6-1-2016

Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication

Tom Cooney
Louisiana State University

Milán Mosonyi
Universitat Autònoma de Barcelona

Mark M. Wilde
Louisiana State University

Recommended Citation

Cooney, T., Mosonyi, M., & Wilde, M. (2016). Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication. *Communications in Mathematical Physics, 344* (3), 797-829. https://doi.org/10.1007/s00220-016-2645-4

Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication

Tom Cooney* Milán Mosonyi†‡ Mark M. Wilde*§

February 29, 2016

Abstract

This paper studies the difficulty of discriminating between an arbitrary quantum channel and a "replacer" channel that discards its input and replaces it with a fixed state. The results obtained here generalize those known in the theory of quantum hypothesis testing for binary state discrimination. We show that, in this particular setting, the most general adaptive discrimination strategies provide no asymptotic advantage over non-adaptive tensor-power strategies. This conclusion follows by proving a quantum Stein's lemma for this channel discrimination setting, showing that a constant bound on the Type I error leads to the Type II error decreasing to zero exponentially quickly at a rate determined by the maximum relative entropy registered between the channels. The strong converse part of the lemma states that any attempt to make the Type II error decay to zero at a rate faster than the channel relative entropy implies that the Type I error necessarily converges to one. We then refine this latter result by identifying the optimal strong converse exponent for this task. As a consequence of these results, we can establish a strong converse theorem for the quantum-feedback-assisted capacity of a channel, sharpening a result due to Bowen. Furthermore, our channel discrimination result demonstrates the asymptotic optimality of a non-adaptive tensor-power strategy in the setting of quantum illumination, as was used in prior work on the topic. The sandwiched Rényi relative entropy is a key tool in our analysis. Finally, by combining our results with recent results of Hayashi and Tomamichel, we find a novel operational interpretation of the mutual information of a quantum channel $\mathcal{N}$ as the optimal type II error exponent when discriminating between a large number of independent instances of $\mathcal{N}$ and an arbitrary "worst-case" replacer channel chosen from the set of all replacer channels.

1 Introduction

Quantum channel discrimination is a natural extension of a basic problem in quantum hypothesis testing, that of distinguishing between the possible states of a quantum system. In the case of binary state discrimination, it is given *a priori* that a quantum system is in one of two states ρ or σ, and

*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA

†Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Barcelona), Spain

‡Mathematical Institute, Budapest University of Technology and Economics, Egry József u 1., Budapest, 1111 Hungary

§Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

the goal is to identify in which state it is by performing a quantum measurement. We say that ρ is the null hypothesis and σ is the alternative hypothesis. A natural extension of this problem occurs in the independent and identically distributed (i.i.d.) setting. Here, the discriminator is provided with n quantum systems in the state $\rho^{\otimes n}$ or $\sigma^{\otimes n}$, and the task is to apply a binary measurement $\{Q_n, I^{\otimes n} - Q_n\}$ on these n systems, with $0 \leq Q_n \leq I^{\otimes n}$, to determine which state he possesses. One is then concerned with two kinds of error probabilities:

$$\alpha_n(Q_n) \equiv \mathrm{Tr}\left\{(I^{\otimes n} - Q_n)\rho^{\otimes n}\right\}, \tag{1.1}$$

the probability of incorrectly rejecting the null hypothesis, the Type I error, and

$$\beta_n(Q_n) \equiv \mathrm{Tr}\left\{Q_n \sigma^{\otimes n}\right\}, \tag{1.2}$$

the probability of incorrectly rejecting the alternative hypothesis, the Type II error. Of course, it is generally impossible to find a quantum measurement such that both of these errors are equal to zero simultaneously, so one instead studies the asymptotic behaviour of α_n and β_n as $n \to \infty$, expecting there to be a trade-off between minimising α_n and minimising β_n.

In asymmetric hypothesis testing, one fixes a constraint on the Type I error, say, and then seeks to minimise the Type II error. When a constant threshold ε is imposed on the Type I error, the optimal Type II error is given by

$$\beta_\varepsilon(\rho\|\sigma) \equiv \min\{\beta(Q) : 0 \leq Q \leq I,\, \alpha(Q) \leq \varepsilon\}. \tag{1.3}$$

The central result in the asymptotic setting is the quantum Stein's lemma, due to Hiai and Petz [25] and Ogawa and Nagaoka [41]. The direct part of the lemma states that for any constant bound on the Type I error, there exists a sequence of measurements $\{Q_n, I^{\otimes n} - Q_n\}$ that meets this constraint and is such that the Type II error decreases to zero exponentially fast with a decay exponent given by the quantum relative entropy $D(\rho\|\sigma)$, defined as [54, 25]

$$D(\rho\|\sigma) \equiv \left\{ \begin{array}{cc} \mathrm{Tr}\left\{\rho\left[\log\rho - \log\sigma\right]\right\} & \text{if } \mathrm{supp}\,(\rho) \subseteq \mathrm{supp}\,(\sigma) \\ +\infty & \text{otherwise} \end{array} \right. . \tag{1.4}$$

In the above and throughout the paper, we take the logarithm to be base two. Furthermore, the strong converse part of the lemma states that any attempt to make the Type II error decay to zero with a decay exponent larger than the relative entropy will result in the Type I error converging to one in the large n limit [41]. The direct and the strong converse parts can be succinctly written as

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma), \qquad \forall\varepsilon \in (0,1). \tag{1.5}$$

That is, for any threshold value $\varepsilon \in (0,1)$, the optimal Type II error decays exponentially fast in the number of copies, and the decay rate is equal to the relative entropy.

It is easy to see that the negative logarithm of the optimal Type II error,

$$D_H^\varepsilon(\rho\|\sigma) \equiv -\log\beta_\varepsilon(\rho\|\sigma), \tag{1.6}$$

is non-negative and monotonic non-increasing under completely positive trace-preserving maps. Thus, it can be considered as a "generalized divergence" or "generalized relative entropy" and it was named "hypothesis testing relative entropy" in [55]. With this notation, Stein's lemma (1.5) can be reformulated as

$$\lim_{n\to\infty}\frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n})=D(\rho\|\sigma),\qquad \forall\varepsilon\in(0,1). \tag{1.7}$$

As a refinement of the quantum Stein's lemma, one can study the optimal Type I error given that the Type II error decays with a given exponential speed. One is then interested in the asymptotics of the optimal Type I error

$$\alpha_{n,r}\equiv\alpha_{2^{-nr}}(\rho^{\otimes n}\|\sigma^{\otimes n})\equiv\min\left\{\alpha_n(Q_n):\,0\le Q_n\le I,\,\beta_n(Q_n)\le 2^{-nr}\right\}, \tag{1.8}$$

with $r>0$ a constant. In the "direct domain," when $r<D(\rho\|\sigma)$, $\alpha_{n,r}$ also decays with an exponential speed, as was shown in [40]. The exact decay rate is determined by the quantum Hoeffding bound theorem [21, 39, 2] as

$$\lim_{n\to\infty}-\frac{1}{n}\log\alpha_{n,r}=H_r(\rho\|\sigma)\equiv\sup_{0<\alpha<1}\frac{\alpha-1}{\alpha}(r-D_\alpha(\rho\|\sigma)), \tag{1.9}$$

where D_α is a quantum Rényi relative entropy, to be defined later, and $H_r(\rho\|\sigma)$ is the Hoeffding divergence of ρ and σ. On the other hand, in the "strong converse domain," when $r>D(\rho\|\sigma)$, $\alpha_{n,r}$ goes to 1 exponentially fast [41, 38]. The rate of this convergence has been determined in [20, pages 80-81] in terms of the limit of post-measurement Rényi relative entropies. A "single-letter" expression has been obtained recently in [36] as

$$\lim_{n\to\infty}-\frac{1}{n}\log(1-\alpha_{n,r})=H_r^*(\rho\|\sigma)\equiv\sup_{\alpha>1}\frac{\alpha-1}{\alpha}(r-\widetilde{D}_\alpha(\rho\|\sigma)), \tag{1.10}$$

where $\tilde{D}_\alpha$ is an alternative version of the quantum Rényi relative entropy [37, 56], and $H_r^*(\rho\|\sigma)$ is the Hoeffding anti-divergence. Note that it is unique to the quantum case that one requires a Rényi relative entropy for the strong converse domain which is different from that used in the direct domain (however, these Rényi relative entropies coincide when ρ and σ commute, i.e., the classical case).

The results in (1.9) and (1.10) give a complete understanding of the trade-off between the two error probabilities in the asymptotics. Note that the quantum Stein's lemma can also be recovered from (1.9) and (1.10) in the limit $r\to D(\rho\|\sigma)$. We remark that there are other ways of refining our understanding of the quantum Stein's lemma, as established recently in [32, 52].

The objectives of channel discrimination are very similar to those of state discrimination; what makes the problem different is the complexity of the available discrimination strategies. In the general setup we have a quantum channel with input system A and output system B, and we know that the channel is described by either $\mathcal{N}_1$ or $\mathcal{N}_2$, where $\mathcal{N}_1$ and $\mathcal{N}_2$ are completely positive trace-preserving (CPTP) maps. We assume that we can use the channel several times, consecutive uses are independent, and the properties of the channel do not change with time. Thus, n uses of the channel are described by either $\mathcal{N}_1^{\otimes n}$ or $\mathcal{N}_2^{\otimes n}$. A non-adaptive discrimination strategy for n uses of the channel consists of feeding an input state $\psi_{R_nA^n}$ into the n-fold tensor-product channel, and then performing a binary measurement $\{Q_n, I-Q_n\}$ on the output, which is either $\mathcal{N}_1^{\otimes n}(\psi_{R_nA^n})\equiv(\mathrm{id}_{R_n}\otimes\mathcal{N}_1^{\otimes n})(\psi_{R_nA^n})$ or $\mathcal{N}_2^{\otimes n}(\psi_{R_nA^n})\equiv(\mathrm{id}_{R_n}\otimes\mathcal{N}_2^{\otimes n})(\psi_{R_nA^n})$. Here, R_n is an ancilla system on which the channel acts trivially as the identity map id_{R_n}. When an adaptive strategy is used, the output of the first k uses of the channel can be used to prepare the input for the $(k+1)$-th use; see Figure 1 for a pictorial explanation and Section 2.1 for a precise definition.

For any discrimination strategy S_n, let $\rho_n(S_n)$ and $\sigma_n(S_n)$ denote the output of the n-fold product channel depending on whether the channel is equal to $\mathcal{N}_1$ or $\mathcal{N}_2$. In analogy with (1.1)-(1.2), one can define the Type I and the Type II errors as

$$\alpha_n(S_n) \equiv \mathrm{Tr}\{(I - Q_n)\rho_n(S_n)\}, \quad \text{(Type I)} \qquad \beta_n(S_n) \equiv \mathrm{Tr}\{Q_n\sigma_n(S_n)\}, \quad \text{(Type II)}, \tag{1.11}$$

where $\{Q_n, I - Q_n\}$ is the measurement part of the strategy. It is then natural to consider the optimal error probabilities

$$\beta^x_\varepsilon(\mathcal{N}_1^{\otimes n}\|\mathcal{N}_2^{\otimes n}) \equiv \inf\{\beta_n(S_n) :\ \alpha_n(S_n) \le \varepsilon\}, \qquad \text{and} \tag{1.12}$$

$$\alpha^x_{n,r} \equiv \alpha^x_{2^{-nr}}(\mathcal{N}_1^{\otimes n}\|\mathcal{N}_2^{\otimes n}) \equiv \inf\left\{\alpha_n(S_n) :\ \beta_n(S_n) \le 2^{-nr}\right\}, \tag{1.13}$$

where x denotes the set of allowed discrimination strategies and the optimisations are over all strategies in the class x. Here, we will consider $x = \mathrm{ad}$ for adaptive and $x = \mathrm{pr}$ for product strategies. The latter are all non-adaptive strategies with an input state $\psi_{R_nA^n} = \psi_{RA}^{\otimes n}$, where ψ_{RA} is an arbitrary state on A and some ancilla R. Obviously, if only product strategies are allowed ($x = \mathrm{pr}$), then the optimal rates of these error probabilities are given by the corresponding channel divergences as

$$\lim_{n\to+\infty} -\frac{1}{n}\log\beta^x_\varepsilon(\mathcal{N}_1^{\otimes n}\|\mathcal{N}_2^{\otimes n}) = D(\mathcal{N}_1\|\mathcal{N}_2) \equiv \sup_{\psi_{RA}} D(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})), \tag{1.14}$$

$$\lim_{n\to+\infty} -\frac{1}{n}\log\alpha^x_{n,r} = H_r(\mathcal{N}_1\|\mathcal{N}_2) \equiv \sup_{\psi_{RA}} H_r(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})), \tag{1.15}$$

$$\lim_{n\to+\infty} -\frac{1}{n}\log(1-\alpha^x_{n,r}) = H^*_r(\mathcal{N}_1\|\mathcal{N}_2) \equiv \inf_{\psi_{RA}} H^*_r(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})), \tag{1.16}$$

according to the previously explained results on state discrimination. Note that in (1.16) an infimum is taken; the reason is that, in the strong converse domain, the goal is to minimise the exponent of the success probability. The Hoeffding (anti-)divergences can also be expressed as

$$H_r(\mathcal{N}_1\|\mathcal{N}_2) = \sup_{0<\alpha<1} \frac{\alpha-1}{\alpha}(r - D_\alpha(\mathcal{N}_1\|\mathcal{N}_2)), \tag{1.17}$$

$$H^*_r(\mathcal{N}_1\|\mathcal{N}_2) = \sup_{1<\alpha} \frac{\alpha-1}{\alpha}(r - \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2)), \tag{1.18}$$

where $D_\alpha(\mathcal{N}_1\|\mathcal{N}_2)$ and $\widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2)$ are the channel Rényi relative entropies:

$$D_\alpha(\mathcal{N}_1\|\mathcal{N}_2) \equiv \sup_{\psi_{RA}} D_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})), \tag{1.19}$$

$$\widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) \equiv \sup_{\psi_{RA}} \widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})). \tag{1.20}$$

The optimizations in (1.14)–(1.16) and (1.19)–(1.20) are taken over all possible bipartite states ψ_{RA} with an arbitrary ancilla system R.

For adaptive strategies, the relations (1.14)-(1.16) are not expected to hold for arbitrary channels. For instance, the results of [19] provide some evidence in this direction. (See [14] as well for related results and more general conclusions.) There are various classes of channels, however, for which (1.14)-(1.16) hold; for these channels, adaptive strategies do not offer any benefit over product strategies. For instance, Hayashi showed (1.14)-(1.16) with $x = \mathrm{ad}$ for any pair of classical channels [22].

Another extreme case is when both $\mathcal{N}_1$ and $\mathcal{N}_2$ are replacer channels, i.e., there exist states ρ, σ such that $\mathcal{N}_1(\cdot) = \mathcal{R}_\rho \equiv \mathrm{Tr}\{\cdot\}\rho$ and $\mathcal{N}_2(\cdot) = \mathcal{R}_\sigma \equiv \mathrm{Tr}\{\cdot\}\sigma$. Obviously, in this case all the channel divergences are equal to the corresponding divergences of the two states; e.g., $D_\alpha(\mathcal{R}_\rho \| \mathcal{R}_\sigma) = D_\alpha(\rho\|\sigma)$, etc. It is also heuristically clear that adaptive strategies do not offer any benefit over product strategies, and the channel discrimination problem reduces to the state discrimination problem between ρ and σ, described before. Two of our main results, Theorems 1 and 2 yield as a special case a mathematically precise argument for these heuristics in the case of (1.14) and (1.16).

A natural intermediate step towards determining the error exponents of the general quantum channel discrimination problem is to allow one of the channels to be arbitrary, while keeping the other channel a replacer channel. This setup interpolates between the fully understood case of state discrimination and the still open problem of general quantum channel discrimination. Here we consider the setup in which the first channel is arbitrary and the second channel is a replacer channel. We prove (1.14) (Stein's lemma) in Section 4.1, and show in Section 4.2 that the strong converse exponent is given as in (1.16) for adaptive strategies ($x = \mathrm{ad}$). As for now, we leave the optimality part of (1.15) open for $x = \mathrm{ad}$.

As a consequence of these results, in Section 5 we can establish a strong converse theorem for the quantum-feedback-assisted capacity of a channel, which is the capacity of a quantum channel for transmitting classical information with the assistance of a noiseless quantum feedback from receiver to sender. Our result here strengthens that of Bowen's [9]. We also make a connection between our results and quantum illumination [34] in Section 2.3. Finally, in Section 4.3, we discuss how to combine the recent results in [24] with ours to obtain a quantum Stein's lemma in a setting more general than that considered in either paper. This gives a novel operational interpretation of the mutual information of a quantum channel, different from that already found in entanglement- and quantum-feedback-assisted communication [7, 28, 9]. We also discuss an open question regarding the characterization of the strong converse exponent in this more general setting.

2 Summary of results

2.1 Quantum Stein's lemma in adaptive channel discrimination

Our first result is a generalization of the quantum Stein's lemma in (1.7) to the setting of adaptive quantum channel discrimination. In particular, we study the difficulty of discriminating between an arbitrary quantum channel $\mathcal{N}$ and a "replacer" channel $\mathcal{R}$ that discards its input and replaces it with a fixed state σ. An important physical realization of this problem is in quantum illumination [34, 50] (discussed more in Section 2.3). We show that a tensor-power strategy is optimal in this case, so that there is no need to consider the most general adaptive strategy (at least in the asymptotic regime). This can be seen as a quantum Stein's lemma for this task; if one optimises the Type II error under the constraint that the Type I error is less than some fixed constant $\varepsilon \in (0, 1)$, then the optimal Type II error probability cannot decrease to zero exponentially faster than a rate determined by the relative entropy. Otherwise, the Type I error necessarily converges to one. It is straightforward to employ the direct part of the established quantum Stein's lemma from [25] in

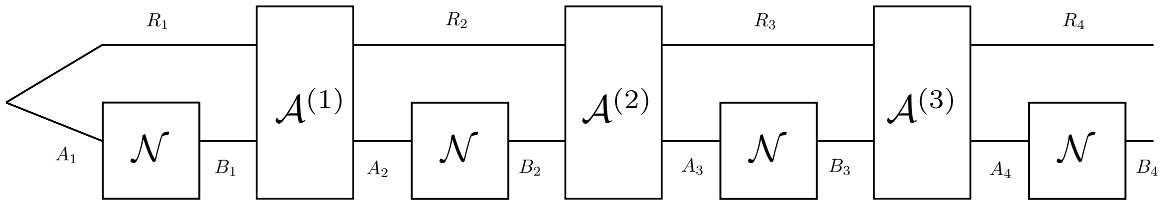

Figure 1: A four-round adaptive discrimination strategy applied to the channel $\mathcal{N}$.

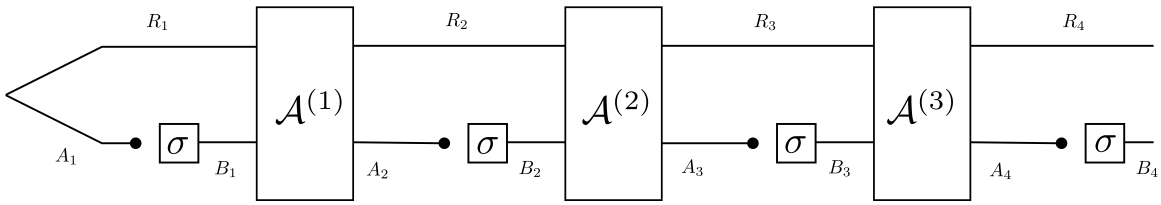

Figure 2: A four-round adaptive discrimination strategy applied to the replacer channel $\mathcal{R}$.

order to establish the direct part for our setting.

In more detail, the most general adaptive discrimination strategy is depicted in Figures 1 and 2. It consists of a choice of input state $\rho_{R_1A_1}$, a sequence $\{\mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}\}_{i\in\{1,\ldots,n-1\}}$ of adaptive quantum channels, and finally a quantum measurement $\{Q_{R_nB_n}, I_{R_nB_n}-Q_{R_nB_n}\}$ to decide which channel was applied. Let $\tau_{R_nB_n}$ denote the output state at the end of the adaptive discrimination strategy (before the final measurement $\{Q_{R_nB_n}, I_{R_nB_n}-Q_{R_nB_n}\}$ is performed) when the channel being applied is $\mathcal{R}$, and let $\rho_{R_nB_n}$ denote the output state at the end of the adaptive discrimination strategy when the channel being applied is $\mathcal{N}$. Let $D^{\varepsilon}_{H,\text{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{\otimes n})$ denote the "adaptive hypothesis testing relative entropy," which generalizes (1.6) by allowing for an optimization over all possible adaptive strategies used to discriminate between $\mathcal{N}^{\otimes n}$ and $\mathcal{R}^{\otimes n}$. We define it formally as follows:

$$D^{\varepsilon}_{H,\text{ad}}\left(\mathcal{N}^{\otimes n}\middle\|\mathcal{R}^{\otimes n}\right) \equiv -\log \beta^{\text{ad}}_{\varepsilon}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{\otimes n}) = -\log \inf \operatorname{Tr}\{Q_{R_nB_n}\tau_{R_nB_n}\}, \tag{2.1}$$

where the infimum is over all measurement operators $Q_{R_nB_n}$ subject to $0 \leq Q_{R_nB_n} \leq I_{R_nB_n}$ and

$$\operatorname{Tr}\{Q_{R_nB_n}\rho_{R_nB_n}\} \geq 1-\varepsilon, \tag{2.2}$$

all preparation states $\rho_{R_1A_1}$ subject to $\rho_{R_1A_1} \geq 0$ and $\operatorname{Tr}\{\rho_{R_1A_1}\} = 1$, and all adaptive quantum channels

$$\left\{\mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}\right\}_{i\in\{1,\ldots,n-1\}}. \tag{2.3}$$

We can now state our first main result:

Theorem 1 *Let $\varepsilon \in (0,1)$ be a fixed constant. Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be an arbitrary quantum channel and let $\mathcal{R} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be the replacer quantum channel $\mathcal{R}(X_A) = \operatorname{Tr}\{X_A\}\sigma_B$, for* *some fixed density operator σ_B. Then the channel version of Stein's lemma,* (1.14) *holds, i.e.,*

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_\varepsilon^{\mathrm{ad}}\left(\mathcal{N}^{\otimes n}\middle\|\mathcal{R}^{\otimes n}\right) = \lim_{n\to\infty}\frac{1}{n}D_{H,\mathrm{ad}}^{\varepsilon}\left(\mathcal{N}^{\otimes n}\middle\|\mathcal{R}^{\otimes n}\right) \tag{2.4}$$

$$= \sup_{\psi_{RA}} D\left(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B\right) = D(\mathcal{N}\|\mathcal{R}) \tag{2.5}$$

for any $\varepsilon\in(0,1)$. It suffices to take system R isomorphic to system A in the above optimization.

This theorem clearly generalizes the quantum Stein's lemma in (1.7). It implies that a tensor-power discrimination strategy is optimal—allowing for an adaptive strategy yields no asymptotic improvement. That is, one should simply prepare n copies of the bipartite state ψ_{RA} optimizing (2.5), send each A system through each channel use (creating the state $[\mathcal{N}_{A\to B}(\psi_{RA})]^{\otimes n}$ or $[\mathcal{R}_{A\to B}(\psi_{RA})]^{\otimes n}$), and finally perform a collective measurement on all systems R^nB^n to decide which channel was applied.

2.2 The strong converse exponent for adaptive channel discrimination

Next, we refine our analysis by identifying the strong converse exponent for the task of discriminating between an arbitrary quantum channel $\mathcal{N}$ and a replacer channel $\mathcal{R}$. It is easy to see (by considering $\varepsilon\to 0$) that Theorem 1 implies that for any rate $r<D(\mathcal{N}\|\mathcal{R})$, there exists a sequence of non-adaptive strategies, along which the type I error goes to zero, and the type II error vanishes exponentially fast, with a rate at least r. This is usually referred to as the direct part of Stein's lemma. Moreover, it also implies that the strong converse property holds, i.e., for any sequence of adaptive srategies, if the type II error vanishes exponentially with a rate $r>D(\mathcal{N}\|\mathcal{R})$, then the type I error goes to 1 (this can be seen by taking $\varepsilon\to 1$). Our aim is to determine the speed of convergence of the type I error to 1 in the strong converse domain, for any decay rate $r>D(\mathcal{N}\|\mathcal{R})$ of the type II errors. As it turns out, this convergence is also exponential, and hence our aim is to determine the exact values of the strong converse exponents:

$$\underline{\mathrm{sc}}(r) \equiv \inf\left\{\liminf_{n\to+\infty} -\frac{1}{n}\log\mathrm{Tr}\left\{Q_{R_nB_n}\rho_{R_nB_n}\right\} : \liminf_{n\to+\infty} -\frac{1}{n}\log\mathrm{Tr}\left\{Q_{R_nB_n}\tau_{R_nB_n}\right\} > r\right\}, \tag{2.6}$$

$$\overline{\mathrm{sc}}(r) \equiv \inf\left\{\limsup_{n\to+\infty} -\frac{1}{n}\log\mathrm{Tr}\left\{Q_{R_nB_n}\rho_{R_nB_n}\right\} : \liminf_{n\to+\infty} -\frac{1}{n}\log\mathrm{Tr}\left\{Q_{R_nB_n}\tau_{R_nB_n}\right\} > r\right\}, \tag{2.7}$$

where the infimum is over all sequences of adaptive measurement strategies, specified by measurement operators $Q_{R_nB_n}$ subject to $0\le Q_{R_nB_n}\le I_{R_nB_n}$, preparation states $\rho_{R_1A_1}$, and adaptive quantum channels

$$\mathcal{A}_{[n]} \equiv \left\{\mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}\right\}_{i\in\{1,\ldots,n-1\}}. \tag{2.8}$$

We establish the following theorem:

Theorem 2 *Let $\mathcal{N}:\mathcal{B}(\mathcal{H}_A)\to\mathcal{B}(\mathcal{H}_B)$ be an arbitrary quantum channel, and let $\mathcal{R}:\mathcal{B}(\mathcal{H}_A)\to\mathcal{B}(\mathcal{H}_B)$ be the replacer quantum channel $\mathcal{R}(X)=\mathrm{Tr}\{X\}\sigma_B$, for some fixed density operator σ_B.*

For any $r > D(\mathcal{N}\|\mathcal{R})$,

$$\underline{\mathrm{sc}}(r) = \overline{\mathrm{sc}}(r) = \lim_{n\to+\infty} -\frac{1}{n}\log(1-\alpha^{\mathrm{ad}}_{n,r}) \tag{2.9}$$

$$= \sup_{\alpha>1}\inf_{\psi_{RA}} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B)\right] \tag{2.10}$$

$$= \inf_{\psi_{RA}}\sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B)\right] \tag{2.11}$$

$$= \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R})\right], \tag{2.12}$$

where $\alpha^{\mathrm{ad}}_{n,r}$ is defined in (1.13)*, the infima are taken over all possible bipartite states ψ_{RA} with an arbitrary ancilla system R; in particular,* (1.16) *holds. Moreover, the same identities hold when the infima are restricted to pure states ψ_{RA} with R being a fixed copy of A.*

Remark 3 *When $D(\mathcal{N}\|\mathcal{R}) = +\infty$, the above statement is empty. On the other hand, when $D(\mathcal{N}\|\mathcal{R})$ is finite, then Theorem 2 also holds for $0 < r \le D(\mathcal{N}\|\mathcal{R})$ in a trivial way. Indeed, by Theorem 1, if $r \le D(\mathcal{N}\|\mathcal{R})$ then the operational quantities in* (2.9) *are equal to* 0*, and so is* (2.12)*, since $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}) \ge D(\mathcal{N}\|\mathcal{R})$ for every $\alpha > 1$, according to Lemma 10.*

2.3 Connection to quantum illumination

Our results have implications for the theory of quantum illumination, which we discuss briefly here. Building on prior work in [47, 48], Lloyd *et al.* show how the use of entangled photons can provide a significant improvement over unentangled light when detecting the presence of an object [34, 50]. The goal in quantum illumination is to determine whether a distant object is present or not by employing quantum light along with a quantum detection strategy. It is sensible and traditional [34, 50] to take the object not being present as the null hypothesis and the object being present as the alternative hypothesis.

In the usual scenario, the transmitter and receiver are in the same location. The protocol begins with the transmitter sending a signal mode that is entangled with an idler mode still in the possession of the transmitter. Let $|\psi\rangle_{SI}$ denote the state of the signal and idler mode. If the object is not present (the null hypothesis), then the signal mode is lost and is replaced by a thermal state θ_S, so that the joint state becomes $\theta_S\otimes\psi_I$. Clearly, this is an instance of the replacer channel. If the object is present (the alternative hypothesis), then the signal beam is reflected off the object and returns to the transmitter. The resulting state is described by $(\mathcal{N}_S\otimes\mathrm{id}_I)(\psi_{SI})$, where $\mathcal{N}_S$ describes the noise characteristics of the reflection channel. This protocol is performed n times with the receiver storing either the state $[\theta_S\otimes\psi_I]^{\otimes n}$ or $[(\mathcal{N}_S\otimes\mathrm{id}_I)(\psi_{SI})]^{\otimes n}$. The receiver finally performs a collective measurement on all of the systems in order to decide whether the object is present. Thus, we have a quantum channel discrimination problem in which one seeks to distinguish between a replacer channel and a noisy channel. However, our results do not apply to this setting if one takes the null and alternative hypotheses in the natural way suggested above.

An alternative scenario is that in which the transmitter and receiver are in different locations. It is technologically more challenging to take advantage of quantum illumination in this setting, due to the fact that the transmitter and receiver need to share and store entanglement over a potentially large distance. Nevertheless, this is the setting to which our results apply. Given that the null hypothesis in this setting corresponds to the object not being present, the channel applied to the transmitted mode will be $\mathcal{N}$, which characterizes the optical loss in the transmission. Since the alternative hypothesis in this setting corresponds to the object being present and such an object will reflect the light incident on it, the signal beam does not make it to the receiving end and the receiver instead detects thermal noise, so that the channel applied to the transmitted mode is the replacer channel $\mathcal{R}$. Thus, the Type I and Type II errors for this setting correspond to our setting described in the previous sections.

Implicit in prior analyses on quantum illumination is the assumption that a tensor-power, non-adaptive strategy is optimal. Our results support this assumption (at least in the particular setting of asymmetric hypothesis testing described above) by showing that no asymptotic advantage is provided by instead using an adaptive strategy for quantum channel discrimination.[1] It remains an open question to determine if a tensor-power, non-adaptive strategy is optimal in the symmetric hypothesis testing setting considered in [34, 50].

2.4 Strong converse theorem for quantum-feedback-assisted communication

There is a well-known connection between hypothesis testing and channel coding, first recognized by Blahut [8], and this connection also holds for quantum channels. The direct part of the channel coding theorem (i.e., the Holevo-Schumacher-Westmoreland theorem) [26, 49] can be obtained from the direct part of Stein's lemma, as shown in [23, 42].

One consequence of Theorem 1 is a strong converse theorem for the quantum-feedback-assisted classical capacity of a quantum channel. In prior work, Bowen proved that a noiseless quantum feedback channel does not increase the entanglement-assisted capacity [6, 7, 28] of a noisy channel, by proving a weak converse for its quantum-feedback-assisted capacity [9]. That is, Bowen proved that the quantum-feedback-assisted capacity of a channel $\mathcal{N}$ is equal to its entanglement-assisted capacity, denoted by

$$I(\mathcal{N}) \equiv \sup_{\psi_{RA}} \inf_{\sigma_B} D\left(\mathcal{N}_{A\to B}(\psi_{RA}) \| \psi_R \otimes \sigma_B\right). \tag{2.13}$$

However, Bowen's result did not exclude the possibility of a trade-off between the communication rate and the error probability; our strong converse theorem shows that no such trade-off is possible in the asymptotic limit of many channel uses. A strong converse theorem in this context states that for any coding scheme, which seeks to transmit at a rate strictly higher than the capacity of the channel, the probability of successful decoding decays to zero exponentially fast in the number of channel uses. So our result sharpens Bowen's [9], strengthens the main result of [18], and generalizes [45, Theorem 7] to the quantum case. The approach taken is inspired by that used by Nagaoka [38], who derived the strong converse theorem for any memoryless quantum channel from the monotonicity of the Rényi relative entropies. Polyanskiy and Verdú [45] later generalised this approach to show how a bound on the success probability could be derived from any relative-entropy-like quantity that satisfies certain natural properties. This approach has already been used to prove several strong converse theorems for quantum channels [31, 56, 18, 53]; here we shall use the sandwiched Rényi relative entropy [37, 56].

[1] Strictly speaking, the results in our paper apply to finite-dimensional systems, whereas the quantum illumination protocols apply to infinite-dimensional, albeit finite-energy, systems. Given that our analysis never has any dimension dependence, this suggests that it should be possible to extend our results to infinite-dimensional systems with energy constraints.

With the proof of Theorem 1 in hand, it requires only a little extra effort to prove a strong converse for the quantum-feedback-assisted capacity of a quantum channel (the capacity when unlimited use of a noiseless quantum feedback channel from receiver to sender is allowed).

Theorem 4 *Let p_{succ} denote the success probability of any rate R quantum-feedback-assisted communication code for a channel $\mathcal{N}$ that uses it $n \geq 1$ times. The following bound holds*

$$p_{\text{succ}} \leq 2^{-n \sup_{\alpha>1}\left(\frac{\alpha-1}{\alpha}\right)\left(R-\widetilde{I}_{\alpha}(\mathcal{N})\right)}, \tag{2.14}$$

where $\widetilde{I}_{\alpha}(\mathcal{N})$ is the sandwiched Rényi mutual information of the channel $\mathcal{N}$, defined in (3.27)*. As a consequence of this bound, we can conclude a strong converse: for any sequence of quantum-feedback-assisted codes for a channel $\mathcal{N}$ with rate $R > I(\mathcal{N})$, the success probability decays exponentially to zero as $n \to \infty$.*

Note that the second statement in Theorem 4 has in fact already been proved in [5, Section IV-E1], via the channel simulation technique. However, our new contribution here is to provide the bound in (2.14) on the strong converse exponent, in addition to providing an arguably more direct proof of the theorem. It remains an open question to determine if the strong converse exponent bound in (2.14) is optimal (i.e., if there exists a quantum-feedback-assisted communication scheme achieving this exponent in the strong converse regime).

3 Rényi relative entropies

For two Hilbert spaces $\mathcal{H}, \mathcal{K}$, let $\mathcal{B}(\mathcal{H}, \mathcal{K})$ denote the set of bounded linear operators from $\mathcal{H}$ to $\mathcal{K}$. When $\mathcal{K} = \mathcal{H}$, we use the shorthand notation $\mathcal{B}(\mathcal{H})$. We restrict ourselves to finite-dimensional Hilbert spaces throughout this paper. The Schatten α-norm of an operator X is defined as

$$\|X\|_{\alpha} \equiv \operatorname{Tr}\{(\sqrt{X^{*}X})^{\alpha}\}^{1/\alpha}, \tag{3.1}$$

for $\alpha \geq 1$. Let $\mathcal{B}(\mathcal{H})_{+}$ denote the subset of positive semi-definite operators; we often simply say that an operator is "positive" if it is positive semi-definite. We also write $X \geq 0$ if $X \in \mathcal{B}(\mathcal{H})_{+}$. An operator ρ is in the set $\mathcal{S}(\mathcal{H})$ of density operators if $\rho \in \mathcal{B}(\mathcal{H})_{+}$ and $\operatorname{Tr}\{\rho\} = 1$. We denote by $\mathcal{B}(\mathcal{H})_{++}$ and $\mathcal{S}(\mathcal{H})_{++}$ the set of positive definite operators and states on $\mathcal{H}$, respectively.

The tensor product of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ is denoted by $\mathcal{H}_A \otimes \mathcal{H}_B$. Given a bipartite density operator $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we write $\rho_A = \operatorname{Tr}_B\{\rho_{AB}\}$ for the reduced density operator on system A. A linear map $\mathcal{N}_{A\to B} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is positive if $\mathcal{N}_{A\to B}(\sigma_A) \in \mathcal{B}(\mathcal{H}_B)_{+}$ whenever $\sigma_A \in \mathcal{B}(\mathcal{H}_A)_{+}$. Let id_A denote the identity map acting on a system A. A linear map $\mathcal{N}_{A\to B}$ is completely positive if the map $\mathrm{id}_R \otimes \mathcal{N}_{A\to B}$ is positive for a reference system R of arbitrary size. A linear map $\mathcal{N}_{A\to B}$ is trace-preserving if $\operatorname{Tr}\{\mathcal{N}_{A\to B}(\tau_A)\} = \operatorname{Tr}\{\tau_A\}$ for all input operators $\tau_A \in \mathcal{B}(\mathcal{H}_A)$. If a linear map is completely positive and trace-preserving (CPTP), we say that it is a quantum channel or quantum operation. A positive operator-valued measure (POVM) is a set $\{\Lambda^m\}$ of positive operators such that $\sum_m \Lambda^m = I$.

The quantum Rényi relative entropy of order $\alpha \in [0,1) \cup (1,\infty)$ between two non-zero positive semidefinite operators ρ and σ is given by [43]

$$D_{\alpha}(\rho\|\sigma) \equiv \begin{cases} \frac{1}{\alpha-1}\log \frac{1}{\operatorname{Tr}\rho}\operatorname{Tr}\left\{\rho^{\alpha}\sigma^{1-\alpha}\right\} & \text{if } \rho \not\perp \sigma \text{ and } (\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma) \text{ or } \alpha \in [0,1)\,) \\ +\infty & \text{otherwise.} \end{cases}, \tag{3.2}$$

with the support conditions established in [51]. Here and henceforth we use the convention that powers of a positive semidefinite operator X are taken only on its support, i.e., if $x_1, \ldots, x_r$ are the strictly positive eigenvalues of X with corresponding spectral projections $P_1, \ldots, P_r$, then $X^t \equiv \sum_{i=1}^r x_i^t P_i$ for every $t \in \mathbb{R}$. In particular, X^0 denotes the projection onto the support of X.

Recently, the sandwiched Rényi relative entropy [37, 56] was introduced. It is defined for $\alpha \in (0,1) \cup (1,\infty)$ as follows:

$$\widetilde{D}_\alpha(\rho\|\sigma) \equiv \begin{cases} \frac{1}{\alpha-1} \log \left[\frac{1}{\operatorname{Tr}\rho} \operatorname{Tr}\left\{\left(\sigma^{(1-\alpha)/2\alpha} \rho \sigma^{(1-\alpha)/2\alpha}\right)^\alpha\right\}\right] & \text{if } \rho \not\perp \sigma \text{ and } (\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma) \\ & \text{or } \alpha \in (0,1)\ ) \\ +\infty & \text{otherwise} \end{cases}. \tag{3.3}$$

It is known [35, 36] that for any fixed ρ, σ,

$$\alpha \mapsto D_\alpha(\rho\|\sigma) \quad \text{and} \quad \alpha \mapsto \widetilde{D}_\alpha(\rho\|\sigma) \quad \text{are monotone increasing,} \tag{3.4}$$

and in the limit $\alpha \to 1$, they both give the relative entropy [37, 56]:

$$\lim_{\alpha\to1} \widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma) \equiv D_1(\rho\|\sigma). \tag{3.5}$$

The Rényi relative entropies have several desirable properties which justify viewing them as distinguishability measures. In particular, $\widetilde{D}_\alpha(\rho\|\sigma)$ satisfies the following data-processing inequality for $\alpha \in [1/2,1) \cup (1,\infty)$ [17, 4, 37, 56, 36]:

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha\left(\mathcal{N}(\rho)\,\|\mathcal{N}(\sigma)\right), \tag{3.6}$$

where $\mathcal{N}$ is a CPTP map. A similar inequality holds for $D_\alpha(\rho\|\sigma)$ when $\alpha \in [0,1) \cup (1,2]$ [43].

The following simple lemma relates the hypothesis testing relative entropy to the sandwiched Rényi relative entropy. The idea for its proof goes back to [25, 38, 41].

Lemma 5 *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be such that* $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$. *For any $Q \in \mathcal{B}(\mathcal{H})$ such that $0 \leq Q \leq I$, and any $\alpha > 1$,*

$$-\log \operatorname{Tr} Q\sigma \leq \widetilde{D}_\alpha(\rho\|\sigma) - \frac{\alpha}{\alpha-1} \log \operatorname{Tr} Q\rho. \tag{3.7}$$

In particular, for any $\alpha > 1$ and any $\varepsilon \in (0,1)$,

$$D_H^\varepsilon(\rho\|\sigma) \leq \widetilde{D}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha-1} \log\left(\frac{1}{1-\varepsilon}\right). \tag{3.8}$$

Proof. Let $p \equiv \operatorname{Tr}\{Q\rho\}$ and $q \equiv \operatorname{Tr}\{Q\sigma\}$. By the monotonicity of the sandwiched Rényi relative entropy for $\alpha > 1$, we find that

$$\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha\left((p,1-p)\,\|\,(q,1-q)\right) \tag{3.9}$$

$$= \frac{1}{\alpha-1} \log\left[p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\right] \tag{3.10}$$

$$\geq \frac{1}{\alpha-1} \log\left[p^\alpha q^{1-\alpha}\right] \tag{3.11}$$

$$= \frac{\alpha}{\alpha-1} \log p - \log q, \tag{3.12}$$

from which (3.7) follows. The statement in (3.8) follows by optimizing over all Q such that $\mathrm{Tr}\{Q\rho\} \geq 1-\varepsilon$. ■

Recall the definition of the channel Rényi relative entropies in (1.19)–(1.20). Let A' be a copy of A, let $e_1, \ldots, e_d$ be an orthonormal basis in A, and define $|\Gamma_{A'A}\rangle \equiv |\sum_{i=1}^d e_i \otimes e_i\rangle$, $\Gamma_{A'A} \equiv |\Gamma_{A'A}\rangle\langle\Gamma_{A'A}|$. Then we have the following:

Lemma 6 *Let A' be a copy of A. For any system R and any pure state ψ_{RA}, there exists a state $\rho_{A'}$ on A' such that for any two channels $\mathcal{N}_1, \mathcal{N}_2$ from A to some system B, and any $\alpha > 0$, we have*

$$\widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = \widetilde{D}_\alpha\Big(\mathcal{N}_1\left(\rho_{A'}^{1/2}|\Gamma_{A'A}\rangle\langle\Gamma_{A'A}|\rho_{A'}^{1/2}\right)\Big\|\mathcal{N}_2\left(\rho_{A'}^{1/2}|\Gamma_{A'A}\rangle\langle\Gamma_{A'A}|\rho_{A'}^{1/2}\right)\Big) \tag{3.13}$$

$$= \widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\Big\|\rho_{A'}^{1/2}\mathcal{N}_2(\Gamma_{A'A})\rho_{A'}^{1/2}\Big). \tag{3.14}$$

Moreover, the same identities hold for D_α.

We give a proof of Lemma 6 in Appendix A.

Lemma 7 *Let $\mathcal{N}_1, \mathcal{N}_2$ be quantum channels from system A to system B. For every $\alpha \in [1/2, +\infty)$, the channel Rényi relative entropies can be written as*

$$\widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = \sup\Big\{\widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) : \psi_{RA} \text{ state on } RA\text{, where } R \text{ is arbitrary}\Big\} \tag{3.15}$$

$$= \sup\Big\{\widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) : \psi_{RA} \text{ pure state on } RA\text{, where } R \cong A\Big\} \tag{3.16}$$

$$= \sup\Big\{\widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}^{1/2}\mathcal{N}_2(\Gamma_{A'A})\rho_{A'}^{1/2}\Big) : \rho_{A'} \text{ state on } A'\text{, where } A' \cong A\Big\}. \tag{3.17}$$

Analogous formulas hold for $D_\alpha(\mathcal{N}_1\|\mathcal{N}_2)$ in (1.19) and $\alpha \in [0,2]$.

Proof. According to [17], $\widetilde{D}_\alpha$ is jointly quasi-convex for $\alpha \in [1/2, +\infty)$, and by [1, 33, 43], the same holds for D_α and $\alpha \in [0,2]$. Hence, the optimizations in (1.19)–(1.20) can be restricted to pure states, and the rest of the proof is immediate from Lemma 6. ■

When the second channel is a replacer channel, the sandwiched channel Rényi relative entropy has a special representation as explained below. This will be key to our approach of obtaining strong converse bounds.

A quantum channel $\mathcal{N}_{A\to B}$ induces a map from $L_1(\mathcal{B}(\mathcal{H}_A)) \to L_\alpha(\mathcal{B}(\mathcal{H}_B))$, where $L_\alpha(\mathcal{B}(\mathcal{H})))$ denotes the space $\mathcal{B}(\mathcal{H})$ together with the Schatten α-norm $\|X\|_\alpha$. The space $L_\alpha(\mathcal{B}(\mathcal{H}))$ has a canonical operator space structure [44], a certain sequence of norms on the spaces $M_n(L_\alpha(\mathcal{B}(\mathcal{H})))$:

$$\|Y\|_{M_n(L_\alpha(\mathcal{B}(\mathcal{H})))} \equiv \sup_{A,B\in M_n} \frac{\|(A\otimes I_{\mathcal{H}})Y(B\otimes I_{\mathcal{H}})\|_\alpha}{\|A\|_{2\alpha}\|B\|_{2\alpha}}. \tag{3.18}$$

One can then define the completely bounded $(1\to\alpha)$-norm of $\mathcal{N} : L_1(\mathcal{B}(\mathcal{H}_A)) \to L_\alpha(\mathcal{B}(\mathcal{H}_B))$ as

$$\sup_n \|\mathrm{id}_n \otimes \mathcal{N}\|_{1\to\alpha} \equiv \sup_n \sup_Y \frac{\|(\mathrm{id}_n \otimes \mathcal{N})(Y)\|_{M_n(L_\alpha(\mathcal{B}(\mathcal{H}_B)))}}{\|Y\|_{M_n(L_1(\mathcal{B}(\mathcal{H}_A)))}}. \tag{3.19}$$

For our purposes, it will be more useful to write the completely bounded $(1 \to \alpha)$-norm of a quantum channel $\mathcal{N}$ as

$$\|\mathcal{N}\|_{\mathrm{CB},1\to\alpha} = \sup_{X\in\mathcal{B}(\mathcal{H}_{A'}\otimes\mathcal{H}_A)_+} \frac{\|(\mathrm{id}\otimes\mathcal{N})(X)\|_\alpha}{\|\operatorname{Tr}_A\{X\}\|_\alpha} = \sup_{|\psi\rangle\in\mathcal{H}_{A'}\otimes\mathcal{H}_A} \frac{\|(\mathrm{id}\otimes\mathcal{N})(|\psi\rangle\langle\psi|)\|_\alpha}{\|\operatorname{Tr}_A\{|\psi\rangle\langle\psi|\}\|_\alpha}, \tag{3.20}$$

where A' is any system with dimension at least that of A; in particular, A' can be taken to be a fixed copy of A. This follows from [13] and Eq. (8) of [29], where these norms have already been considered in the context of quantum information theory. The above representation of the completely bounded $(1 \to \alpha)$-norm will prove useful later due to the following connection between the sandwiched Rényi relative entropy and the Schatten α-norm:

$$\widetilde{D}_\alpha(\rho\|\sigma) = \frac{\alpha}{\alpha-1}\log\left\|\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha, \tag{3.21}$$

where ρ and σ are density operators. Throughout, Θ_X denotes the map

$$\Theta_X(Y) \equiv X^{1/2}YX^{1/2}, \tag{3.22}$$

where X is a positive operator.

The following lemma is from [18]; for readers' convenience, we give a detailed proof in Appendix A.

Lemma 8 *Let $\mathcal{N} = \mathcal{N}_{A\to B}$ be a quantum channel and $\mathcal{R}_{\sigma_B}(\cdot) \equiv \operatorname{Tr}\{\cdot\}\sigma_B$ be a replacer channel with some fixed state σ_B. For every $\alpha \in (1, +\infty)$,*

$$\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) = \sup_{\psi_{RA}} \widetilde{D}_\alpha\left(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B\right) = \frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha}. \tag{3.23}$$

We will also use the following Rényi mutual information quantities, originally defined in [56, 4, 18]. For every bipartite state ρ_{RB}, and every $\alpha \in (0, +\infty)$, let

$$I_\alpha(R;B)_\rho \equiv \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)} D_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B), \tag{3.24}$$

$$\widetilde{I}_\alpha(R;B)_\rho \equiv \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)} \widetilde{D}_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B). \tag{3.25}$$

These quantities appeared in the direct and strong converse exponents of [24]. We also define the channel Rényi mutual informations. For any CPTP map $\mathcal{N}_{A\to B} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, let

$$I_\alpha(\mathcal{N}) \equiv \sup_{\psi_{RA}\in\mathcal{S}(\mathcal{H}_{RA})} I_\alpha(R;B)_\omega, \tag{3.26}$$

$$\widetilde{I}_\alpha(\mathcal{N}) \equiv \sup_{\psi_{RA}\in\mathcal{S}(\mathcal{H}_{RA})} \widetilde{I}_\alpha(R;B)_\omega, \tag{3.27}$$

where $\omega_{RB} \equiv \mathcal{N}_{A\to B}(\psi_{RA})$.

Lemma 9 *Let A' be a copy of A. Then*

$$I_\alpha(\mathcal{N}) = \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} I_\alpha(R;B)_\omega, \qquad \alpha \in [0,2], \tag{3.28}$$

$$\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} I_\alpha(R;B)_\omega, \qquad \alpha \in [1/2,+\infty), \tag{3.29}$$

where $\omega_{RB} \equiv \mathcal{N}_{A\to B}\left(\rho_R^{1/2}|\Gamma_{RA}\rangle\langle\Gamma_{RA}|\rho_R^{1/2}\right)$.

Proof. According to [17], $\widetilde{D}_\alpha$ is monotone non-increasing under partial trace for $\alpha \in [1/2, +\infty)$, and by [43], the same holds for D_α and $\alpha \in [0, 2]$. Hence, by taking purifications of ψ_{RA} in (3.26) and (3.27), the values can only increase. Thus, the optimizations in (3.26) and (3.27) can be restricted to pure states. Using Lemma 6 with $\mathcal{N}_1 = \mathcal{N}$ and $\mathcal{N}_2 = \mathcal{R}_{\sigma_B}$, the assertions follow. ■

Note that for $\alpha = 1$, the above quantities are defined using the relative entropy $D = D_1$, and we have $I_1(R; B)_\rho = \widetilde{I}_1(R; B)_\rho \equiv I(R; B)_\rho = D(\rho_{RB} \| \rho_R \otimes \rho_B)$, and $I_1(\mathcal{N}) = \widetilde{I}_1(\mathcal{N}) = I(\mathcal{N})$, where $I(\mathcal{N})$ is defined in (2.13). We will need the following extensions of (3.4)–(3.5):

Lemma 10 *(i) For any two channels $\mathcal{N}_1, \mathcal{N}_2$, $D_\alpha(\mathcal{N}_1 \| \mathcal{N}_2)$ and $\widetilde{D}_\alpha(\mathcal{N}_1 \| \mathcal{N}_2)$ are monotone increasing in α, and*

$$\lim_{\alpha \to 1} \widetilde{D}_\alpha(\mathcal{N}_1 \| \mathcal{N}_2) = \lim_{\alpha \to 1} D_\alpha(\mathcal{N}_1 \| \mathcal{N}_2) = D(\mathcal{N}_1 \| \mathcal{N}_2). \tag{3.30}$$

(ii) For every bipartite state ρ_{RB}, $I_\alpha(R; B)_\rho$ and $\widetilde{I}_\alpha(R; B)_\rho$ are monotone increasing in α, and

$$\lim_{\alpha \to 1} I_\alpha(R; B)_\rho = \lim_{\alpha \to 1} \widetilde{I}_\alpha(R; B)_\rho = I(R; B)_\rho. \tag{3.31}$$

(iii) For every channel $\mathcal{N}$, $I_\alpha(\mathcal{N})$ and $\widetilde{I}_\alpha(\mathcal{N})$ are monotone increasing in α, and

$$\lim_{\alpha \to 1} I_\alpha(\mathcal{N}) = \lim_{\alpha \to 1} \widetilde{I}_\alpha(\mathcal{N}) = I(\mathcal{N}). \tag{3.32}$$

Proof. See Appendix A. ■

The channel Rényi mutual informations also have the following geometric interpretation, as the "distance" of the channel from the set of all replacer channels, where the "distance" is measured by the channel Rényi divergences. See Section 4.3 for the relevance of this geometric picture.

Lemma 11 *For every channel $\mathcal{N}_{A \to B}$, and every $\alpha \in [1/2, +\infty)$,*

$$\widetilde{I}_\alpha(\mathcal{N}) = \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} \widetilde{D}_\alpha(\mathcal{N} \| \mathcal{R}_{\sigma_B}). \tag{3.33}$$

Proof. See Appendix A. ■

4 The strong converse theorem for adaptive quantum channel discrimination

4.1 Quantum Stein's lemma for adaptive channel discrimination

This section provides a proof of Theorem 1. In the setting of this theorem, we seek to distinguish between an arbitrary quantum channel $\mathcal{N}$ and a "replacer" channel $\mathcal{R}$ that maps all states ω_A to a fixed state σ, i.e., $\mathcal{R}(\omega_A) = \mathrm{Tr}\{\omega_A\}\sigma_B$. We allow the preparation of an arbitrary input state $\rho_{R_1 A_1} = \tau_{R_1 A_1}$, 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, an adaptive operation $\mathcal{A}^{(i)}$ is applied to the registers R_i and B_i, yielding a quantum state $\rho_{R_{i+1}A_{i+1}}$ or $\tau_{R_{i+1}A_{i+1}}$ in registers $R_{i+1}A_{i+1}$, depending on whether the channel is equal to $\mathcal{N}$ or $\mathcal{R}$. That is,

$$\rho_{R_{i+1}A_{i+1}} \equiv \mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}(\rho_{R_iB_i}), \qquad \rho_{R_iB_i} \equiv \mathcal{N}_{A_i\to B_i}(\rho_{R_iA_i}) \tag{4.1}$$

$$\tau_{R_{i+1}A_{i+1}} \equiv \mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}(\tau_{R_iB_i}), \qquad \tau_{R_iB_i} \equiv \mathcal{R}_{A_i\to B_i}(\tau_{R_iA_i}) \tag{4.2}$$

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 $\{Q_{R_nB_n}, I_{R_nB_n} - Q_{R_nB_n}\}$ is performed on the systems R_nB_n to decide which channel was applied. Such a general protocol is depicted in Figures 1 and 2. Note that since $\mathcal{R}$ is a replacer channel, we can write

$$\tau_{R_iB_i} = \tau_{R_i} \otimes \sigma_{B_i}, \qquad 1 \le i \le n. \tag{4.3}$$

Recall the hypothesis testing relative entropy $D^\varepsilon_H(\rho\|\sigma)$ from (1.6) and the "adaptive hypothesis testing relative entropy" $D^\varepsilon_{H,\text{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{\otimes n})$ from (2.1). So $D^\varepsilon_H(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n})$ denotes the hypothesis testing relative entropy in which there is a fixed initial state $\rho_{R_1A_1}$ and fixed adaptive maps $\{\mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}\}_{i\in\{1,\dots,n-1\}}$.

Clearly, we have that

$$\liminf_{n\to\infty} \frac{1}{n} D^\varepsilon_{H,\text{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{\otimes n}) \ge \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B), \tag{4.4}$$

by employing a tensor-power strategy with no adaptation (i.e., we can simply invoke the direct part of the usual quantum Stein's lemma). In more detail, the initial state of this strategy is the optimal ψ_{RA} in (4.4) and each map $\mathcal{A}^{(i)}_{R_iB_i\to R_{i+1}A_{i+1}}$ simply prepares the state ψ_{RA} at the input of the $i+1$st channel while acting as the identity map on the i states $(\mathcal{N}_{A\to B}(\psi_{RA}))^{\otimes i}$ or $(\mathcal{R}_{A\to B}(\psi_{RA}))^{\otimes i}$ (so the strategy is non-adaptive). After the nth channel has acted, the discriminator performs a binary collective measurement on the state $(\mathcal{N}_{A\to B}(\psi_{RA}))^{\otimes n}$ or $(\mathcal{R}_{A\to B}(\psi_{RA}))^{\otimes n}$ to decide which channel was applied. So the lower bound in (4.4) follows directly from the state discrimination result in (1.7).

The more interesting part is to show that this strategy is asymptotically optimal, i.e., that

$$\limsup_{n\to\infty} \frac{1}{n} D^\varepsilon_{H,\text{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{\otimes n}) \le \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B). \tag{4.5}$$

Since this inequality is trivial when $\sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B) = +\infty$, we assume the contrary for the rest. We start by bounding the adaptive hypothesis testing relative entropy in terms of the sandwiched Rényi relative entropy.

Throughout this section, the parameter α is assumed to be strictly larger than one and we fix some constant $\varepsilon \in (0,1)$. We fix some input state $\rho_{R_1A_1}$ and an adaptive strategy $(\mathcal{A}^{(1)},\cdots,\mathcal{A}^{(n-1)})$. Lemma 5 implies that

$$D^\varepsilon_H(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n}) \le \widetilde{D}_\alpha(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n}) + \frac{\alpha}{\alpha-1}\log\left(\frac{1}{1-\varepsilon}\right). \tag{4.6}$$

We now focus on the $\widetilde{D}_\alpha$ term. Let Θ_ω denote the completely positive map $\Theta_\omega(X) = \omega^{1/2} X \omega^{1/2}$ that conjugates X by a positive operator $\omega^{1/2}$. From (3.21), it follows that

$$\begin{aligned}
&\widetilde{D}_\alpha(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n}) \\
&= \frac{\alpha}{\alpha-1}\log\left\|(\tau_{R_n}\otimes\sigma_{B_n})^{\frac{1-\alpha}{2\alpha}}\,\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\,(\tau_{R_n}\otimes\sigma_{B_n})^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha &(4.7)\\
&= \frac{\alpha}{\alpha-1}\log\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\left(\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_nA_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha. &(4.8)
\end{aligned}$$

Let us focus on the expression inside the logarithm:

$$\begin{aligned}
&\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\left(\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_nA_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha \\
&= \frac{\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\left(\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_nA_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right)\right\|_\alpha}{\left\|\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha}\cdot\left\|\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha &(4.9)\\
&\le \left(\sup_{X_{R_nA_n}\ge 0}\frac{\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)(X_{R_nA_n})\right\|_\alpha}{\|X_{R_n}\|_\alpha}\right)\cdot\left\|\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha &(4.10)\\
&= \left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha}\cdot\left\|\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha. &(4.11)
\end{aligned}$$

The equality in (4.11) follows from the characterisation of the completely bounded $(1\to\alpha)$-norm given in (3.20). Rewriting this inequality in terms of the sandwiched Rényi relative entropy, we have that

$$\begin{aligned}
&\widetilde{D}_\alpha(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n}) \\
&\le \frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha} + \frac{\alpha}{\alpha-1}\log\left\|\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\rho_{R_n}\tau_{R_n}^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha &(4.12)\\
&= \frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha} + \widetilde{D}_\alpha(\rho_{R_n}\|\tau_{R_n}) &(4.13)\\
&\le \frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha} + \widetilde{D}_\alpha(\mathcal{N}_{A_{n-1}\to B_{n-1}}(\rho_{R_{n-1}A_{n-1}})\|\tau_{R_{n-1}}\otimes\sigma_{B_{n-1}}), &(4.14)
\end{aligned}$$

where the last inequality follows from monotonicity of the sandwiched Rényi relative entropy under the map $\mathrm{Tr}_{A_n}\circ\mathcal{A}^{(n-1)}_{R_{n-1}B_{n-1}\to R_nA_n}$.

Note that we are now left with the quantity $\widetilde{D}_\alpha(\mathcal{N}_{A_{n-1}\to B_{n-1}}(\rho_{R_{n-1}A_{n-1}})\|\tau_{R_{n-1}}\otimes\sigma_{B_{n-1}})$, which corresponds to applying the first $n-1$ rounds of the adaptive discrimination process. We can thus iterate the above argument through all n steps of the adaptive strategy. Noting that $\rho_{R_1} = \tau_{R_1}$, and thus $\widetilde{D}_\alpha(\rho_{R_1}\|\tau_{R_1}) = 0$, we obtain the bound

$$\widetilde{D}_\alpha\left(\mathcal{N}_{A_n\to B_n}(\rho_{R_nA_n})\|\tau_{R_n}\otimes\sigma_{B_n}\right) \leq n\cdot\frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right\|_{\mathrm{CB},1\to\alpha} \tag{4.15}$$

$$= n\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}), \tag{4.16}$$

where (4.16) follows from Lemma 8. This bound is independent of any particular adaptive strategy used for discriminating these channels. Thus, we can conclude that

$$\frac{1}{n}D^{\varepsilon}_{H,\mathrm{ad}}\left(\mathcal{N}^{\otimes n}\middle\|\mathcal{R}^{\otimes n}\right) \leq \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}) + \frac{1}{n}\cdot\frac{\alpha}{\alpha-1}\log\left(\frac{1}{1-\varepsilon}\right). \tag{4.17}$$

Taking the limsup as $n\to\infty$, we get the ε-independent bound

$$\limsup_{n\to\infty}\frac{1}{n}D^{\varepsilon}_{H,\mathrm{ad}}\left(\mathcal{N}^{\otimes n}\middle\|\mathcal{R}^{\otimes n}\right) \leq \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}). \tag{4.18}$$

Taking now the infimum over $\alpha > 1$, the assertion follows due to Lemma 10.

4.2 The strong converse exponent for adaptive channel discrimination

Having just proven a quantum Stein's lemma for adaptive channel discrimination, it is then natural to study the trade-off between error probabilities, when we impose the condition that the Type II error probability has exponential decay rate r for

$$r > \sup_{\psi_{RA}} D(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B). \tag{4.19}$$

One expects the Type I error to tend to one exponentially quickly. Building on the above results, we identify the strong converse exponent for the channel discrimination problem (where, as before, we assume that the alternative hypothesis is a replacer channel). Our result generalizes the quantum state discrimination result from [36, Theorem IV.10]. The notation is the same as in the previous section; in particular, $\rho_{R_nB_n}$ and $\tau_{R_nB_n}$ are as in (4.1) and (4.2), respectively. Recall the definitions of $\underline{\mathrm{sc}}(r)$ and $\overline{\mathrm{sc}}(r)$ from (2.6)–(2.7), and the definition of $\alpha^{A}_{n,r}$ from (1.13). We will need the following lemma:

Lemma 12 *Let $\mathcal{N}$ be a quantum channel from system A to system B, and $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$. The following are equivalent:*

(i) For every $k\in\mathbb{N}$, every system R, and every $\psi_{RA^k}\in\mathcal{S}(\mathcal{H}_{RA^k})$, $\operatorname{supp}\mathcal{N}^{\otimes k}(\psi_{RA^k}) \subseteq \operatorname{supp}\psi_R\otimes\sigma_B^{\otimes k}$.

(ii) For every $\rho_A\in\mathcal{S}(\mathcal{H}_A)$, $\operatorname{supp}\mathcal{N}(\rho_A)\subseteq\operatorname{supp}\sigma_B$.

(iii) $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) < +\infty$ for all $\alpha\geq 1$.

(iv) $\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) < +\infty$ for some $\alpha\geq 1$.

Proof. (i)$\Longrightarrow$(ii) is trivial (by taking $\mathcal{H}_R = \mathbb{C}$ and $k = 1$). By Lemma 7,

$$\widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) = \sup_{\rho_{A'}\in\mathcal{S}(\mathcal{H}_{A'})} \widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}\otimes\sigma_B\Big).$$

(ii)$\Longrightarrow$(i) because $\operatorname{supp}\mathcal{N}^{\otimes k}(\psi_{RA^k}) \subseteq \operatorname{supp}\psi_R \otimes \bigotimes_{i=1}^k \mathcal{N}(\psi_{A_i}) \subseteq \operatorname{supp}\psi_R\otimes\sigma_B^{\otimes k}$, which follows by iterating the general inclusion $\operatorname{supp}\omega_{CD} \subseteq \operatorname{supp}\omega_C\otimes\omega_D$ (see, e.g., [46, Appendix B.4]) and applying (ii). If (ii) is satisfied then $\rho_{A'} \mapsto \widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}\otimes\sigma_B\Big)$ is a continuous finite-valued function on the compact set $\mathcal{S}(\mathcal{H}_{A'})$, and hence its supremum is finite, proving (iii). The implication (iii)$\Longrightarrow$(iv) is trivial. Finally, (iv)$\Longrightarrow$(ii) by applying the definition of $\widetilde{D}_\alpha$. ■

Proof of Theorem 2. The statement is empty when $D(\mathcal{N}\|\mathcal{R}) = +\infty$, and hence for the rest we assume the contrary.

We begin by proving the optimality part

$$\underline{\mathrm{sc}}(r) \geq \liminf_{n\to+\infty} -\frac{1}{n}\log(1-\alpha_{n,r}^{\mathrm{ad}}) \geq \sup_{\alpha>1}\inf_{\psi_{RA}} \frac{\alpha-1}{\alpha}\Big[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B)\Big]. \tag{4.20}$$

Note that if S_n, $n\in\mathbb{N}$, is a sequence of adaptive strategies such that $\liminf_{n\to\infty} -\frac{1}{n}\log\beta_n(S_n) > r$ then for all large enough n, $\beta_n(S_n) \leq 2^{-nr}$, and thus $1-\alpha_n(S_n) \leq 1-\alpha_{n,r}^{\mathrm{ad}}$, which yields the first inequality in (4.20).

To prove the second inequality in (4.20), consider the output states $\rho_{R_nB_n}, \tau_{R_nB_n}$, and the test $Q_{R_nB_n}$, at the end of the adaptive discrimination strategy. By Lemma 5 and (4.16), we get

$$\frac{1}{n}\log\operatorname{Tr}\{Q_{R_nB_n}\rho_{R_nB_n}\} \leq \frac{\alpha-1}{\alpha}\left[\frac{1}{n}\log\operatorname{Tr}\{Q_{R_nB_n}\tau_{R_nB_n}\} + \sup_{\psi_{RA}}\widetilde{D}_\alpha(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))\right]. \tag{4.21}$$

Taking the supremum of both sides of (4.21) over all strategies such that the Type II error is at most 2^{-nr}, we obtain

$$\frac{1}{n}\log(1-\alpha_{n,r}^{\mathrm{ad}}) \leq \frac{\alpha-1}{\alpha}\left[-r + \sup_{\psi_{RA}}\widetilde{D}_\alpha(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))\right], \tag{4.22}$$

which yields the second inequality in (4.20).

We now establish the achievability part

$$\limsup_{n\to+\infty} -\frac{1}{n}(1-\alpha_{n,r}^{\mathrm{ad}}) \leq \overline{\mathrm{sc}}(r) \leq \sup_{\alpha>1}\inf_{\psi_{RA}} \frac{\alpha-1}{\alpha}\Big[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma_B)\Big]. \tag{4.23}$$

The first inequality follows the same way as the first inequality in (4.20). Let R be an arbitrary system. According to Theorem IV.10 and Remark IV.11 in [36], for every state $\psi_{RA}\in\mathcal{S}(\mathcal{H}_R\otimes\mathcal{H}_A)$ and every $r'>0$, there exists a sequence of tests $Q_{R_nB_n}, n\geq 1$, such that

$$\limsup_{n\to+\infty}\frac{1}{n}\log\operatorname{Tr}\left\{Q_{R_nB_n}\mathcal{R}(\psi_{RA})^{\otimes n}\right\} \leq -r', \qquad \text{and} \tag{4.24}$$

$$\liminf_{n\to+\infty}\frac{1}{n}\log\operatorname{Tr}\left\{Q_{R_nB_n}\mathcal{N}(\psi_{RA})^{\otimes n}\right\} \geq -H_{r'}^*(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA})). \tag{4.25}$$

Thus,

$$\overline{\mathrm{sc}}(r) \leq \inf_{r'>r} H^*_{r'}(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA})). \tag{4.26}$$

From the definition (1.10) of the Hoeffding anti-divergence, it is clear that $r \mapsto H^*_r(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))$ is a monotone increasing convex function on $(0, +\infty)$. Moreover, Lemma IV.9 in [36] implies that $H^*_r(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))$ is finite for every $r > 0$. Thus, $r \mapsto H^*_r(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))$ is continuous on $(0, +\infty)$, and (4.26) yields

$$\overline{\mathrm{sc}}(r) \leq H^*_r(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA})). \tag{4.27}$$

Since this is true for every ψ_{RA}, we finally get

$$\overline{\mathrm{sc}}(r) \leq \inf_{\psi_{RA}} \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}(\psi_{RA})\|\mathcal{R}(\psi_{RA}))\right]. \tag{4.28}$$

The last step is to show that the RHS of (4.20) and (4.28) are equal to each other. First, note that the RHS of (4.20) can be written as

$$\sup_{\alpha>1}\inf_{\rho_{A'}} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}\otimes\sigma_B\Big)\right],$$

where the infimum is taken over $\mathcal{S}(\mathcal{H}_{A'})$ with $A' \cong A$, due to Lemma 7. Moreover, the RHS of (4.28) can be trivially upper bounded by

$$\inf_{\rho_{A'}}\sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}\otimes\sigma_B\Big)\right]$$

(see the proof of Lemma 7 in Appendix A). Next, define

$$F(\alpha, \rho_{A'}) \equiv (\alpha-1)\widetilde{D}_\alpha\Big(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}\otimes\sigma_B\Big) \tag{4.29}$$

for $\alpha > 1$ and $\rho_{A'} \in \mathcal{S}(\mathcal{H}_{A'})$. Introducing the new variable $u \equiv \frac{\alpha-1}{\alpha}$, we have to show that

$$\sup_{0<u<1}\inf_{\rho_{A'}} f(u,\rho_{A'}) = \inf_{\rho_{A'}}\sup_{0<u<1} f(u,\rho_{A'}), \tag{4.30}$$

where

$$f(u,\rho_{A'}) \equiv ur - \widetilde{F}(u,\rho_{A'}), \qquad \widetilde{F}(u,\rho_{A'}) \equiv (1-u)F\left(\frac{1}{1-u}, \rho_{A'}\right). \tag{4.31}$$

By Lemmas 3 and 4 in [18], $\rho_{A'} \mapsto F(u,\rho_{A'})$ is concave, and hence $\rho_{A'} \mapsto f(u,\rho_{A'})$ is convex, for any fixed $u \in (0,1)$. On the other hand, $u \mapsto F(u,\rho_{A'})$ is convex by Corollary 3.11 in [36], and Lemma 13 below yields that $u \mapsto \widetilde{F}(u,\rho_{A'})$ is also convex, which in turn implies the concavity of $u \mapsto f(u,\rho_{A'})$ for any fixed $\rho_{A'}$. By assumption, $D(\mathcal{N}\|\mathcal{R}) < +\infty$, and taking into account Lemma 12, it is easy to see that $\rho_{A'} \mapsto f(u,\rho_{A'})$ is continuous for any $u \in (0,1)$. Since the state space of $\mathcal{H}_A$ is compact, the Kneser-Fan minimax theorem [30, 16] yields (4.30). ■

Lemma 13 *Let $f : (0,1) \to \mathbb{R}$ be a convex function. Then*

$$\tilde{f} : u \mapsto (1-u)f\left(\frac{1}{1-u}\right)$$

is convex as well.

Proof. Since f is convex, it can be written as the supremum of affine functions, i.e., $f(x) = \sup_i\{a_i x + b_i\}$ for some $a_i, b_i \in \mathbb{R}$, and thus

$$\tilde{f}(u) = (1-u)\sup_i\left\{a_i\frac{1}{1-u} + b_i\right\} = \sup_i\left\{a_i + b_i(1-u)\right\}.$$

As a supremum of affine functions, $\tilde{f}$ is convex. ∎

Remark 14 *It is not too difficult to see that Theorem 1 can be reformulated the following way:*

[Direct part] For every $r < D(\mathcal{N}\|\mathcal{R})$, there exists a sequence of adaptive strategies such that the type I error goes to 0 *and the type II error decays exponentially with a rate at least r.*

[Strong converse part] For every $r > D(\mathcal{N}\|\mathcal{R})$, and any sequence of adaptive strategies such that the type II error decays exponentially with a rate at least r, the type I error goes to 1*.*

As we have seen, the direct part is an immediate consequence of Stein's lemma for state discrimination. For the proof of the strong converse part and for the proof of the optimality part of Theorem 2, we followed the same argument of first using the monotonicity of the Rényi relative entropies under measurements and then applying (4.15)*. In fact, one could first prove the optimality part of Theorem 2 and obtain the optimality part of Theorem 1 from it in the limit $r \searrow D(\mathcal{N}\|\mathcal{R})$. Indeed, Lemma 10 implies that for any $r > D(\mathcal{N}\|\mathcal{R})$, there exists an $\alpha > 1$ such that $r > \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R})$, and hence the RHS of* (4.20) *is strictly positive, from which the strong converse part of the channel Stein's lemma is immediate.*

4.3 Related results

Hayashi and Tomamichel recently published their independently obtained results about a hypothesis testing scenario somewhat similar to ours [24]. Both our paper and theirs generalise the task of binary state discrimination but in different and not directly comparable directions. They consider the problem of composite hypothesis testing, where the null hypothesis is the presence of a fixed bipartite state and the alternative hypothesis is the presence of a product state that shares one marginal with the null hypothesis. Considered as a channel discrimination problem, the null hypothesis is that the i.i.d. channel $\mathcal{N}_1^{\otimes n}$ is applied to the A systems of the input, where the input state is restricted to be a fixed tensor-power state of the form $\psi_{RA}^{\otimes n}$. The alternative hypothesis is that a general "worst-case" replacer channel is applied to the A systems, which leads to an output $\psi_R^{\otimes n} \otimes \sigma_{B^n}$, where σ_{B^n} could be any state on the B systems. Not only do they allow for this more general alternative hypothesis, but they also determine both the direct and the strong converse exponents in their scenario. On the other hand, one has to note that when the above result is considered as a channel discrimination problem, allowing only the tensor powers of one fixed state as an input is extremely restrictive. In contrast, our results do allow for more general input states and for the adaptive strategies that distinguish the problem of quantum channel discrimination from binary state discrimination.

While the results of the two papers go in quite different directions, there is also a natural combination of them, which enables us to obtain a Stein's lemma with strong converse for the following channel discrimination problem with composite alternative hypothesis. For every $n \in \mathbb{N}$, the null hypothesis is that the channel is $\mathcal{N}^{\otimes n}$, where $\mathcal{N}$ is a fixed channel, and the alternative hypothesis is that the channel belongs to the set $\mathcal{R}^{(n)} \equiv \{\mathcal{R}_{\sigma_n} : \sigma_n \in \Sigma_n\}$, where

$$\{\sigma^{\otimes n} : \sigma \in \mathcal{S}(\mathcal{H}_B)\} \subseteq \Sigma_n \subseteq \mathcal{S}(\mathcal{H}_B^{\otimes n}). \tag{4.32}$$

For Stein's lemma, one is interested in the asymptotics of the optimal Type II error

$$\beta_\varepsilon^x \equiv \beta_\varepsilon^x(\mathcal{N}^{\otimes n}\|\mathcal{R}^{(n)}) \equiv \inf\left\{\sup_{\sigma_n\in\Sigma_n}\beta_n(S_n|\sigma_n)\,:\,\alpha_n(S_n)\le\varepsilon\right\}, \tag{4.33}$$

where the infimum is over all strategies in the class x with Type I error below ε. Combining Theorem 11 in [24] and Theorem 1 in this paper, we obtain the following:

Theorem 15 *In the above setting, for every $\varepsilon \in (0,1)$,*

$$\lim_{n\to+\infty} -\frac{1}{n}\log\beta_\varepsilon^{\mathrm{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{(n)}) = \lim_{n\to+\infty} -\frac{1}{n}\log\beta_\varepsilon^{\mathrm{pr}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{(n)}) = \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)} D(\mathcal{N}\|\mathcal{R}_{\sigma_B}) \tag{4.34}$$

$$= I(\mathcal{N}), \tag{4.35}$$

where $D(\mathcal{N}\|\mathcal{R}_{\sigma_B})$ is the channel relative entropy (1.14)*, and $I(\mathcal{N})$ is the channel mutual information* (2.13)*.*

Proof. Just as in (4.45)–(4.48) (see below), we have

$$\beta_\varepsilon^{\mathrm{ad}} \le \beta_\varepsilon^{\mathrm{pr}} = \inf_{\psi_{RA}}\inf_{Q_n}\left\{\sup_{\sigma_n\in\Sigma_n}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\}\,:\,\mathrm{Tr}\left\{(I_n-Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\}\le\varepsilon\right\} \tag{4.36}$$

$$\le \inf_{Q_n}\left\{\sup_{\sigma_n\in\Sigma_n}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\}\,:\,\mathrm{Tr}\left\{(I_n-Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\}\le\varepsilon\right\} \tag{4.37}$$

$$\equiv \beta_\varepsilon(\psi_{RA}), \tag{4.38}$$

where Q_n runs over all $Q_n \in \mathcal{B}(\mathcal{H}_{RA}^{\otimes n})_+$ such that $Q_n \le I_n$, and the second inequality holds for every ψ_{RA}. By [24, Theorem 11], for any ψ_{RA} and any rate r, there exists a sequence of binary measurements $(Q_n, I_n - Q_n)$, for which

$$\sup_{\sigma_n\in\mathcal{S}(\mathcal{H}_B^{\otimes n})}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\}\le 2^{-nr}, \tag{4.39}$$

$$\limsup_{n\to+\infty}\frac{1}{n}\log\mathrm{Tr}\left\{(I_n-Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\} = -\sup_{\alpha\in(0,1)}\frac{\alpha-1}{\alpha}\left[r - I_\alpha(R;B)_{\mathcal{N}(\psi)}\right]. \tag{4.40}$$

By Lemma 10, the RHS of (4.40) is strictly negative for every $r < I(R;B)_{\mathcal{N}(\psi)}$, and hence

$$\limsup_{n\to+\infty}\frac{1}{n}\log\beta_\varepsilon(\psi_{RA}) \le -I(R;B)_{\mathcal{N}(\psi)}. \tag{4.41}$$

When combined with (4.38), this yields

$$\limsup_{n\to+\infty}\frac{1}{n}\log\beta_\varepsilon^{\mathrm{ad}} \le \limsup_{n\to+\infty}\frac{1}{n}\log\beta_\varepsilon^{\mathrm{pr}} \le \inf_{\psi_{RA}} -I(R;B)_{\mathcal{N}(\psi)} = -I(\mathcal{N}). \tag{4.42}$$

Suppose now that

$$\liminf_{n\to+\infty}\frac{1}{n}\log\beta_\varepsilon^{\mathrm{ad}} < -r \tag{4.43}$$

for some $r \in \mathbb{R}$. For every $\sigma \in \mathcal{S}(\mathcal{H}_B)$, $\beta_\varepsilon^{\mathrm{ad}} = \beta_\varepsilon^{\mathrm{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}^{(n)}) \geq \beta_\varepsilon^{\mathrm{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}_\sigma^{\otimes n})$. Hence the assumption yields that $\liminf_{n\to+\infty}\frac{1}{n}\log\beta_\varepsilon^{\mathrm{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}_\sigma^{\otimes n}) < -r$, and by Theorem 1 this is only possible if $r \leq D(\mathcal{N}\|\mathcal{R}_\sigma)$. Since this is true for every $\sigma \in \mathcal{S}(\mathcal{H}_B)$, we finally get that $r \leq \inf_{\sigma\in\mathcal{S}(\mathcal{H}_B)} D(\mathcal{N}\|\mathcal{R}_\sigma) = I(\mathcal{N})$, completing the proof of (4.34).

The equality of (4.35) and (4.34) is due to Lemma 11. ■

It is now natural to ask whether the exact strong converse exponent can be determined for this problem, analogously to Theorem 2. Below we give lower and upper bounds for the strong converse exponent. We conjecture that these bounds in fact coincide, and thus give the exact strong converse exponent; indeed, this could be proved if one could justify interchanging the order of infima and suprema in (4.52) and (4.56) below.

The problem can be formulated as follows. For any adaptive discrimination strategy S_n, and any $\sigma_n \in \mathcal{S}(\mathcal{H}_B^{\otimes n})$, let $\alpha_n(S_n)$ and $\beta_n(S_n|\sigma_n)$ be the Type I and Type II error probabilities for discriminating between $\mathcal{N}^{\otimes n}$ and $\mathcal{R}_{\sigma_n}$, as given in (1.11). We consider the optimal Type I error

$$\alpha_{n,r}^x \equiv \alpha_{2^{-nr}}^x(\mathcal{N}^{\otimes n}\|\mathcal{R}^{(n)}) \equiv \inf\left\{\alpha_n(S_n):\ \sup_{\sigma_n\in\Sigma_n}\beta_n(S_n|\sigma_n) \leq 2^{-nr}\right\}, \tag{4.44}$$

where x denotes the set of allowed discrimination strategies and the optimisation is over all strategies in the class x. As before, we take $x = \mathrm{pr}$ and $x = \mathrm{ad}$, for product and adaptive strategies, respectively. We have

$$\begin{aligned}
\alpha_{n,r}^{\mathrm{pr}} &= \inf_{\psi_{RA}}\inf_{Q_n}\left\{\mathrm{Tr}\left\{(I_n - Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\}:\ \sup_{\sigma_n\in\Sigma_n}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\} \leq 2^{-nr}\right\} &(4.45)\\
&\leq \inf_{\psi_{RA}}\inf_{Q_n}\left\{\mathrm{Tr}\left\{(I_n - Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\}:\ \sup_{\sigma_n\in\mathcal{S}(\mathcal{H}^{\otimes n})}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\} \leq 2^{-nr}\right\} &(4.46)\\
&\leq \inf_{Q_n}\left\{\mathrm{Tr}\left\{(I_n - Q_n)\left(\mathcal{N}_{A\to B}(\psi_{RA})\right)^{\otimes n}\right\}:\ \sup_{\sigma_n\in\mathcal{S}(\mathcal{H}^{\otimes n})}\mathrm{Tr}\left\{Q_n(\psi_R^{\otimes n}\otimes\sigma_n)\right\} \leq 2^{-nr}\right\} &(4.47)\\
&\equiv \alpha_{n,r}(\psi_{RA}), &(4.48)
\end{aligned}$$

where Q_n runs over all $Q_n \in \mathcal{B}(\mathcal{H}_{RA}^{\otimes n})_+$ such that $Q_n \leq I_n$, and the second inequality holds for every ψ_{RA}. Applying now the results of [24], we get that

$$\begin{aligned}
\limsup_{n\to+\infty} -\frac{1}{n}\log(1-\alpha_{n,r}^{\mathrm{ad}}) &\leq \limsup_{n\to+\infty} -\frac{1}{n}\log(1-\alpha_{n,r}^{\mathrm{pr}}) &(4.49)\\
&\leq \inf_{\psi_{RA}}\limsup_{n\to+\infty} -\frac{1}{n}\log(1-\alpha_{n,r}(\psi_{RA})) &(4.50)\\
&= \inf_{\psi_{RA}}\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left[r - \inf_{\sigma\in\mathcal{S}(\mathcal{H}_B)}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma)\right] &(4.51)\\
&= \inf_{\psi_{RA}}\sup_{\alpha>1}\sup_{\sigma\in\mathcal{S}(\mathcal{H}_B)}\frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma)\right], &(4.52)
\end{aligned}$$

where (4.51) is due to [24, Theorem 13]. On the other hand,

$$\alpha_{n,r}^{\mathrm{ad}} \geq \inf\left\{\alpha_n(S_n):\ \beta_n(S_n|\sigma^{\otimes n}) \leq 2^{-nr}\right\} = \alpha_{2^{-nr}}^{\mathrm{ad}}(\mathcal{N}^{\otimes n}\|\mathcal{R}_\sigma^{\otimes n}), \qquad \sigma\in\mathcal{S}(\mathcal{H}_B), \tag{4.53}$$

and hence

$$\liminf_{n\to+\infty} -\frac{1}{n}\log(1-\alpha^{\mathrm{ad}}_{n,r}) \geq \sup_{\sigma\in\mathcal{S}(\mathcal{H}_B)} \liminf_{n\to+\infty} -\frac{1}{n}\log(1-\alpha^{\mathrm{ad}}_{2^{-nr}}(\mathcal{N}^{\otimes n}\|\mathcal{R}_\sigma^{\otimes n}) \tag{4.54}$$

$$= \sup_{\sigma\in\mathcal{S}(\mathcal{H}_B)} \sup_{\alpha>1} \inf_{\psi_{RA}} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma)\right] \tag{4.55}$$

$$= \sup_{\sigma\in\mathcal{S}(\mathcal{H}_B)} \inf_{\psi_{RA}} \sup_{\alpha>1} \frac{\alpha-1}{\alpha}\left[r - \widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{RA})\|\psi_R\otimes\sigma)\right] \tag{4.56}$$

where the two equalities are due to Theorem 2. If one had joint concavity in the variables $\alpha > 1$ and σ, then one could interchange the optima and show that (4.52) and (4.56) are equal to each other, obtaining strong converse exponents for this channel discrimination problem. However, it remains unclear to us if the joint concavity holds or more generally if the exchange is possible.

Remark 16 *Theorem 15 gives an operational interpretation to the channel mutual information $I(\mathcal{N})$, and its geometric representation given in Lemma 11. If* (4.52) *and* (4.56) *could be shown to be equal, that would give an analogous operational interpretation to the channel Rényi mutual informations $\widetilde{I}_\alpha(\mathcal{N})$ and their geometric representation in Lemma 11, for every $\alpha > 1$.*

5 Strong converse for quantum-feedback-assisted classical communication

In this section, we give a detailed proof of Theorem 4, which identifies a strong converse exponent for quantum-feedback-assisted communication and states that a strong converse theorem holds for the quantum-feedback-assisted classical capacity of a quantum channel.

In an n-round feedback-assisted protocol $\mathcal{P}_n$, Alice and Bob initially share an entangled state on Alice's system X_0 and Bob's system B'_0. If Alice wants to transmit message $m \in \{1,\ldots,M_n\}$, where $M_n \in \mathbb{N}$ is the number of messages, she applies a quantum channel $\mathcal{E}^1_m$ with output system A'_1A_1 to her part of the entangled state; the result is a state $\rho^m_{A'_1A_1B'_0} = \tau^m_{A'_1A_1B'_0}$ on systems $A'_1A_1B'_0$, where A_1 is sent over the channel to Bob, with an output in system B_1, while A'_1 is kept at Alice's side for possible later use. After this, Bob may apply a quantum channel $\mathcal{D}^1$ on $B_1B'_0$ with an output on $X_1B'_1$, of which system X_1 contains the feedback information, that is sent back to Alice, while B'_1 is kept at Bob's side for possible later use. This procedure is repeated n times, as depicted in Figure 3 (with $n = 3$). At each round, an encoding channel $\mathcal{E}^i_m : A'_{i-1}X_{i-1} \to A'_iA_i$ corresponding to the same fixed message m is applied, but the $\mathcal{E}^i_m$ may be different channels for different i's. At the end of the protocol, the $\mathcal{D}^n$ channel is a POVM on $B_nB'_{n-1}$ with outcomes in $\{1,\ldots,M_n\}$, specified by the POVM elements $\{D^m_{B_nB'_{n-1}}\}_{m=1}^{M_n}$. In the last round, A'_n can be taken one-dimensional, since whatever information may be stored there does not influence the outcome of the final measurement on Bob's systems. We assume for simplicity that the feedback channel is noiseless, although it is not necessary to do so; indeed, we are looking for an upper bound on the success probability, and noisy feedback can only decrease the success probability.

For every stage of the communication process, let ρ^m with the appropriate labels denote the state obtained from $\rho^m_{A'_1A_1B'_0}$ by the action of all channels $\mathcal{E}^i_m, \mathcal{N}^i, \mathcal{D}^i$ up to that stage; e.g., $\rho^m_{A'_1B_1B'_0} = \mathcal{N}_{A_1\to B_1}(\rho^m_{A'_1A_1B'_0})$, etc. Similarly, let τ^m with the appropriate labels denote the state obtained from $\tau^m_{A'_1A_1B'_0}$ up to a certain stage of the process, where all uses of $\mathcal{N}$ are replaced by $\mathcal{R}_\sigma$ for some fixed

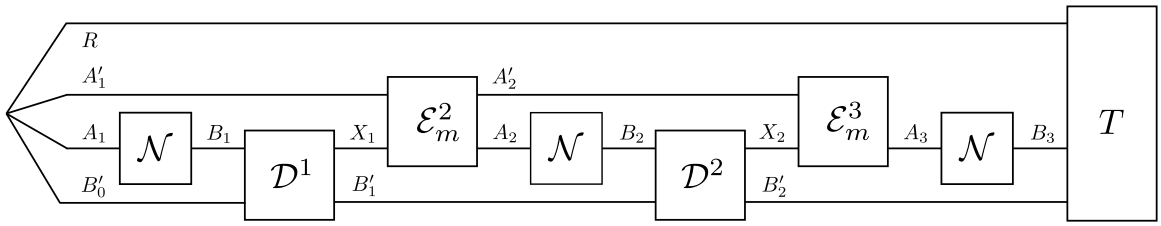

Figure 3: A general quantum feedback-assisted communication protocol for the channel $\mathcal{N}$.

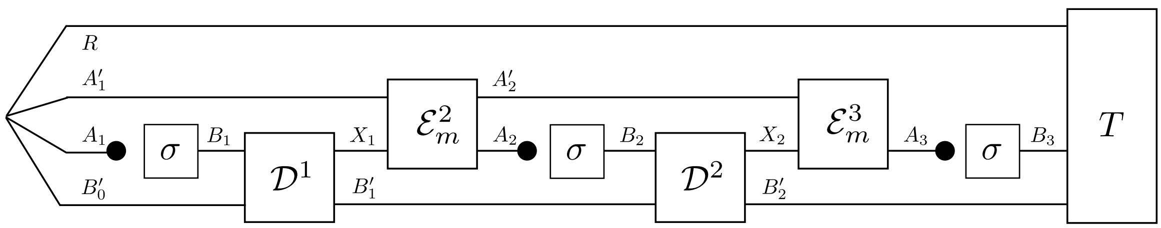

Figure 4: A general quantum feedback-assisted communication protocol, with a replacer channel instead of the channel $\mathcal{N}$.

state σ; see Figure 4. Moreover, we introduce an auxiliary system R with an orthonormal basis $\{|m\rangle_R\}_{m=1}^{M_n}$, and define

$$\rho_{RL} \equiv \frac{1}{M_n}\sum_{m=1}^{M_n} |m\rangle\langle m|_R \otimes \rho_L^m, \qquad \tau_{RL} \equiv \frac{1}{M_n}\sum_{m=1}^{M_n} |m\rangle\langle m|_R \otimes \tau_L^m, \tag{5.1}$$

where L is any set of indices that can occur at a certain stage of the communication process, and

$$T_{RB_nB'_{n-1}} \equiv \sum_{m=1}^{M_n} |m\rangle\langle m|_R \otimes D^m_{B_nB'_{n-1}}, \tag{5.2}$$

such that $0 \leq T_{RB_nB'_{n-1}} \leq I_{RB_nB'_{n-1}}$. For every $1 \leq i \leq n$, we define $\mathcal{E}^i : RA'_{i-1}X_{i-1} \to RA'_iA_i$ as

$$\mathcal{E}^i\left(\sum_{m=1}^{M_n} |m\rangle\langle m|_R \otimes \psi_{A'_{i-1}X_{i-1}}\right) \equiv \sum_{m=1}^{M_n} |m\rangle\langle m|_R \otimes \mathcal{E}^i_m(\psi_{A'_{i-1}X_{i-1}}). \tag{5.3}$$

If the outcome of the final measurement $\mathcal{D}^n$ is m' then Bob concludes that the message m' was sent. The success probability $p_{\rm succ}(\mathcal{P}_n)$ of the protocol is given by

$$p_{\rm succ}(\mathcal{P}_n) \equiv \frac{1}{M_n}\sum_{m=1}^{M_n} \operatorname{Tr}\left\{D^m_{B_nB'_{n-1}}\rho^m_{B_nB'_{n-1}}\right\} = \operatorname{Tr}\left\{T_{RB_nB'_{n-1}}\rho_{RB_nB'_{n-1}}\right\}. \tag{5.4}$$

Note that for every round k, $\tau^m_{B_kB'_{k-1}}$ is independent of m, and we have

$$\tau_{RB_kB'_{k-1}} = \tau_R \otimes \sigma_{B_k} \otimes \tau_{B'_{k-1}}, \qquad \text{where} \qquad \tau_R = \frac{1}{M_n} I_R. \tag{5.5}$$

This is because all information about the identity of the message is kept at Alice's side all through the protocol, as one can easily see in Figure 4. Hence,

$$\operatorname{Tr}\left\{T_{RB_nB'_{n-1}}\tau_{RB_nB'_{n-1}}\right\} = \frac{1}{M_n}\sum_{m=1}^{M_n}\operatorname{Tr}\left\{D^m_{B_nB'_{n-1}}(\sigma_{B_n}\otimes\tau_{B'_{n-1}})\right\} = \frac{1}{M_n}. \tag{5.6}$$

Now we can apply Nagaoka's method [38], and use the monotonicity of $\widetilde{D}_\alpha$ to get

$$\widetilde{D}_\alpha(\rho_{RB_nB'_{n-1}}\|\tau_{RB_nB'_{n-1}}) \geq \frac{1}{\alpha-1}\log\left[p_{\text{succ}}(\mathcal{P}_n)^\alpha\left(\frac{1}{M_n}\right)^{1-\alpha}\right] \tag{5.7}$$

$$= \frac{\alpha}{\alpha-1}\log p_{\text{succ}}(\mathcal{P}_n) + \log M_n. \tag{5.8}$$

We will use the same iterative method as in Section 4.1 to complete the proof of Theorem 4. For every $k > 1$,

$$\widetilde{D}_\alpha(\rho_{RA'_kB_kB'_{k-1}}\|\tau_{RA'_kB_kB'_{k-1}}) \tag{5.9}$$

$$= \widetilde{D}_\alpha(\mathcal{N}_{A_k\to B_k}(\rho_{RA'_kA_kB'_{k-1}})\|\tau_{RA'_kB'_{k-1}}\otimes\sigma_{B_k}) \tag{5.10}$$

$$= \frac{\alpha}{\alpha-1}\log\left\|\left(\Theta_{\sigma_{B_k}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_k\to B_k}\right)\left(\tau^{\frac{1-\alpha}{2\alpha}}_{RA'_kB'_{k-1}}\rho_{RA'_kA_kB'_{k-1}}\tau^{\frac{1-\alpha}{2\alpha}}_{RA'_kB'_{k-1}}\right)\right\|_\alpha \tag{5.11}$$

$$\leq \frac{\alpha}{\alpha-1}\log\sup_{X_{RA'_kA_kB'_{k-1}}\geq 0}\frac{\left\|\left(\Theta_{\sigma_{B_k}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_k\to B_k}\right)X_{RA'_kA_kB'_{k-1}}\right\|_\alpha}{\left\|X_{RA'_kB'_{k-1}}\right\|_\alpha}\left\|\tau^{\frac{1-\alpha}{2\alpha}}_{RA'_kB'_{k-1}}\rho_{RA'_kB'_{k-1}}\tau^{\frac{1-\alpha}{2\alpha}}_{RA'_kB'_{k-1}}\right\|_\alpha \tag{5.12}$$

$$= \frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma_{B_k}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_k\to B_k}\right\|_{\text{CB},1\to\alpha} + \widetilde{D}_\alpha(\rho_{RA'_kB'_{k-1}}\|\tau_{RA'_kB'_{k-1}}). \tag{5.13}$$

Now, if $k = 1$ then $\rho_{RA'_kB'_{k-1}} = \tau_{RA'_kB'_{k-1}}$ by definition, and the last term above is zero. Otherwise we can upper bound the last term above as

$$\widetilde{D}_\alpha(\rho_{RA'_kB'_{k-1}}\|\tau_{RA'_kB'_{k-1}}) \leq \widetilde{D}_\alpha(\rho_{RA'_{k-1}B_{k-1}B'_{k-2}}\|\tau_{RA'_{k-1}B_{k-1}B'_{k-2}}), \tag{5.14}$$

where the inequality is due to the monotonicity of $\widetilde{D}_\alpha$ under $\operatorname{Tr}_{A_k}\circ\mathcal{E}^k\circ\mathcal{D}^{k-1}$.

Using the above steps iteratively, we finally get

$$\frac{\alpha}{\alpha-1}\log p_{\text{succ}}(\mathcal{P}_n) + \log M_n \leq \widetilde{D}_\alpha(\rho_{RB_nB'_{n-1}}\|\tau_{RB_nB'_{n-1}}) \tag{5.15}$$

$$\leq n\frac{\alpha}{\alpha-1}\log\left\|\Theta_{\sigma^{\frac{1-\alpha}{2\alpha}}}\circ\mathcal{N}\right\|_{\text{CB},1\to\alpha} \tag{5.16}$$

$$= n\sup_{\psi_{\hat{A}A}}\widetilde{D}_\alpha(\mathcal{N}_{A\to B}(\psi_{\hat{A}A})\|\psi_{\hat{A}}\otimes\sigma_B), \tag{5.17}$$

where the last identity is due to (3.23), and the supremum is over all pure states on $\hat{A}A$, where $\hat{A}$ is a copy of A. Since this is true for every $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$, we get

$$\frac{\alpha}{\alpha-1}\frac{1}{n}\log p_{\text{succ}}(\mathcal{P}_n) + \frac{1}{n}\log M_n \leq \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)}\sup_{\psi_{\hat{A}A}}\widetilde{D}_\alpha\left(\mathcal{N}_{A\to B}(\psi_{\hat{A}A})\|\psi_{\hat{A}}\otimes\sigma_B\right) = \widetilde{I}_\alpha(\mathcal{N}). \tag{5.18}$$

Hence, if $\mathcal{P}_n$ is a feedback-assisted coding scheme such that $\frac{1}{n}\log M_n \geq R$ then

$$\frac{1}{n}\log p_{\text{succ}}(\mathcal{P}_n) \leq -\sup_{\alpha>1}\frac{\alpha-1}{\alpha}\left(R-\widetilde{I}_\alpha(\mathcal{N})\right), \tag{5.19}$$

where we used that (5.18) holds for every $\alpha > 1$. This proves (2.14) of Theorem 4.

By (5.19), the success probability goes to zero exponentially fast for any rate $R > R_{\text{min}} \equiv \inf_{\alpha>1}\widetilde{I}_\alpha(\mathcal{N})$. By the monotonicity of the Rényi relative entropies in α, $\inf_{\alpha>1}\widetilde{I}_\alpha(\mathcal{N}) = \lim_{\alpha\searrow 1}\widetilde{I}_\alpha(\mathcal{N})$, and the latter is equal to $I_1(\mathcal{N}) = I(\mathcal{N})$, due to Lemma 10. This proves the last assertion of Theorem 4.

6 Conclusion

This paper establishes a quantum Stein's lemma and identifies the strong converse exponent when discriminating an arbitrary channel from the replacer channel. The conclusion is that a tensor-power, non-adaptive strategy is optimal in this regime. This result has implications in the physical setting of quantum illumination, as described in Section 2.3. We have also proven that a strong converse theorem holds in the setting of quantum-feedback-assisted communication, strengthening a weak converse result due to Bowen [9]. This strong converse theorem also strengthens the main result of [18], in which a bound on the strong converse exponent was established for the entanglement-assisted communication setting. We show here that this same bound holds in the more general quantum-feedback-assisted communication setting. We also briefly discussed how to combine our results in adaptive channel discrimination with those of Hayashi and Tomamichel from [24] to obtain a quantum Stein's lemma in a more general setting than that considered in either paper. It remains an open question to determine the strong converse exponent for this more general setting.

There are several other open questions to consider going forward from here. First, is the strong converse exponent bound in (2.14) optimal? That is, does there exist an entanglement-assisted communication protocol that achieves this bound? Encouraging for us here is that a full solution is known for the classical version of this problem [3, 15, 12]. Next, can we say anything about the direct domain for either the adaptive channel discrimination setting or the quantum-feedback-assisted communication setting? Any results obtained in the latter setting would be a counterpart to the error exponents found in [10, 27, 21] for classical communication. Finally, would the conclusions of this paper extend to the setting of symmetric hypothesis testing? That is, would it be possible to show that non-adaptive strategies suffice here?

Acknowledgements—We thank Nilanjanna Datta, Manish K. Gupta, Bhaskar Roy Bardhan, and Marco Tomamichel for insightful discussions on these topics. We thank David Ding for feedback on the manuscript. TC and MMW acknowledge support from the Department of Physics and Astronomy at LSU, from the NSF under Award No. CCF-1350397, and from the DARPA Quiness Program through US Army Research Office Award No. W31P4Q-12-1-0019. MM acknowledges support from the European Research Council Advanced Grant "IRQUAT", the Spanish MINECO (Project No. FIS2013-40627-P), the Generalitat de Catalunya CIRIT (Project No. 2014 SGR 966), the Hungarian Research Grant OTKA-NKFI K104206, and the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763.

A Channel divergences

Proof of Lemma 6. We only prove the assertion for $\widetilde{D}_\alpha$, as the proof for D_α goes the same way.

For every system R, $|\Gamma_{A'A}\rangle$ defines an isomorphism between $\mathcal{H}_R \otimes \mathcal{H}_A$ and $\mathcal{B}(\mathcal{H}_{A'}, \mathcal{H}_R)$, under which $X \in \mathcal{B}(\mathcal{H}_{A'}, \mathcal{H}_R)$ corresponds to $(X \otimes I_A)|\Gamma_{A'A}\rangle = \sum_i (Xe_i) \otimes e_i$. Given a pure state ψ_{RA}, it can be written as $\psi_{RA} = |\psi_{RA}\rangle\langle\psi_{RA}|$, with $|\psi_{RA}\rangle = (X \otimes I_A)|\Gamma_{A'A}\rangle$, where $X \in \mathcal{B}(\mathcal{H}_{A'}, \mathcal{H}_R)$, and $\operatorname{Tr} X^*X = 1$. Thus, for any channel $\mathcal{N}_{A\to B}$,

$$\mathcal{N}_{A\to B}(\psi_{RA}) = \mathcal{N}\left((X \otimes I_A)\Gamma_{A'A}(X^* \otimes I_A)\right) = (X \otimes I_A)\mathcal{N}\left(\Gamma_{A'A}\right)(X^* \otimes I_A). \tag{A.1}$$

Let $V : \mathcal{H}_{A'} \to \mathcal{H}_R$ be a partial isometry such that $X = V|X|$. Then it is easy to see that

$$\begin{aligned}
&\widetilde{D}_\alpha\left(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})\right) && \text{(A.2)}\\
&= \widetilde{D}_\alpha\left((X \otimes I_A)\left(\mathcal{N}_1(\Gamma_{A'A})\right)(X^* \otimes I_A)\|(X \otimes I_A)\left(\mathcal{N}_2(\Gamma_{A'A})\right)(X^* \otimes I_A)\right) && \text{(A.3)}\\
&= \widetilde{D}_\alpha\left((V|X| \otimes I_A)\left(\mathcal{N}_1(\Gamma_{A'A})\right)(|X|V^* \otimes I_A)\|(V|X| \otimes I_A)\left(\mathcal{N}_2(\Gamma_{A'A})\right)(|X|V^* \otimes I_A)\right) && \text{(A.4)}\\
&= \widetilde{D}_\alpha\left((|X| \otimes I_A)\left(\mathcal{N}_1(\Gamma_{A'A})\right)(|X| \otimes I_A)\|(|X| \otimes I_A)\left(\mathcal{N}_2(\Gamma_{A'A})\right)(|X| \otimes I_A)\right) && \text{(A.5)}\\
&= \widetilde{D}_\alpha\left(\rho_{A'}^{1/2}\mathcal{N}_1(\Gamma_{A'A})\rho_{A'}^{1/2}\|\rho_{A'}^{1/2}\mathcal{N}_2(\Gamma_{A'A})\rho_{A'}^{1/2}\right) && \text{(A.6)}\\
&= \widetilde{D}_\alpha\left(\mathcal{N}_1\left(\rho_{A'}^{1/2}|\Gamma_{A'A}\rangle\langle\Gamma_{A'A}|\rho_{A'}^{1/2}\right)\Big\|\mathcal{N}_2\left(\rho_{A'}^{1/2}|\Gamma_{A'A}\rangle\langle\Gamma_{A'A}|\rho_{A'}^{1/2}\right)\right), && \text{(A.7)}
\end{aligned}$$

where $\rho_{A'} := |X|^2 = X^*X$. The equality of the last two expressions follow from the fact that the channels only act on the A system. This completes the proof.

Proof of Lemma 8. Let Θ denote the conjugation by $\sigma_B^{\frac{1-\alpha}{2\alpha}}$. With the notations in the proof of Lemma 6, every pure state $\psi_{A'A}$ can be written as $\psi_{A'A} = |\psi_{A'A}\rangle\langle\psi_{A'A}|$, $|\psi_{A'A}\rangle = (X \otimes I)|\Gamma_{A'A}\rangle$. Then $\psi_{A'} = XX^*$, and hence

$$\begin{aligned}
&\widetilde{D}_\alpha\left(\mathcal{N}(\psi_{A'A})\|\mathcal{R}_{\sigma_B}(\psi_{A'A})\right) = \widetilde{D}_\alpha\left(\mathcal{N}_{A\to B}(\psi_{A'A})\|\psi_{A'} \otimes \sigma_B\right) && \text{(A.8)}\\
&= \frac{1}{\alpha-1}\log\operatorname{Tr}\left\{\left[(\psi_{A'} \otimes \sigma_B)^{\frac{1-\alpha}{2\alpha}}\left(\mathcal{N}_{A\to B}(\psi_{A'A})\right)(\psi_{A'} \otimes \sigma_B)^{\frac{1-\alpha}{2\alpha}}\right]^\alpha\right\} && \text{(A.9)}\\
&= \frac{1}{\alpha-1}\log\operatorname{Tr}\left\{\left[\left((XX^*)^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)(X \otimes I)\left(\mathcal{N}(\Gamma_{A'A})\right)(X^* \otimes I)\left((XX^*)^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]^\alpha\right\}. && \text{(A.10)}
\end{aligned}$$

Let $X = V|X|$ for some unitary V; then $XX^* = V|X|^2V^*$, and $(XX^*)^{\frac{1-\alpha}{2\alpha}}X = V|X|^{\frac{1-\alpha}{\alpha}}V^*V|X| = V|X|^{\frac{1}{\alpha}}$. Thus

$$\begin{aligned}
&\widetilde{D}_\alpha\left(\mathcal{N}_{A\to B}(\psi_{A'A})\|\psi_{A'} \otimes \sigma_B\right)\\
&= \frac{1}{\alpha-1}\log\operatorname{Tr}\left\{\left[(V \otimes I_B)\left(|X|^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\left(\mathcal{N}(\Gamma_{A'A})\right)\left(|X|^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)(V^* \otimes I_B)\right]^\alpha\right\} && \text{(A.11)}\\
&= \frac{1}{\alpha-1}\log\operatorname{Tr}\left\{\left[\left(|X|^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\left(\mathcal{N}(\Gamma_{A'A})\right)\left(|X|^{\frac{1}{\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}}\right)\right]^\alpha\right\} && \text{(A.12)}\\
&= \frac{1}{\alpha-1}\log\operatorname{Tr}\left\{\left[\left(|X|^{\frac{1}{\alpha}} \otimes I_B\right)(\Theta \circ \mathcal{N})(\Gamma_{A'A})\left(|X|^{\frac{1}{\alpha}} \otimes I_B\right)\right]^\alpha\right\} && \text{(A.13)}\\
&= \frac{\alpha}{\alpha-1}\log\left\|\left(Y^{\frac{1}{2\alpha}} \otimes I_B\right)(\Theta \circ \mathcal{N})(\Gamma_{A'A})\left(Y^{\frac{1}{2\alpha}} \otimes I_B\right)\right\|_\alpha, && \text{(A.14)}
\end{aligned}$$

where we used the notation $Y \equiv X^*X \in \mathcal{B}(\mathcal{H}_{A'})$. Hence, optimizing (A.11) over all bipartite pure states $\psi_{A'A}$ is equivalent to optimizing (A.14) over all $Y \in \mathcal{B}(\mathcal{H}_{A'})_+$ such that $\mathrm{Tr}\{Y\} = 1$, i.e., all states Y on $\mathcal{H}_{A'}$. The latter yields

$$
\begin{aligned}
&\sup_{Y \in \mathcal{S}(\mathcal{H}_{A'})} \left\| \left(Y^{\frac{1}{2\alpha}} \otimes I_B\right) (\Theta \circ \mathcal{N})(\Gamma_{A'A}) \left(Y^{\frac{1}{2\alpha}} \otimes I_B\right) \right\|_\alpha \\
&= \sup_{Y \in \mathcal{B}(\mathcal{H}_{A'})_+ \setminus \{0\}} \frac{1}{(\mathrm{Tr}\{Y\})^{\frac{1}{\alpha}}} \left\| \left(Y^{\frac{1}{2\alpha}} \otimes I_B\right) (\Theta \circ \mathcal{N})(\Gamma_{A'A}) \left(Y^{\frac{1}{2\alpha}} \otimes I_B\right) \right\|_\alpha && \text{(A.15)} \\
&= \sup_U \sup_{Y \in \mathcal{B}(\mathcal{H}_{A'})_+ \setminus \{0\}} \frac{1}{(\mathrm{Tr}\{Y\})^{\frac{1}{\alpha}}} \left\| \left(UY^{\frac{1}{2\alpha}} \otimes I_B\right) (\Theta \circ \mathcal{N})(\Gamma_{A'A}) \left(Y^{\frac{1}{2\alpha}} U^* \otimes I_B\right) \right\|_\alpha && \text{(A.16)} \\
&= \sup_{Z \in \mathcal{B}(\mathcal{H}_{A'}) \setminus \{0\}} \frac{1}{(\mathrm{Tr}\{(Z^*Z)^\alpha\})^{\frac{1}{\alpha}}} \left\| (Z \otimes I_B) (\Theta \circ \mathcal{N})(\Gamma_{A'A}) (Z^* \otimes I_B) \right\|_\alpha && \text{(A.17)} \\
&= \sup_{|z\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_A \setminus \{0\}} \frac{\|(\Theta \circ \mathcal{N})|z\rangle\langle z|\|_\alpha}{\|\mathrm{Tr}_{A'}\{|z\rangle\langle z|\}\|_\alpha} = \|\Theta \circ \mathcal{N}\|_{\mathrm{CB},1\to\alpha}\,, && \text{(A.18)}
\end{aligned}
$$

where the first supremum in (A.16) is taken over all unitaries U on $\mathcal{H}_{A'}$, and the second equality in (A.18) is due to (3.20). This finishes the proof. ■

To prove Lemma 10, we will need the following minimax theorem from [35, Corollary A2]:

Lemma 17 *Let X be a compact topological space, Y be a subset of the real line, and let $f : X \times Y \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a function. Assume that*

(i) $f(., y)$ is lower semicontinuous for every $y \in Y$, and

(ii) $f(x, .)$ is monotonic increasing for every $x \in X$, or $f(x, .)$ is monotonic decreasing for every $x \in X$.

Then

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \inf_{x \in X} f(x, y). \tag{A.19}$$

It is easy to see that for any fixed ρ, σ, $\varepsilon \mapsto D_\alpha(\rho\|\sigma + \varepsilon I)$ and $\varepsilon \mapsto \widetilde{D}_\alpha(\rho\|\sigma + \varepsilon I)$ are monotone decreasing on $(0, +\infty)$, and

$$D_\alpha(\rho\|\sigma) = \sup_{\varepsilon > 0} D_\alpha(\rho\|\sigma + \varepsilon I), \qquad \widetilde{D}_\alpha(\rho\|\sigma) = \sup_{\varepsilon > 0} \widetilde{D}_\alpha(\rho\|\sigma + \varepsilon I); \tag{A.20}$$

for these latter relations see, e.g. [35] and [37]. Since for any fixed $\varepsilon > 0$, $(\rho, \sigma) \mapsto D_\alpha(\rho\|\sigma + \varepsilon I)$ and $(\rho, \sigma) \mapsto \widetilde{D}_\alpha(\rho\|\sigma + \varepsilon I)$ are continuous, we get that

$$(\rho, \sigma) \mapsto D_\alpha(\rho\|\sigma) \quad \text{and} \quad (\rho, \sigma) \mapsto \widetilde{D}_\alpha(\rho\|\sigma) \quad \text{are lower semicontinuous.} \tag{A.21}$$

Now we are ready to prove Lemma 10.

Proof of Lemma 10. The assertions about monotonicity are obvious from the definitions and (3.4).

(i) We only prove the assertion for $\widetilde{D}_\alpha$, as the proof for D_α goes exactly the same way. Let $\mathcal{N}_1, \mathcal{N}_2 : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be channels. By the monotonicity (3.4), we have

$$\lim_{\alpha\nearrow 1} \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = \sup_{\alpha\in(0,1)} \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) \tag{A.22}$$

$$= \sup_{\alpha\in(0,1)} \sup_{\psi_{RA}} \widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.23}$$

$$= \sup_{\psi_{RA}} \sup_{\alpha\in(0,1)} \widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.24}$$

$$= \sup_{\psi_{RA}} D(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.25}$$

$$= D(\mathcal{N}_1\|\mathcal{N}_2). \tag{A.26}$$

Note that for any $\alpha > 1$, $\widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = +\infty \iff D(\mathcal{N}_1\|\mathcal{N}_2) = +\infty$, and hence for the rest we assume that all these quantities are finite, since otherwise $\lim_{\alpha\searrow 1} \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = D(\mathcal{N}_1\|\mathcal{N}_2)$ is trivial. Let $\mathcal{S}$ denote the set of pure states on $\mathcal{H}_{A'A}$, where A' is a copy of A; then $\mathcal{S}$ is a compact set, and $\psi \mapsto \widetilde{D}_\alpha(\mathcal{N}_1(\psi)\|\mathcal{N}_2(\psi))$ is lower semicontinuous on $\mathcal{S}$ by (A.21), while for a fixed $\psi \in \mathcal{S}$, $\alpha \mapsto \widetilde{D}_\alpha(\mathcal{N}_1(\psi)\|\mathcal{N}_2(\psi))$ is monotone increasing (3.4). Using now Lemma 17, we get

$$\lim_{\alpha\searrow 1} \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) = \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}_1\|\mathcal{N}_2) \tag{A.27}$$

$$= \inf_{\alpha>1} \sup_{\psi\in\mathcal{S}} \widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.28}$$

$$= \sup_{\psi\in\mathcal{S}} \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.29}$$

$$= \sup_{\psi\in\mathcal{S}} D(\mathcal{N}_1(\psi_{RA})\|\mathcal{N}_2(\psi_{RA})) \tag{A.30}$$

$$= D(\mathcal{N}_1\|\mathcal{N}_2). \tag{A.31}$$

(ii) We only prove the assertion for $I_\alpha(R;B)_\rho$, as the proof for $\widetilde{I}_\alpha(R;B)_\rho$ goes exactly the same way. First, we have

$$\lim_{\alpha\searrow 1} I_\alpha(R;B)_\rho = \inf_{\alpha>1} I_\alpha(R;B)_\rho = \inf_{\alpha>1} \inf_{\sigma_B} D_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B) \tag{A.32}$$

$$= \inf_{\sigma_B} \inf_{\alpha>1} D_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B) \tag{A.33}$$

$$= \inf_{\sigma_B} D(\rho_{RB}\|\rho_R\otimes\sigma_B) = I(R;B)_\rho. \tag{A.34}$$

Next, note that by (A.21), $D(\rho_{RB}\|\rho_R\otimes\sigma_B)$ is lower semicontinuous in σ_B on the compact set $\mathcal{S}(\mathcal{H}_B)$, and it is monotone increasing in α. Hence, by Lemma 17,

$$\lim_{\alpha\nearrow 1} I_\alpha(R;B)_\rho = \sup_{\alpha\in(0,1)} I_\alpha(R;B)_\rho = \sup_{\alpha\in(0,1)} \inf_{\sigma_B} D_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B) \tag{A.35}$$

$$= \inf_{\sigma_B} \sup_{\alpha\in(0,1)} D_\alpha(\rho_{RB}\|\rho_R\otimes\sigma_B) \tag{A.36}$$

$$= \inf_{\sigma_B} D(\rho_{RB}\|\rho_R\otimes\sigma_B) = I(R;B)_\rho. \tag{A.37}$$

(iii) We only prove the assertion for $\widetilde{I}_\alpha(\mathcal{N})$, as the proof for $I_\alpha(\mathcal{N})$ goes exactly the same way. First,

$$\lim_{\alpha\nearrow 1}\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\alpha\in(0,1)}\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\alpha\in(0,1)}\sup_{\psi_{RA}}\widetilde{I}_\alpha(R;B)_{\mathcal{N}(\psi)} = \sup_{\psi_{RA}}\sup_{\alpha\in(0,1)}\widetilde{I}_\alpha(R;B)_{\mathcal{N}(\psi)} \tag{A.38}$$

$$= \sup_{\psi_{RA}}\widetilde{I}(R;B)_{\mathcal{N}(\psi)} = I(\mathcal{N}). \tag{A.39}$$

Next, let $\hat{A}$ be a fixed copy of A. Note that $\psi_{\hat{A}A} \mapsto \inf_{\sigma_B\in S(\mathcal{H})_{++}}\widetilde{D}_\alpha\left(\mathcal{N}_{A\to B}(\psi_{\hat{A}A})\|\psi_{\hat{A}}\otimes\sigma_B\right) = \widetilde{I}_\alpha(\hat{A};B)_{\mathcal{N}(\psi)}$ is the infimum of continuous functions, and hence it is upper semi-continuous on the compact set of pure states on $\mathcal{H}_{\hat{A}A}$. On the other hand, it is monotone in α by (3.4), and hence we can use Lemma 17 and (A.32)–(A.34) to obtain

$$\lim_{\alpha\searrow 1}\widetilde{I}_\alpha(\mathcal{N}) = \inf_{\alpha>1}\widetilde{I}_\alpha(\mathcal{N}) = \inf_{\alpha>1}\sup_{\psi_{\hat{A}A}}\widetilde{I}_\alpha(R;B)_{\mathcal{N}(\psi)} = \sup_{\psi_{\hat{A}A}}\inf_{\alpha>1}\widetilde{I}_\alpha(R;B)_{\mathcal{N}(\psi)} = \sup_{\psi_{\hat{A}A}}I(R;B)_{\mathcal{N}(\psi)} = I(\mathcal{N}). \tag{A.40}$$

■

Proof of Lemma 11. Let $\hat{A}$ be a copy of A. By Lemma 9, we have

$$\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})}\inf_{\sigma_B\in S(\mathcal{H}_B)_{++}}\widetilde{D}_\alpha\left(\rho_{\hat{A}}^{1/2}\mathcal{N}_{A\to B}(\Gamma_{\hat{A}A})\rho_{\hat{A}}^{1/2}\middle\|\rho_{\hat{A}}\otimes\sigma_B\right). \tag{A.41}$$

Let $\Gamma^{\mathcal{N}} \equiv \mathcal{N}_{A\to B}(\Gamma_{\hat{A}A})$. According to [18, Lemma 3], the Rényi divergence in (A.41) can be written as

$$\frac{1}{\alpha-1}\log s(\alpha)\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B), \tag{A.42}$$

where

$$\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) = s(\alpha)\operatorname{Tr}\left\{\left([\Gamma^{\mathcal{N}}]^{1/2}\left(\rho_{\hat{A}}^{\frac{1}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\right)[\Gamma^{\mathcal{N}}]^{1/2}\right)^\alpha\right\}, \tag{A.43}$$

and $s(\alpha) := -1$ for $\alpha\in(0,1)$, and $s(\alpha) := 1$ for $\alpha > 1$. For $\alpha\in[1/2,1)$, $x\mapsto x^{\frac{1-\alpha}{\alpha}}$ is operator concave on $\mathbb{R}_+$, and $X\mapsto \operatorname{Tr}\{X^\alpha\}$ is monotone increasing and concave on positive semidefinite operators. Thus $\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B)$ is convex in σ_B. Note that $\rho_{\hat{A}}\mapsto\rho_{\hat{A}}\otimes I_B$ is affine, and applying Theorem 1.1 in [11], with $p := \frac{1}{\alpha}, q = 1$, $B = I_{\hat{A}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}[\Gamma^{\mathcal{N}}]^{1/2}$, to the quantity (1.3) in [11], we get that $\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B)$ is concave in $\rho_{\hat{A}}$. Similarly, for $\alpha > 1$, $x\mapsto x^{\frac{1-\alpha}{\alpha}}$ is operator convex on $\mathbb{R}_{++}$, and $X\mapsto\operatorname{Tr}\{X^\alpha\}$ is monotone increasing and convex on positive semidefinite operators. Thus $\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B)$ is convex in σ_B, and again by Theorem 1.1 in [11], it is concave in $\rho_{\hat{A}}$. Hence, we can use the Kneser-Fan minimax theorem [16, 30] to obtain

$$\widetilde{I}_\alpha(\mathcal{N}) = \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})} \inf_{\sigma_B\in S(\mathcal{H}_B)_{++}} \frac{1}{\alpha-1}\log s(\alpha)\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) \tag{A.44}$$

$$= \frac{1}{\alpha-1}\log s(\alpha) \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})} \inf_{\sigma_B\in S(\mathcal{H}_B)_{++}} \widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) \tag{A.45}$$

$$= \frac{1}{\alpha-1}\log s(\alpha) \inf_{\sigma_B\in S(\mathcal{H}_B)_{++}} \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})} \widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) \tag{A.46}$$

$$= \inf_{\sigma_B\in S(\mathcal{H}_B)_{++}} \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})} \frac{1}{\alpha-1}\log s(\alpha)\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) \tag{A.47}$$

$$= \inf_{\sigma_B\in S(\mathcal{H}_B)} \sup_{\rho_{\hat{A}}\in\mathcal{S}(\mathcal{H}_{\hat{A}})} \frac{1}{\alpha-1}\log s(\alpha)\widetilde{Q}_\alpha(\rho_{\hat{A}},\sigma_B) \tag{A.48}$$

$$= \inf_{\sigma_B\in S(\mathcal{H}_B)} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) \tag{A.49}$$

for every $\alpha \in [1/2, +\infty) \setminus \{1\}$. The case $\alpha = 1$ follows by

$$I(\mathcal{N}) = I_1(\mathcal{N}) = \inf_{\alpha>1} \widetilde{I}_\alpha(\mathcal{N}) = \inf_{\alpha>1} \inf_{\sigma_B\in S(\mathcal{H}_B)} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) = \inf_{\sigma_B\in S(\mathcal{H}_B)} \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}\|\mathcal{R}_{\sigma_B}) \tag{A.50}$$

$$= \inf_{\sigma_B\in S(\mathcal{H}_B)} \widetilde{D}(\mathcal{N}\|\mathcal{R}_{\sigma_B}), \tag{A.51}$$

where the second and the last identities are due to Lemma 10. ∎

References

[1] Tsuyoshi Ando. Convexity of certain maps on positive definite matrices and applications to Hadamard products. *Linear Algebra and its Applications*, 26:203–241, 1979.

[2] K. M. R. Audenaert, M. Nussbaum, A. Szkola, and F. Verstraete. Asymptotic error rates in quantum hypothesis testing. *Communications in Mathematical Physics*, 279:251–283, 2008. arXiv:0708.4282.

[3] U. Augustin. Noisy channels. Habilitation thesis, Universitat Erlangen-Nurnberg, West Germany, September 1978.

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

[5] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, May 2014. arXiv:0912.5537.

[6] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081–3084, October 1999. arXiv:quant-ph/9904023.

[7] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48:2637–2655, October 2002. arXiv:quant-ph/0106052.

[8] Richard Blahut. Hypothesis testing and information theory. *IEEE Transactions on Information Theory*, 20(4):405–417, July 1974.

[9] Garry Bowen. Quantum feedback channels. *IEEE Transactions on Information Theory*, 50:2429–2433, October 2004. arXiv:quant-ph/0209076.

[10] Marat V. Burnashev and Alexander S. Holevo. On reliability function of quantum communication channel. *Problems of Information Transmission*, 34:97–107, 1998. arXiv:quant-ph/9703013.

[11] Eric A. Carlen and Elliot H. Lieb. A Minkowski type trace inequality and strong subadditivity of quantum entropy II: convexity and concavity. *Letters in Mathematical Physics*, 83(2):107–126, February 2008. arXiv:0710.4167.

[12] Imre Csiszar and Janos Korner. Feedback does not affect the reliability function of a DMC at rates above capacity. *IEEE Transactions on Information Theory*, 28(1):92–93, January 1982.

[13] Igor Devetak, Christopher King, Marius Junge, and Mary Beth Ruskai. Multiplicativity of completely bounded p-norms implies a new additivity result. *Communications in Mathematical Physics*, 266(1):37–63, August 2006. arXiv:quant-ph/0506196.

[14] Runyao Duan, Yuan Feng, and Mingsheng Ying. Perfect distinguishability of quantum operations. *Physical Review Letters*, 103(21):210501, November 2009. arXiv:0908.0119.

[15] Gunther Dueck and Janos Korner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Transactions on Information Theory*, 25(1):82–85, January 1979.

[16] Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42–47, 1953.

[17] 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.

[18] Manish Gupta and Mark M. Wilde. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. *Communications in Mathematical Physics*, 334(2):867–887, March 2015. arXiv:1310.7028.

[19] Aram W. Harrow, Avinatan Hassidim, Debbie Leung, and John Watrous. Adaptive versus non-adaptive strategies for quantum channel discrimination. *Physical Review A*, 81(3):032339, March 2010. arXiv:0909.0256.

[20] Masahito Hayashi. *Quantum Information Theory: An Introduction.* Springer, 2006.

[21] 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.

[22] Masahito Hayashi. Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Transactions on Information Theory*, 55(8):3807–3820, August 2009. arXiv:0804.0686.

[23] 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.

[24] Masahito Hayashi and Marco Tomamichel. Correlation detection and an operational interpretation of the Rényi mutual information. August 2014. arXiv:1408.6894.

[25] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99–114, December 1991.

[26] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998.

[27] Alexander S. Holevo. Reliability function of general classical-quantum channel. *IEEE Transactions on Information Theory*, 46(6):2256–2261, September 2000. arXiv:quant-ph/9907087.

[28] Alexander S. Holevo. On entanglement assisted classical capacity. *Journal of Mathematical Physics*, 43(9):4326–4333, September 2002. arXiv:quant-ph/0106075.

[29] Anna Jenčová. A relation between completely bounded norms and conjugate channels. *Communications in Mathematical Physics*, 266(1):65–70, August 2006. arXiv:quant-ph/0601071.

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

[31] 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.

[32] Ke Li. Second order asymptotics for quantum hypothesis testing. *Annals of Statistics*, 42(1):171–189, February 2014. arXiv:1208.1400.

[33] Elliot H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11:267–288, December 1973.

[34] Seth Lloyd. Enhanced sensitivity of photodetection via quantum illumination. *Science*, 321(5895):1463–1465, September 2008. arXiv:0803.2022.

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

[36] Milán Mosonyi and Tomohiro Ogawa. Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. *Communications in Mathematical Physics*, 334(3):1617–1648, March 2015. arXiv:1309.3228.

[37] 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.

[38] 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.

[39] Hiroshi Nagaoka. The converse part of the theorem for quantum Hoeffding bound. November 2006. arXiv:quant-ph/0611289.

[40] Tomohiro Ogawa and Masahito Hayashi. On error exponents in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 50(6):1368–1372, June 2004. arXiv:quant-ph/0206151.

[41] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, November 2000. arXiv:quant-ph/9906090.

[42] Tomohiro Ogawa and Hiroshi Nagaoka. Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Transactions on Information Theory*, 53(6):2261–2266, June 2007. arXiv:quant-ph/0208139.

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

[44] Gilles Pisier. Non-commutative vector valued L_p-spaces and completely p-summing maps. *Astérisque*, 247, 1998.

[45] Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation*, pages 1327–1333, September 2010.

[46] Renato Renner. *Security of Quantum Key Distribution.* PhD thesis, ETH Zurich, September 2005. arXiv:quant-ph/0512258.

[47] Massimiliano F. Sacchi. Entanglement can enhance the distinguishability of entanglement-breaking channels. *Physical Review A*, 72(1):014305, July 2005. arXiv:quant-ph/0505174.

[48] Massimiliano F. Sacchi. Optimal discrimination of quantum operations. *Physical Review A*, 71(6):062340, June 2005. arXiv:quant-ph/0505183.

[49] Benjamin Schumacher and Michael Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

[50] Si-Hui Tan, Baris I. Erkmen, Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, Stefano Pirandola, and Jeffrey H. Shapiro. Quantum illumination with Gaussian states. *Physical Review Letters*, 101(25):253601, December 2008. arXiv:0810.0534.

[51] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, December 2009. arXiv:0811.1221.

[52] Marco Tomamichel and Masahito Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, November 2013. arXiv:1208.1478.

[53] Marco Tomamichel, Mark M. Wilde, and Andreas Winter. Strong converse rates for quantum communication. June 2014. arXiv:1406.2946.

[54] Hisaharu Umegaki. Conditional expectation in an operator algebra. *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962.

[55] Ligong Wang and Renato Renner. One-shot classical-quantum capacity and hypothesis testing. *Physical Review Letters*, 108(20):200501, May 2012. arXiv:1007.5456.

[56] 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.