

Sampling complexity of quantum channel discrimination

Lei Li¹, Zhe Ji¹, Qing-Wen Wang², Shu-Qian Shen¹ and Ming Li¹

¹College of Science, China University of Petroleum, Qingdao 266580, China

²College of Science, Shanghai University, Shanghai 200444, China

E-mail: lileiupc@126.com

Received 26 November 2024, revised 3 March 2025

Accepted for publication 3 March 2025

Published 13 June 2025



CrossMark

Abstract

In this paper, we study the problem of sampling complexity for channel discrimination with respect to two different strategies: product strategy and adaptive strategy. We first formally introduce the definitions of the sampling complexity of the channels under the framework of hypothesis testing, wherein the goal is to determine the minimum number of samples needed to reach a desired error probability. We then establish the lower and upper bounds on the sampling complexity of the symmetric, asymmetric, and error exponent hypothesis testing settings. We show that, by imposing product strategy on testing, the bounds are always characterized by the generalized channel divergence, while with adaptive strategy, the bounds are characterized by the amortized channel divergence. Finally, we analyze two concrete examples, and obtain that the adaptive strategy can not lead to an advantage to the problem of determining the sampling complexity for classical-quantum channels, which can bring advantages for generalized amplitude damping channels.

Keywords: quantum channel, adaptive strategy, quantum discrimination, quantum complexity

(Some figures may appear in colour only in the online journal)

1. Introduction

Distinguishing between two quantum statistical objects is a central task in quantum information theory. So mathematical methods underlying distinguishability have been studied for a long time [1–3] and are now highly developed under the general framework of quantum hypothesis testing [4, 5]. Briefly speaking, the hypothesis testing theory means that, with sufficient statistical evidence, one can determine which hypothesis should be rejected in favor of the other and if the null hypothesis is accepted, one can form alternative hypotheses to test against the null hypothesis in future experiments. Quantum hypothesis testing is an important topic in the sense that it has wide applications to quantum communication [6, 7], and quantum computation [8].

The most basic setting is that of state discrimination where we are given a quantum state that is either in the state ρ or in the state σ . The goal of the distinguisher, who does not know *a priori* which state was prepared, is to determine the identity of the unknown state. Here we consider the independent and identically distributed (IID) case: given n

quantum systems in the state $\rho^{\otimes n}$ or $\sigma^{\otimes n}$, the task is to apply a binary measurement $\{Q_n, I^{\otimes n} - Q_n\}$ with $0 \leq Q_n \leq I^{\otimes n}$, to determine which state one possesses. Then two kinds of error probabilities occur:

$$\alpha_n(Q_n) = \text{Tr}\{(I^{\otimes n} - Q_n)\rho^{\otimes n}\}, \quad (\text{type I}) \quad (1)$$

$$\beta_n(Q_n) = \text{Tr}\{Q_n\sigma^{\otimes n}\}, \quad (\text{type II}). \quad (2)$$

When a constant threshold ε is imposed on the type I error, the optimal type II error is given by

$$\beta_\varepsilon(\rho||\sigma) = \min\{\beta(Q): 0 \leq Q \leq I, \alpha(Q) \leq \varepsilon\}. \quad (3)$$

The quantum stein's lemma states that [4, 5, 9, 10]

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}||\sigma^{\otimes n}) \\ = \lim_{n \rightarrow \infty} -\frac{1}{n} \log \beta_\varepsilon(\rho^{\otimes n}||\sigma^{\otimes n}) = D(\rho||\sigma), \end{aligned} \quad (4)$$

where $D_H^\varepsilon(\rho||\sigma) = -\log \inf\{\text{Tr}(Q\sigma): 0 \leq Q \leq I, \text{Tr}(Q\rho) \geq 1 - \varepsilon\}$ denotes the quantum hypothesis testing relative entropy and $D(\rho||\sigma)$, the standard quantum relative entropy. It is also remarkable that the quantum stein's lemma provides a

rigorous operational interpretation for the quantum relative entropy.

Although research in quantum hypothesis testing has largely focused on quantum states, there are various situations in which quantum channels play the role of the main objects of study. The task of channel discrimination is quite different from that of state discrimination due to its nature as dynamic resources [6, 11–16]. Unlike state discrimination, the dynamic feature of quantum channels raises challenging difficulties in determining the optimal discrimination scheme as we have to handle the additional optimization of the input states and the non-IID structure of the testing states [17]. In addition, we need to consider what role the discrimination strategies play in the quantum channel discrimination. Despite inherent mathematical links between the channel and state discrimination problems, the discrimination of channels is more complicated, due to the additional degrees of freedom introduced by the adaptive strategies. Several works have been dedicated to study the potential advantages of adaptive strategies over nonadaptive strategies in channel discrimination [18, 19]. It is known that asymptotically the exponential error rate for classical-quantum channel discrimination is not improved by adaptive strategies [11, 16, 20]. However, there indeed exists separation between adaptive and nonadaptive symmetric hypothesis testing exponents for quantum channels, and the concrete example is provided in [20]. For quantum-quantum channels, i.e., channels having quantum input and quantum outputs, it is shown that adaptive strategies provide an advantage in the non-asymptotic regime for discrimination in the symmetric Chernoff setting [18, 21, 22]. Therefore, a deeper understanding and analysis on hypothesis testing of quantum channels is imperative.

It is well known in information theory literature [23, 24] that the error probability decreases exponentially fast in the number of samples provided, that is, extra samples are helpful for decreasing the error probability when trying to determine the unknown objects. So it is desirable to consider the natural problem that given a constant upper bound on the desired error probability, how the distinguisher determines the value of x with as few samples as possible (minimize n) while meeting the error probability constraint. This notion is often called sampling complexity in the literature of algorithms and machine learning [25, 26]. It is useful because it indicates how long one needs to wait in order to make a decision with a desired performance [27].

In this paper, we systematically study the sampling complexity of channel discrimination under the framework of quantum hypothesis testing. Here, we take two discrimination strategies into consideration. The product strategy for n uses of the channel consists of feeding an input state into the n -fold tensor product channel. The adaptive strategy refers to the way of sending the initial state to the given channel followed by applying an update channel. We provide the lower and upper bound on the sampling complexity with respect to the symmetric, asymmetric and error exponent settings. These bounds are characterized by various channel divergences defined below. Two particular examples are considered. We find the amortized channel divergence collapses into

generalized channel divergence for the classical-quantum channel, which indicates that in such a case, the lower and upper bounds appeared in the main theorems for two different strategies are completely equal. Thus, the adaptive strategy can not lead to an advantage to the problem of determining the sampling complexity for classical-quantum channels. For generalized amplitude damping channels, the numerical simulation shows that there is always a gap between amortized channel divergence and generalized channel divergence. That is to say, the adaptive strategy can bring advantages over the product strategy in the task of determining the sampling complexity for generalized amplitude damping channels. This partially clarifies the long-standing question of the usefulness of adaptive strategies in quantum channel discrimination protocols.

2. Preliminaries

2.1. Basic notation

Throughout, we denote the finite dimensional Hilbert space by H_A, H_B corresponding to physical system A, B . Quantum states of system A are denoted by $\rho_A \in \mathcal{S}(A)$. A quantum channel $\mathcal{N}_{A \rightarrow B}$ is a completely positive (CP) and trace-preserving (TP) linear map from operators acting on H_A to operators acting on H_B . Let Θ denote a quantum superchannel that takes quantum channel $\mathcal{N}_{A \rightarrow B}$ to $\mathcal{M}_{C \rightarrow D}$. Then, for any quantum superchannel Θ , it can be physically realized as follows [28]:

$$\begin{aligned} \mathcal{M}_{C \rightarrow D} &= \Theta(\mathcal{N}_{A \rightarrow B}) \\ &= \mathcal{V}_{BR \rightarrow D} \circ (\mathcal{N}_{A \rightarrow B} \otimes \mathcal{I}_R) \circ \mathcal{U}_{C \rightarrow AR}, \end{aligned} \quad (5)$$

where $\mathcal{U}_{C \rightarrow AR}$ is a pre-processing channel, system E corresponds to environmental system, and $\mathcal{V}_{BR \rightarrow D}$ is a post-processing channel.

Let ρ and σ be two quantum states. The quantum relative entropy is defined as:

$$D(\rho||\sigma) = \text{Tr}(\rho(\log \rho - \log \sigma)), \quad (6)$$

where Tr denotes the trace and \log is the logarithm with base 2 throughout the paper. The Petz–Rényi divergence is defined for $\alpha \in (0, 1) \cup (1, \infty)$ as [29, 30]:

$$D_\alpha(\rho||\sigma) = \frac{1}{\alpha - 1} \log \text{Tr}[\rho^\alpha \sigma^{1-\alpha}]. \quad (7)$$

Recently, the Sandwiched Rényi relative entropy was introduced in [31]:

$$\tilde{D}_\alpha(\rho||\sigma) = \frac{1}{\alpha - 1} \log \text{Tr}\{(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha\}, \quad (8)$$

when the support of ρ is contained in the support of σ and it is equal to $+\infty$ otherwise. It is known [32] that in the limit $\alpha \rightarrow 1$ and $\alpha \rightarrow +\infty$,

$$\lim_{\alpha \rightarrow 1} \tilde{D}_\alpha(\rho||\sigma) = \lim_{\alpha \rightarrow 1} D_\alpha(\rho||\sigma) = D(\rho||\sigma), \quad (9)$$

and

$$\lim_{\alpha \rightarrow +\infty} \tilde{D}_\alpha(\rho||\sigma) = D_{\max}(\rho||\sigma) = \inf\{\lambda: \rho \leq 2^\lambda \sigma\}. \quad (10)$$

In particular, $D_\alpha(\rho||\sigma)$ and $\tilde{D}_\alpha(\rho||\sigma)$ were shown to obey the desirable monotonicity inequality for all $\alpha \in [0, 1) \cup (1, +2)$ and $\alpha \in [\frac{1}{2}, 1) \cup (1, +\infty)$ [33], respectively. The Uhlmann, fidelity between the two quantum states ρ and σ is defined as $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$, where $\|\cdot\|_1$ denotes the trace norm. Obviously, we have that $\tilde{D}_{\frac{1}{2}}(\rho||\sigma) = -2 \log F(\rho, \sigma)$.

The divergence between the quantum states can be naturally extended to quantum channels. The main idea is to quantify the maximum divergence between the outputs from the channels.

Definition 1. (Generalized channel divergence) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$, we define the generalized channel divergence induced by state divergence D as [34]

$$\begin{aligned} \mathbf{D}(\mathcal{N}||\mathcal{M}) \\ = \sup_{\rho_{RA} \in \mathcal{S}(RA)} D(\mathcal{N}_{A \rightarrow B}(\rho_{RA})||\mathcal{M}_{A \rightarrow B}(\rho_{RA})). \end{aligned} \quad (11)$$

Note that as a consequence of purification, data processing, and the Schmidt decomposition, it follows that the supremum can be restricted to be with respect to pure states.

In order to explore the largest distinguishability that can be realized between the channels, the amortized channel divergence is defined as follows.

Definition 2. (Amortized channel divergence) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$, we define the amortized channel divergence induced by state divergence D as [16]

$$\begin{aligned} \mathbf{D}^A(\mathcal{N}||\mathcal{M}) \\ = \sup_{\rho_{RA}, \sigma_{RA} \in \mathcal{S}(RA)} [D(\mathcal{N}_{A \rightarrow B}(\rho_{RA})||\mathcal{M}_{A \rightarrow B}(\sigma_{RA})) - D(\rho_{RA}||\sigma_{RA})]. \end{aligned} \quad (12)$$

Note that due to the supremum over the input state, the above two quantities are difficult to calculate in general. However, for specific cases, we can evaluate them numerically or via semi-definite program (SDP). When we choose the state divergence as the max-relative entropy, amortized channel divergence collapses into generalized channel divergence. Then $\mathbf{D}_{\max}^A(\mathcal{N}||\mathcal{M}) = \mathbf{D}_{\max}(\mathcal{N}||\mathcal{M}) = \mathbf{D}_{\max}(\mathcal{N}(\Phi_{RA})||\mathcal{M}(\Phi_{RA}))$ with Φ_{RA} being the maximal entanglement state, which is a SDP and therefore efficiently computable [16]. Another extreme case [6] is when both channels are replacer channels, general divergences are equal to the corresponding divergences of the two states. Similarly, based on the fidelity of the quantum state, we can define the fidelity of the two quantum channels \mathcal{N} and \mathcal{M} as follows.

Definition 3. (Fidelity of channel) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$, we define the fidelity of channel induced by state fidelity as

$$\begin{aligned} \mathbf{F}(\mathcal{N}||\mathcal{M}) \\ = \inf_{\rho_{RA} \in \mathcal{S}(RA)} F(\mathcal{N}_{A \rightarrow B}(\rho_{RA}), \mathcal{M}_{A \rightarrow B}(\rho_{RA})). \end{aligned} \quad (13)$$

The idea behind the fidelity of the quantum channel is to allow for the arbitrary input state to be used to reduce the degree of approximation between the two channels. Intuitively, the definition quantifies the worst-case behavior of the channels for information preservation. We discuss several basic properties of channel fidelity in [appendix A](#), which may be of independent interest.

2.2. Strategies for quantum channel discrimination

The problem of quantum channel discrimination is made mathematically precise by the following hypothesis testing problems for quantum channels. Given two quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ acting on an input system A and an output system B , there are two fundamental different classes of available strategies as follows [6].

(1) Product strategy: in a product strategy the testing state is created by choosing a sequence of input states $\varphi \in \mathcal{S}(RA)$ and sending them to the unknown channel \mathcal{G} individually. The generated testing state is given by $\mathcal{G}^{\otimes n}(\varphi^{\otimes n})$. The class of all product strategies is denoted as product strategies (PRO).

(2) Adaptive strategy: in an adaptive strategy, we first choose an initial state $\varphi_n \in \mathcal{S}(R_1 A_1)$ and send it to one copy of the channel \mathcal{G} followed by applying an update channel \mathcal{P}_1 . After this, we apply another copy of the channel \mathcal{G} and then an update channel \mathcal{P}_2 . Then, repeat this process n times. We get the final testing state $\mathcal{G} \circ \mathcal{P}_{n-1} \circ \dots \circ \mathcal{P}_2 \circ \mathcal{G} \circ \mathcal{P}_1 \circ \mathcal{G}(\varphi_n)$ with $\mathcal{P}_i \in CPTP(R_i B_i \rightarrow R_{i+1} A_i)$. The class of all adaptive strategies is denoted as adaptive strategies (ADP). For more details about adaptive strategy, please refer to the literature [16].

It is clear that $\text{PRO} \subseteq \text{ADP}$.

2.3. Setting of quantum hypothesis testing for quantum channel

In this section, we describe the information-theoretic setting for non-asymptotic quantum channel discrimination. For any discrimination strategy S_n , let $\rho_n(S_n)$ and $\sigma_n(S_n)$ denote the output state generated by n uses of the channel depending on whether it is \mathcal{N} or \mathcal{M} . $\{Q_n, I - Q_n\}$ is the measurement part of the strategy. In analogy with (1) and (2), one can define the type I and II error as:

$$\alpha_n(S_n, Q_n) = \text{Tr}\{(I - Q_n)\rho_n(S_n)\}, \quad (\text{type } I) \quad (14)$$

$$\beta_n(S_n, Q_n) = \text{Tr}\{Q_n \sigma_n(S_n)\}, \quad (\text{type } II). \quad (15)$$

The discrimination task is then to decide which hypothesis is true based on the data drawn from an optimal positive operator-valued measure (POVM) which leads to the minimal error probability, as detailed below. The setting of the quantum hypothesis testing can be classified as symmetric and asymmetric depending on whether the two types of decision errors are treated equally.

In the symmetric setting, the two types of errors are treated equally, where the purpose is to minimize the average (Bayesian) of two error probabilities weighted by the prior probabilities [35]. Given *a priori* probability p that the first channel \mathcal{N} is selected, the non-asymptotic expected error

probability is defined as

$$p_{e\{Q_n\}}(p, \mathcal{N}, \mathcal{M}, n) = p\alpha_n(S_n, Q_n) + (1-p)\beta_n(S_n, Q_n). \quad (16)$$

Then the distinguisher can minimize the error probability expression in (16) over all possible measurements. The Helstrom–Holevo theorem [2, 3] states that the optimal error probability of hypothesis testing is as follows:

$$p_e(p, \mathcal{N}, \mathcal{M}, n) = \inf_{Q_n} p_{e\{Q_n\}}(p, \mathcal{N}, \mathcal{M}, n) = \frac{1}{2}(1 - \|p\rho_n(S_n) - (1-p)\sigma_n(S_n)\|_1). \quad (17)$$

In asymmetric hypothesis testing, the goal of the distinguisher is to minimize the probability of the second kind of error subject to a constraint on the probability of the first kind of error. Given a fix $\varepsilon \in (0, 1)$, we are interested in the following non-asymptotic optimization problem:

$$\zeta(\varepsilon, \mathcal{N}, \mathcal{M}, n) = \inf_{Q_n} \{\beta_n(S_n, Q_n) | \alpha_n(S_n, Q_n) \leq \varepsilon\}. \quad (18)$$

As a refinement of asymmetric hypothesis testing, the error exponent studies that the type II error probability is constrained to decrease exponentially with exponent $r > 0$. We are then concerned with the error exponent behavior of the type I error under this constraint. This is defined as the non-asymptotic quantity

$$\xi(r, \mathcal{N}, \mathcal{M}, n) = \inf_{Q_n} \{\alpha_n(S_n, Q_n) | \beta_n(S_n, Q_n) \leq 2^{-m}\}. \quad (19)$$

3. Sampling complexity for quantum channel hypothesis testing

3.1. Definitions of sampling complexity

Note that the above quantities (17)–(19) are somewhat different from the traditional definitions in the literature [35–37], where they are called Chernoff setting, Stein setting and Hoeffding setting, respectively. Here we are no longer interested in the asymptotic formulas of these quantities. The goal of this paper is to determine the minimum value of n (i.e., minimum number of samples) needed to meet a fixed error probability constraint. These formulations of the hypothesis testing problem are more consistent with the notion of the runtime of a probabilistic or quantum algorithm, for which one typically finds that the runtime depends on the desired error. Indeed, if there are procedures for preparing the various states involved in a given hypothesis testing problem, with fixed runtimes, then the sampling complexity corresponds to the total amount of time one must wait to prepare n samples to achieve a desired error probability [27].

In this section, we formally introduce the definitions of the sampling complexity of quantum hypothesis testing in the following.

Definition 4. (Sampling complexity of (17)) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ and $\varepsilon \in (0, 1)$, we define the sample complexity induced by (17) as

$$n^*(\varepsilon, p, \mathcal{N}, \mathcal{M}, n) = \inf\{n: p_e(p, \mathcal{N}, \mathcal{M}, n) \leq \varepsilon\}. \quad (20)$$

Definition 5. (Sampling complexity of (18)) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ and $\varepsilon, \delta \in (0, 1)$, we define the sample complexity induced by (18) as

$$n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) = \inf\{n: \zeta(\varepsilon, \mathcal{N}, \mathcal{M}, n) \leq \delta\}. \quad (21)$$

Definition 6. (Sampling complexity of (19)) Given quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$, $\varepsilon \in (0, 1)$, and $r > 0$, we define the sample complexity induced by (19) as

$$n^*(\varepsilon, r, \mathcal{N}, \mathcal{M}, n) = \inf\{n: \xi(r, \mathcal{N}, \mathcal{M}, n) \leq \varepsilon\}. \quad (22)$$

Indeed, the definitions of sampling complexity above have more explicit expressions, which can be in favor of proceeding to the development of our main results in the next section.

Remark 1. (Equivalent expressions for sampling complexity) By recalling the quantities defined in the above definitions (20–22), we rewrite the sampling complexity in the following ways:

$$\begin{aligned} n^*(\varepsilon, p, \mathcal{N}, \mathcal{M}, n) &= \inf\{n: \frac{1}{2}(1 - \|p\rho_n(S_n) - (1-p)\sigma_n(S_n)\|_1) \leq \varepsilon\} \\ &= \inf\{n: \|p\rho_n(S_n) - (1-p)\sigma_n(S_n)\|_1 \geq 1 - 2\varepsilon\}. \end{aligned} \quad (23)$$

$$\begin{aligned} n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) &= \inf_{Q_n} \{\beta_n(S_n, Q_n) \leq \delta | \alpha_n(S_n, Q_n) \leq \varepsilon\}. \end{aligned} \quad (24)$$

$$\begin{aligned} n^*(\varepsilon, r, \mathcal{N}, \mathcal{M}, n) &= \inf_{Q_n} \{\alpha_n(S_n, Q_n) \leq \varepsilon | \beta_n(S_n, Q_n) \leq 2^{-m}\}. \end{aligned} \quad (25)$$

It is clear that calculating the exact value of $n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n)$, $n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n)$ and $n^*(\varepsilon, r, \mathcal{N}, \mathcal{M}, n)$ is very difficult due to the optimization process. Therefore, here we aim to find the upper and lower bounds on these sampling complexities, which are easier to compute. Before proceeding with stating our results, we discuss the conditions for when the channels become perfectly distinguishable. By employing the idea of [21], we say that the channels \mathcal{N} and \mathcal{M} are perfectly distinguishable if and only if $\exists \varphi_{RA} \in \mathcal{S}(A)$ with $|R| = |A|$, such that $\text{supp}(\mathcal{N}(\varphi_{RA})) \cap \text{supp}(\mathcal{M}(\varphi_{RA})) = \emptyset$. In such case, it is obvious that the sampling complexities of the above hypothesis testing become trivial, that is, $n^* = 1$.

3.2. Sampling complexity for product strategy

In what follows, we begin to consider the sampling complexity of three hypothesis testing with product strategy, which is much simpler than that of an adaptive case. In order

to get the lower and upper bound of sampling complexity, we recall some useful lemmas in the Appendix B.

Theorem 1. Let $\varepsilon, p, \mathcal{N}, \mathcal{M}$ be as stated in definition 4, and we choose the product channel discrimination strategy. Then the following bounds hold

$$\frac{\log \frac{\varepsilon}{p(1-p)}}{2 \log \mathbf{F}(\mathcal{N} \parallel \mathcal{M})} \leq n^*(\varepsilon, p, \mathcal{N}, \mathcal{M}, n) \leq \left\lceil \inf_{s \in (0,1)} \frac{\log \frac{\varepsilon}{p^s(1-p)^{1-s}}}{(s-1)\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil. \quad (26)$$

Proof. We begin by establishing the left hand side of inequality in (26).

$$\begin{aligned} \varepsilon &\geq \frac{1}{2}(1 - \|p\rho_n - (1-p)\sigma_n\|_1) \\ &\geq \frac{1}{2}(1 - \sqrt{1 - 4p(1-p)F^2(\rho_n, \sigma_n)}) \\ &\geq p(1-p)F^2(\rho_n, \sigma_n) \\ &= p(1-p)F(\mathcal{N}(\varphi), \mathcal{M}(\varphi))^{2n} \\ &\geq p(1-p)\mathbf{F}(\mathcal{N}, \mathcal{M})^{2n}. \end{aligned} \quad (27)$$

The second inequality follows from lemma B1 in Appendix B. The third inequality from $\frac{1}{2}(1 - \sqrt{1-x}) \geq \frac{x}{4}$ for all $x \in [0, 1]$. The first equality follows from the multiplicativity of fidelity, and the last inequality is from definition 3.

For the other direction, let s^* be the optimal value of $\left\lceil \inf_{s \in (0,1)} \frac{\log \frac{\varepsilon}{p^s(1-p)^{1-s}}}{(s-1)\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil$. Then for all n satisfying $n \geq \frac{\log \frac{\varepsilon}{p^{s^*}(1-p)^{1-s^*}}}{(s^*-1)\mathbf{D}_{s^*}(\mathcal{N} \parallel \mathcal{M})}$, we have

$$n(s^* - 1)\mathbf{D}_{s^*}(\mathcal{N} \parallel \mathcal{M}) \leq \log \frac{\varepsilon}{p^{s^*}(1-p)^{1-s^*}}, \quad (28)$$

$$\log p^{s^*}(1-p)^{1-s^*} + n(s^* - 1)\mathbf{D}_{s^*}(\mathcal{N} \parallel \mathcal{M}) \leq \log \varepsilon, \quad (29)$$

$$\text{Tr}[p\rho_n]^{s^*} [(1-p)\sigma_n]^{1-s^*} \leq \varepsilon. \quad (30)$$

(30) follows from the additivity of Petz-Rényi divergence, and definition 1. Then by using lemma B2, we find that

$$\frac{1}{2}(1 - \|p\rho_n - (1-p)\sigma_n\|_1) \leq \varepsilon. \quad (31)$$

Therefore, we complete the proof of the right hand side of (26) due to the fact that the optimal sample complexity can not exceed this choice. \square

Remark 2. In fact, the lower bound can be rewritten as $-\frac{\log \frac{\varepsilon}{p(1-p)}}{\mathbf{D}_{\frac{1}{2}}(\mathcal{N} \parallel \mathcal{M})}$ due to the definition of sandwiched Rényi relative entropy.

Theorem 2. Let $\delta, \varepsilon, \mathcal{N}, \mathcal{M}$ be as stated in definition 5, and we choose the product strategy. Then the following

bounds hold

$$\sup_{s>1} \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{1-\varepsilon}\right)}{\tilde{\mathbf{D}}_s(\mathcal{N} \parallel \mathcal{M})} \leq n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) \leq \left\lceil \inf_{s \in (0,1)} \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right)}{\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil. \quad (32)$$

Proof. We begin to prove the lower bound in (32). Suppose that there exists a scheme such that (18) holds. Then for all $s > 1$, we have

$$\begin{aligned} \log \frac{1}{\delta} &\leq -\log \beta_n \\ &\leq \tilde{\mathbf{D}}_s(\rho_n \parallel \sigma_n) + \frac{s}{s-1} \log \frac{1}{1-\varepsilon} \\ &= n\tilde{\mathbf{D}}_s(\mathcal{N}(\varphi), \mathcal{M}(\varphi)) + \frac{s}{s-1} \log \frac{1}{1-\varepsilon} \\ &\leq n\tilde{\mathbf{D}}_s(\mathcal{N}, \mathcal{M}) + \frac{s}{s-1} \log \frac{1}{1-\varepsilon}. \end{aligned} \quad (33)$$

The second inequality follows from lemma B4. The equality follows from the additivity of sandwiched Rényi entropy. The third inequality follows from the definition of generalized channel divergence (11).

Now we prove the upper bound of (32). Setting that $n = \left\lceil \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right)}{\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil$, then we have

$$n \geq \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right)}{\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})}, \quad (34)$$

which can be written as

$$n\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M}) \geq \log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right). \quad (35)$$

Together with the additivity of the Petz-Rényi divergence and lemma B5, we have

$$\beta_n(S_n) \leq \delta, \quad (36)$$

under the constrain $\alpha_1 \leq \varepsilon$, and for all $s \in (0, 1)$. From definition 5, since the optimal sample complexity can not exceed this choice, we thus conclude the following upper bound on sampling complexity

$$n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) \leq \left\lceil \inf_{s \in (0,1)} \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right)}{\mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil. \quad (37)$$

\square

Theorem 3. Let $\varepsilon, r, \mathcal{N}, \mathcal{M}$ be as state in definition 6 and we choose the product strategy. Then for all $s \in (0, 1)$,

- (1) if $r > \mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})$, there is no scheme that can distinguish the two channels \mathcal{N} and \mathcal{M} .
- (2) if $r \leq \mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})$, we have

$$n^*(\varepsilon, r, \mathcal{N}, \mathcal{M}, n) \leq \left\lceil \inf_{s \in (0,1)} \frac{\left(\frac{s}{1-s}\right) \log \varepsilon}{r - \mathbf{D}_s(\mathcal{N} \parallel \mathcal{M})} \right\rceil. \quad (38)$$

Proof. The upper bound is shown using a quantum Neyman–Pearson test [43] for comparing ρ_n with σ_n . Let us fix any $s \in (0, 1)$. Let λ_n be real numbers to be specified later and define the sequence of tests:

$$Q_n = \{\rho_n \geq 2^{\lambda_n} \sigma_n\}, \quad (39)$$

where $\{\rho_n \geq 2^{\lambda_n} \sigma_n\}$ is the projector onto the subspace spanned by eigenvectors corresponding to non-negative eigenvalues of $\rho_n - 2^{\lambda_n} \sigma_n$. Then we have

$$\begin{aligned} \beta_n &= \text{Tr}(Q_n \sigma_n) \\ &\leq 2^{-s\lambda_n} \text{Tr} \rho_n^s \sigma_n^{1-s} \\ &= 2^{-s\lambda_n} 2^{(s-1)D_s(\rho_n \parallel \sigma_n)} \\ &= 2^{-s\lambda_n + n(s-1)D_s(\mathcal{N}(\varphi) \parallel \mathcal{M}(\varphi))}. \end{aligned} \quad (40)$$

The inequality follows from lemma B6. We find that the choice

$$\lambda_n = \frac{nr + n(s-1)D_s(\mathcal{N}(\varphi) \parallel \mathcal{M}(\varphi))}{s}, \quad (41)$$

achieves the desired bound $\beta_n \leq 2^{-nr}$. On the other hand, using lemma B6, we also find

$$\begin{aligned} \alpha_n &= \text{Tr}((I - Q_n)\rho_n) \\ &\leq 2^{(1-s)\lambda_n} \text{Tr} \rho_n^s \sigma_n^{1-s} \\ &= 2^{(1-s)[\lambda_n - nD_s(\mathcal{N}(\varphi) \parallel \mathcal{M}(\varphi))]} \end{aligned} \quad (42)$$

Substituting the last line of (42) for λ_n , we find that, if $r > D_s(\mathcal{N} \parallel \mathcal{M})$,

$$\begin{aligned} &2^{\frac{1-s}{s}[nr - nD_s(\mathcal{N}(\varphi) \parallel \mathcal{M}(\varphi))]} \\ &\geq 2^{\frac{1-s}{s}[nr - nD_s(\mathcal{N} \parallel \mathcal{M})]} > \varepsilon. \end{aligned} \quad (43)$$

Thus, there is no scheme that can distinguish the two channels \mathcal{N} and \mathcal{M} .

If $r \leq D_s(\mathcal{N} \parallel \mathcal{M})$, and choosing the optimal state φ in (11), for any $n \geq \frac{(\frac{s}{1-s}) \log \varepsilon}{r - D_s(\mathcal{N} \parallel \mathcal{M})}$, we have

$$\begin{aligned} \alpha_n &\leq 2^{\frac{1-s}{s}[nr - nD_s(\mathcal{N}(\varphi) \parallel \mathcal{M}(\varphi))]} \\ &= 2^{\frac{1-s}{s}[nr - nD_s(\mathcal{N} \parallel \mathcal{M})]} \leq \varepsilon. \end{aligned} \quad (44)$$

By definition 6, this concludes our proof of (38) since the optimal sampling complexity can not exceed this choice. \square

The result is interesting, since it is consistent with the strong converse property of the state discrimination task, i.e., if the type II error vanishes exponentially with a rate larger than the relative entropy of two quantum states, then the type I error goes to 1. Here our result shows that the conclusion also applies to the case of distinguishing between two channels. If $r > D_s(\mathcal{N} \parallel \mathcal{M})$, there is no scheme that can distinguish two given channels.

3.3. Sampling complexity for adaptive strategy

Next, we refine our analysis by identifying the sampling complexity for the task of adaptive discrimination between the two arbitrary channels \mathcal{N} and \mathcal{M} . It is well known that adaptive strategies can give an advantage over nonadaptive strategies for the discrimination of classical and quantum channels in the finite, non-asymptotic regime. However, [16]

shows that for classical-quantum channels, asymptotically the exponential error rate for discrimination is not improved by adaptive strategies. Therefore, it is expected that the analysis for adaptiveness reveals the role of adaptive methods in various types of quantum information processing [20]. Here, we study the specific topic of sampling complexity of the adaptive discrimination of quantum channels, i.e, what the total amount of time one must wait to prepare n samples to achieve a desired (exponential) error probability. It turns out that in adaptive setting, the bounds of sampling complexity are characterized mainly based on the amortized quantum divergence of quantum channels.

Theorem 12. *Let $\varepsilon, p, \mathcal{N}, \mathcal{M}$ be as stated in definition 4, and we choose the adaptive channel discrimination strategy. Then the following lower bounds hold*

$$\begin{aligned} &-\frac{\log \frac{\varepsilon}{p(1-p)}}{\tilde{D}_1^A(\mathcal{N} \parallel \mathcal{M})} \leq n^*(\varepsilon, p, \mathcal{N}, \mathcal{M}, n) \\ &\leq \left[\inf_{s \in (0,1)} \frac{\log \frac{\varepsilon}{p^s(1-p)^{1-s}}}{(s-1)\tilde{D}_s^A(\mathcal{N} \parallel \mathcal{M})} \right]. \end{aligned} \quad (45)$$

Proof. We follow the proof strategy from theorem 8, combined with considering the Meta-converse of the adaptive channel discrimination. Here we only present the proof of the lower bound. The upper bound is given by following the same line. By using lemma B1 again, we have

$$\begin{aligned} \varepsilon &\geq \frac{1}{2}(1 - \|p\rho_n - (1-p)\sigma_n\|_1) \\ &\geq \frac{1}{2}(1 - \sqrt{1 - 4p(1-p)F^2(\rho_n, \sigma_n)}) \\ &\geq p(1-p)F^2(\rho_n, \sigma_n). \end{aligned} \quad (46)$$

Taking the logarithm on both sides of (46), this can be rewritten as

$$\begin{aligned} -\log \frac{\varepsilon}{p(1-p)} &\leq -2 \log F(\rho_n, \sigma_n) \\ &= \tilde{D}_1^A(\rho_n, \sigma_n) \\ &\leq n\tilde{D}_1^A(\mathcal{N} \parallel \mathcal{M}). \end{aligned} \quad (47)$$

The last inequality follows from lemma C1 (Meta-converse). \square

Remark 13. *Theorem 8 and theorem 12 establish the lower and upper bounds of sampling complexity for the same symmetric setting, with the product strategy and adaptive strategy respectively. Obviously, the differences of the bounds mainly exist in two kinds of channel divergences. Note that $\tilde{D}^A(\mathcal{N} \parallel \mathcal{M}) \geq D(\mathcal{N} \parallel \mathcal{M})$ [16]. Therefore, the lower and upper bounds in theorem 8 are usually larger than that in theorem 12, which indicates that the sampling complexity with product strategy lies in a relatively larger interval (both left and right) than that with adaptive strategy. The observation also applies to the following asymmetry and error exponent settings. In other words, if amortized channel divergence is strictly larger than the generalized divergence, the adaptive strategy usually needs fewer samples than the*

product strategy for the same error exponent, which explicitly means that the adaptive strategy shows its advantage in solving the sampling complexity problem of channel discrimination.

Theorem 5. Let $\delta, \varepsilon, \mathcal{N}, \mathcal{M}$ be as stated in definition 5, and we choose the adaptive strategy. Then the following bounds hold

$$\begin{aligned} \sup_{p>1} \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{1-\varepsilon}\right)}{\tilde{D}_s^A(\mathcal{N}||\mathcal{M})} &\leq n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) \\ &\leq \left[\inf_{s \in (0,1)} \frac{\log\left(\frac{1}{\delta}\right) - \frac{s}{s-1} \log\left(\frac{1}{\varepsilon}\right)}{D_s^A(\mathcal{N}||\mathcal{M})} \right]. \end{aligned} \quad (48)$$

Proof. The proof follows from the same line with theorem 10. \square

As shown in definition 5, the quantity $n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n)$ is closely related to the asymmetric hypothesis testing-quantum Stein's Lemma. We are then establishing a new lower bound by employing the original proof of quantum Stein's lemma for quantum states.

Theorem 6. Let $\delta, \varepsilon, \mathcal{N}, \mathcal{M}$ be as stated in definition 5, and we choose the adaptive strategy. Then the following bound holds

$$n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) \geq \frac{(\varepsilon - 1) \log \delta - h_2(\varepsilon)}{D^A(\mathcal{N}||\mathcal{M})}, \quad (49)$$

where $h_2(\varepsilon) = -\varepsilon \log \varepsilon - (1 - \varepsilon) \log(1 - \varepsilon)$ denotes the binary entropy.

Proof. Let $p = \text{Tr}\{Q_n \rho_n(S_n)\}$ and $q = \text{Tr}\{Q_n \sigma_n(S_n)\}$ be the final decision probabilities corresponding to measurement Q_n . We can modify the constraint $\alpha_n(S_n, Q_n) \leq \varepsilon$ to $\alpha_n(S_n, Q_n) = \varepsilon$, whereas the type II error probability can only decrease under this modification [16]. Then by using of the lemma C1 (Meta-converse) again, we find that

$$\begin{aligned} nD^A(\mathcal{N}||\mathcal{M}) &\geq D(p||q) \\ &= (1-p) \log \frac{1-p}{1-q} + p \log \frac{p}{q} \\ &= \alpha_n(S_n, Q_n) \log \frac{\alpha_n(S_n, Q_n)}{1 - \beta_n(S_n, Q_n)} \\ &\quad + (1 - \alpha_n(S_n, Q_n)) \log \frac{1 - \alpha_n(S_n, Q_n)}{\beta_n(S_n, Q_n)} \\ &\geq -h_2(\varepsilon) - (1 - \varepsilon) \log \beta_n(S_n, Q_n). \end{aligned} \quad (50)$$

By arranging the above equation, we have

$$\frac{nD^A(\mathcal{N}||\mathcal{M}) + h_2(\varepsilon)}{\varepsilon - 1} \leq \log \beta_n(S_n, Q_n) \leq \log \delta, \quad (51)$$

from which we follow that $n^*(\delta, \varepsilon, \mathcal{N}, \mathcal{M}, n) \geq \frac{(\varepsilon - 1) \log \delta - h_2(\varepsilon)}{D^A(\mathcal{N}||\mathcal{M})}$. \square

The analogous results can be given for the sampling complexity of the error exponent case.

Theorem 7. Let $\varepsilon, r, \mathcal{N}, \mathcal{M}$ be as state in definition 6 and we choose the adaptive strategy. Then for all $s \in (0, 1)$,

(1) if $r > D_s^A(\mathcal{N}||\mathcal{M})$, there is no scheme that can distinguish two channels \mathcal{N} and \mathcal{M} .

(2) if $r \leq D_s^A(\mathcal{N}||\mathcal{M})$, we have

$$n^*(\varepsilon, r, \mathcal{N}, \mathcal{M}, n) \leq \left\lceil \inf_{s \in (0,1)} \frac{\left(\frac{s}{1-s}\right) \log \varepsilon}{r - D_s^A(\mathcal{N}||\mathcal{M})} \right\rceil. \quad (52)$$

Proof. The proof follows from the same line with theorem 11. What needs to be slightly modified is to use lemma C1. \square

Remark 4. We should note that, in [16], it is found that for classical-quantum channel discrimination, asymptotically the exponential error rate is not improved by adaptive strategy. This is because the asymptotical rate is characterized by the amortized quantum relative entropy of quantum channels which collapses into quantum relative entropy of states for such a case. So in the following examples, we will study the difference of channel divergences for two interesting classes of channels, classical-quantum channel and generalized amplitude damping channel to show how discrimination strategies affect our framework of sampling complexity.

4. Examples

In this section, we detail two examples to illustrate how to evaluate the channel divergences that appear in the bounds in previous theorems.

Example 1. Firstly, we consider classical-quantum channels that represent as

$$\mathcal{N}_{X \rightarrow B} = \sum_x \langle x | \cdot | x \rangle \nu_B^x,$$

$$\mathcal{M}_{X \rightarrow B} = \sum_x \langle x | \cdot | x \rangle \mu_B^x,$$

where $\{|x\rangle\}$ is an orthonormal basis and $\{\nu_B^x\}$ and $\{\mu_B^x\}$ are sets of states. Due to the special structure of such channels, it is shown that [16]

$$D_s^A(\mathcal{N}||\mathcal{M}) = \max_x D_s(\nu_B^x || \mu_B^x), \quad \alpha \in [0, 2], \quad (53)$$

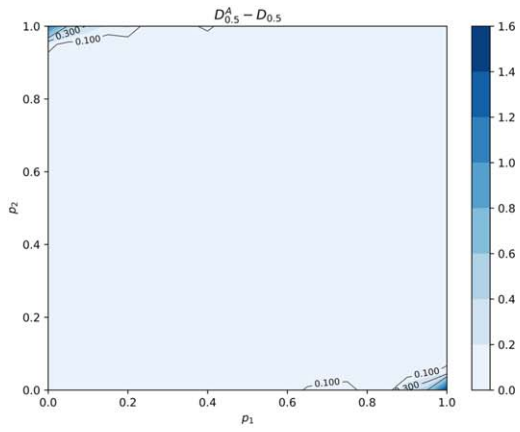
and

$$\tilde{D}_s^A(\mathcal{N}||\mathcal{M}) = \max_x \tilde{D}_s(\nu_B^x || \mu_B^x), \quad \alpha \geq \frac{1}{2}. \quad (54)$$

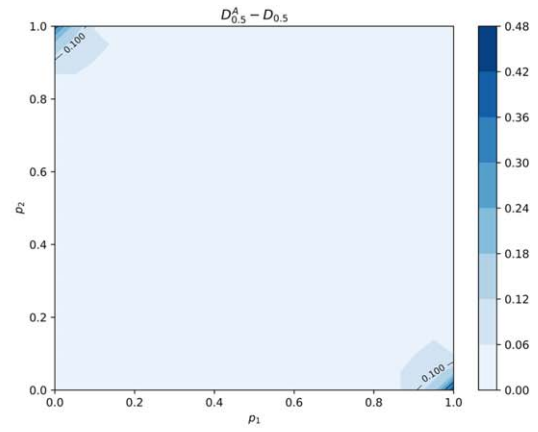
Combining with the above theorems, (53) and (54) imply that the sampling complexity n falls within the same interval for three hypothesis testing settings. This observation shows that for classical-quantum channels, the adaptive strategy does not lead to an advantage in determining the sampling complexity.

Example 2. The generalized amplitude damping channel with parameters (η, p) is a qubit channel. A set of Kraus operators for it is defined as:

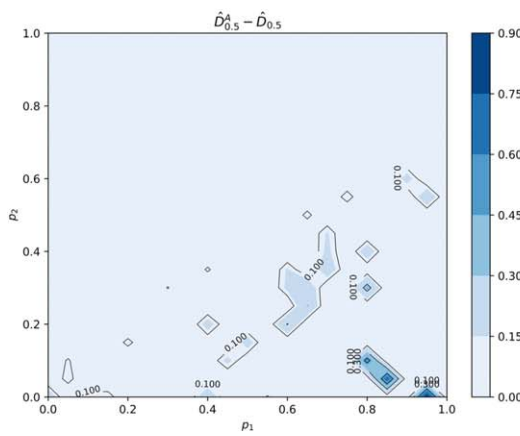
$$A_1 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{\eta} \end{pmatrix}, A_2 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{1-\eta} \\ 0 & 0 \end{pmatrix}, \quad (55)$$



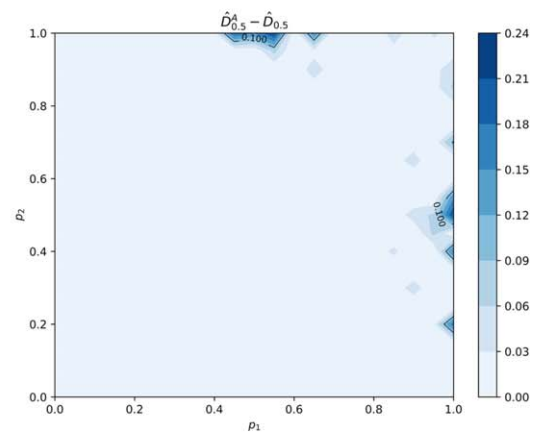
(a) $\eta_1 = 0.1, \eta_2 = 0.5$



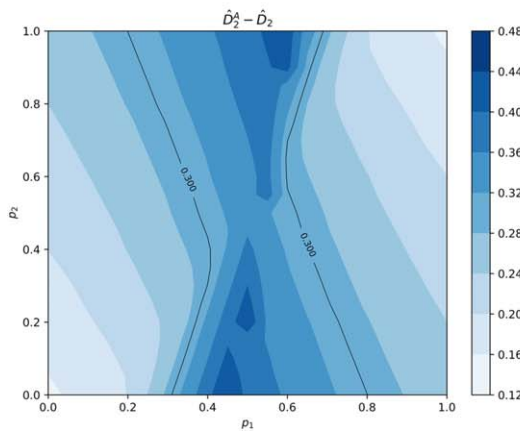
(b) $\eta_1 = \eta_2 = 0.5$



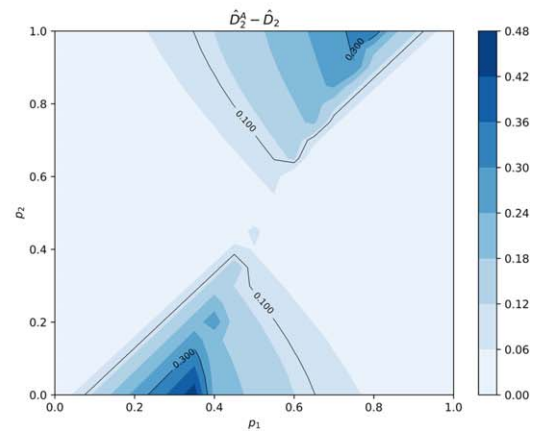
(c) $\eta_1 = 0.1, \eta_2 = 0.5$



(d) $\eta_1 = \eta_2 = 0.5$



(e) $\eta_1 = 0.9, \eta_2 = 0.7$



(f) $\eta_1 = \eta_2 = 0.6$

Figure 1. (a) and (b) show the difference between $D_{\frac{1}{2}}^A(\mathcal{N}||\mathcal{M})$ and $D_{\frac{1}{2}}(\mathcal{N}||\mathcal{M})$. (c) and (d) show the difference between $\hat{D}_{\frac{1}{2}}^A(\mathcal{N}||\mathcal{M})$ and $\hat{D}_{\frac{1}{2}}(\mathcal{N}||\mathcal{M})$. (e) and (f) show the difference between $\hat{D}_2^A(\mathcal{N}||\mathcal{M})$ and $\hat{D}_2(\mathcal{N}||\mathcal{M})$. We vary the parameters p_1 and p_2 and fix the parameters η .

$$\begin{aligned} A_3 &= \sqrt{1-p} \begin{pmatrix} \sqrt{\eta} & 0 \\ 0 & 1 \end{pmatrix}, \\ A_4 &= \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{1-\eta} & 0 \end{pmatrix}, \end{aligned} \quad (56)$$

where $p \in [0, 1]$ is the dissipation to the environment and $\eta \in [0, 1]$ is related to how much the input qubit mixes with the environment qubit. Let \mathcal{N} be the generalized amplitude damping channel with parameters (η_1, p_1) , and let \mathcal{M} be the generalized amplitude damping channel with parameters (η_2, p_2) . Very differently from Example 1, it is impossible to evaluate $D_s^A(\mathcal{N}||\mathcal{M})$ and $D^A(\mathcal{N}||\mathcal{M})$ analytically for two generalized amplitude damping channels due to multiple parameters that appear during the optimization process. Now we carry out the corresponding numerical simulations for them in Python qiskit package. We adopt the approach proposed in [34]. Since the generalized amplitude damping channel is covariant with respect to I and Z , it suffices to restrict the optimization to input state as $|\psi(z)\rangle_{RA} = \sqrt{z}|00\rangle + \sqrt{1-z}|11\rangle$. Thus we find that

$$\begin{aligned} D_s^A(\mathcal{N}||\mathcal{M}) &= \max_{z_1, z_2 \in [0, 1]} [D_s(\mathcal{N}(|\psi(z_1)\rangle_{RA}) \\ &\quad ||\mathcal{M}(|\psi(z_2)\rangle_{RA})) - D_s(|\psi(z_1)\rangle_{RA} || |\psi(z_2)\rangle_{RA})], \end{aligned} \quad (57)$$

and

$$\begin{aligned} D_s(\mathcal{N}||\mathcal{M}) &= \max_{z \in [0, 1]} D_s(\mathcal{N}(|\psi(z)\rangle_{RA}) || \mathcal{M}(|\psi(z)\rangle_{RA})). \end{aligned} \quad (58)$$

The differences between the various amortized channel divergence and generalized channel divergence are shown in figure 1.

The gap between the amortized channel divergences and the generalized channel divergences displays the advantage of the adaptive strategy over product strategy in determining the sampling complexity for discrimination the generalized amplitude damping channels.

5. Conclusion

This paper is to clarify the problem of the sampling complexity of channel discrimination. That is to say, what is the smallest number of samples needed to ensure that the distinguisher can figure out the unknown channels with a fixed error probability. We have introduced various versions of channel divergences and the definitions of sample complexity of symmetric setting, asymmetric setting, and error exponent setting. As our main results, we have presented the lower and upper bounds of the sampling complexity of channel discrimination. For the product strategy, we have found that these bounds are always characterized in terms of generalized channel divergences. For the adaptive strategy, the bounds are characterized in terms of the amortized channel divergences. Finally, we demonstrate our results by two important classes of channels. On the one hand, for classical-quantum channels, the adaptive strategy can not take advantage over the product strategy in the setting of determining sampling complexity by using of the previous results [16]. On the other hand, for the generalized amplitude

damping channels, we conclude that the adaptive strategy does take advantage over the product strategy in the setting of determining the sampling complexity by numerical simulation. We believe that our results confirm the advantage of adaptive strategy and separation from the product strategy for some particular channels.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant Nos. 24CX03003A, 23CX03011A, 22CX03005A), the Natural Science Foundation of Shandong Province (Grant Nos. ZR2023MA025, ZR2021LL2002), and from the NSFC (Grant Nos. 11775306, 11571220).

Appendix A Properties of the channel fidelity

Lemma A.1 *Let $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ be quantum channels. Then we can restrict the infimum to pure states $\varphi_{RA} \in \mathcal{S}(RA)$ in definition 3, i.e.,*

$$\begin{aligned} F(\mathcal{N}, \mathcal{M}) &= \inf_{\varphi_{RA} \in \mathcal{S}(RA)} F(\mathcal{N}_{A \rightarrow B}(\varphi_{RA}), \mathcal{M}_{A \rightarrow B}(\varphi_{RA})). \end{aligned} \quad (A1)$$

Proof. Suppose that $\rho_{RA} = \sum_i p_i |\varphi_i\rangle_{RA} \langle \varphi_i|$ is an arbitrary ensemble decomposition of mixed state ρ_{RA} . We find that

$$\begin{aligned} &F(\mathcal{N}_{A \rightarrow B}(\rho_{RA}), \mathcal{M}_{A \rightarrow B}(\rho_{RA})) \\ &= F(\mathcal{N}_{A \rightarrow B}(\sum_i p_i |\varphi_i\rangle_{RA} \langle \varphi_i|), \\ &\quad \mathcal{M}_{A \rightarrow B}(\sum_i p_i |\varphi_i\rangle_{RA} \langle \varphi_i|)) \\ &= F(\sum_i p_i \mathcal{N}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|), \\ &\quad \sum_i p_i \mathcal{M}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|)) \\ &\geq \sum_i p_i F(\mathcal{N}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|), \\ &\quad \mathcal{M}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|)) \\ &\geq \sum_i p_i \min_i F(\mathcal{N}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|), \\ &\quad \mathcal{M}_{A \rightarrow B}(|\varphi_i\rangle_{RA} \langle \varphi_i|)) \\ &= F(\mathcal{N}_{A \rightarrow B}(\varphi_{RA}), \mathcal{M}_{A \rightarrow B}(\varphi_{RA})), \end{aligned} \quad (A2)$$

where φ_{RA} is the optimal pure state in the optimization (A1).

The second equality follows from the linear property of quantum channel. The first inequality follows from the joint concavity of fidelity.

Lemma A.2 *(Data processing) Let $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ be quantum channels. Let Θ be a superchannel as described in (5). Then the following inequality holds*

$$F(\mathcal{N}, \mathcal{M}) \leq F(\Theta(\mathcal{N}), \Theta(\mathcal{M})). \quad (A3)$$

Proof. Let $\mathcal{F}_{C \rightarrow D} = \Theta(\mathcal{N})$ and $\mathcal{G}_{C \rightarrow D} = \Theta(\mathcal{M})$ be the channels that are output from superchannel Θ . Then we have

$$\begin{aligned} & F(\mathcal{F}(\varphi_{EC}), \mathcal{G}(\varphi_{EC})) \\ &= F(\mathcal{V}_{BR \rightarrow D} \circ (\mathcal{N}_{A \rightarrow B} \otimes \mathcal{I}_R) \circ \mathcal{U}_{C \rightarrow AR}(\varphi_{EC}), \\ & \quad \mathcal{V}_{BR \rightarrow D} \circ (\mathcal{M}_{A \rightarrow B} \otimes \mathcal{I}_R) \circ \mathcal{U}_{C \rightarrow AR}(\varphi_{EC})) \\ &= F(\mathcal{V}_{BR \rightarrow D} \circ (\mathcal{N}_{A \rightarrow B} \otimes \mathcal{I}_R)(\eta_{ARE}), \\ & \quad \mathcal{V}_{BR \rightarrow D} \circ (\mathcal{M}_{A \rightarrow B} \otimes \mathcal{I}_R)(\eta_{ARE})) \\ &\geq F((\mathcal{N}_{A \rightarrow B} \otimes \mathcal{I}_R)(\eta_{ARE}), (\mathcal{M}_{A \rightarrow B} \otimes \mathcal{I}_R)(\eta_{ARE})) \\ &\geq F(\mathcal{N}, \mathcal{M}), \end{aligned} \quad (\text{A4})$$

where $\eta_{ARE} = \mathcal{U}_{C \rightarrow AR}(\varphi_{EC})$. The first inequality follows from the data processing property of the state fidelity. The second inequality follows from the fact that η_{ARE} is a particular state, but the fidelity of the channel involves an optimization over all input state.

Lemma A.3 (Joint concavity) Let $\mathcal{N}_{A \rightarrow B}^i$ and $\mathcal{M}_{A \rightarrow B}^i$ be the quantum channels for all $i \in \mathcal{I}$, and let p_i be a probability distribution over the same index set. Then we have

$$F\left(\sum_i p_i \mathcal{N}^i, \sum_i p_i \mathcal{M}^i\right) \geq \sum_i p_i F(\mathcal{N}^i, \mathcal{M}^i). \quad (\text{A5})$$

Proof. Let $\bar{\mathcal{N}} = \sum_i p_i \mathcal{N}^i$ and $\bar{\mathcal{M}} = \sum_i p_i \mathcal{M}^i$. Next, let φ be the optimal state for the definition $F(\bar{\mathcal{N}}, \bar{\mathcal{M}})$, that is $F(\bar{\mathcal{N}}, \bar{\mathcal{M}}) = F(\bar{\mathcal{N}}(\varphi), \bar{\mathcal{M}}(\varphi))$. Then we have

$$\begin{aligned} F(\bar{\mathcal{N}}, \bar{\mathcal{M}}) &= F(\bar{\mathcal{N}}(\varphi), \bar{\mathcal{M}}(\varphi)) \\ &= F\left(\sum_i p_i \mathcal{N}^i(\varphi), \sum_i p_i \mathcal{M}^i(\varphi)\right) \\ &\geq \sum_i p_i F(\mathcal{N}^i(\varphi), \mathcal{M}^i(\varphi)) \\ &\geq \sum_i p_i F(\mathcal{N}^i, \mathcal{M}^i). \end{aligned} \quad (\text{A6})$$

The first inequality follows from the concavity of state fidelity. The second inequality follows from the definition of the channel fidelity, where we should take infimum over all pure state.

Appendix B

Lemma B.1 (Generalized Fuchs–van de Graaf inequality) For positive semi-definite operator A and B , it holds that [38, 39]

$$\|A - B\|_F^2 + 4\|\sqrt{A}\sqrt{B}\|_1^2 \leq (\text{Tr}(A + B))^2 \quad (\text{A7})$$

Lemma B.2 (Quantum Chernoff bound) For positive semi-definite operator A and B , it holds that [40]

$$\frac{1}{2}[\text{Tr}(A + B) - \|A - B\|_1] \leq Q_{\min}(A\|B), \quad (\text{A8})$$

where $Q_{\min}(A\|B) = \min_{\alpha \in [0,1]} \text{Tr}[A^\alpha B^{1-\alpha}]$.

Definition B.3 (Chernoff divergence) For quantum states ρ and σ , the Chernoff divergence is defined as [40]

$$C(\rho\|\sigma) = -\log Q_{\min}(\rho\|\sigma). \quad (\text{A9})$$

Lemma B.4 Let ρ and σ be quantum states. Then for $\alpha \in (1, \infty)$ and $\varepsilon \in (0, 1)$, it holds that [41]

$$-\log \beta_1 \leq \tilde{D}_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha - 1} \log \frac{1}{1 - \varepsilon} \quad (\text{A10})$$

under the constrain $\alpha_1 \leq \varepsilon$.

Lemma B.5 Let ρ and σ be quantum states. Then for $\alpha \in (0, 1)$ and $\varepsilon \in (0, 1)$, it holds that [42]

$$-\log \beta_1 \geq D_\alpha(\rho\|\sigma) + \frac{\alpha}{\alpha - 1} \log \frac{1}{\varepsilon} \quad (\text{A11})$$

under the constrain $\alpha_1 \leq \varepsilon$.

Lemma B.6 [40] Let L and K be positive semi-definite and $s \in (0, 1)$. Then,

$$\begin{aligned} \text{Tr}[L^s K^{1-s}] &\geq \text{Tr}[K\{L \geq K\}] \\ &+ \text{Tr}[L\{L < K\}], \end{aligned} \quad (\text{A12})$$

where $\{L \geq K\} = I - \{L < K\}$ is the projector onto the subspace spanned by eigenvectors corresponding to non-negative eigenvalues of $L - K$.

Appendix C

We define amortized channel divergence as a measure of the distinguishability of two quantum channels, in the sense that, in contrast to the generalized channel divergence defined above, we consider two different states ρ_{RA} and ρ_{RA} that can be input to the channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$, in order to explore the largest distinguishability that can be realized between the channels. However, it is sensible to subtract off the initial distinguishability of the states ρ_{RA} and ρ_{RA} from the final distinguishability of the channel output states $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ from a resource-theoretic perspective, since these initial states themselves could have some distinguishability. The amortized channel divergence has many attractive properties [16], such as data-processing inequality, joint convexity, and so on. In particular, the following lemma named as meta-converse, which was initially used to recover particular converse statements for quantum channel discrimination, plays a central role in our proof of the sample complexity with adaptive strategy.

Lemma C.1 (Meta-converse) [16] Let $p = \text{Tr}\{Q_n \rho_n(S_n)\}$ and $q = \text{Tr}\{Q_n \sigma_n(S_n)\}$ be the final decision probabilities corresponding to measurement Q_n . Then for adaptive channel discrimination as introduced in section 2.2, and any faithful general divergence D , we have

$$D(p\|q) \leq D(\rho_n(S_n)\|\sigma_n(S_n)) \leq nD^A(\mathcal{N}\|\mathcal{M}). \quad (\text{A13})$$

References

- [1] Hayashi M 2006 *Quantum Information: An Introduction* (Berlin: Springer)
- [2] Helstrom C W 1967 Detection theory and quantum mechanics *Inf. Control* **10** 254

- [3] Helstrom C W 1969 Quantum detection and estimation theory *J. Stat. Phys.* **1** 231
- [4] Hiai F and Petz D 1991 The proper formula for relative entropy and its asymptotics in quantum probability *Commun. Math. Phys.* **143** 99
- [5] Ogawa T and Nagaoka H 2000 Strong converse and Stein lemma in quantum hypothesis testing *IEEE Trans. Inf. Theory* **46** 2428
- [6] Cooney T, Mosonyi M and Wilde M M 2016 Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication *Commun. Math. Phys.* **344** 797–829
- [7] Wang L and Renner R 2012 One-shot Classical-quantum capacity and hypothesis testing *Phys. Rev. Lett.* **108** 200501
- [8] Bae J and Kwek L-C 2015 Quantum state discrimination and its applications *J. Phys. A: Math. Theor.* **48** 083001
- [9] Hayashi M 2002 Optimal sequence of quantum measurements in the sense of Steins lemma in quantum hypothesis testing *J. Phys. A: Math. Gen.* **35** 10759
- [10] Nagaoka H and Hayashi M 2007 An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses *IEEE Trans. Inf. Theory* **53** 534–49
- [11] Hayashi M 2009 Discrimination of two channels by adaptive methods and its application to quantum system *IEEE Trans. Inf. Theory* **55** 3807–20
- [12] Piani M and Watrous J 2009 All entangled states are useful for channel discrimination *Phys. Rev. Lett.* **102** 250501
- [13] Yuan H and Fung C F 2017 Fidelity and Fisher information on quantum channels *New J. Phys.* **19** 113039
- [14] Wang X and Wilde M M 2019 Resource theory of asymmetric distinguishability for quantum channels *Phys. Rev. Res.* **1** 033169
- [15] Fang K, Fawzi O, Renner R and Sutter D 2020 Chain rule for the quantum relative entropy *Phys. Rev. Lett.* **124** 100501
- [16] Wilde M M, Berta M, Hirche C and Kaur E 2020 Amortized channel divergence for asymptotic quantum channel discrimination *Lett. Math. Phys.* **110** 2277–336
- [17] Fang K, Gour G and Wang X 2022 Towards the ultimate limits of quantum channel discrimination arXiv:2110.14842
- [18] Harrow A W, Hassidim A, Leung D W and Watrous J 2010 Adaptive versus nonadaptive strategies for quantum channel discrimination *Phys. Rev. A* **81** 032339
- [19] Krawiec A, Pawela L and Pucha Z 2020 Discrimination of POVMs with rank-one effects *Quantum Inf. Process.* **19** 428
- [20] Salek F, Hayashi M and Winter A 2022 Usefulness of adaptive strategies in asymptotic quantum channel discrimination *Phys. Rev. A* **105** 022419
- [21] Duan R, Feng Y and Ying M 2009 Perfect distinguishability of quantum operations *Phys. Rev. Lett.* **103** 210501
- [22] Puzzioli D and Watrous J 2017 Ancilla dimension in quantum channel discrimination *Ann. Henri Poincaré* **18** 1153
- [23] Li K 2016 Discriminating quantum states: the multiple Chernoff distance *Annal. Stat.* **44** 1661
- [24] Tomamichel M and Hayashi M 2013 A hierarchy of information quantities for finite block length analysis of quantum tasks *IEEE Trans. Inf. Theory* **59** 7693
- [25] Arunachalam S and DeWolf R 2018 Optimal quantum sample complexity of learning algorithms *J. Mach. Learn. Res.* <https://jmlr.org/papers/volume19/18-195/18-195.pdf> **19** 1–36
- [26] Canonne C L 2022 Topics and techniques in distribution testing: A biased but representative sample *Found. Trends Comm. Inf. Theory* **19** 1032
- [27] Cheng H, Datta N, Liu N, Nuradha T, Salzman R and Wilde M M 2024 Sample complexity of quantum hypothesis testing arXiv:2403.17868
- [28] Chiribella G, Mauro D, Ariano G and Perinotti P 2008 Transforming quantum operations: quantum supermaps *Europhys. Lett.* **83** 30004
- [29] Petz D 1985 Quasi-entropies for states of a von Neumann algebra *Publ. Res. Instit. Math. Sci.* **21** 787–800
- [30] Petz D 1986 Quasi-entropies for finite quantum systems *Rep. Math. Phys.* **23** 57–65
- [31] Müller-Lennert M, Dupuis F, Szehr O, Fehr S and Tomamichel M 2013 On quantum Renyi entropies: A new generalization and some properties *J. Math. Phys.* **54** 122203
- [32] Wilde M M, Winter A and Yang D 2014 Strong converse exponents for the feedback-assisted classical capacity of entanglement-breaking channels *Commun. Math. Phys.* **331** 593 622
- [33] Müller-Lennert M, Dupuis F, Szehr O, Fehr S and Tomamichel M 2013 On quantum renyi entropies: a new generalization and some properties *J. Math. Phys.* **54** 122203
- [34] Leditzky F, Kaur E, Datta N and Wilde M M 2018 Approaches for approximate additivity of the holevo information of quantum channels *Phys. Rev. A* **97** 012332
- [35] Audenaert K M R, Nussbaum M, Szkola A and Verstraete F 2008 Asymptotic error rates in quantum hypothesis testing *Commun. Math. Phys.* **279** 251
- [36] Hayashi M 2007 Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding *Phys. Rev. A* **76** 062301
- [37] Nagaoka H 2006 The converse part of the theorem for quantum Hoeffding bound arXiv:quant-ph/0611289
- [38] Fuchs C and van de Graaf J 1999 Cryptographic distinguishability measures for quantum mechanical states *IEEE Trans. Inf. Theory* **45** 1216
- [39] Audenaert K M 2014 Comparisons between quantum state distinguishability measures *Quantum Inf. Comput.* **14** 31
- [40] Audenaert K M R, Calsamiglia J, Muñoz-Tapia R, Bagan E, Masanes L, Acin A and Verstraete F 2007 Discriminating states: the quantum Chernoff bound *Phys. Rev. Lett.* **98** 160501
- [41] Mosonyi M and Ogawa T 2014 Quantum hypothesis testing and the operational interpretation of the quantum Renyi relative entropies *Commun. Math. Phys.* **334** 1617
- [42] Audenaert K M R, Nussbaum M, Szkola A and Verstraete F 2008 Asymptotic error rates in quantum hypothesis testing *Commun. Math. Phys.* **279** 251
- [43] Hayashi M and Tomamichel M 2016 Correlation Detection and an Operational Interpretation of the Rényi Mutual Information *J. Math. Phys.* **57** 102201