 \le i < n$ on the left-hand side, and for every $1 \le i \le n$ on the right-hand side. Finally, a quantum measurement $\{Q_{B_nR_n}, 1_{B_nR_n} - Q_{B_nR_n}\}$ is performed on the systems B_nR_n to decide which channel was applied. The outcome Q corresponds to a final decision that the channel is $\mathcal{N}$, while the outcome $1-Q$ corresponds to a final decision that the channel is $\mathcal{M}$. We define the final decision probabilities

$$p := \operatorname{Tr}\left[Q_{B_nR_n}\rho_{B_nR_n}\right], \quad q := \operatorname{Tr}\left[Q_{B_nR_n}\tau_{B_nR_n}\right]. \tag{8}$$

[1]Throughout the paper, we employ the convention that inverses are to be understood as generalized inverses.

Figure 1 depicts such a protocol for channel discrimination when the channel $\mathcal{N}$ or $\mathcal{M}$ is called three times.[2] We use the simplifying notation $\{Q, \mathcal{A}\}$ to identify a particular strategy using channels $\{\mathcal{A}^{(i)}_{B_iR_i\to A_{i+1}R_{i+1}}\}_i$ and a final measurement $\{Q_{B_nR_n}, 1_{B_nR_n} - Q_{B_nR_n}\}$. For simplicity, this shorthand also includes the preparation of the initial state $\rho_{A_1R_1} = \tau_{A_1R_1}$, which can be understood as arising from the action of an initial channel $\mathcal{A}^{(0)}_{B_0R_0\to A_1R_1}$ for which the input systems B_0 and R_0 are trivial. This naturally gives rise to respective the type I and type II error probabilities:

$$\alpha_n(\{Q,\mathcal{A}\}) := \operatorname{tr}\big[(1_{B_nR_n} - Q_{B_nR_n})\rho_{B_nR_n}\big], \tag{9}$$
$$\beta_n(\{Q,\mathcal{A}\}) := \operatorname{tr}\big[Q_{B_nR_n}\tau_{B_nR_n}\big]. \tag{10}$$

For asymmetric hypothesis testing in the sense of Stein, we minimize the type II error probability, under the constraint that the type I error probability does not exceed a constant $\varepsilon \in (0,1)$. We are then interested in characterizing

$$D_h^{\varepsilon,n}(\mathcal{N}\|\mathcal{M}) := \sup_{\{Q,\mathcal{A}\}}\left\{-\frac{1}{n}\log\beta_n(\{Q,\mathcal{A}\})\,\middle|\,\alpha_n(\{Q,\mathcal{A}\}) \le \varepsilon\right\} \tag{11}$$

in the asymptotic limit $n \to \infty$. The strong converse exponent is a refinement of the asymmetric hypothesis testing quantity discussed above. For $r > 0$, we are interested in characterizing

$$H_r^n(\mathcal{N}\|\mathcal{M}) := \inf_{\{Q,\mathcal{A}\}}\left\{-\frac{1}{n}\log(1-\alpha_n(\{Q,\mathcal{A}\}))\,\middle|\,\beta_n(\{Q,\mathcal{A}\}) \le 2^{-rn}\right\} \tag{12}$$

in the asymptotic limit $n \to \infty$. The interpretation is that the type II error probability is constrained to tend to zero exponentially fast at a rate $r > 0$, but then if r is too large, the type I error probability will necessarily tend to one exponentially fast, and we are interested in the exact rate of exponential convergence. Note that this strong converse exponent is only non-trivial if r is sufficiently large.

Similarly one can also study the symmetric Chernoff setting, where one is interested in minimizing the total error probability of guessing incorrectly — as done in [16].

IV. Amortized Distinguishability of Channels

We say that a function $\mathbf{D} : \mathcal{S}(A) \times \mathcal{S}(A) \to \mathbb{R} \cup \{+\infty\}$ is a divergence if for quantum states ρ_A, σ_A and quantum channels $\mathcal{N}_{A\to B}$, we have the monotonicity

$$\mathbf{D}(\rho_A\|\sigma_A) \ge \mathbf{D}(\mathcal{N}_{A\to B}(\rho_A)\|\mathcal{N}_{A\to B}(\sigma_A)). \tag{13}$$

From this, one then defines a channel divergence as a measure for the distinguishability of two quantum channels [17].

[2]Another kind of channel discrimination strategy often considered in the literature is a parallel discrimination strategy, in which a state ρ_{A^nR} is prepared, either the tensor-power channel $\mathcal{N}^{\otimes n}_{A\to B}$ or $\mathcal{M}^{\otimes n}_{A\to B}$ is applied, and then a joint measurement is performed on the systems B^nR. As noted in [8], a parallel channel discrimination strategy of the channels $\mathcal{N}$ and $\mathcal{M}$ is a special case of an adaptive channel discrimination strategy.

Definition IV.1. Let $\mathbf{D}$ be a divergence and $\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}$ quantum channels. The channel divergence of $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ is defined as

$$\mathbf{D}(\mathcal{N}\|\mathcal{M}) := \sup_{\rho\in\mathcal{S}(AR)} \mathbf{D}(\mathcal{N}_{A\to B}(\rho_{AR})\|\mathcal{M}_{A\to B}(\rho_{AR})). \tag{14}$$

We now define the amortized channel divergence as another measure of the distinguishability of two quantum channels. The idea behind this measure is to consider two different states ρ_{AR} and σ_{AR} that can be input to the channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, in order to explore the largest distinguishability that can be realized between the channels. However, from a resource-theoretic perspective, these initial states themselves could have some distinguishability, and so it is sensible to subtract off the initial distinguishability of the states ρ_{AR} and σ_{AR} from the final distinguishability of the channel output states $\mathcal{N}_{A\to B}(\rho_{AR})$ and $\mathcal{M}_{A\to B}(\sigma_{AR})$.

Definition IV.2. Let $\mathbf{D}$ be a divergence and $\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}$ quantum channels. We define the amortized channel divergence as[3]

$$\mathbf{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) := \sup_{\rho,\sigma\in\mathcal{S}(AR)} \mathbf{D}(\mathcal{N}_{A\to B}(\rho_{AR})\|\mathcal{M}_{A\to B}(\sigma_{AR})) - \mathbf{D}(\rho_{AR}\|\sigma_{AR}). \tag{15}$$

Based on the concept of amortized channel divergences, we now develop a meta-converse for quantum channel discrimination. Consider a channel discrimination protocol as introduced in Section III, with final decision probabilities p and q, as given in (8). Conceptually, the statement of the lemma is that the distinguishability of the final decision probabilities p and q at the end of a channel discrimination protocol, in which the channels are called n times, is limited by n times the amortized channel divergence of the two channels.

Lemma IV.3. *Let $\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}$ be quantum channels. Then, we have for any quantum channel discrimination protocol and faithful divergence that*

$$\mathbf{D}(p\|q) \le n\cdot\mathbf{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) \tag{16}$$

where $\mathbf{D}(p\|q) := \mathbf{D}(\zeta(p)\|\zeta(q))$ *with* $\zeta(p) := p|0\rangle\langle 0| + (1-p)\,|1\rangle\langle 1|$.

Proof. Let $\{\mathrm{Q}, \mathcal{A}\}$ denote a protocol for discrimination of the channels $\mathcal{N}$ and $\mathcal{M}$, and let p and q denote the final decision probabilities. Consider that

$$\begin{aligned}
&\mathbf{D}(p\|q)\\
&\le \mathbf{D}(\rho_{B_nR_n}\|\tau_{B_nR_n}) && (17)\\
&\le \mathbf{D}(\rho_{B_nR_n}\|\tau_{B_nR_n}) - \mathbf{D}(\rho_{A_1R_1}\|\tau_{A_1R_1}) && (18)\\
&= \mathbf{D}(\rho_{B_nR_n}\|\tau_{B_nR_n}) - \mathbf{D}(\rho_{A_1R_1}\|\tau_{A_1R_1})\\
&\quad + \sum_{i=2}^{n}\Big(\mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i}) - \mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i})\Big) && (19)\\
&= \mathbf{D}(\rho_{B_nR_n}\|\tau_{B_nR_n}) - \mathbf{D}(\rho_{A_1R_1}\|\tau_{A_1R_1})\\
&\quad + \sum_{i=2}^{n}\Big(\mathbf{D}\big(\mathcal{A}^{(i-1)}_{B_{i-1}R_{i-1}\to A_iR_i}(\rho_{B_{i-1}R_{i-1}})\\
&\qquad\qquad \big\|\mathcal{A}^{(i-1)}_{B_{i-1}R_{i-1}\to A_iR_i}(\tau_{B_{i-1}R_{i-1}})\big)\\
&\qquad - \mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i})\Big) && (20)
\end{aligned}$$

The first inequality follows from monotonicity under the final measurement $\{Q_{B_nR_n}, 1_{B_nR_n} - Q_{B_nR_n}\}$, and the second inequality follows from the assumption of faithfulness together with the fact that the initial states are equal, i.e., $\rho_{A_1R_1} = \tau_{A_1R_1}$. Continuing, we have that

$$\begin{aligned}
&\text{Eq. (20)}\\
&\le \mathbf{D}(\rho_{B_nR_n}\|\tau_{B_nR_n}) - \mathbf{D}(\rho_{A_1R_1}\|\tau_{A_1R_1})\\
&\quad + \sum_{i=1}^{n-1}\mathbf{D}(\rho_{B_iR_i}\|\tau_{B_iR_i}) - \sum_{i=2}^{n}\mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i}) && (21)\\
&= \sum_{i=1}^{n}\Big(\mathbf{D}(\rho_{B_iR_i}\|\tau_{B_iR_i}) - \mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i})\Big) && (22)\\
&= \sum_{i=1}^{n}\Big(\mathbf{D}(\mathcal{N}_{A\to B}(\rho_{A_iR_i})\|\mathcal{M}_{A\to B}(\tau_{A_iR_i}))\\
&\qquad - \mathbf{D}(\rho_{A_iR_i}\|\tau_{A_iR_i})\Big) && (23)\\
&\le n\cdot\mathbf{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) && (24)
\end{aligned}$$

where the first inequality follows from monotonicity with respect to the channel $\mathcal{A}^{(i-1)}_{B_{i-1}R_{i-1}\to A_iR_i}$. ☐

V. Classical-Quantum Channel Discrimination

For non-adaptive protocols, when we restrict the input states to be product states — but still allow for a quantum memory system R — it directly follows from Stein's lemma for quantum state discrimination [1], [2] that the optimal asymptotic error exponent for $\varepsilon \in (0,1)$ is given by the quantum relative entropy divergence $D(\mathcal{N}\|\mathcal{M})$, as observed in [18]. This obviously also gives an achievability bound for the adaptive setting. In the following, we are interested in converse bounds for the adaptive setting.

Proposition V.1. *Let $\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}$ be quantum channels. Then, we have for $n\in\mathbb{N}$ and $\varepsilon\in[0,1)$ that*

$$D_h^{\varepsilon,n}(\mathcal{N}\|\mathcal{M}) \le \frac{1}{1-\varepsilon}\left(D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) + \frac{h_2(\varepsilon)}{n}\right) \tag{25}$$

where $h_2(\varepsilon)$ denotes the binary entropy.

Note that this bound is *a priori* an unbounded optimization problem and that it is in general unclear if the amortized quantity $D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M})$ can be achieved by an adaptive protocol.

Proof of Proposition V.1. Standard arguments as in the proof of quantum Stein's lemma [1] give

$$D(p\|q) = (1-p)\log\frac{1-p}{1-q} + p\log(p/q) \tag{26}$$

[3] In contrast to Definition IV.1 the supremum cannot be restricted to pure states only (in general). Moreover, there is a priori no dimension bound on the system R.

$$= \alpha_n(\{Q,\mathcal{A}\}) \log \frac{\alpha_n(\{Q,\mathcal{A}\})}{1-\beta_n(\{Q,\mathcal{A}\})} + (1-\alpha_n(\{Q,\mathcal{A}\})) \log \frac{1-\alpha_n(\{Q,\mathcal{A}\})}{\beta_n(\{Q,\mathcal{A}\})} \tag{27}$$

$$= \varepsilon \log \frac{\varepsilon}{1-\beta_n(\{Q,\mathcal{A}\})} + (1-\varepsilon) \log \frac{1-\varepsilon}{\beta_n(\{Q,\mathcal{A}\})} \tag{28}$$

$$= -h_2(\varepsilon) - \varepsilon \log(1-\beta_n(\{Q,\mathcal{A}\})) - (1-\varepsilon) \log \beta_n(\{Q,\mathcal{A}\}) \tag{29}$$

$$\geq -h_2(\varepsilon) - (1-\varepsilon) \log \beta_n(\{Q,\mathcal{A}\}). \tag{30}$$

The claim follows by rearranging the above together with Lemma IV.3 for the relative entropy channel divergence. □

In the following we consider classical-quantum channels

$$\mathcal{N}_{X\to B}(\cdot) = \sum_x \langle x| \cdot |x\rangle \nu_B^x,\ \mathcal{M}_{X\to B}(\cdot) = \sum_x \langle x| \cdot |x\rangle \mu_B^x \tag{31}$$

where $\{|x\rangle\}_x$ is an orthonormal basis and $\{\nu_B^x\}_x$ and $\{\mu_B^x\}_x$ are sets of quantum states. We find that the optimal asymptotic classical-quantum channel discrimination protocol for the Stein setting is to pick the best possible input and then to apply a tensor-power strategy.

Theorem V.2. *Let $\mathcal{N}_{X\to B}, \mathcal{M}_{X\to B}$ be classical-quantum channels. Then, we have that*

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} D_h^{\varepsilon,n}(\mathcal{N}\|\mathcal{M}) = \max_x D(\nu_B^x\|\mu_B^x). \tag{32}$$

This result implies that adaptive strategies, quantum memories, and entangled inputs do not improve the Stein exponent. Note that this slightly extends the classical case as well, in the sense that it was previously not resolved if quantum memories could be helpful in the asymptotic case.

The achievability part of Theorem V.2 follows directly by employing a product-state discrimination strategy [1], [2]. Therefore, it remains to show the converse direction. We know that the amortized quantum relative entropy divergence $D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M})$ provides a weak converse rate (Proposition V.1), with the only missing step being to evaluate that quantity. The following lemma then immediately implies Theorem V.2.

Lemma V.3. *Let $\mathcal{N}_{X\to B}, \mathcal{M}_{X\to B}$ be classical-quantum channels. Then, we have the amortization collapse*

$$D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) = \max_x D(\nu^x\|\mu^x), \tag{33}$$

$$D_\alpha^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) = \max_x D_\alpha(\nu^x\|\mu^x) \text{ for } \alpha \in [0,2], \tag{34}$$

$$\widetilde{D}_\alpha^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) = \max_x \widetilde{D}_\alpha(\nu^x\|\mu^x) \text{ for } \alpha \geq \frac{1}{2}. \tag{35}$$

Proof. We give a proof for the relative entropy and refer to [16, Lem. 26] for the Rényi divergences. By picking $\rho_{AR} = \sigma_{AR} = |x\rangle\langle x|_A \otimes |x\rangle\langle x|_R$, we get $D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) \geq D(\nu_B^x\|\mu_B^x)$. Since this holds for all x, we conclude that

$$D^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) \geq \max_x D(\nu_B^x\|\tau_B^x). \tag{36}$$

For the other direction, consider for states ρ_{AR} and σ_{AR} that

$$\mathcal{N}_{A\to B}(\rho_{AR}) = \sum_x p_x \nu_B^x \otimes \rho_R^x, \tag{37}$$

$$\mathcal{M}_{A\to B}(\sigma_{AR}) = \sum_x q_x \mu_B^x \otimes \sigma_R^x, \tag{38}$$

where $p_x\rho_R^x := \langle x|_A \rho_{AR} |x\rangle_A$ and $q_x\sigma_R^x := \langle x|_A \sigma_{AR} |x\rangle_A$, with p_x and q_x probability distributions and $\{\rho_R^x\}_x$ and $\{\sigma_R^x\}_x$ sets of quantum states. Then, we have from the monotonicity of the relative entropy under channels that

$$D(\mathcal{N}_{A\to B}(\rho_{AR})\|\mathcal{M}_{A\to B}(\sigma_{AR})) - D(\rho_{AR}\|\sigma_{AR})$$

$$\leq D\left(\sum_x p_x \nu_B^x \otimes \rho_R^x \middle\| \sum_x q_x \mu_B^x \otimes \sigma_R^x\right) - D\left(\sum_x p_x |x\rangle\langle x|_X \otimes \nu_B^x \otimes \rho_R^x \middle\| \sum_x q_x |x\rangle\langle x|_X \otimes \nu_B^x \otimes \sigma_R^x\right) \tag{39}$$

$$\leq D\left(\sum_x p_x |x\rangle\langle x|_X \otimes \nu_B^x \otimes \rho_R^x \middle\| \sum_x q_x |x\rangle\langle x|_X \otimes \mu_B^x \otimes \sigma_R^x\right) - D\left(\sum_x p_x |x\rangle\langle x|_X \otimes \nu_B^x \otimes \rho_R^x \middle\| \sum_x q_x |x\rangle\langle x|_X \otimes \nu_B^x \otimes \sigma_R^x\right) \tag{40}$$

$$= \sum_x p_x \Bigg(\operatorname{Tr}\left[(\nu_B^x \otimes \rho_R^x) \log (q_x \nu_B^x \otimes \sigma_R^x)\right] - \operatorname{Tr}\left[(\nu_B^x \otimes \rho_R^x) \log (q_x \mu_B^x \otimes \sigma_R^x)\right] \Bigg) \tag{41}$$

$$= \sum_x p_x \operatorname{Tr}\left[(\nu_B^x \otimes \rho_R^x)\left((\log \nu_B^x - \log \mu_B^x) \otimes 1_R\right)\right] \tag{42}$$

$$= \sum_x p_x D(\nu_B^x\|\mu_B^x) \tag{43}$$

$$\leq \max_x D(\nu_B^x\|\mu_B^x). \tag{44}$$

This concludes the proof. □

The strong converse exponent for the discrimination of classical-quantum channels is derived along similar lines.

Theorem V.4. *Let $\mathcal{N}_{X\to B}, \mathcal{M}_{X\to B}$ be classical-quantum channels. Then, for $r > 0$ we have that*

$$\lim_{n\to\infty} H_r^n(\mathcal{N}\|\mathcal{M}) = \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left(r - \max_x \widetilde{D}_\alpha(\nu^x\|\mu^x)\right). \tag{45}$$

The optimality part follows from the amortization collapse of the sandwiched Rényi divergences in Lemma V.3 together with a standard strong converse exponent argument [16, Prop. 20]. The achievability part requires a minimax argument [16, Thm. 27]. As a corollary, we then get the strong variant of Stein's lemma for classical-quantum channels.

Corollary V.5. *Let $\mathcal{N}_{X\to B}, \mathcal{M}_{X\to B}$ be classical-quantum channels. Then, for $\varepsilon \in (0,1)$ we have that*

$$\lim_{n\to\infty} D_h^{\varepsilon,n}(\mathcal{N},\mathcal{M}) = \max_x D(\nu^x\|\mu^x). \tag{46}$$

Proof. In Theorem V.4, if $r > \max_x D(\nu_B^x \| \mu_B^x)$, then by the fact that $\max_x \widetilde{D}_\alpha(\nu_B^x \| \mu_B^x)$ is monotone increasing with α and the continuity

$$\lim_{\alpha \to 1} \max_x \widetilde{D}_\alpha(\nu_B^x \| \mu_B^x) = \max_x D(\nu_B^x \| \mu_B^x), \tag{47}$$

there exists $\alpha > 1$ such that $r > \max_x \widetilde{D}_\alpha(\nu_B^x \| \mu_B^x)$, and so $\lim_{n\to\infty} H_r^n(\mathcal{N} \| \mathcal{M}) > 0$, implying that the Type I error probability tends to one exponentially fast. □

In contrast to the weak and strong Stein's lemma (Theorem V.2 and Corollary V.5), we cannot conclude that the strong converse exponent in Theorem V.4 is achieved by picking the best possible input element x, but we instead have to consider distributions over the input alphabet (see [11, Sect. IV] for an extended discussion). This is similar to the classical case and Hayashi in fact gives an explicit example where considering only one input element x is not sufficient [11, Section IV]. He then shows that, in the classical case, it suffices to optimize with respect to probability distributions that are strictly positive on just two elements [11, Theorem 3].

Finally, the amortization collapses from Lemma V.3 can also be employed to make statements about the symmetric Chernoff setting [16, Sect. VI.C-D]. However, the tight information-theoretic characterization remains open even for classical-quantum channels.[4]

VI. Outlook

In order to derive upper bounds on the power of adaptive quantum channel discrimination protocols, we introduced a framework based on the concept of amortized channel divergence. As our main result we then established the strong Stein's lemma for classical-quantum channels. We regard our work as an initial step towards a plethora of open questions surrounding quantum channel discrimination and refer to the discussion in [16, Sect. V.II] for more examples with tight single-letter characterizations.

We have to leave open the general question of whether adaptive protocols improve the exponential error rate for quantum channel discrimination in the asymmetric Stein setting — as they do in the symmetric Chernoff setting [10], [9]. A first step in this direction would be to look at an intermediate strategy in which a state $\rho_{A^n R}$ is prepared, either the tensor-power channel $\mathcal{N}_{A\to B}^{\otimes n}$ or $\mathcal{M}_{A\to B}^{\otimes n}$ is applied, and then a joint measurement is performed on the systems $B^n R$. We emphasise that it is not even known whether this setting offers an asymptotic advantage compared to a tensor-power strategy with input $\rho_{AR}^{\otimes n}$. The question might be thought of as determining if

$$\frac{1}{n} D\left(\mathcal{N}^{\otimes n} \middle\| \mathcal{M}^{\otimes n}\right) \overset{?}{\to} D\left(\mathcal{N} \| \mathcal{M}\right) \tag{48}$$

holds for all quantum channels — whereas our work implies the limit for classical-quantum channels. Now, note that if we restrict the quantum memory system R to be one-dimensional, then the Hastings counterexamples to the minimal output entropy conjecture [19], applied to the setting involving a replacer channel, immediately give a separation to the tensor-product strategy. This suggests that for a non-trivial quantum memory R, there are some deep entropic additivity questions that remain to be explored.

Acknowledgements: We thank Fernando Brandão, Gilad Gour, Milan Mosonyi, Giacomo de Palma, and Andreas Winter for discussions related to the topic of this paper.

[4] Already for entanglement-breaking channels, the straightforward Chernoff bound conjecture is known not to hold [9].

References

[1] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, 1991.

[2] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428–2433, 2000.

[3] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *The Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, 2009.

[4] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating states: The quantum Chernoff bound," *Physical Review Letters*, vol. 98, no. 16, p. 160501, 2007.

[5] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," 2006, arXiv:quant-ph/0611289.

[6] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, no. 1, pp. 251–283, 2008.

[7] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," *Communications in Mathematical Physics*, vol. 334, no. 3, pp. 1617–1648, 2015.

[8] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory effects in quantum channel discrimination," *Physical Review Letters*, vol. 101, no. 18, p. 180501, 2008.

[9] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, "Adaptive versus nonadaptive strategies for quantum channel discrimination," *Physical Review A*, vol. 81, no. 3, p. 032339, 2010.

[10] R. Duan, Y. Feng, and M. Ying, "Perfect distinguishability of quantum operations," *Physical Review Letters*, vol. 103, no. 21, p. 210501, 2009.

[11] M. Hayashi, "Discrimination of two channels by adaptive methods and its application to quantum system," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3807–3820, 2009.

[12] H. Umegaki, "Conditional expectations in an operator algebra IV," *Kodai Mathematical Seminar Reports*, vol. 14, no. 2, pp. 59–85, 1962.

[13] D. Petz, "Quasi-entropies for states of a von Neumann algebra," *Publications RIMS Kyoto University*, vol. 21, pp. 787–800, 1985.

[14] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: a new generalization and some properties," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122203, 2013.

[15] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, 2014.

[16] M. Berta, C. Hirche, E. Kaur, and M. M. Wilde, "Amortized channel divergence for asymptotic quantum channel discrimination," 2018, arXiv:1808.01498.

[17] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, "Approaches for approximate additivity of the Holevo information of quantum channels," *Physical Review A*, vol. 97, no. 1, p. 012332, 2018.

[18] T. Cooney, M. Mosonyi, and M. M. Wilde, "Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication," *Communications in Mathematical Physics*, vol. 344, no. 3, pp. 797–829, 2016.

[19] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Physics*, vol. 5, p. 255, 2009.