Letters in Mathematical Physics (2024) 114:76
https://doi.org/10.1007/s11005-024-01821-z

On the optimal error exponents for classical and quantum antidistinguishability

Hemant K. Mishra[1] · Michael Nussbaum[2] · Mark M. Wilde[1]

Received: 7 September 2023 / Revised: 8 May 2024 / Accepted: 13 May 2024 /
Published online: 5 June 2024

Abstract

The concept of antidistinguishability of quantum states has been studied to investigate foundational questions in quantum mechanics. It is also called quantum state elimination, because the goal of such a protocol is to guess which state, among finitely many chosen at random, the system is not prepared in (that is, it can be thought of as the first step in a process of elimination). Antidistinguishability has been used to investigate the reality of quantum states, ruling out ψ-epistemic ontological models of quantum mechanics (Pusey et al. in Nat Phys 8(6):475–478, 2012). Thus, due to the established importance of antidistinguishability in quantum mechanics, exploring it further is warranted. In this paper, we provide a comprehensive study of the optimal error exponent—the rate at which the optimal error probability vanishes to zero asymptotically—for classical and quantum antidistinguishability. We derive an exact expression for the optimal error exponent in the classical case and show that it is given by the multivariate classical Chernoff divergence. Our work thus provides this divergence with a meaningful operational interpretation as the optimal error exponent for antidistinguishing a set of probability measures. For the quantum case, we provide several bounds on the optimal error exponent: a lower bound given by the best pairwise Chernoff divergence of the states, a single-letter semi-definite programming upper bound, and lower and upper bounds in terms of minimal and maximal multivari-

Dedicated to the memory of Mary Beth Ruskai. She was an important foundational figure in the field of quantum information, and her numerous seminal research contributions and reviews, including [25, 37, 49], have inspired many quantum information scientists.

✉ Hemant K. Mishra
hemant.mishra@cornell.edu

Michael Nussbaum
nussbaum@math.cornell.edu

Mark M. Wilde
wilde@cornell.edu

1 School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14850, USA

2 Department of Mathematics, Cornell University, Ithaca, NY 14850, USA

ate quantum Chernoff divergences. It remains an open problem to obtain an explicit expression for the optimal error exponent for quantum antidistinguishability.

Keywords Antidistinguishability · Multivariate Chernoff divergence · Hellinger transform · Asymptotic error exponent · Extended max-relative entropy

Mathematics Subject Classification 81P45 · 62H15 · 62B10 · 81P18 · 62H05 · 62H15

Contents

1 Introduction

Quantum state discrimination is a fundamental component of quantum information science, playing a key role in quantum computing [4], quantum communication [3, 20], and quantum key distribution [8]. The state discrimination or distinguishability task is to infer the actual state of a quantum system by applying a quantum measurement to the system. More formally, consider a quantum system prepared in one of the quantum states $\rho_1, \ldots, \rho_r$. A quantum measurement is specified by a positive operator-valued measure $\{M_1, \ldots, M_r\}$ with output i indicating ρ_i as the true state of the system with success probability $\mathrm{Tr}[M_i \rho_i]$, as given by the Born rule [10].

The task that we consider here is in a sense opposite to the aforementioned task of distinguishability, and it is thus called *antidistinguishability* of quantum states or *quantum state elimination* [5, 7, 12, 19, 31, 33, 46, 47]. In particular, for the task of antidistinguishability, we are interested in designing a measurement whose outcome corresponds to a state that is not the actual state of the quantum system. In the classical version of the antidistinguishability problem, quantum states are replaced by probability measures on a measurable space, and the task is to rule out one of the probability measures upon observing i.i.d. (independent and identically distributed) data that is not produced by the probability measure.

As an illustrative example in the classical case, suppose that one of three possible dice is tossed, a red one with probability distribution p_R, a green one with probability distribution p_G, or a blue one with probability distribution p_B. The task is then, after observing a sample, to output "not red" if the green or blue die is tossed, "not green" if the red or blue die is tossed, and "not blue" if the red or green die is tossed. It is also of interest to consider the antidistinguishability task when the same colored die is tossed multiple times, leading to several samples that one can use to arrive at a conclusion.

To the best of our knowledge, an analysis of the asymptotics of the error probability of antidistinguishability is missing in the literature for both cases, classical as well as quantum, and it is this scenario that we consider in our paper.

1.1 Contributions

In this paper, we provide a comprehensive study of the optimal error exponent—the rate at which the optimal error probability vanishes to zero asymptotically—for classical and quantum antidistinguishability. Our contributions are as follows:

- We derive an exact expression for the optimal error exponent in the classical case and show that it is given by the multivariate classical Chernoff divergence (Theorem 6). Our work thus provides this multivariate divergence with a meaningful operational interpretation as the optimal error exponent for antidistinguishing a set of probability measures.
- We provide several bounds on the optimal error exponent in the quantum case:
    - lower bound given by the best pairwise Chernoff divergence of the states (Theorem 11),
    - lower and upper bounds in terms of minimal and maximal multivariate quantum Chernoff divergences (Theorem 17), and
    - single-letter semi-definite programming upper bound (Theorem 19).
- We also provide an upper bound on the optimal error probability of antidistinguishing an ensemble of quantum states in terms of the pairwise optimal error probabilities of the states, and consequently, we deduce that the given quantum states are perfectly antidistinguishable if at least two of them are orthogonal to each other (Theorem 8).

- As a contribution of independent interest and auxiliary to Theorem 19, we establish several fundamental properties of the extended max-relative entropy, a quantity of interest originally defined in [62].

It remains an intriguing open problem to determine an explicit expression for the optimal error exponent in the quantum case.

1.2 Literature review

Let us briefly review some prior contributions to the topic of antidistinguishability. We note here that quantum state discrimination is equivalent to finding a size-$(r-1)$ subset of $\{\rho_1, \ldots, \rho_r\}$ such that none of the states in the subset is the true state of the system; thus, the task is equivalent to what is called *quantum* $(r-1)$*-state exclusion*. A generalization of this task is *quantum* m*-state exclusion* for $1 \leq m \leq r-1$, which aims at detecting a size-m subset of $\{\rho_1, \ldots, \rho_r\}$ such that none of the states in the subset is the true state of the system [47]. Quantum 1-state exclusion is therefore the same as antidistinguishability of quantum states.

The concept of antidistinguishability has been studied to investigate foundational questions in quantum mechanics [5, 12, 33, 46]. For example, it was used in [46] to investigate the reality of quantum states, ruling out ψ-epistemic ontological models of quantum mechanics. It was also used in studying quantum communication complexity [19], in deriving noncontextuality inequalities [31], and has applications in quantum cryptography [11]. Thus, due to the established importance of antidistinguishability in quantum mechanics, exploring it further is warranted. There have been a number of works that determine algebraic conditions on a set of quantum states such that perfect antidistinguishability is possible. A sufficient condition for perfect antidistinguishability of pure states [22] is that if some positive linear combination of the pure states is a projection with a "special" kernel, then the states are *antidistinguishable*. In the same paper, a necessary and sufficient condition for antidistinguishability of pure states was given, which demands the existence of projections satisfying three non-trivial conditions. Very recently, a necessary and sufficient condition for non-antidistinguishability of general quantum states was given in [47], which also demands the existence of a Hermitian matrix with positive trace satisfying a set of non-trivial inequalities. Even though the conditions given in the aforementioned works are interesting and insightful, verifying them is not straightforward. One of the consequences of our work is that we provide a simple sufficient condition for perfect antidistinguishability of quantum states (Theorem 8).

1.3 Paper organization

The organization of our paper is as follows: In Sect. 2, we state some definitions and provide a brief mathematical background of relevant topics covered in our paper. We start Sect. 3 by building a theory of classical antidistinguishability, where we introduce the notions of optimal error probability and optimal error exponent. We then derive an explicit expression for the optimal error exponent in the classical case, and we show that it is given by the multivariate classical Chernoff divergence (Theorem 6).

The following sections deal with the optimal error exponent in the quantum case. We begin Sect. 4 by providing an upper bound on the optimal error probability of antidistinguishing an ensemble of quantum states in terms of the pairwise optimal error probabilities of the states (Theorem 8), and we then use this result to derive a lower bound on the optimal error exponent (Theorem 11). Next, we provide both lower and upper bounds on the optimal error exponent in terms of minimal and maximal multivariate quantum Chernoff divergences in Sect. 5 (Theorem 17). Lastly, in Sect. 6, we derive a single-letter semi-definite programming upper bound on the optimal error exponent (Theorem 19). Appendices A through I contain mathematical proofs of various claims made throughout the paper.

2 Mathematical background

2.1 Antidistinguishability of probability measures

Let $P_1, \ldots, P_r$ be probability measures on a measurable space $(\Omega, \mathcal{A})$, where $\mathcal{A}$ is a σ-algebra on the set Ω. Set $[r] \coloneqq \{1, \ldots, r\}$. Let $\eta_1, \ldots, \eta_r$ be strictly positive real numbers such that $\sum_{i\in[r]} \eta_i = 1$. Throughout the paper, we call

$$\mathcal{E}_{\mathrm{cl}} \coloneqq \{(\eta_i, P_i) : i \in [r]\} \tag{1}$$

an ensemble of probability measures on the measurable space $(\Omega, \mathcal{A})$. Let μ be the dominating measure

$$\mu \coloneqq \sum_{i\in[r]} \eta_i P_i, \tag{2}$$

and $p_1, \ldots, p_r$ the induced densities

$$p_i \coloneqq \frac{\mathbf{d}P_i}{\mathbf{d}\mu}, \qquad i \in [r], \tag{3}$$

which are given by the Radon–Nikodým theorem [6].

The problem of distinguishability, i.e., identifying the correct probability density p_i based on i.i.d. (independent and identically distributed) data, has been well studied. This problem is as follows: Suppose that i is sampled with probability η_i, and then, n i.i.d. samples are selected according to the product measure $P_i^{\otimes n}$. The task is to identify the correct value of i based on the n i.i.d. samples observed. It is known that the maximum likelihood method for the identification task is optimal, and the optimal success probability, in the case that $n = 1$, is given by

$$\int \mathbf{d}\mu \; (\eta_1 p_1 \vee \cdots \vee \eta_r p_r) \coloneqq \int \mathbf{d}\mu(\omega) \; \max\{\eta_1 p_1(\omega), \ldots, \eta_r p_r(\omega)\}. \tag{4}$$

Asymptotically, the optimal error probability vanishes to zero exponentially, and the error exponent is known to be equal to the Chernoff divergence for the least favorable pair (p_i, p_j), for $i \neq j$ [35, 50–52, 57].

For the antidistinguishability problem in the classical case, the task is to guess a probability density that is not represented by the observed data. For this problem, no literature is available to the best of our knowledge. A reasonable first idea for selecting a density that is unlikely to be the true one is to choose the one such that $\eta_i p_i(\omega)$ is minimum if ω is observed. This corresponds to a *minimum likelihood principle*. In what follows, we discuss this idea more formally.

A deterministic decision rule for the antidistinguishability problem is a function

$$\delta : \Omega \to \{\mathbf{e}_i : i \in [r]\}, \tag{5}$$

where $\mathbf{e}_i$ is the ith standard unit vector in $\mathbb{R}^r$, such that $\delta(\omega) = \mathbf{e}_i$ means that we indicate p_i to be our guess for the density that is not the true one. More generally, we can admit a randomized decision rule, along the following lines:

$$\delta : \Omega \to [0, 1]^r, \qquad \sum_{i \in [r]} \delta_i(\omega) = 1. \tag{6}$$

If p_i is the true density, then the antidistinguishability error probability is given by:

$$\int \mathbf{d}\mu(\omega)\ \delta_i(\omega) p_i(\omega), \tag{7}$$

and the total error probability is:

$$\mathrm{Err}_{\mathrm{cl}}(\delta; \mathcal{E}_{\mathrm{cl}}) := \sum_{i \in [r]} \eta_i \int \mathbf{d}\mu(\omega)\ \delta_i(\omega) p_i(\omega) = \int \mathbf{d}\mu(\omega) \sum_{i \in [r]} \delta_i(\omega) \eta_i p_i(\omega). \tag{8}$$

To minimize the above expression, we can minimize the integrand for every ω. Since $\delta_i(\omega)$ is a weight, we should place maximum weight on the smallest of $\eta_i p_i(\omega)$. So, the optimal decision for given ω corresponds to the *minimum likelihood rule*: $\delta^*(\omega) = \mathbf{e}_i$, if $i \in [r]$ is the minimum index such that $\eta_i p_i(\omega) = \min\{\eta_1 p_1(\omega), \ldots, \eta_r p_r(\omega)\}$. The total error probability when using the decision rule δ^* is the optimal error probability, given by

$$\begin{aligned} \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}) := \mathrm{Err}_{\mathrm{cl}}(\delta^*; \mathcal{E}_{\mathrm{cl}}) &= \int \mathbf{d}\mu(\omega)\ \min\{\eta_1 p_1(\omega), \ldots, \eta_r p_r(\omega)\} \\ &= \int \mathbf{d}\mu\ \left(\eta_1 p_1 \wedge \cdots \wedge \eta_r p_r\right). \end{aligned} \tag{9}$$

In the asymptotic treatment of the problem, we consider the *n-fold ensemble* $\mathcal{E}_{\mathrm{cl}}^n := \{(\eta_i, P_i^{\otimes n}) : i \in [r]\}$ on the n-fold measurable space $(\Omega^n, \mathcal{A}^{(n)})$, where Ω^n is the n-fold Cartesian product of Ω and $\mathcal{A}^{(n)}$ is the σ-algebra on Ω^n generated by the n-fold

Cartesian product of $\mathcal{A}$. It then follows that the optimal antidistinguishability error probability, in this case, is

$$\mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) = \int \mathbf{d}\mu^{\otimes n} \ \left(\eta_1 p_1^{\otimes n} \wedge \cdots \wedge \eta_r p_r^{\otimes n}\right). \tag{10}$$

Set $\eta_{\min} := \min\{\eta_1, \ldots, \eta_r\}$ and $\eta_{\max} := \max\{\eta_1, \ldots, \eta_r\}$. It is easy to see that, for all $n \in \mathbb{N}$, we have

$$\eta_{\min} \int \mathbf{d}\mu^{\otimes n} \ \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right) \leq \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq \eta_{\max} \int \mathbf{d}\mu^{\otimes n} \ \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right). \tag{11}$$

This implies that

$$\liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) = \liminf_{n\to\infty} -\frac{1}{n} \ln \int \mathbf{d}\mu^{\otimes n} \ \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right), \tag{12}$$

which is independent of $\eta_1, \ldots, \eta_r$.

Definition 1 The optimal error exponent for antidistinguishing the probability measures of a given ensemble $\mathcal{E}_{\mathrm{cl}} = \{(\eta_i, P_i) : i \in [r]\}$ is defined by

$$\mathrm{E}_{\mathrm{cl}}(P_1, \ldots, P_r) := \liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n). \tag{13}$$

Remark 1 We note that Definition 1 of the optimal error exponent is independent of the choice of dominating measure μ. This is because the development in (4)–(12) is independent of the choice of the probability measure μ dominating $P_1, \ldots, P_r$. Indeed, if μ' is an arbitrary probability measure dominating $P_1, \ldots, P_r$, then μ' also dominates μ. Let $\nu := \frac{\mathbf{d}\mu}{\mathbf{d}\mu'}$. We then have

$$p_i' := \frac{\mathbf{d}P_i}{\mathbf{d}\mu'} = \frac{\mathbf{d}P_i}{\mathbf{d}\mu} \cdot \frac{\mathbf{d}\mu}{\mathbf{d}\mu'} = p_i \nu, \quad \text{for all } i \in [r]. \tag{14}$$

Consequently, the quantity in (4) is given by

$$\int \mathbf{d}\mu \ (\eta_1 p_1 \vee \cdots \vee \eta_r p_r) = \int \mathbf{d}\mu' \ (\eta_1 p_1 \vee \cdots \vee \eta_r p_r)\, \nu \tag{15}$$

$$= \int \mathbf{d}\mu' \ (\eta_1 p_1 \nu \vee \cdots \vee \eta_r p_r \nu) \tag{16}$$

$$= \int \mathbf{d}\mu' \ \left(\eta_1 p_1' \vee \cdots \vee \eta_r p_r'\right). \tag{17}$$

Similarly, the remaining quantities in (4)–(12) can be shown to be independent of the choice of μ. See [54, p. 233]. We will see later, in Theorem 6, that the limit inferior on the left-hand side of (13) is actually a limit.

2.2 Multivariate classical Chernoff divergence

The Hellinger transform, as it is known in the literature on probability and statistics, plays an important role in our work. The quantity seems to have been defined first in [38, page 189], and the term "Hellinger transform" was perhaps first used in [32], followed by several works in the area of probability and statistics. See [16, 18, 27, 36, 56–59] and references therein. See also [30, Section 3.3] for a historical discussion.

We recall the definition of the Hellinger transform below. Let $\mathcal{E}_{\mathrm{cl}} = \{(\eta_i, P_i) : i \in [r]\}$ be an ensemble of probability measures on a measurable space $(\Omega, \mathcal{A})$. Let μ be the dominating measure defined by (2), and let $p_1, \ldots, p_r$ be the induced probability densities given by (3). Let $\mathbb{S}_r$ denote the unit simplex in $\mathbb{R}^r$:

$$\mathbb{S}_r := \left\{ \mathbf{s} \in [0, 1]^r : \mathbf{s} = (s_1, \ldots, s_r), \sum_{i \in [r]} s_i = 1 \right\}. \tag{18}$$

Definition 2 The Hellinger transform of the probability measures $P_1, \ldots, P_r$ is a function on the unit simplex, defined as

$$\mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) := \int \mathbf{d}\mu \; p_1^{s_1} \cdots p_r^{s_r}, \qquad \text{for all } \mathbf{s} := (s_1, \ldots, s_r) \in \mathbb{S}_r. \tag{19}$$

Here we use the convention $0^0 = 0$.

Remark 2 Some authors use the convention $0^0 = 1$ when defining the Hellinger transform [56, Definition 5.10], while others define it only in the interior of the unit simplex [36, Definition 1.87, p. 49]. Note that our definition is in contrast to the former, and it is consistent with first defining $\mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r)$ on the interior of the unit simplex $\mathbb{S}_r$ (so that $s_i > 0$ for all $i \in [r]$) and then on the boundary as follows:

$$\mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) = \lim_{\varepsilon \searrow 0} \mathbb{H}_{(1-\varepsilon)\mathbf{s}+\varepsilon\mathbf{u}}(P_1, \ldots, P_r), \tag{20}$$

where $\mathbf{u} := (1/r, \ldots, 1/r) \in \mathbb{R}^r$.

Also, we emphasize that the Hellinger transform given in Definition 2, and hence, our further analysis of the classical antidistinguishability error probability is independent of the choice of the dominating measure μ. This easily follows by similar arguments as in Remark 1. See [54, Chapter 3, Section 9, Lemma 3] for a proof in the case of $r = 2$.

The Hellinger transform given in Definition 2 is continuous on $\mathbb{S}_r$. Indeed, we have $P_i \leq \eta_i^{-1}\mu$, implying that $p_i \leq \eta_i^{-1}$ for all $i \in [r]$. This gives $\prod_{i\in[r]}(p_i + 1)$ as an integrable upper bound on $\prod_{i\in[r]} p_i^{s_i}$ for all $(s_1, \ldots, s_r) \in \mathbb{S}_r$. Thus, for every $\mathbf{s} := (s_1, \ldots, s_r) \in \mathbb{S}_r$ and for every sequence $(\mathbf{s}^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{S}_r$ with $\mathbf{s}^{(n)} := (s_1^{(n)}, \ldots, s_r^{(n)})$ and $\lim_{n\to\infty} \mathbf{s}^{(n)} = \mathbf{s}$, we have

$$\lim_{n\to\infty} \prod_{i\in[r]} (p_i(\omega))^{s_i^{(n)}} = \prod_{i\in[r]} (p_i(\omega))^{s_i}, \qquad \text{for all } \omega \in \Omega. \tag{21}$$

Note that the existence of the limit in (21) is due to our convention $0^0 = 0$. By the Lebesgue-dominated convergence theorem, we then have $\lim_{n\to\infty} \mathbb{H}_{\mathbf{s}^{(n)}}(P_1, \ldots, P_r) = \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r)$, thereby proving continuity of the Hellinger transform on $\mathbb{S}_r$.

In general, the Hellinger transform is a measure of closeness or affinity among several probability distributions. It is easy to see that

$$0 \leq \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) \leq 1, \tag{22}$$

which follows from Hölder's inequality [56, Lemma 53.3]. As the value of $\mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r)$ gets close to zero, the distance among the measures increases in some sense [16].

The following quantity plays an important role in our paper.

Definition 3 We define the multivariate Chernoff divergence of the probability measures $P_1, \ldots, P_r$ by

$$\xi_{\mathrm{cl}}(P_1, \ldots, P_r) := -\ln \inf_{\mathbf{s}\in\mathbb{S}_r} \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r), \tag{23}$$

where $\mathbb{H}_{\mathbf{s}}$ is defined in (19).

The divergence can be viewed as a generalization of the classical Chernoff divergence, the latter being a special case of the former for $r = 2$ [13]. One of the main results of our paper is that the optimal error exponent for antidistinguishing an ensemble of probability measures is equal to their multivariate Chernoff divergence (Theorem 6).

2.3 Quantum states, channels, and measurements

A quantum system is associated with a complex Hilbert space $\mathcal{H}$. We focus exclusively on systems with finite-dimensional Hilbert spaces in this paper. Let $\dim(\mathcal{H})$ denote the dimension of $\mathcal{H}$. We denote every element of $\mathcal{H}$ using the *ket* notation as $|\psi\rangle$, $|\phi\rangle$, etc., and every element of its dual using the *bra* notation as $\langle\psi|$, $\langle\phi|$, etc. The notations go well with the natural action of a dual element $\langle\psi|$ on a vector $|\phi\rangle$ in terms of the inner product of the two vectors: $\langle\psi|(|\phi\rangle) = \langle\psi|\phi\rangle$.

A quantum state of a system is identified by a density operator ρ, which is a self-adjoint, positive semi-definite operator of unit trace acting on $\mathcal{H}$. A pure state is given by a state vector $|\psi\rangle \in \mathcal{H}$ whose corresponding density operator is $|\psi\rangle\langle\psi|$. The set of density operators forms a convex set with pure states as the extreme points. Let $\mathcal{D}(\mathcal{H})$ denote the set of density operators, $\mathcal{L}(\mathcal{H})$ the space of linear operators acting on $\mathcal{H}$, and $\mathcal{L}_+(\mathcal{H})$ the set of positive semi-definite operators acting on $\mathcal{H}$. We shall use the notation $\mathcal{D}$ for the set of density operators whenever the underlying Hilbert space is clear from the context. A quantum channel $\mathcal{N}$, between two quantum systems represented by Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, is a completely positive, trace-preserving linear map from $\mathcal{L}(\mathcal{H})$ to $\mathcal{L}(\mathcal{K})$. In particular, for all $\rho \in \mathcal{D}(\mathcal{H})$, we have that $\mathcal{N}(\rho) \in \mathcal{D}(\mathcal{K})$.

A quantum measurement is described by a positive operator-valued measure (POVM) $\mathscr{M} = \{M_1, \ldots, M_r\}$, which is a finite set of positive semi-definite operators whose sum is the identity operator, i.e.,

$$M_i \geq 0 \text{ for all } i \in [r], \qquad \sum_{i\in[r]} M_i = \mathbb{I}, \tag{24}$$

where $\mathbb{I}$ is the identity operator acting on $\mathcal{H}$.

The projection onto the support of an operator A is denoted by $\operatorname{supp}(A)$, its absolute value is denoted by $|A| := \sqrt{A^\dagger A}$, and its positive part by $A_+ := \frac{1}{2}(A + |A|)$. For two Hermitian operators A and B, we use the notation

$$A \wedge B := \frac{1}{2}(A + B - |A - B|), \tag{25}$$

in analogy with $\min(a, b) = \frac{1}{2}(a + b - |a - b|) \equiv a \wedge b$ for $a, b \in \mathbb{R}$.

2.4 Antidistinguishability of quantum states

Suppose that a quantum system is prepared in one of the quantum states $\rho_1, \ldots, \rho_r$ with *a priori* probability distribution $\eta_1, \ldots, \eta_r$ such that $\eta_i > 0$ for all $i \in [r]$. Throughout the paper, we call $\{(\eta_i, \rho_i) : i \in [r]\}$ an ensemble of quantum states over a Hilbert space $\mathcal{H}$ and denote it by $\mathcal{E}$. Antidistinguishability of the states, as realized by a POVM $\mathscr{M} = \{M_1, \ldots, M_r\}$, can be described as follows: "the measurement outcome i occurring corresponds to a guess that the true state of the system is not ρ_i." Thus, if ρ_i is the true state of the system, then $\operatorname{Tr}[M_i\rho_i]$ is the error probability for making an incorrect guess. The average error probability of antidistinguishability, for a fixed POVM $\mathscr{M}$, is then given by:

$$\mathrm{Err}(\mathscr{M}; \mathcal{E}) := \sum_{i\in[r]} \eta_i \operatorname{Tr}[M_i \rho_i]. \tag{26}$$

We are interested in determining the optimal antidistinguishability error probability, which is optimized over all possible measurements:

$$\mathrm{Err}(\mathcal{E}) := \inf_{\mathscr{M}} \mathrm{Err}(\mathscr{M}; \mathcal{E}), \tag{27}$$

where the infimum is taken over all POVMs of the form $\mathscr{M} = \{M_1, \ldots, M_r\}$ acting on $\mathcal{H}$.

The quantum states are said to be perfectly antidistinguishable if there exists a quantum measurement whose outputs always correspond to a false state of the system; i.e., there exists a POVM $\mathscr{M}$ such that $\mathrm{Err}(\mathscr{M}; \mathcal{E}) = 0$. In general, an ensemble of quantum states may not be antidistinguishable, which means, for such an ensemble $\mathcal{E}$, that $\mathrm{Err}(\mathscr{M}; \mathcal{E}) > 0$ for every POVM $\mathscr{M}$. For instance, two non-orthogonal quantum states are not perfectly antidistinguishable [33].

In the asymptotic treatment of the antidistinguishability problem for a given ensemble $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$, we consider the n-fold ensemble $\mathcal{E}^n := \{(\eta_i, \rho_i^{\otimes n}) : i \in [r]\}$. The optimal error probability of antidistinguishability for $\mathcal{E}^n$ is by definition given as

$$\mathrm{Err}(\mathcal{E}^n) = \inf_{\mathscr{M}^{(n)}} \sum_{i\in[r]} \eta_i \operatorname{Tr}\left[M_i^{(n)} \rho_i^{\otimes n}\right], \tag{28}$$

where the infimum is taken over the set of POVMs $\mathscr{M}^{(n)} = \{M_1^{(n)}, \ldots, M_r^{(n)}\}$ acting on the n-fold tensor product Hilbert space $\mathcal{H}^{\otimes n}$. Similar to what we discussed around (11), we find for all $n \in \mathbb{N}$ that

$$\eta_{\min} \inf_{\mathscr{M}^{(n)}} \sum_{i\in[r]} \operatorname{Tr}\left[M_i^{(n)} \rho_i^{\otimes n}\right] \leq \mathrm{Err}(\mathcal{E}^n) \leq \eta_{\max} \inf_{\mathscr{M}^{(n)}} \sum_{i\in[r]} \operatorname{Tr}\left[M_i^{(n)} \rho_i^{\otimes n}\right], \tag{29}$$

which implies that

$$\liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^n) = \liminf_{n\to\infty} -\frac{1}{n} \ln \inf_{\mathscr{M}^{(n)}} \sum_{i\in[r]} \operatorname{Tr}\left[M_i^{(n)} \rho_i^{\otimes n}\right], \tag{30}$$

the latter being independent of $\eta_1, \ldots, \eta_r$.

Definition 4 The optimal error exponent for antidistinguishing the states of a quantum ensemble $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$ is defined by

$$\mathrm{E}(\rho_1, \ldots, \rho_r) := \liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^n). \tag{31}$$

2.5 Quantum Chernoff divergence

Here we briefly recall known results for distinguishability of two or more states; a key quantity for this purpose is as follows:

Definition 5 The quantum Chernoff divergence between two states ρ_1 and ρ_2 is defined as:

$$\xi(\rho_1, \rho_2) := -\ln \inf_{s\in[0,1]} \mathrm{Tr}[\rho_1^s \rho_2^{1-s}]. \tag{32}$$

If $\rho_1 = |\psi\rangle\langle\psi|$ and $\rho_2 = |\phi\rangle\langle\phi|$ are pure states, then

$$\xi(\rho_1, \rho_2) = -\ln |\langle\psi|\phi\rangle|^2. \tag{33}$$

The quantum Chernoff divergence between two states is known to be the optimal error exponent in distinguishing them [1, 45], i.e.,

$$\lim_{n\to\infty} -\frac{1}{n} \ln \left(\mathrm{Tr}[\eta_1\rho_1^{\otimes n} \wedge \eta_2\rho_2^{\otimes n}]\right) = \lim_{n\to\infty} -\frac{1}{n} \ln \left(\frac{1}{2}\left[1 - \left\|\eta_1\rho_1^{\otimes n} - \eta_2\rho_2^{\otimes n}\right\|_1\right]\right) = \xi(\rho_1, \rho_2), \tag{34}$$

where we have used the well-known fact that the optimal error probability in distinguishing $\rho_1^{\otimes n}$ from $\rho_2^{\otimes n}$ is equal to

$$\mathrm{Tr}[\eta_1\rho_1^{\otimes n} \wedge \eta_2\rho_2^{\otimes n}] = \frac{1}{2}\left[1 - \left\|\eta_1\rho_1^{\otimes n} - \eta_2\rho_2^{\otimes n}\right\|_1\right], \tag{35}$$

with $\rho_1^{\otimes n}$ prepared with probability η_1 and $\rho_2^{\otimes n}$ with probability η_2 [20, 24].

It is known more generally that the optimal error exponent in distinguishing the ensemble $\{(\eta_i, \rho_i^{\otimes n}) : i \in [r]\}$ is equal to the minimum pairwise Chernoff divergence [34].

3 Optimal error exponent for classical antidistinguishability

In this section, we first show that the optimal error exponent for antidistinguishing an ensemble of probability measures (13) is equal to the multivariate Chernoff divergence of the probability measures. We then compare the multivariate Chernoff divergence with the pairwise Chernoff divergence, showing that the former can be strictly greater than the latter for every pair of the probability measures.

3.1 Multivariate Chernoff divergence as the optimal error exponent

The following theorem is the main result of this section. Some aspects of the proof presented below follow the development given in the appendix of [45].

Theorem 6 *Consider an ensemble $\mathcal{E}_{\rm cl} = \{(\eta_i, P_i) : i \in [r]\}$ of probability measures on a measurable space $(\Omega, \mathcal{A})$, where $\eta_i > 0$ for all $i \in [r]$. The optimal error exponent for antidistinguishing the probability measures is given by their multivariate Chernoff divergence, i.e.,*

$$\mathrm{E}_{\rm cl}(P_1, \ldots, P_r) = \lim_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}_{\rm cl}(\mathcal{E}_{\rm cl}^n) = \xi_{\rm cl}(P_1, \ldots, P_r), \tag{36}$$

where recalling (18), (19), and (23), the multivariate classical Chernoff divergence $\xi_{\rm cl}$ is defined as

$$\xi_{\rm cl}(P_1, \ldots, P_r) := -\ln \inf_{\mathbf{s}\in\mathbb{S}_r} \int \mathbf{d}\mu \; p_1^{s_1} \cdots p_r^{s_r}. \tag{37}$$

Remark 3 Note that we defined the optimal error exponent in terms of the limit inferior in Definition 1. However, Theorem 6 demonstrates that the limit exists and is equal to the multivariate Chernoff divergence.

Proof of Theorem 6 Let μ be the dominating measure given by (2) and $p_1, \ldots, p_r$ the induced densities defined in (3). Let D be the intersection of the supports of the densities $p_1, \ldots, p_r$:

$$D := \{\omega \in \Omega : p_i(\omega) > 0, \ \forall i \in [r]\}. \tag{38}$$

For all $\mathbf{s} := (s_1, \ldots, s_r) \in \mathbb{S}_r$, the following equality holds by employing the convention $0^0 = 0$:

$$\mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) = \int_D \mathbf{d}\mu \; p_1^{s_1} \cdots p_r^{s_r}. \tag{39}$$

We also have from (10) and the definition (38) that

$$\mathrm{Err}_{\rm cl}(\mathcal{E}_{\rm cl}^n) = \int_{D^n} \mathbf{d}\mu^{\otimes n} \; \left(\eta_1 p_1^{\otimes n} \wedge \cdots \wedge \eta_r p_r^{\otimes n}\right). \tag{40}$$

Throughout the proof, we only work with D defined in (38). The case $\mu(D) = 0$ is trivial, since the antidistinguishability error probability is equal to zero in this case. So, we assume henceforth that $\mu(D) > 0$.

From (40), we get

$$\mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq \int_{D^n} \mathbf{d}\mu^{\otimes n} \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right). \tag{41}$$

Let $\mathbf{s} \in \mathbb{S}_r$ be arbitrary. We can write the right-hand side of (41) as

$$\int_{D^n} \mathbf{d}\mu^{\otimes n} \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right) = \int_{D^n} \mathbf{d}\mu^{\otimes n} \, (p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n})^{s_1} \cdots (p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n})^{s_r} \tag{42}$$

$$\leq \int_{D^n} \mathbf{d}\mu^{\otimes n} \, (p_1^{\otimes n})^{s_1} \cdots (p_r^{\otimes n})^{s_r}. \tag{43}$$

The expression on the right-hand side of (43) has a product structure. Indeed, we have that

$$\begin{aligned}
&\int_{D^n} \mathbf{d}\mu^{\otimes n} \, (p_1^{\otimes n})^{s_1} \cdots (p_r^{\otimes n})^{s_r} \\
&\quad = \int_{D^n} \mathbf{d}\mu^{\otimes n}(\omega_1, \ldots, \omega_n) \, (p_1^{\otimes n}(\omega_1, \ldots, \omega_n))^{s_1} \cdots (p_r^{\otimes n}(\omega_1, \ldots, \omega_n))^{s_r} \\
&\quad = \int_{D^n} \prod_{k\in[n]} \mathbf{d}\mu(\omega_k) \left(\prod_{k\in[n]} p_1(\omega_k)\right)^{s_1} \cdots \left(\prod_{k\in[n]} p_r(\omega_k)\right)^{s_r} && (44) \\
&\quad = \int_{D^n} \prod_{k\in[n]} \mathbf{d}\mu(\omega_k) \prod_{k\in[n]} p_1^{s_1}(\omega_k) \cdots p_r^{s_r}(\omega_k) && (45) \\
&\quad = \int_{D^n} \prod_{k\in[n]} \left(\mathbf{d}\mu(\omega_k) \, p_1^{s_1}(\omega_k) \cdots p_r^{s_r}(\omega_k)\right) && (46) \\
&\quad = \prod_{k\in[n]} \int_{D} \mathbf{d}\mu(\omega_k) \, (p_1^{s_1} \cdots p_r^{s_r})(\omega_k) && (47) \\
&\quad = \left(\int_{D} \mathbf{d}\mu \; p_1^{s_1} \cdots p_r^{s_r}\right)^n && (48) \\
&\quad = \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r)^n. && (49)
\end{aligned}$$

From (41), (43), and (49), we thus get

$$\mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r)^n, \qquad \text{for all } \mathbf{s} \in \mathbb{S}_r. \tag{50}$$

This implies, for all $n \in \mathbb{N}$, that

$$-\frac{1}{n} \ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \geq -\ln \inf_{\mathbf{s}\in\mathbb{S}_r} \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) = \xi_{\mathrm{cl}}(P_1, \ldots, P_r). \tag{51}$$

Therefore, we get

$$\liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}^n_{\mathrm{cl}}) \geq \xi_{\mathrm{cl}}(P_1, \ldots, P_r). \tag{52}$$

This proves the achievability part of the optimal error exponent.

To prove the optimality part, we apply multivariable calculus and the law of large numbers. For this purpose, let us parameterize the unit simplex of $\mathbb{R}^r$ by the corner of the standard unit cube of $\mathbb{R}^{r-1}$, defined as

$$\mathbb{T}_r := \left\{ \mathbf{t} \in [0,1]^{r-1} : \mathbf{t} := (t_1, \ldots, t_{r-1}), \sum_{i\in[r-1]} t_i \leq 1 \right\}. \tag{53}$$

The unit simplex (18) can be expressed as:

$$\mathbb{S}_r = \left\{ \left(t_1, \ldots, t_{r-1}, 1 - \sum_{i\in[r-1]} t_i \right) : (t_1, \ldots, t_{r-1}) \in \mathbb{T}_r \right\}. \tag{54}$$

Using the new parameterization, let us denote the elements of $\mathbb{S}_r$ by $\mathbf{s_t} := (t_1, \ldots, t_{r-1}, 1 - \sum_{i\in[r-1]} t_i)$ for $\mathbf{t} := (t_1, \ldots, t_{r-1}) \in \mathbb{T}_r$. The Hellinger transform of $P_1, \ldots, P_r$ can then be expressed as the following function on $\mathbb{T}_r$:

$$\mathrm{H}(\mathbf{t}) := \mathbb{H}_{\mathbf{s_t}}(P_1, \ldots, P_r), \qquad \text{for all } \mathbf{t} \in \mathbb{T}_r. \tag{55}$$

Thus, the multivariate Chernoff divergence of $P_1, \ldots, P_r$ has the form

$$\xi_{\mathrm{cl}}(P_1, \ldots, P_r) = \sup_{\mathbf{t}\in\mathbb{T}_r} -\ln \mathrm{H}(\mathbf{t}). \tag{56}$$

In what follows, using the reparametrized Hellinger transform (55), we define an exponential family of densities $p_{\mathbf{t}}$ with $\mathbf{t} \in \mathbb{T}_r$, as given in (61), which enables us to express each $p_i^{\otimes n}$ in terms of $p_{\mathbf{t}}^{\otimes n}$ for all $n \in \mathbb{N}$ as given in (82). This then allows for the use of the law of large numbers to deduce a family of upper bounds on the asymptotic error exponent, given by $-\min_{1\leq i\leq r} \gamma_i(\mathbf{t})$ for the non-corner points in $\mathbb{T}_r$, as defined later on in (80). Lastly, we use multivariable calculus to prove that there exists a non-corner point $\mathbf{t}^*$ such that $\ln \mathrm{H}(\mathbf{t}^*) = \min_{1\leq i\leq r} \gamma_i(\mathbf{t}^*)$. This implies that the multivariate Chernoff divergence is the optimal bound for the asymptotic error rate.

For every $\mathbf{t} \in \mathbb{T}_r$, let us express $\mathrm{H}(\mathbf{t})$ in an *exponential-integral* form as follows:

$$\mathrm{H}(\mathbf{t}) = \int_D \mathbf{d}\mu \; p_1^{t_1} \cdots p_{r-1}^{t_{r-1}} p_r^{1-\sum_{i\in[r-1]} t_i} \tag{57}$$

$$= \int_D \mathbf{d}\mu \; (p_1/p_r)^{t_1} \cdots (p_{r-1}/p_r)^{t_{r-1}} p_r \tag{58}$$

$$= \int_D \mathbf{d}\mu \; p_r \exp\left(\sum_{i \in [r-1]} t_i \ln{(p_i/p_r)} \right) \tag{59}$$

$$= \int_D \mathbf{d}\mu \; p_r \exp\left(\sum_{i \in [r-1]} t_i q_i \right), \tag{60}$$

where $q_i := \ln(p_i/p_r)$. The assumption $\mu(D) > 0$ implies that $\mathrm{H}(\mathbf{t}) > 0$ for all $\mathbf{t} \in \mathbb{T}_r$. This allows us to define an exponential family of densities on D with respect to μ for $\mathbf{t} \in \mathbb{T}_r$ given by

$$p_{\mathbf{t}}(\omega) := \frac{1}{\mathrm{H}(\mathbf{t})} p_r(\omega) \exp\left(\sum_{i \in [r-1]} t_i q_i(\omega) \right) \qquad \text{for all } \omega \in D. \tag{61}$$

Define a function $\mathrm{K} : \mathbb{T}_r \to \mathbb{R}$ by

$$\mathrm{K}(\mathbf{t}) := \ln \mathrm{H}(\mathbf{t}). \tag{62}$$

Let $\mathbf{e}_1, \ldots, \mathbf{e}_{r-1}$ denote the standard unit vectors in $\mathbb{R}^{r-1}$, and let $\mathbb{T}_r^\circ$ denote the interior of $\mathbb{T}_r$ which is given by

$$\mathbb{T}_r^\circ := \left\{ (t_1, \ldots, t_{r-1}) \in (0, 1)^{r-1} : \sum_{i \in [r-1]} t_i < 1 \right\}. \tag{63}$$

We note that the set $\mathbb{T}_r^\circ$ represents a parametrization of the interior $\mathbb{S}_r^\circ$ of the unit simplex $\mathbb{S}_r$. By Theorem 2.64 of [53], we know that H is a smooth function on $\mathbb{T}_r^\circ$; also its partial derivatives are given for $\mathbf{t} \in \mathbb{T}_r^\circ$ and $i \in [r-1]$ by

$$\partial_i \mathrm{H}(\mathbf{t}) := \lim_{h \to 0} \frac{\mathrm{H}(\mathbf{t} + h\mathbf{e}_i) - \mathrm{H}(\mathbf{t})}{h} \tag{64}$$

$$= \int_D \mathbf{d}\mu \; q_i p_r \exp\left(\sum_{j \in [r-1]} t_j q_j \right) \tag{65}$$

$$= \mathrm{H}(\mathbf{t}) \mathbb{E}_{\mathbf{t}} [q_i], \tag{66}$$

where $\mathbb{E}_{\mathbf{t}}$ is the expectation under the density $p_{\mathbf{t}}$. We know that the Hellinger transform is a continuous function taking only positive values on $\mathbb{S}_r$. This implies that K is a real-valued continuous function on $\mathbb{T}_r$. Additionally, the smoothness of H on $\mathbb{T}_r^\circ$ implies the smoothness of K on $\mathbb{T}_r^\circ$. From (66), we have that

$$\partial_i \mathrm{K}(\mathbf{t}) = \frac{1}{\mathrm{H}(\mathbf{t})} \partial_i \mathrm{H}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}} [q_i], \qquad \text{for all } i \in [r-1], \; \mathbf{t} \in \mathbb{T}_r^\circ. \tag{67}$$

Also, K is a convex function on $\mathbb{T}_r^\circ$. Indeed, for $\mathbf{s}, \mathbf{s}' \in \mathbb{S}_r^\circ$ and all $t \in (0,1)$, we have

$$\int_D \mathbf{d}\mu \ p_1^{(1-t)s_1+ts_1'} \cdots p_r^{(1-t)s_r+ts_r'} = \int_D \mathbf{d}\mu \ \left(p_1^{s_1} \cdots p_r^{s_r}\right)^{1-t} \left(p_1^{s_1'} \cdots p_r^{s_r'}\right)^t \tag{68}$$

$$\leq \left(\int_D \mathbf{d}\mu \ p_1^{s_1} \cdots p_r^{s_r}\right)^{1-t} \left(\int_D \mathbf{d}\mu \ p_1^{s_1'} \cdots p_r^{s_r'}\right)^t \tag{69}$$

due to Hölder's inequality. The convexity of K then follows by taking the logarithm on both sides of (69). By continuity, K is convex on $\mathbb{T}_r$. Let $\mathbb{T}_r^1$ denote the set

$$\mathbb{T}_r^1 := \left\{ (t_1, \ldots, t_{r-1}) \in \mathbb{T}_r : \sum_{i \in [r-1]} t_i < 1 \right\}. \tag{70}$$

We call $\mathbb{T}_r^1$ the set of non-corner points of $\mathbb{T}_r$. It is easy to see that $\mathbb{T}_r^\circ \subset \mathbb{T}_r^1$. For all $\mathbf{t} \in \mathbb{T}_r^1$ and $i \in [r-1]$, the limit

$$\partial_i^+ \mathrm{K}(\mathbf{t}) := \lim_{h \searrow 0} \frac{\mathrm{K}(\mathbf{t} + h\mathbf{e}_i) - \mathrm{K}(\mathbf{t})}{h} \tag{71}$$

exists in $\mathbb{R} \cup \{-\infty\}$ (Lemma 22). Observe that for $\mathbf{t} \in \mathbf{T}_r^\circ$, we have $\partial_i \mathrm{K}(\mathbf{t}) = \partial_i^+ \mathrm{K}(\mathbf{t})$ for all $i \in [r-1]$. It is shown in Lemma 23 that for $\mathbf{t} \in \mathbb{T}_r^1$ and $i \in [r-1]$, the expectation value $\mathbb{E}_{\mathbf{t}}[q_i]$ exists in $\mathbb{R} \cup \{-\infty\}$ and satisfies

$$\partial_i^+ \mathrm{K}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}}[q_i]. \tag{72}$$

Define a set

$$\mathbb{T}_{r,f}^1 := \left\{ \mathbf{t} \in \mathbb{T}_r^1 : \partial_i^+ \mathrm{K}(\mathbf{t}) \neq -\infty, \ \forall i \in [r-1] \right\}. \tag{73}$$

Note that $\mathbb{T}_r^\circ \subset \mathbb{T}_{r,f}^1$.

Using the definition (61) of the density $p_{\mathbf{t}}$ on D for $\mathbf{t} := (t_1, \ldots, t_{r-1}) \in \mathbb{T}_{r,f}^1$ and $i \in [r]$, we have

$$\ln \frac{p_i}{p_{\mathbf{t}}} = \ln p_i - \ln p_{\mathbf{t}} \tag{74}$$

$$= \ln p_i - \sum_{j \in [r-1]} t_j q_j - \ln p_r + \ln \mathrm{H}(\mathbf{t}) \tag{75}$$

$$= \ln \frac{p_i}{p_r} - \sum_{j \in [r-1]} t_j q_j + \mathrm{K}(\mathbf{t}) \tag{76}$$

$$= q_i - \sum_{j \in [r-1]} t_j q_j + \mathrm{K}(\mathbf{t}), \tag{77}$$

where q_r is the zero function on D. We write (77) in a more compact form as

$$\ln \frac{p_i}{p_{\mathbf{t}}} = \sum_{j\in[r-1]} (\delta_{ij} - t_j)q_j + \mathrm{K}(\mathbf{t}), \qquad \text{for all } i \in [r], \quad \mathbf{t} \in \mathbb{T}^1_{r,f}. \tag{78}$$

Here δ_{ij} is the Kronecker delta (taking the value 1 if $i = j$, and 0 otherwise). By taking the expectation on both sides of (78) under the density $p_{\mathbf{t}}$ and then using (72), we get

$$\gamma_i(\mathbf{t}) \coloneqq \mathbb{E}_{\mathbf{t}}\left[\ln \frac{p_i}{p_{\mathbf{t}}}\right] = \sum_{j\in[r-1]} (\delta_{ij} - t_j)\mathbb{E}_{\mathbf{t}}[q_j] + \mathrm{K}(\mathbf{t}) \tag{79}$$

$$= \sum_{j\in[r-1]} (\delta_{ij} - t_j)\partial^+_j \mathrm{K}(\mathbf{t}) + \mathrm{K}(\mathbf{t}) \tag{80}$$

for all $i \in [r]$ and $\mathbf{t} \in \mathbb{T}^1_{r,f}$. We can write (80) in a more compact form as

$$\gamma_i(\mathbf{t}) = \begin{cases} \partial^+_i \mathrm{K}(\mathbf{t}) - \mathbf{t}^T \nabla^+ \mathrm{K}(\mathbf{t}) + \mathrm{K}(\mathbf{t}), & i \in [r-1], \\ -\mathbf{t}^T \nabla^+ \mathrm{K}(\mathbf{t}) + \mathrm{K}(\mathbf{t}), & i = r, \end{cases} \qquad \text{for all } \mathbf{t} \in \mathbb{T}^1_{r,f}, \tag{81}$$

where $\nabla^+ \mathrm{K}(\mathbf{t}) \coloneqq (\partial^+_1 \mathrm{K}(\mathbf{t}), \ldots, \partial^+_{r-1} \mathrm{K}(\mathbf{t}))^T$.

Let $\omega^n \coloneqq (\omega_1, \ldots, \omega_n) \in D^n$ and $\mathbf{t} \in \mathbb{T}^1_{r,f}$ be arbitrary. We have that

$$p_i^{\otimes n}(\omega^n) = \left(\prod_{j\in[n]} \frac{p_i}{p_{\mathbf{t}}}(\omega_j) \right) p_{\mathbf{t}}^{\otimes n}(\omega^n) = \exp\Big(n G^{(i)}_{\mathbf{t},n}(\omega^n)\Big)\, p_{\mathbf{t}}^{\otimes n}(\omega^n), \tag{82}$$

where

$$G^{(i)}_{\mathbf{t},n}(\omega^n) \coloneqq \frac{1}{n} \sum_{j\in[n]} \ln \frac{p_i}{p_{\mathbf{t}}}(\omega_j), \qquad \text{for all } i \in [r]. \tag{83}$$

Let $P_{\mathbf{t}}^{\otimes n}$ be the product measure corresponding to the density $p_{\mathbf{t}}^{\otimes n}$ on D^n, and let $\mathbb{E}^n_{\mathbf{t}}$ be the pertaining expectation. By the definition in (79), we then have that

$$\mathbb{E}^n_{\mathbf{t}}\left[G^{(i)}_{\mathbf{t},n}\right] = \gamma_i(\mathbf{t}), \qquad \text{for all } i \in [r]. \tag{84}$$

Since $G^{(i)}_{\mathbf{t},n}$ is an i.i.d. average, the law of large numbers [6] implies that for arbitrary $\delta > 0$, there exists $n_\delta \in \mathbb{N}$ such that the probability of the event

$$U_{n,\delta} \coloneqq \{\omega^n \in D^n : \forall i \in [r],\ G^{(i)}_{\mathbf{t},n}(\omega^n) \geq \gamma_i(\mathbf{t}) - \delta\} \tag{85}$$

satisfies

$$P_{\mathbf{t}}^{\otimes n}(U_{n,\delta}) \geq 1-\delta, \qquad \text{for all } n \geq n_\delta. \tag{86}$$

The development in (82)–(86) implies that, for all $n \geq n_\delta$,

$$\begin{aligned}
\mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) &= \int_{D^n} \mathbf{d}\mu^{\otimes n} \ \left(\eta_1 p_1^{\otimes n} \wedge \cdots \wedge \eta_r p_r^{\otimes n}\right) &(87)\\
&\geq \eta_{\min} \int_{D^n} \mathbf{d}\mu^{\otimes n} \ \left(p_1^{\otimes n} \wedge \cdots \wedge p_r^{\otimes n}\right) &(88)\\
&= \eta_{\min} \int_{D^n} \mathbf{d}\mu^{\otimes n} \ \left(\exp\left(nG_{\mathbf{t},n}^{(1)}\right) \wedge \cdots \wedge \exp\left(nG_{\mathbf{t},n}^{(r)}\right)\right) p_{\mathbf{t}}^{\otimes n} &(89)\\
&= \eta_{\min}\, \mathbb{E}_{\mathbf{t}}^n\left[\exp\left(nG_{\mathbf{t},n}^{(1)}\right) \wedge \cdots \wedge \exp\left(nG_{\mathbf{t},n}^{(r)}\right)\right] &(90)\\
&\geq \eta_{\min}\, \mathbb{E}_{\mathbf{t}}^n\left[\mathbf{1}_{U_{n,\delta}}\left(\exp\left(nG_{\mathbf{t},n}^{(1)}\right) \wedge \cdots \wedge \exp\left(nG_{\mathbf{t},n}^{(r)}\right)\right)\right] &(91)\\
&\geq \eta_{\min}\, P_{\mathbf{t}}^{\otimes n}(U_{n,\delta}) \exp\left(n \min_{1\leq i\leq r} (\gamma_i(\mathbf{t})-\delta)\right) &(92)\\
&\geq \eta_{\min}(1-\delta) \exp\left(n \min_{1\leq i\leq r} \gamma_i(\mathbf{t}) - n\delta\right). &(93)
\end{aligned}$$

Here $\mathbf{1}_{U_{n,\delta}}$ denotes the indicator function of the set $U_{n,\delta}$. Therefore, we have that

$$-\frac{1}{n}\ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq -\frac{\ln(\eta_{\min}(1-\delta))}{n} - \left(\min_{1\leq i\leq r} \gamma_i(\mathbf{t}) - \delta\right), \qquad \text{for all } n \geq n_\delta. \tag{94}$$

By taking the limit superior as $n \to \infty$ on both sides of (94) and then the limit $\delta \to 0$, we thus get

$$\limsup_{n\to\infty} -\frac{1}{n}\ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq -\min_{1\leq i\leq r} \gamma_i(\mathbf{t}), \qquad \text{for all } \mathbf{t} \in \mathbb{T}_{r,f}^1. \tag{95}$$

Recall from (56) and the fact $\mathrm{K}(\mathbf{t}) = \ln \mathrm{H}(\mathbf{t})$, our goal is to show that

$$\limsup_{n\to\infty} -\frac{1}{n}\ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}^n) \leq \sup_{\mathbf{t}\in\mathbb{T}_r} -\mathrm{K}(\mathbf{t}). \tag{96}$$

In view of (95), it suffices to show that for some $\mathbf{t}^* \in \mathbb{T}_{r,f}^1$, the following holds

$$\min_{1\leq i\leq r} \gamma_i(\mathbf{t}^*) \geq \mathrm{K}(\mathbf{t}^*). \tag{97}$$

We now argue that such a $\mathbf{t}^*$ exists. Since K is a continuous function on the compact set $\mathbb{T}_r$, there exists $\mathbf{t}^* := (t_1^*, \ldots, t_{r-1}^*) \in \mathbb{T}_r$ that minimizes K over $\mathbb{T}_r$, i.e.,

$$\mathrm{K}(\mathbf{t}^*) = \min_{\mathbf{t} \in \mathbb{T}_r} \mathrm{K}(\mathbf{t}). \tag{98}$$

Consider the following two cases.

Case A Suppose $\mathbf{t}^* \in \mathbb{T}_r^1$. Choose arbitrary $i \in [r-1]$. If $t_i^* = 0$, then by convexity, continuity of K, and the fact that $\mathbf{t}^*$ is a minimizer, we have $\partial_i^+ \mathrm{K}(\mathbf{t}^*) \geq 0$ (see Lemma 22). Else we have $0 < t_i^* < 1$ and the first-order necessary condition for a minimizer implies $\partial_i^+ \mathrm{K}(\mathbf{t}^*) = 0$. Combining these, we get $\partial_i^+ \mathrm{K}(\mathbf{t}^*) \geq 0$ for all $i \in [r-1]$ and hence $\mathbf{t}^* \in \mathbb{T}_{r,f}^1$, and $\mathbf{t}^{*T} \nabla^+ \mathrm{K}(\mathbf{t}^*) = 0$. From (81), we thus get

$$\gamma_i(\mathbf{t}^*) = \begin{cases} \partial_i^+ \mathrm{K}(\mathbf{t}^*) + \mathrm{K}(\mathbf{t}^*), & i \in [r-1], \\ \mathrm{K}(\mathbf{t}^*), & i = r. \end{cases} \tag{99}$$

This implies that the inequality (97) holds for the minimizer $\mathbf{t}^*$.

Case B Suppose $\mathbf{t}^* \in \mathbb{T}_r \backslash \mathbb{T}_r^1$, i.e., $t_1^* + \cdots + t_{r-1}^* = 1$. For some $i \in [r-1]$, we have $t_i^* > 0$. According to the current parameterization of the unit simplex given in (54), $\mathbf{t}^*$ corresponds to the vector $(t_1^*, \ldots, t_{r-1}^*, 0)$ in $\mathbb{S}_r$. We reparameterize the unit simplex $\mathbb{S}_r$ as

$$\mathbf{s}_{\mathbf{u}} = \left(u_1, \ldots, u_{i-1}, 1 - \sum_{j \in [r-1]} u_j, u_i, \ldots, u_{r-1} \right), \qquad \mathbf{u} \in \mathbb{T}_r. \tag{100}$$

In the reparameterized problem, the corresponding minimizer $\mathbf{u}^*$ of K satisfies $\mathbf{s}_{\mathbf{u}^*} = (t_1^*, \ldots, t_{r-1}^*, 0)$, which implies

$$1 - \sum_{j \in [r-1]} u_j^* = t_i^* > 0. \tag{101}$$

This reduces the problem to Case A, which implies that (97) holds.

Combining the above two cases, we conclude that (97) holds for the minimizer $\mathbf{t}^*$ and this completes the proof. □

3.2 Multivariate Chernoff divergence versus pairwise Chernoff divergences

Identifying the true probability measure out of the given r probability measures is the same as eliminating all the remaining $r-1$ false probability measures. As such, general intuition says that, upon observing i.i.d. data, it is easier to eliminate a false probability measure than to identify the true probability measure. This also means that the optimal error exponent of classical antidistinguishability should be greater than that of multiple classical hypothesis testing, the former being the multivariate classical Chernoff divergence and the latter being the minimum of the pairwise Chernoff divergences of the probability measures [50] (see also [57, Theorem 4.2] and [35, 51, 52]). Indeed, for any two indices $i, j \in [r]$ define a subset of $\mathbb{S}_r$:

$$\mathbb{S}_r^{(i,j)} = \{\mathbf{s} \in \mathbb{S}_r : \mathbf{s} := (s_1, \ldots, s_r), s_i + s_j = 1\}. \tag{102}$$

By definition, we have

$$\begin{aligned}
\xi_{\text{cl}}(P_1, \ldots, P_r) &\geq -\ln \inf_{\mathbf{s}\in\mathbb{S}_r^{(i,j)}} \mathbb{H}_{\mathbf{s}}(P_1, \ldots, P_r) &(103)\\
&= -\ln \inf_{s\in[0,1]} \int \mathbf{d}\mu \; p_i^s p_j^{(1-s)} &(104)\\
&= \xi_{\text{cl}}(P_i, P_j), &(105)
\end{aligned}$$

where $\xi_{\text{cl}}(P_i, P_j)$ is the Chernoff divergence of the probability measures P_i and P_j. This gives

$$\xi_{\text{cl}}(P_1, \ldots, P_r) \geq \max_{i<j} \xi_{\text{cl}}(P_i, P_j) \geq \min_{i<j} \xi_{\text{cl}}(P_i, P_j). \tag{106}$$

The following example illustrates an instance for which the first inequality in (106) is strict.

Example 7 Consider a uniform ensemble $\mathcal{E}_{\text{cl}} = \{(1/3, P_1), (1/3, P_2), (1/3, P_3)\}$ of probability measures on a discrete space $\Omega = \{x, y, z\}$ whose densities with respect to the counting measure μ are given by

$$p_1 = \frac{1}{2}\mathbf{1}_{\{x,y\}}, \qquad p_2 = \frac{1}{2}\mathbf{1}_{\{x,z\}}, \qquad p_3 = \frac{1}{3}\mathbf{1}_{\Omega}. \tag{107}$$

We have for $\omega^n \in \Omega^n$,

$$(p_1^{\otimes n} \wedge p_2^{\otimes n} \wedge p_3^{\otimes n})(\omega^n) = \begin{cases} \dfrac{1}{3^n}, & \text{if } \omega^n = (\underbrace{x, \ldots, x}_{n \text{ times}}), \\ 0, & \text{otherwise.} \end{cases} \tag{108}$$

By the minimum likelihood principle, we thus get

$$\begin{aligned}
\text{Err}_{\text{cl}}(\mathcal{E}_{\text{cl}}^n) &= \frac{1}{3} \int \mathbf{d}\mu^{\otimes n} \; \left(p_1^{\otimes n} \wedge p_2^{\otimes n} \wedge p_3^{\otimes n}\right) &(109)\\
&= \frac{1}{3} \cdot \mu^{\otimes n}(\{(\underbrace{x, \ldots, x}_{n \text{ times}})\}) \cdot \frac{1}{3^n} &(110)
\end{aligned}$$

$$= \frac{1}{3} \cdot \mu(\{x\})^n \cdot \frac{1}{3^n} \tag{111}$$
$$= \frac{1}{3} \cdot 1 \cdot \frac{1}{3^n} \tag{112}$$
$$= \frac{1}{3^{n+1}}. \tag{113}$$

This gives the optimal error exponent

$$\mathrm{E}_{\mathrm{cl}}(P_1, P_2, P_3) = \liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}^n_{\mathrm{cl}}) = \ln 3. \tag{114}$$

We now compute the pairwise Chernoff divergences of the probability measures as follows.

$$\xi_{\mathrm{cl}}(P_1, P_2) = -\ln \inf_{s\in[0,1]} \int \mathbf{d}\mu\ p_1^s p_2^{(1-s)} \tag{115}$$
$$= -\ln \inf_{s\in[0,1]} \int_{\{x\}} \mathbf{d}\mu\ \frac{1}{2^s} \frac{1}{2^{(1-s)}} \tag{116}$$
$$= -\ln \inf_{s\in[0,1]} \frac{1}{2^s} \frac{1}{2^{(1-s)}} \tag{117}$$
$$= -\ln \left(\frac{1}{2}\right) \tag{118}$$
$$= \ln 2. \tag{119}$$

Also,

$$\xi_{\mathrm{cl}}(P_1, P_3) = -\ln \inf_{s\in[0,1]} \int \mathbf{d}\mu\ p_1^s p_3^{(1-s)} \tag{120}$$
$$= -\ln \inf_{s\in[0,1]} \int_{\{x,y\}} \mathbf{d}\mu\ \frac{1}{2^s} \frac{1}{3^{(1-s)}} \tag{121}$$
$$= -\ln \left[\mu(\{x, y\}) \cdot \frac{1}{3} \cdot \inf_{s\in[0,1]} \left(\frac{3}{2}\right)^s\right] \tag{122}$$
$$= -\ln \left[2 \cdot \frac{1}{3} \cdot 1\right] \tag{123}$$
$$= \ln(3/2). \tag{124}$$

By similar arguments, we get $\xi_{\mathrm{cl}}(P_2, P_3) = \ln(3/2)$. This implies

$$\max\{\xi_{\mathrm{cl}}(P_1, P_2), \xi_{\mathrm{cl}}(P_1, P_3), \xi_{\mathrm{cl}}(P_2, P_3)\} = \ln 2. \tag{125}$$
$$\min\{\xi_{\mathrm{cl}}(P_1, P_2), \xi_{\mathrm{cl}}(P_1, P_3), \xi_{\mathrm{cl}}(P_2, P_3)\} = \ln(3/2). \tag{126}$$

From (114), (125), and (126), we have

$$\mathrm{E}_{\mathrm{cl}}(P_1, P_2, P_3) > \max\{\xi_{\mathrm{cl}}(P_1, P_2), \xi_{\mathrm{cl}}(P_1, P_3), \xi_{\mathrm{cl}}(P_2, P_3)\} \tag{127}$$
$$> \min\{\xi_{\mathrm{cl}}(P_1, P_2), \xi_{\mathrm{cl}}(P_1, P_3), \xi_{\mathrm{cl}}(P_2, P_3)\}. \tag{128}$$

4 Achievable error exponent for quantum antidistinguishability

4.1 One-shot case

Observe that the "antidistinguishability problem" between any two states ρ_1 and ρ_2 is the same as the state discrimination problem. Indeed, if we say that "ρ_1 is not the true state," then we are saying "ρ_2 is the true state." Using this observation, we obtain an upper bound on the optimal error probability of antidistinguishing the states of a given quantum ensemble by considering "special" POVMs that focus on pairs of states, as expounded upon in the proof of the following theorem:

Theorem 8 *Consider a quantum ensemble $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$. An upper bound on the optimal error probability of antidistinguishing the states of the ensemble is given by:*

$$\mathrm{Err}(\mathcal{E}) \leq \min_{1\leq i<j\leq r} \mathrm{Tr}[\eta_i \rho_i \wedge \eta_j \rho_j] \tag{129}$$
$$= \min_{1\leq i<j\leq r} \frac{1}{2}\left(\eta_i + \eta_j - \left\|\eta_i \rho_i - \eta_j \rho_j\right\|_1\right). \tag{130}$$

In particular, if at least two states in $\rho_1, \ldots, \rho_r$ are mutually orthogonal, then $\mathrm{Err}(\mathcal{E}) = 0$.

Proof Given two fixed indices $i, j \in [r]$, let $\Xi_r^{(i,j)}$ denote the set of POVMs $\mathscr{M} = \{M_1, \ldots, M_r\}$ such that $M_k = 0$ if $k \notin \{i, j\}$. For such POVMs, we have $M_i + M_j = \mathbb{I}$ and

$$\mathrm{Err}(\mathscr{M}; \mathcal{E}) = \eta_i \,\mathrm{Tr}[M_i \rho_i] + \eta_j \,\mathrm{Tr}\big[(\mathbb{I} - M_i)\rho_j\big] \tag{131}$$
$$= \eta_i \,\mathrm{Tr}[M_i \rho_i] + \eta_j - \eta_j \,\mathrm{Tr}\big[M_i \rho_j\big] \tag{132}$$
$$= \eta_j - \mathrm{Tr}\big[M_i \left(\eta_j \rho_j - \eta_i \rho_i\right)\big]. \tag{133}$$

By taking the infimum over $\Xi_r^{(i,j)}$ on both sides of (133), we get

$$\inf_{\mathscr{M}\in\Xi_r^{(i,j)}} \mathrm{Err}(\mathscr{M}; \mathcal{E}) = \eta_j - \sup_{0\leq M_i\leq \mathbb{I}} \mathrm{Tr}\big[M_i \left(\eta_j \rho_j - \eta_i \rho_i\right)\big], \tag{134}$$

where the supremum on the right-hand side of (134) is taken over every positive semi-definite operator M_i such that $0 \leq M_i \leq \mathbb{I}$. The supremum is attained by the *Helstrom–Holevo measurement* [20, 24] given by $M_i = \mathrm{supp}(\eta_j\rho_j - \eta_i\rho_i)_+$ (note the order of i and j). We thus get

$$\inf_{\mathscr{M}\in\Xi_r^{(i,j)}} \mathrm{Err}(\mathscr{M};\mathcal{E}) = \eta_j - \mathrm{Tr}\Big[\big(\eta_j\rho_j - \eta_i\rho_i\big)_+\Big] \tag{135}$$

$$= \mathrm{Tr}\big[\eta_j\rho_j\big] - \mathrm{Tr}\big[\big(\eta_j\rho_j - \eta_i\rho_i + |\eta_j\rho_j - \eta_i\rho_i|\big)/2\big] \tag{136}$$

$$= \mathrm{Tr}\big[\big(\eta_i\rho_i + \eta_j\rho_j - |\eta_i\rho_i - \eta_j\rho_j|\big)/2\big] \tag{137}$$

$$= \mathrm{Tr}\big[\eta_i\rho_i \wedge \eta_j\rho_j\big] \tag{138}$$

$$= \frac{1}{2}\left(\eta_i + \eta_j - \left\|\eta_i\rho_i - \eta_j\rho_j\right\|_1\right). \tag{139}$$

It is clear that the optimal antidistinguishability error probability satisfies

$$\mathrm{Err}(\mathcal{E}) \leq \inf_{\mathscr{M}\in\Xi_r^{(i,j)}} \mathrm{Err}(\mathscr{M};\mathcal{E}) = \mathrm{Tr}\big[\eta_i\rho_i \wedge \eta_j\rho_j\big], \qquad \text{for all } 1 \leq i < j \leq r. \tag{140}$$

By combining (138)–(140), we thus get the upper bound on the optimal antidistinguishability error probability stated in the theorem. □

The expression on the right-hand side of (129) can be further simplified for pure states. This is a consequence of the following identity (see Proposition 24 in Appendix B):

$$\| |\varphi\rangle\langle\varphi| - |\zeta\rangle\langle\zeta| \|_1^2 = (\langle\varphi|\varphi\rangle + \langle\zeta|\zeta\rangle)^2 - 4\,|\langle\zeta|\varphi\rangle|^2, \tag{141}$$

which holds for vectors $|\varphi\rangle$ and $|\zeta\rangle$, as well as Theorem 1 of [1] which states that for all positive semi-definite operators A, B and all $0 \leq s \leq 1$, we have

$$\mathrm{Tr}[A \wedge B] \leq \mathrm{Tr}\, A^s B^{1-s}. \tag{142}$$

Corollary 9 *If the quantum states in Theorem 8 are pure, i.e., given by $\rho_i = |\psi_i\rangle\langle\psi_i|$, then we have*

$$\mathrm{Err}(\mathcal{E}) \leq \min_{1\leq i<j\leq r} \frac{\eta_i + \eta_j}{2}\left(1 - \sqrt{1 - \frac{4\eta_i\eta_j|\langle\psi_i|\psi_j\rangle|^2}{(\eta_i+\eta_j)^2}}\right) \tag{143}$$

$$\leq \frac{1}{2}\min_{1\leq i<j\leq r} |\langle\psi_i|\psi_j\rangle|^2. \tag{144}$$

Proof Applying (130) and (141), we find that for all $1 \leq i < j \leq r$,

$$\mathrm{Err}(\mathcal{E}) \leq \frac{1}{2}\left(\eta_i + \eta_j - \left\| \eta_i |\psi_i\rangle\langle\psi_i| - \eta_j |\psi_j\rangle\langle\psi_j| \right\|_1\right) \tag{145}$$

$$= \frac{1}{2}\left(\eta_i + \eta_j - \sqrt{(\eta_i + \eta_j)^2 - 4\eta_i\eta_j |\langle\psi_i|\psi_j\rangle|^2}\right) \tag{146}$$

$$= \frac{\eta_i + \eta_j}{2}\left(1 - \sqrt{1 - \frac{4\eta_i\eta_j|\langle\psi_i|\psi_j\rangle|^2}{(\eta_i + \eta_j)^2}}\right). \tag{147}$$

This proves the inequality (143). By (142), we get that for all $1 \leq i < j \leq r$ and $0 \leq s \leq 1$,

$$\mathrm{Tr}\left[\eta_i |\psi_i\rangle\langle\psi_i| \wedge \eta_j |\psi_j\rangle\langle\psi_j|\right] \leq \eta_i^s \eta_j^{1-s}\, \mathrm{Tr}\left[|\psi_i\rangle\langle\psi_i|\psi_j\rangle\langle\psi_j|\right] = \eta_i^s \eta_j^{1-s} |\langle\psi_i|\psi_j\rangle|^2. \tag{148}$$

Since (148) holds for all $s \in [0, 1]$, we get

$$\mathrm{Tr}\left[\eta_i |\psi_i\rangle\langle\psi_i| \wedge \eta_j |\psi_j\rangle\langle\psi_j|\right] \leq (\eta_i \wedge \eta_j)|\langle\psi_i|\psi_j\rangle|^2 \leq \frac{1}{2}|\langle\psi_i|\psi_j\rangle|^2. \tag{149}$$

The desired inequality (144) thus follows by using the inequality (149) in (129). □

The sufficient condition for perfect antidistinguishability given in Theorem 8 is not a necessary condition, even in the simple case of commuting states. This is illustrated in the following example.

Example 10 Consider states ρ_1, ρ_2, and ρ_3 diagonalizable in a common eigenbasis $\{|1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|\}$, given by

$$\rho_1 = \frac{1}{2}\left(|1\rangle\langle 1| + |2\rangle\langle 2|\right), \tag{150}$$

$$\rho_2 = \frac{1}{2}\left(|1\rangle\langle 1| + |3\rangle\langle 3|\right), \tag{151}$$

$$\rho_3 = \frac{1}{2}\left(|2\rangle\langle 2| + |3\rangle\langle 3|\right). \tag{152}$$

Consider a POVM $\mathscr{M} = \{M_1, M_2, M_3\}$ given by

$$M_1 = |3\rangle\langle 3|, \tag{153}$$

$$M_2 = |2\rangle\langle 2|, \tag{154}$$

$$M_3 = |1\rangle\langle 1|. \tag{155}$$

The POVM $\mathscr{M}$ antidistinguishes the states perfectly because $\mathrm{Tr}[M_i \rho_i] = 0$ for $i \in [3]$. However, no pair of states are mutually orthogonal to each other.

4.2 Asymptotic case

As a consequence of Theorem 8, we arrive at a lower bound on the optimal error exponent, as stated in the following theorem.

Theorem 11 *Consider a quantum ensemble $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$. A lower bound on the optimal error exponent for antidistinguishing the states of the ensemble is given by the maximum of the pairwise Chernoff divergence of the states; i.e., we have*

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \geq \max_{1 \leq i < j \leq r} \xi(\rho_i, \rho_j). \tag{156}$$

Proof By Theorem 8, we have

$$\mathrm{Err}(\mathcal{E}^n) \leq \min_{1 \leq i < j \leq r} \mathrm{Tr}[\eta_i \rho_i^{\otimes n} \wedge \eta_j \rho_j^{\otimes n}]. \tag{157}$$

By combining (157) with (34), we get the desired inequality in (156). □

Let us recall from Example 7 that the inequality in (156) can be strict in some cases.

Corollary 12 *If the quantum states in Theorem 11 are pure, given by $\rho_i = |\psi_i\rangle\!\langle\psi_i|$, then we have*

$$\mathrm{E}(|\psi_1\rangle\!\langle\psi_1|, \ldots, |\psi_r\rangle\!\langle\psi_r|) \geq \max_{1 \leq i < j \leq r} -\ln |\langle\psi_i|\psi_j\rangle|^2. \tag{158}$$

Proof It follows directly from (144). □

5 Bounds on the optimal error exponent for quantum antidistinguishability from multivariate quantum Chernoff divergences

In this section, we begin by introducing the general concept of multivariate quantum Chernoff divergences, and after that, we employ this concept in order to obtain bounds on the optimal error exponent for quantum antidistinguishability. The reasoning used here is inspired by similar reasoning used for distinguishability problems between two states [23, 26, 39–41, 44].

5.1 Multivariate quantum Chernoff divergences

Definition 13 Let $r \geq 2$ be an integer. We call a function $\boldsymbol{\xi} : \mathcal{D}^r \to [0, \infty]$ a multivariate quantum Chernoff divergence if it satisfies the following properties:

1. Data processing: for states $\rho_1, \ldots, \rho_r$ and a channel $\mathcal{N}$,

$$\boldsymbol{\xi}(\rho_1, \ldots, \rho_r) \geq \boldsymbol{\xi}(\mathcal{N}(\rho_1), \ldots, \mathcal{N}(\rho_r)), \tag{159}$$

2. Reduction to the multivariate classical Chernoff divergence for commuting states: if the states $\rho_1, \ldots, \rho_r$ commute, then

$$\boldsymbol{\xi}(\rho_1, \ldots, \rho_r) = \xi_{\mathrm{cl}}(P_1, \ldots, P_r), \tag{160}$$

where ξ_{cl} is defined in (23), $P_1, \ldots, P_r$ are probability measures on $[\dim(\mathcal{H})]$,

$$P_\ell(X) := \sum_{i \in X} \lambda_{\ell,i}, \qquad \text{for } X \subseteq [\dim(\mathcal{H})], \tag{161}$$

given by a spectral decomposition of the states in a common eigenbasis

$$\rho_\ell = \sum_{i \in [\dim(\mathcal{H})]} \lambda_{\ell,i} |i\rangle\langle i|, \qquad \text{for } \ell \in [r]. \tag{162}$$

As stated above, all multivariate quantum Chernoff divergences agree on commuting states and are equal to the multivariate classical Chernoff divergence of the corresponding probability measures induced by the states in their common eigenbasis. If $\rho_1, \ldots, \rho_r$ are commuting states, then we denote their divergence by $\xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r)$. In this case, it is easy to verify that

$$\xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r) = -\ln \inf_{\mathbf{s} \in \mathbb{S}_r} \sum_{i \in [\dim(\mathcal{H})]} \left(\prod_{\ell \in [r]} \lambda_{\ell,i}^{s_\ell} \right). \tag{163}$$

As a first starting point, let us explicitly note that the optimal error exponent in (31) is itself a multivariate quantum Chernoff divergence.

Proposition 14 *The optimal error exponent* $\mathrm{E} : \mathcal{D}^r \to [0, \infty]$ *defined by* (31) *is a multivariate quantum Chernoff divergence.*

Proof See Appendix C. □

Let us note that other multivariate quantum Chernoff divergences can be constructed from the multivariate log-Euclidean divergence, as discussed in Remark 4, as well as by means of the multivariate quantum Rényi divergences proposed in [17, 42]. In what follows, we discuss some other constructions of multivariate quantum Chernoff divergences.

We say that a multivariate quantum Chernoff divergence $\xi_{\min}$ is minimal if it is a lower bound on any other multivariate quantum Chernoff divergence; i.e., for every multivariate quantum Chernoff divergence $\boldsymbol{\xi}$, we have

$$\xi_{\min}(\rho_1, \ldots, \rho_r) \leq \boldsymbol{\xi}(\rho_1, \ldots, \rho_r), \qquad \text{for all } (\rho_1, \ldots, \rho_r) \in \mathcal{D}^r. \tag{164}$$

A minimal multivariate quantum Chernoff divergence is unique by definition, and it can be obtained as an optimization over *quantum-to-classical* or *measurement* channels as presented in Proposition 15.

Let $\mathcal{K}$ be a complex Hilbert space of dimension t with an orthonormal basis $\{|1\rangle, \ldots, |t\rangle\}$. Associated with a POVM $\{M_1, \ldots, M_t\}$ acting on the Hilbert space $\mathcal{H}$ is a channel $\mathcal{M}$, known as a measurement channel, which has the following action on an input state $\rho \in \mathcal{D}(\mathcal{H})$:

$$\mathcal{M}(\rho) = \sum_{\omega \in [t]} \mathrm{Tr}[M_\omega \rho] |\omega\rangle\langle\omega|. \tag{165}$$

The action of the measurement channel on any given states $\rho_1, \ldots, \rho_r$ produces the commuting states $\mathcal{M}(\rho_1), \ldots, \mathcal{M}(\rho_r)$. This induces probability measures $P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}$ on the discrete space $\Omega = [t]$, defined by

$$P_i^{\mathcal{M}}(X) := \sum_{x \in X} \mathrm{Tr}[M_x \rho_i], \qquad \text{for } X \subseteq \Omega. \tag{166}$$

It can be easily verified that the optimal error probability of antidistinguishing the commuting states $\mathcal{M}(\rho_1), \ldots, \mathcal{M}(\rho_r)$ is equal to that of antidistinguishing the corresponding probability measures $P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}$. See (C7)–(C14) in Appendix C.

Proposition 15 *The minimal multivariate quantum Chernoff divergence is given by*

$$\xi_{\min}(\rho_1, \ldots, \rho_r) = \sup_{\mathcal{M}} \xi_{\mathrm{cl}}(P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}), \tag{167}$$

where the supremum is taken over all measurement channels $\mathcal{M}$ with a t-dimensional classical output space for all $t \in \mathbb{N}$ and each probability measure $P_i^{\mathcal{M}}$ is defined in (166).

Proof See Appendix D. □

Similar to the definition of minimal multivariate quantum Chernoff divergence, we can define the maximal multivariate quantum Chernoff divergence. We say that a multivariate quantum Chernoff divergence $\xi_{\max}$ is maximal if it is an upper bound on any other multivariate quantum Chernoff divergence, i.e., for any multivariate quantum Chernoff divergence $\boldsymbol{\xi}$, we have

$$\xi_{\max}(\rho_1, \ldots, \rho_r) \geq \boldsymbol{\xi}(\rho_1, \ldots, \rho_r), \qquad \text{for } (\rho_1, \ldots, \rho_r) \in \mathcal{D}^r. \tag{168}$$

A maximal multivariate quantum Chernoff divergence is unique by definition, and it can be obtained as an optimization over *classical-to-quantum* or *preparation* channels as given in Proposition 16.

We can view any probability measure P on the discrete space $\Omega = [t]$ as a quantum state in $\mathcal{K}$ with the fixed eigenbasis $\{|1\rangle\langle 1|, \ldots, |t\rangle\langle t|\}$, i.e.,

$$P \equiv \sum_{\omega \in \Omega} P(\{\omega\})|\omega\rangle\langle\omega|. \tag{169}$$

A quantum channel $\mathcal{P} : \mathcal{L}(\mathcal{K}) \to \mathcal{L}(\mathcal{H})$ is said to prepare a state $\rho \in \mathcal{D}(\mathcal{H})$ from a probability measure P if it satisfies $\mathcal{P}(P) = \rho$ and is called a preparation channel or classical–to–quantum channel (see [61, Section 4.6.5] for a review of classical–to–quantum channels).

Proposition 16 *The maximal multivariate quantum Chernoff divergence is given by:*

$$\xi_{\max}(\rho_1, \ldots, \rho_r) = \inf_{(\mathcal{P}, \{P_i\}_{i \in [r]})} \{\xi_{\mathrm{cl}}(P_1, \ldots, P_r) : \mathcal{P}(P_i) = \rho_i \ \ \textit{for all } i \in [r]\}, \tag{170}$$

where the infimum involves preparation channels $\mathcal{P}$ with a t-dimensional classical input system, for all $t \in \mathbb{N}$, as well as probability measures $\{P_1, \ldots, P_r\}$ of the form in (169).

Proof See Appendix E. □

5.2 Bounds on the optimal error exponent for quantum antidistinguishability

The optimal error exponent for quantum antidistinguishability can be bounded from above and below by the minimal and the maximal multivariate quantum Chernoff divergences, respectively, as stated in the following theorem.

Theorem 17 *Let $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$ be a quantum ensemble. We have*

$$\xi_{\min}(\rho_1, \dots, \rho_r) \le \mathrm{E}(\rho_1, \dots, \rho_r) \le \xi_{\max}(\rho_1, \dots, \rho_r), \tag{171}$$

where $\xi_{\min}$ and $\xi_{\max}$ are given by (167) and (170), respectively. Additionally, the bounds in (171) can be strengthened through regularization as

$$\sup_{\ell \in \mathbb{N}} \frac{1}{\ell} \xi_{\min}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}) \le \mathrm{E}(\rho_1, \dots, \rho_r) \le \inf_{\ell \in \mathbb{N}} \frac{1}{\ell} \xi_{\max}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}). \tag{172}$$

Proof We know from Proposition 14 that the optimal error exponent is a multivariate quantum Chernoff divergence, which, along with (164) and (168), justifies the inequalities in (171).

We know from Lemma 25 in Appendix F that

$$\mathrm{E}(\rho_1, \dots, \rho_r) = \frac{1}{\ell} \mathrm{E}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}) \qquad \text{for all } \ell \in \mathbb{N}. \tag{173}$$

Substituting the above equality into (171) gives

$$\frac{1}{\ell} \xi_{\min}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}) \le \mathrm{E}(\rho_1, \dots, \rho_r) \le \frac{1}{\ell} \xi_{\max}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}) \qquad \text{for all } \ell \in \mathbb{N}, \tag{174}$$

which implies the inequalities (172). □

We note that in the upper bound in (172), the infimum over $\ell \in \mathbb{N}$ can be replaced with the limit $\ell \to \infty$:

$$\inf_{\ell \in \mathbb{N}} \frac{1}{\ell} \xi_{\max}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}) = \lim_{\ell \to \infty} \frac{1}{\ell} \xi_{\max}(\rho_1^{\otimes \ell}, \dots, \rho_r^{\otimes \ell}). \tag{175}$$

See Appendix G. It is open to determine whether the supremum over $\ell \in \mathbb{N}$ in the lower bound in (172) can be replaced with the limit $\ell \to \infty$, if the limit exists.

It is known from [44, Corollary III.8] and [26, Corollary 4] (see also [40, Section 9.3]) that when $r = 2$, the following equality holds

$$\sup_{\ell \in \mathbb{N}} \frac{1}{\ell} \xi_{\min}(\rho_1^{\otimes \ell}, \rho_2^{\otimes \ell}) = \widetilde{\xi}(\rho_1, \rho_2) := \sup_{s \in (0,1)} \left[- \ln \widetilde{Q}_s(\rho_1, \rho_2) \right], \tag{176}$$

where

$$\widetilde{Q}_s(\rho_1, \rho_2) := \begin{cases} \mathrm{Tr}\left[\left(\rho_2^{(1-s)/2s} \rho_1 \rho_2^{(1-s)/2s} \right)^s \right] & : s \in [1/2, 1) \\ \mathrm{Tr}\left[\left(\rho_1^{s/2(1-s)} \rho_2 \rho_1^{s/2(1-s)} \right)^{1-s} \right] & : s \in (0, 1/2) \end{cases}. \tag{177}$$

Since the optimal error exponent is known in this case to be $\xi(\rho_1, \rho_2)$, which is defined in (32), and it is also known from [14, Lemma 3] that

$$\xi(\rho_1, \rho_2) \geq \widetilde{\xi}(\rho_1, \rho_2), \tag{178}$$

where the inequality is strict if ρ_1 and ρ_2 are invertible and do not commute (see [21, Theorem 2.1]), it follows that the lower bound in (172) cannot be optimal in general.

It is also known from [39, 41] that when $r = 2$, we have

$$\inf_{\ell \in \mathbb{N}} \frac{1}{\ell} \xi_{\max}(\rho_1^{\otimes \ell}, \rho_2^{\otimes \ell}) \geq \widehat{\xi}(\rho_1, \rho_2) := \sup_{s \in (0,1)} - \ln \widehat{Q}_s(\rho_1, \rho_2), \tag{179}$$

where

$$\widehat{Q}_s(\rho_1, \rho_2) := \mathrm{Tr}\Big[\rho_2 \left(\rho_2^{-1/2} \tilde{\rho}_1 \rho_2^{-1/2}\right)^s\Big]. \tag{180}$$

Here $\tilde{\rho}_1$ is the absolutely continuous part of ρ_1 with respect to ρ_2 [2], and the negative power of ρ_2 is taken in on its support (see also [29, Proposition 66]). Since the optimal error exponent is known in this case to be $\xi(\rho_1, \rho_2)$ given in (32), and it is also known from [39, 41] that

$$\xi(\rho_1, \rho_2) \leq \widehat{\xi}(\rho_1, \rho_2), \tag{181}$$

where the inequality is strict if ρ_1 and ρ_2 are invertible and do not commute (see [23, Theorem 4.3]), it follows that the upper bound in (172) cannot be the tightest possible upper bound in general.

6 Single-letter semi-definite programming upper bound on the optimal error exponent for antidistinguishability

In this section, we derive a single-letter semi-definite programming upper bound on the optimal error exponent. Let us begin by recalling that the minimum error probability of antidistinguishability of an ensemble $\mathcal{E} := \{(\eta_i, \rho_i) : i \in [r]\}$ can also be expressed in terms of the following primal and dual semi-definite programs [7, Section II] (see also [63, Eq. (III.15)]):

$$\mathrm{Err}(\mathcal{E}) = \inf_{\{M_i\}_{i \in [r]}} \left\{ \sum_{i \in [r]} \eta_i \, \mathrm{Tr}[M_i \rho_i] : M_i \geq 0 \quad \text{for all } i \in [r] \, , \sum_{i \in [r]} M_i = \mathbb{I} \right\} \tag{182}$$

$$= \sup_{Y \in \mathrm{Herm}} \left\{ \mathrm{Tr}[Y] : Y \leq \eta_i \rho_i \quad \text{for all } i \in [r] \right\}, \tag{183}$$

where Herm denotes the set of Hermitian operators. The equality holds as a consequence of Slater's condition; indeed, we see this by noting that $M_i = \mathbb{I}/r$ is strictly feasible for the primal and $Y = 0$ is feasible for the dual. Defining $\eta_{\min} := \min_{i \in [r]} \eta_i$, then it follows that

$$\mathrm{Err}(\mathcal{E}) = \sup_{Y \in \mathrm{Herm}} \left\{ \mathrm{Tr}[Y] : Y \leq \eta_i \rho_i \ \ \forall i \in [r] \right\} \tag{184}$$

$$\geq \sup_{Y \in \text{Herm}} \{\text{Tr}[Y] : Y \leq \eta_{\min}\rho_i \ \ \forall i \in [r]\} \tag{185}$$

$$= \sup_{Y \in \text{Herm}} \{\text{Tr}[\eta_{\min} Y] : \eta_{\min} Y \leq \eta_{\min}\rho_i \ \ \forall i \in [r]\} \tag{186}$$

$$= \eta_{\min} \cdot \sup_{Y \in \text{Herm}} \{\text{Tr}[Y] : Y \leq \rho_i \ \ \forall i \in [r]\} \tag{187}$$

$$\geq \eta_{\min} \kappa(\rho_1, \ldots, \rho_r), \tag{188}$$

where

$$\kappa(\rho_1, \ldots, \rho_r) := \sup_{Y \in \text{Herm}} \{\text{Tr}[Y] : -\rho_i \leq Y \leq \rho_i \ \ \forall i \in [r]\}. \tag{189}$$

The first inequality follows because

$$Y \leq \eta_{\min}\rho_i \ \ \forall i \in [r] \quad \Rightarrow \quad Y \leq \eta_i \rho_i \ \ \forall i \in [r]. \tag{190}$$

The second equality follows because optimizing over all Hermitian Y is equivalent to optimizing over $\eta_{\min} Y$ since $\eta_{\min} > 0$. The third equality follows because $\eta_{\min} Y \leq \eta_{\min}\rho_i \Leftrightarrow Y \leq \rho_i$ and by factoring $\eta_{\min}$ out of the optimization. The final inequality follows because the optimization in the definition of $\kappa(\rho_1, \ldots, \rho_r)$ adds extra constraints.

The main advantage of the κ quantity over the antidistinguishability error probability itself is that it is supermultiplicative, as stated below. For this reason, we can use it to bound the error exponent.

Lemma 18 *For the tuples of states, $(\rho_1, \ldots, \rho_r)$ and $(\sigma_1, \ldots, \sigma_r)$, the following supermultiplicativity inequality holds*

$$\kappa(\rho_1 \otimes \sigma_1, \ldots, \rho_r \otimes \sigma_r) \geq \kappa(\rho_1, \ldots, \rho_r) \cdot \kappa(\sigma_1, \ldots, \sigma_r). \tag{191}$$

Proof Let $Y_\rho, Y_\sigma \in \text{Herm}$ satisfy $-\rho_i \leq Y_\rho \leq \rho_i$ and $-\sigma_i \leq Y_\sigma \leq \sigma_i$ for all $i \in [r]$. Now invoking Lemma 12.35 of [28], we conclude that, for all $i \in [r]$,

$$-\rho_i \otimes \sigma_i \leq Y_\rho \otimes Y_\sigma \leq \rho_i \otimes \sigma_i. \tag{192}$$

It then follows that

$$\text{Tr}[Y_\rho] \cdot \text{Tr}[Y_\sigma] = \text{Tr}[Y_\rho \otimes Y_\sigma] \tag{193}$$

$$\leq \sup_{Y \in \text{Herm}} \{\text{Tr}[Y] : -\rho_i \otimes \sigma_i \leq Y \leq \rho_i \otimes \sigma_i \ \ \forall i \in [r]\} \tag{194}$$

$$= \kappa(\rho_1 \otimes \sigma_1, \ldots, \rho_r \otimes \sigma_r). \tag{195}$$

Since the inequality holds for all Y_ρ and Y_σ satisfying the aforementioned constraints, we conclude (191). □

By applying the supermultiplicativity result inductively, combined with the development in (184)–(188), we conclude the following:

Theorem 19 *For states $\rho_1, \ldots, \rho_r$, the following upper bound holds for the asymptotic error exponent of quantum antidistinguishability:*

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \leq -\ln \kappa(\rho_1, \ldots, \rho_r). \tag{196}$$

Proof Consider that

$$\begin{aligned}
\mathrm{E}(\rho_1, \ldots, \rho_r) &= \liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^n) && (197)\\
&\leq \liminf_{n\to\infty} -\frac{1}{n} \ln \left(\eta_{\min}\kappa(\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n})\right) && (198)\\
&\leq \liminf_{n\to\infty} -\frac{1}{n} \ln \left(\kappa(\rho_1, \ldots, \rho_r)^n\right) && (199)\\
&= -\ln \kappa(\rho_1, \ldots, \rho_r). && (200)
\end{aligned}$$

The first inequality follows from (184)–(188). The second inequality follows from $\liminf -\frac{1}{n} \ln \eta_{\min} = 0$ and Lemma 18 applied inductively. □

The upper bound in (196) can be bounded from above by a quantity expressed in terms of the extended max-relative entropy, defined for a Hermitian operator X and a positive semi-definite operator σ as [62, Eqs. (14)–(16)]:

$$D_{\max}(X\|\sigma) := \ln \inf_{\lambda \geq 0} \{\lambda : -\lambda\sigma \leq X \leq \lambda\sigma\}. \tag{201}$$

If the support of X is not contained in the support of σ, then there is no finite $\lambda \geq 0$ such that the constraints above can be satisfied, and so $D_{\max}(X\|\sigma) = +\infty$ in this case. Also, whenever the support of X is contained in the support of σ, we have $D_{\max}(X\|\sigma) < +\infty$ and in this case,

$$D_{\max}(X\|\sigma) = \ln \left\|\sigma^{-\frac{1}{2}} X \sigma^{-\frac{1}{2}}\right\|_\infty, \tag{202}$$

where the inverse is understood to be taken on the support of σ. In Appendix H, we derive several fundamental properties of the extended max-relative entropy, including monotonicity, data processing, joint quasi-convexity, lower semi-continuity, non-negativity and faithfulness, and additivity, which we think are of independent interest. We also show that

$$D_{\max}(X\|\sigma) = \sup_{\varepsilon>0} D_{\max}(X\|\sigma + \varepsilon I). \tag{203}$$

Theorem 20 *For quantum states $\rho_1, \ldots, \rho_r$, the quantity $\kappa(\rho_1, \ldots, \rho_r)$ is bounded from below in terms of the extended max-relative entropy, as follows:*

$$\kappa(\rho_1, \ldots, \rho_r) \geq \exp\left(-\inf_{\omega \in \mathcal{D}'} \max_{i \in [r]} D_{\max}(\omega \| \rho_i)\right), \tag{204}$$

where $\mathcal{D}' := \left\{\omega : \omega = \omega^\dagger,\ \mathrm{Tr}[\omega] = 1\right\}$ is the set of all Hermitian operators with trace one. Consequently, we have

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \leq \inf_{\omega \in \mathcal{D}'} \max_{i \in [r]} D_{\max}(\omega \| \rho_i) \tag{205}$$

$$= \max_{\{s_i\}_{i \in [r]}} \inf_{\omega \in \mathcal{D}'} \sum_{i \in [r]} s_i D_{\max}(\omega \| \rho_i), \tag{206}$$

where $\{s_i\}_{i \in [r]}$ is a probability distribution.

Proof By the definition (189) and the fact that $Y = 0$ is always feasible for $\kappa(\rho_1, \ldots, \rho_r)$, we conclude that

$$\kappa(\rho_1, \ldots, \rho_r) = \sup_{Y \in \mathrm{Herm}} \{\mathrm{Tr}[Y] : -\rho_i \leq Y \leq \rho_i \ \ \forall i \in [r]\} \tag{207}$$

$$= \sup_{Y \in \mathrm{Herm}: \mathrm{Tr}[Y] \geq 0} \{\mathrm{Tr}[Y] : -\rho_i \leq Y \leq \rho_i \ \ \forall i \in [r]\} \tag{208}$$

$$= \sup_{\lambda \geq 0, \omega \in \mathcal{D}'} \{\mathrm{Tr}[\lambda\omega] : -\rho_i \leq \lambda\omega \leq \rho_i, \quad \forall i \in [r]\} \tag{209}$$

$$= \sup_{\lambda \geq 0, \omega \in \mathcal{D}'} \{\lambda : -\rho_i \leq \lambda\omega \leq \rho_i, \quad \forall i \in [r]\} \tag{210}$$

$$\geq \sup_{\lambda > 0, \omega \in \mathcal{D}'} \{\lambda : -\rho_i \leq \lambda\omega \leq \rho_i, \quad \forall i \in [r]\} \tag{211}$$

$$= \sup_{\lambda > 0, \omega \in \mathcal{D}'} \left\{\lambda : -\frac{1}{\lambda}\rho_i \leq \omega \leq \frac{1}{\lambda}\rho_i, \quad \forall i \in [r]\right\} \tag{212}$$

$$= \sup_{\lambda' > 0, \omega \in \mathcal{D}'} \left\{\frac{1}{\lambda'} : -\lambda'\rho_i \leq \omega \leq \lambda'\rho_i, \quad \forall i \in [r]\right\} \tag{213}$$

$$= \left[\inf_{\lambda' > 0, \omega \in \mathcal{D}'} \{\lambda' : -\lambda'\rho_i \leq \omega \leq \lambda'\rho_i, \quad \forall i \in [r]\}\right]^{-1} \tag{214}$$

$$= \left[\inf_{\omega \in \mathcal{D}'} \exp\left(\max_{i \in [r]} D_{\max}(\omega \| \rho_i)\right)\right]^{-1} \tag{215}$$

$$= \left[\exp\left(\inf_{\omega \in \mathcal{D}'} \max_{i \in [r]} D_{\max}(\omega \| \rho_i)\right)\right]^{-1} \tag{216}$$

$$= \exp\left(-\inf_{\omega \in \mathcal{D}'} \max_{i \in [r]} D_{\max}(\omega \| \rho_i)\right). \tag{217}$$

The equality (207) follows because $Y = 0$ is feasible in (189). The equality (209) follows because for every Hermitian operator Y with positive trace, we can choose $\lambda = \mathrm{Tr}[Y]$ and $\omega = Y/\mathrm{Tr}[Y] \in \mathcal{D}'$ so that $Y = \lambda\omega$; and if $Y = 0$ then we can choose $\lambda = 0$ and $\omega = \mathbb{I}/\dim(\mathcal{H}) \in \mathcal{D}'$ so that $Y = \lambda\omega$. The equality (213) follows from the substitution $\lambda = \frac{1}{\lambda'}$.

The desired inequality (205) is a direct consequence of (196) and (204). Also, we have

$$\inf_{\omega\in\mathcal{D}'} \max_{i\in[r]} D_{\max}(\omega\|\rho_i) = \inf_{\omega\in\mathcal{D}'} \max_{\{s_i\}_{i\in[r]}} \sum_{i\in[r]} s_i D_{\max}(\omega\|\rho_i) \tag{218}$$

$$= \max_{\{s_i\}_{i\in[r]}} \inf_{\omega\in\mathcal{D}'} \sum_{i\in[r]} s_i D_{\max}(\omega\|\rho_i). \tag{219}$$

The first equality follows because the maximum over a finite set can be replaced with a maximum of the expected value of the elements of the set, with the maximum taken over all possible distributions. The second equality follows from an application of Sion's minimax theorem [55]. Indeed, if $\cap_{i\in[r]} \mathrm{supp}(\rho_i) \neq \emptyset$, then the infima in (218) and (219) can be restricted to a smaller set $\mathcal{D}'' := \{\omega \in \mathcal{D}' : \mathrm{supp}(\omega) \subseteq \cap_{i\in[r]} \mathrm{supp}(\rho_i)\}$ so that $\sum_{i\in[r]} s_i D_{\max}(\omega\|\rho_i)$ is finite for all $\omega \in \mathcal{D}''$ and every probability distribution $\{s_i\}_{i\in[r]}$. Also, the objective function $\sum_{i\in[r]} s_i D_{\max}(\omega\|\rho_i)$ is linear and continuous in the probability distribution $\{s_i\}_{i\in[r]}$, and it is lower semi-continuous and quasi-convex in $\omega \in \mathcal{D}''$ (Appendix H). Sion's minimax theorem thus applies and gives the equality (219). In the case when $\cap_{i\in[r]} \mathrm{supp}(\rho_i) = \emptyset$, both the sides of (219) are infinity and the equality holds trivially. □

Remark 4 By replacing the set $\mathcal{D}'$ with $\mathcal{D}$ (the set of density operators) in Theorem 20, we get an interesting (although weaker) upper bound on the optimal error exponent:

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \leq \max_{\{s_i\}_{i\in[r]}} \inf_{\omega\in\mathcal{D}} \sum_{i\in[r]} s_i D_{\max}(\omega\|\rho_i). \tag{220}$$

This upper bound has a resemblance to the following divergence:

$$\max_{\{s_i\}_{i\in[r]}} \inf_{\omega\in\mathcal{D}} \sum_{i\in[r]} s_i D(\omega\|\rho_i) = \max_{\{s_i\}_{i\in[r]}} \left(-\ln \mathrm{Tr}\left[\exp\left(\sum_{i\in[r]} s_i \ln \rho_i \right) \right] \right), \tag{221}$$

where the equality follows whenever each ρ_i is positive definite. Indeed, the only difference between (220) and (221) is the substitution $D_{\max}(\rho\|\sigma) \to D(\rho\|\sigma) := \mathrm{Tr}[\rho(\ln\rho - \ln\sigma)]$, where the latter denotes the standard quantum relative entropy [60]. The equality in (221) was established in Eq. (V.121) and Example V.25 of [42]. See Appendix I for a review of the proof of (221). Finally, note that (221) reduces to the multivariate classical Chernoff divergence when the states in the set $\{\rho_i\}_{i\in[r]}$ commute (have a common eigenbasis). As such, this quantity is a multivariate quantum Chernoff divergence according to Definition 13.

We end the section by deriving another alternative form for the κ quantity.

Proposition 21 *The quantity $\kappa(\rho_1, \ldots, \rho_r)$ can alternatively be written as:*

$$\kappa(\rho_1, \ldots, \rho_r) = \inf_{\substack{Z_{1,i}, Z_{2,i} \geq 0 \\ \forall i \in [r]}} \left\{ \sum_{i\in[r]} \operatorname{Tr}[\left(Z_{1,i} + Z_{2,i}\right) \rho_i] : \mathbb{I} = \sum_{i\in[r]} Z_{2,i} - Z_{1,i} \right\}. \tag{222}$$

Proof We prove this by showing that the expression on the right-hand side of (222) is the dual SDP of $\kappa(\rho_1, \ldots, \rho_r)$ and that the strong duality holds. We derive it as follows:

$$\sup_{Y\in\text{Herm}} \{\operatorname{Tr}[Y] : -\rho_i \leq Y \leq \rho_i \;\; \forall i \in [r]\}$$

$$= \sup_{Y\in\text{Herm}} \left\{ \operatorname{Tr}[Y] + \inf_{Z_{1,i}, Z_{2,i} \geq 0} \left\{ \sum_{i\in[r]} \left(\operatorname{Tr}[Z_{1,i}\,(Y + \rho_i)] + \operatorname{Tr}[Z_{2,i}\,(\rho_i - Y)]\right) \right\} \right\} \tag{223}$$

$$= \sup_{Y\in\text{Herm}} \inf_{Z_{1,i}, Z_{2,i} \geq 0} \left\{ \operatorname{Tr}[Y] + \sum_{i\in[r]} \left(\operatorname{Tr}[Z_{1,i}\,(Y + \rho_i)] + \operatorname{Tr}[Z_{2,i}\,(\rho_i - Y)]\right) \right\} \tag{224}$$

$$= \sup_{Y\in\text{Herm}} \inf_{Z_{1,i}, Z_{2,i} \geq 0} \left\{ \operatorname{Tr}\left[Y \left(\mathbb{I} + \sum_{i\in[r]} (Z_{1,i} - Z_{2,i}) \right) \right] + \sum_{i\in[r]} \operatorname{Tr}[\left(Z_{1,i} + Z_{2,i}\right) \rho_i] \right\} \tag{225}$$

$$\leq \inf_{Z_{1,i}, Z_{2,i} \geq 0} \sup_{Y\in\text{Herm}} \left\{ \operatorname{Tr}\left[Y \left(\mathbb{I} + \sum_{i\in[r]} (Z_{1,i} - Z_{2,i}) \right) \right] + \sum_{i\in[r]} \operatorname{Tr}[\left(Z_{1,i} + Z_{2,i}\right) \rho_i] \right\} \tag{226}$$

$$= \inf_{Z_{1,i}, Z_{2,i} \geq 0} \left\{ \sum_{i\in[r]} \operatorname{Tr}[\left(Z_{1,i} + Z_{2,i}\right) \rho_i] : \mathbb{I} = \sum_{i\in[r]} Z_{2,i} - Z_{1,i} \right\}. \tag{227}$$

Strong duality holds here by picking $Z_{2,i} = 2\mathbb{I}/r$ and $Z_{1,i} = \mathbb{I}/r$ for all $i \in [r]$ in the dual and by picking $Y = 0$ for the primal. □

7 Conclusion

Summary We have solved the classical antidistinguishability problem of finding the optimal error exponent, which we proved to be equal to the multivariate classical Chernoff divergence of the given probability measures. To the best of our knowledge, this result constitutes the first operational interpretation of the divergence involving three or more states. We have also given various upper and lower bounds on the optimal error exponent in the quantum case, while it still remains an open problem to compute its exact expression. In analogy with the classical case, we believe that the quantity that gives the exact error exponent in the quantum case should be called the multivariate quantum Chernoff divergence.

Future directions Recall from [7] that quantum m-state exclusion can be thought of as antidistinguishability of a set of states related to the original set. We leave it as an intriguing open question to determine the optimal asymptotic error exponent for quantum m-state exclusion.

Analogous to the task of antidistinguishing quantum states, one may consider the problem of antidistinguishing an ensemble of quantum channels. In this problem, a quantum channel is chosen randomly from a finite set of quantum channels, with known a priori probability distribution. The antidistinguisher is allowed to pass one share of a bipartite quantum state through the channel, after which both the reference system and the channel output system are measured. Based on the measurement outcome, the antidistinguisher's goal is to rule out a quantum channel other than the selected one. It would be an interesting future work to study the asymptotics of the error rates for antidistinguishing an ensemble of quantum channels.

Appendix A: Expectation values at non-corner points

We begin by stating a known property of convex functions in the lemma below. We include a proof of the statement for the sake of completeness.

Lemma 22 *Let $a > 0$ be arbitrary. Let $f : [0, a] \to \mathbb{R}$ be a convex and continuous function on $[0, a]$, and suppose f is differentiable on $(0, a)$. Then, the one-sided derivative*

$$f'_+(0) := \lim_{t \searrow 0} \frac{f(t) - f(0)}{t} \tag{A1}$$

exists and fulfills

$$f'_+(0) = \lim_{t \searrow 0} f'(t)\,. \tag{A2}$$

Here $f'_+(0)$ is either finite or takes the value $-\infty$; if f takes its minimum value at 0, then $f'_+(0)$ is finite and $f'_+(0) \geq 0$.

Proof The map $t \mapsto (f(t) - f(0))/t$ defined on $(0, a)$ is non-decreasing. See [9, Section 2.1, Exercise 7]). Also, the limit in (A1) exists in $\mathbb{R} \cup \{-\infty\}$ [9, Proposition 3.1.2]. By the Lagrange mean-value theorem, for any $t \in (0, a)$ there exists $u_t \in (0, t)$ such that

$$\frac{f(t) - f(0)}{t} = f'(u_t). \tag{A3}$$

We know that f being convex, its derivative is a non-decreasing function on $(0, a)$. We thus get from (A3) that

$$f'_+(0) = \lim_{t \searrow 0} f'(t), \tag{A4}$$

with a possible value $-\infty$. If f is minimized at 0, then we have $f(t) - f(0) \geq 0$ for all $t \in (0, a)$. It then directly follows from the definition (A1) that $f'_+(0) \geq 0$. □

Lemma 23 *For $\mathbf{t} \in \mathbb{T}_r^1$ and $i \in [r-1]$, the expectation value $\mathbb{E}_{\mathbf{t}}[q_i]$ exists in $\mathbb{R} \cup \{-\infty\}$ and satisfies*

$$\partial_i^+ \mathrm{K}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}}[q_i]. \tag{A5}$$

Proof Recall that $\mathbb{T}_r^1$ is the set of non-corner points of $\mathbb{T}_r$ given by (70). Let $\mathbf{t} \in \mathbb{T}_r^1$. Define a set

$$B_{\mathbf{t}} := \{i \in [r-1] : t_i > 0\}, \tag{A6}$$

and let $B_{\mathbf{t}}^c := [r-1] \backslash B_{\mathbf{t}}$. Let β denote the cardinality of the set $B_{\mathbf{t}}$. We emphasize that if $B_{\mathbf{t}} \neq \emptyset$ so that $\beta \geq 1$, $\mathbf{t}$ corresponds to an interior point of $\mathbb{T}_{\beta+1}$, which is the β-vector obtained by discarding the zero entries of $\mathbf{t}$. This allows us to use properties of the exponential family of densities given in (61). So, if $i \in B_{\mathbf{t}}$ so that $B_{\mathbf{t}} \neq \emptyset$ then by similar arguments as given for (67), it follows that the expectation value $\mathbb{E}_{\mathbf{t}}[q_i]$ exists, and it satisfies $\partial_i \mathrm{K}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}}[q_i]$. It remains to show for $i \in B_{\mathbf{t}}^c$ that $\mathbb{E}_{\mathbf{t}}[q_i]$ exists, and it is equal to $\partial_i^+ \mathrm{K}(\mathbf{t})$. Let us fix an arbitrary index $i \in B_{\mathbf{t}}^c$. Choose a small number $\varepsilon > 0$ such that $\mathbf{t} + h\mathbf{e}_i \in \mathbb{T}_r^1$ for all $h \in [0, \varepsilon]$. The function $h \mapsto \mathrm{K}(\mathbf{t} + h\mathbf{e}_i)$ is continuous, convex on $[0, \varepsilon]$, and it is differentiable on $(0, \varepsilon)$. Lemma 22 thus implies that

$$\partial_i^+ \mathrm{K}(\mathbf{t}) = \lim_{h \searrow 0} \partial_i \mathrm{K}(\mathbf{t} + h\mathbf{e}_i) = \lim_{h \searrow 0} \mathbb{E}_{\mathbf{t}+h\mathbf{e}_i}[q_i]. \tag{A7}$$

Here we used the relation $\partial_i \mathrm{K}(\mathbf{t} + h\mathbf{e}_i) = \mathbb{E}_{\mathbf{t}+h\mathbf{e}_i}[q_i]$ proved earlier. We now claim that $\mathbb{E}_{\mathbf{t}}[q_i]$ exists and satisfies

$$\lim_{h \searrow 0} \mathbb{E}_{\mathbf{t}+h\mathbf{e}_i}[q_i] = \mathbb{E}_{\mathbf{t}}[q_i] \tag{A8}$$

with a possible value of $-\infty$. Indeed, we have

$$\mathbb{E}_{\mathbf{t}+h\mathbf{e}_i}[q_i] = \frac{1}{\mathrm{H}(\mathbf{t} + h\mathbf{e}_i)} \int_D \mathbf{d}\mu \; q_i p_r \exp\left(\sum_{j \in [r-1]} t_j q_j + h q_i \right). \tag{A9}$$

By continuity of H, we have $\mathrm{H}(\mathbf{t} + h\mathbf{e}_i) \to \mathrm{H}(\mathbf{t})$ as $h \searrow 0$. Thus, for (A8) to hold, it suffices to prove that

$$\lim_{h\searrow 0} \int_D \mathbf{d}\mu \; q_i \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j + h q_i\right) = \int_D \mathbf{d}\mu \; q_i \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j\right). \tag{A10}$$

Let $q_i = q_i^+ - q_i^-$, where q_i^+ and q_i^- are non-negative functions with mutually disjoint supports. This gives

$$\int_D \mathbf{d}\mu \; q_i \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j + h q_i\right) = \int_D \mathbf{d}\mu \; q_i^+ \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j + h q_i^+\right) - \int_D \mathbf{d}\mu \; q_i^- \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j - h q_i^-\right). \tag{A11}$$

Both integral terms in the right-hand side of (A11) are finite, because for $h \in (0, \varepsilon)$, the left-hand side is finite. Indeed then $\mathbf{t} + h\mathbf{e}_i$ corresponds to an interior point of $\mathbb{T}_{r-\beta+1}$ so that the properties of an exponential family of densities apply. Consider now the first integral term on the right-hand side of (A11). We have the pointwise monotone convergence on D

$$q_i^+ \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j + h q_i^+\right) \searrow q_i^+ \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j\right) \qquad \text{as } h \searrow 0. \tag{A12}$$

By the monotone convergence theorem, we have

$$\lim_{h\searrow 0} \int_D \mathbf{d}\mu \; q_i^+ \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j + h q_i^+\right) = \int_D \mathbf{d}\mu \; q_i^+ \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j\right) < \infty \tag{A13}$$

where the limit is finite because the integrand is nonnegative. We now consider the second integral term on the right-hand side of (A11). We have the pointwise monotone convergence on D

$$q_i^- \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j - h q_i^-\right) \nearrow q_i^- \, p_r \exp\left(\sum_{j\in[r-1]} t_j q_j\right), \qquad \text{as } h \searrow 0. \tag{A14}$$

By the monotone convergence theorem, we get

$$\lim_{h \searrow 0} \int_D \mathbf{d}\mu \; q_i^- \, p_r \exp\left(\sum_{j \in [r-1]} t_j q_j - h q_i^- \right) = \int_D \mathbf{d}\mu \; q_i^- \, p_r \exp\left(\sum_{j \in [r-1]} t_j q_j \right) \tag{A15}$$

regardless of whether the right-hand integral in (A15) is finite or infinite. The latter point is explicitly stressed in Theorem 16.2 of [6]. By taking the limit $h \searrow 0$ in (A11) and then using (A7), (A13), and (A15), we get

$$\partial_i^+ \mathrm{K}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}}[q_i^+] - \mathbb{E}_{\mathbf{t}}[q_i^-] = \mathbb{E}_{\mathbf{t}}[q_i]. \tag{A16}$$

Since $\mathbb{E}_{\mathbf{t}}[q_i^+]$ is a real number, $\mathbb{E}_{\mathbf{t}}[q_i]$ takes a value in $\mathbb{R} \cup \{-\infty\}$. If $\mathbf{t}$ is a minimizer of K, then by Lemma 22 we have $\partial_i^+ \mathrm{K}(\mathbf{t}) \geq 0$, and hence, $\mathbb{E}_{\mathbf{t}}[q_i]$ is finite. We have thus accomplished that if $\mathbf{t} \in \mathbb{T}_r^1$ is a minimizer of K and $i \in [r-1]$, then the expectation value $\mathbb{E}_{\mathbf{t}}[q_i]$ exists, is finite, and satisfies $\partial_i^+ \mathrm{K}(\mathbf{t}) = \mathbb{E}_{\mathbf{t}}[q_i]$. □

Appendix B: Proof of Equation (141)

Proposition 24 *For arbitrary (not necessarily normalized) vectors $|\varphi\rangle, |\zeta\rangle \in \mathcal{H}$, the following equality holds:*

$$\| |\varphi\rangle\langle\varphi| - |\zeta\rangle\langle\zeta| \|_1^2 = (\langle\varphi|\varphi\rangle + \langle\zeta|\zeta\rangle)^2 - 4 \, |\langle\zeta|\varphi\rangle|^2 \, . \tag{B1}$$

Proof The equality (B1) trivially holds if one of the vectors is zero. So, we assume that both $|\varphi\rangle$ and $|\zeta\rangle$ are nonzero vectors. Define

$$|\varphi'\rangle := \frac{|\varphi\rangle}{\| |\varphi\rangle \|}, \qquad |\zeta'\rangle := \frac{|\zeta\rangle}{\| |\zeta\rangle \|}. \tag{B2}$$

Then, the desired equality is equivalent to

$$\left\| c|\varphi'\rangle\langle\varphi'| - d|\zeta'\rangle\langle\zeta'| \right\|_1^2 = (c+d)^2 - 4cd \left| \langle\zeta'|\varphi'\rangle \right|^2 , \tag{B3}$$

where

$$c := \| |\varphi\rangle \|^2 , \qquad d := \| |\zeta\rangle \|^2 . \tag{B4}$$

Defining $|\varphi^\perp\rangle$ to be the unit vector orthogonal to $|\varphi'\rangle$ in span $\{|\varphi'\rangle, |\zeta'\rangle\}$, we find that

$$|\zeta'\rangle = \cos(\theta)|\varphi'\rangle + \sin(\theta)|\varphi^\perp\rangle, \tag{B5}$$

where

$$\cos(\theta) = \langle\varphi'|\zeta'\rangle. \tag{B6}$$

Then, it follows that

$$
\begin{aligned}
& c|\varphi'\rangle\langle\varphi'| - d|\zeta'\rangle\langle\zeta'| \\
&\quad = c|\varphi'\rangle\langle\varphi'| - d\left(\cos(\theta)|\varphi'\rangle + \sin(\theta)|\varphi^{\perp}\rangle\right)\left(\cos(\theta)\langle\varphi'| + \sin(\theta)\langle\varphi^{\perp}|\right) && \text{(B7)} \\
&\quad = \left[c - d\cos^2(\theta)\right]|\varphi'\rangle\langle\varphi'| - d\sin(\theta)\cos(\theta)|\varphi^{\perp}\rangle\langle\varphi'| \\
&\qquad - d\sin(\theta)\cos(\theta)|\varphi'\rangle\langle\varphi^{\perp}| - d\sin^2(\theta)|\varphi^{\perp}\rangle\langle\varphi^{\perp}|. && \text{(B8)}
\end{aligned}
$$

As a matrix with respect to the basis $\left\{|\varphi'\rangle, |\varphi^{\perp}\rangle\right\}$, the last line has the following form:

$$
\begin{bmatrix} c - d\cos^2(\theta) & -d\sin(\theta)\cos(\theta) \\ -d\sin(\theta)\cos(\theta) & -d\sin^2(\theta) \end{bmatrix}, \tag{B9}
$$

and this matrix has the following eigenvalues:

$$
\lambda_1 = \frac{1}{2}\left(c - d + \sqrt{(c+d)^2 - 4cd\cos^2(\theta)}\right), \tag{B10}
$$

$$
\lambda_2 = \frac{1}{2}\left(c - d - \sqrt{(c+d)^2 - 4cd\cos^2(\theta)}\right). \tag{B11}
$$

Note that $c \geq 0$ and $d \geq 0$. Without loss of generality, suppose that $c \geq d$. Then

$$
\begin{aligned}
0 &\leq 4cd\sin^2(\theta) && \text{(B12)} \\
&= 4cd\left(1 - \cos^2(\theta)\right) && \text{(B13)} \\
\Rightarrow \quad -2cd &\leq 2cd - 4cd\cos^2(\theta) && \text{(B14)} \\
\Rightarrow \quad c^2 - 2cd + d^2 &\leq c^2 + 2cd + d^2 - 4cd\cos^2(\theta) && \text{(B15)} \\
\Rightarrow \quad (c-d)^2 &\leq (c+d)^2 - 4cd\cos^2(\theta) && \text{(B16)} \\
\Rightarrow \quad c - d &\leq \sqrt{(c+d)^2 - 4cd\cos^2(\theta)}. && \text{(B17)}
\end{aligned}
$$

Then, it follows that the square of the trace norm of $c|\varphi'\rangle\langle\varphi'| - d|\zeta'\rangle\langle\zeta'|$ is given by:

$$
\begin{aligned}
& \left\| c|\varphi'\rangle\langle\varphi'| - d|\zeta'\rangle\langle\zeta'| \right\|_1^2 \\
&\quad = (|\lambda_1| + |\lambda_2|)^2 && \text{(B18)} \\
&\quad = \left(\frac{1}{2}\left(c - d + \sqrt{(c+d)^2 - 4cd\cos^2(\theta)}\right) - \frac{1}{2}\left(c - d - \sqrt{(c+d)^2 - 4cd\cos^2(\theta)}\right)\right)^2 && \text{(B19)} \\
&\quad = (c+d)^2 - 4cd\cos^2(\theta), && \text{(B20)}
\end{aligned}
$$

concluding the proof. □

Appendix C: Proof of Proposition 14

To prove the data-processing inequality, let $\mathcal{N}$ be an arbitrary quantum channel. We denote by $\mathcal{N}(\mathcal{E})$ the ensemble $\{(\eta_i, \mathcal{N}(\rho_i)) : i \in [r]\}$, which results from applying the channel $\mathcal{N}$ to each state in $\mathcal{E}$. The optimal antidistinguishability error probability for the ensemble $\mathrm{Err}(\mathcal{E})$ is not more than that for the ensemble $\mathcal{N}(\mathcal{E})$. To see this, let $\mathscr{M} = \{M_1, \ldots, M_r\}$ be an arbitrary POVM. We have

$$\mathrm{Err}(\mathscr{M}; \mathcal{N}(\mathcal{E})) = \sum_{i\in[r]} \eta_i \operatorname{Tr}[M_i \mathcal{N}(\rho_i)] \tag{C1}$$

$$= \sum_{i\in[r]} \eta_i \operatorname{Tr}[\mathcal{N}^\dagger(M_i)\rho_i] \tag{C2}$$

$$\geq \mathrm{Err}(\mathcal{E}). \tag{C3}$$

The inequality (C3) follows because $\{\mathcal{N}^\dagger(M_1), \ldots, \mathcal{N}^\dagger(M_r)\}$ is a POVM. Since (C3) holds for every POVM $\mathscr{M}$, we have

$$\mathrm{Err}(\mathcal{E}) \leq \mathrm{Err}(\mathcal{N}(\mathcal{E})). \tag{C4}$$

Therefore, for all $n \in \mathbb{N}$, we get

$$-\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^n) \geq -\frac{1}{n} \ln \mathrm{Err}(\mathcal{N}(\mathcal{E})^n), \tag{C5}$$

which implies

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \geq \mathrm{E}(\mathcal{N}(\rho_1), \ldots, \mathcal{N}(\rho_r)). \tag{C6}$$

Now, suppose that the states in the given ensemble commute with each other. The following arguments show that the optimal error of antidistinguishing the given states is equal to that of the induced probability measures. Let $P_1, \ldots, P_r$ be the probability measures on the discrete space $[\dim(\mathcal{H})]$ induced by the states in a common eigenbasis as defined in (161), and let $\mathcal{E}_{\mathrm{cl}}$ be the classical ensemble $\{(\eta_i, P_i) : i \in [r]\}$. Suppose $p_1, \ldots, p_r$ are the corresponding densities of the probability measures with respect to the counting measure μ. This gives the following representation of each state:

$$\rho_i = \int_{[\dim(\mathcal{H})]} \mathbf{d}\mu(\omega)\ p_i(\omega) |\omega\rangle\langle\omega|, \qquad i \in [r]. \tag{C7}$$

We have

$$\mathrm{Err}(\mathscr{M}; \mathcal{E}) = \sum_{i\in[r]} \eta_i \operatorname{Tr}[M_i \rho_i] \tag{C8}$$

$$= \sum_{i\in[r]} \eta_i \operatorname{Tr}\left[M_i \left(\int_{[\dim(\mathcal{H})]} \mathbf{d}\mu(\omega)\ p_i(\omega) |\omega\rangle\langle\omega| \right) \right] \tag{C9}$$

$$= \int_{[\dim(\mathcal{H})]} \mathbf{d}\mu(\omega) \sum_{i\in[r]} \langle\omega|M_i|\omega\rangle \eta_i p_i(\omega) \tag{C10}$$

$$= \mathrm{Err}_{\mathrm{cl}}(\delta; \mathcal{E}_{\mathrm{cl}}), \tag{C11}$$

where δ is the decision rule given by $\delta(\omega) := (\langle\omega|M_1|\omega\rangle, \ldots, \langle\omega|M_r|\omega\rangle)$. We note here that for any POVM $\mathscr{M}$, there corresponds a decision rule δ that satisfies (C8)–(C11). Conversely, given any decision rule δ for antidistinguishing the classical ensemble $\mathcal{E}_{\mathrm{cl}}$ there corresponds a POVM $\mathscr{M} = \{M_1, \ldots, M_r\}$, given by

$$M_i := \int_{[\dim(\mathcal{H})]} \mathbf{d}\mu(\omega)\, \delta_i(\omega)|\omega\rangle\langle\omega|, \tag{C12}$$

that satisfies (C8)–(C11). This then implies

$$\inf_{\mathscr{M}} \mathrm{Err}(\mathscr{M}; \mathcal{E}) = \inf_{\delta} \mathrm{Err}(\delta; \mathcal{E}_{\mathrm{cl}}), \tag{C13}$$

where the infima are taken over all POVMs $\mathscr{M}$ and decision rules δ corresponding to the given quantum and classical ensembles, respectively. We have thus proved that

$$\mathrm{Err}(\mathcal{E}) = \mathrm{Err}_{\mathrm{cl}}(\mathcal{E}_{\mathrm{cl}}), \tag{C14}$$

which directly implies

$$\mathrm{E}(\rho_1, \ldots, \rho_r) = \mathrm{E}_{\mathrm{cl}}(P_1, \ldots, P_r). \tag{C15}$$

Appendix D: Proof of Proposition 15

Define a map $\xi' : \mathcal{D}^r \to [0, \infty]$ by

$$\xi'(\rho_1, \ldots, \rho_r) := \sup_{\mathcal{M}} \xi_{\mathrm{cl}}(P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}) \tag{D1}$$

as given on the right-hand side of (167). We first show that ξ' is a lower bound on any multivariate Chernoff divergence. Let $\xi : \mathcal{D}^r \to [0, \infty]$ be any multivariate quantum Chernoff divergence and $\rho_1, \ldots, \rho_r$ be arbitrary quantum states. For any measurement channel $\mathcal{M}$, we have

$$\xi(\rho_1, \ldots, \rho_r) \geq \xi(\mathcal{M}(\rho_1), \ldots, \mathcal{M}(\rho_r)) = \xi_{\mathrm{cl}}(P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}). \tag{D2}$$

Here we used the assumptions that ξ satisfies the data-processing inequality and reduces to the multivariate classical Chernoff divergence for commuting states. Since the inequality (D2) holds for an arbitrary measurement channel $\mathcal{M}$, taking the supremum over $\mathcal{M}$ gives

$$\xi(\rho_1, \ldots, \rho_r) \geq \xi'(\rho_1, \ldots, \rho_r). \tag{D3}$$

We now show that ξ' is a multivariate quantum Chernoff divergence, i.e., it satisfies the data-processing inequality and reduces to the multivariate classical Chernoff divergence for commuting states. Consider a quantum channel $\mathcal{N}$ and any measurement channel $\mathcal{M}$ corresponding to a POVM $\{M_1, \ldots, M_t\}$ on the output Hilbert space of the channel $\mathcal{N}$. Let $\mathcal{M}_{\mathcal{N}}$ be the measurement channel corresponding to the POVM $\{\mathcal{N}^\dagger(M_1), \ldots, \mathcal{N}^\dagger(M_t)\}$. Let $P_1^{\mathcal{M}_{\mathcal{N}}}, \ldots, P_r^{\mathcal{M}_{\mathcal{N}}}$ denote the probability measures induced by $\mathcal{M}_{\mathcal{N}}$ corresponding to the states $\rho_1, \ldots, \rho_r$ as given in the development (165)–(166). Similarly, let $Q_1^{\mathcal{M}}, \ldots, Q_r^{\mathcal{M}}$ denote the probability measures induced by $\mathcal{M}$ corresponding to the states $\mathcal{N}(\rho_1), \ldots, \mathcal{N}(\rho_r)$. Since $\mathrm{Tr}[M_j \mathcal{N}(\rho_i)] = \mathrm{Tr}[\mathcal{N}^\dagger(M_j)(\rho_i)]$ for all i, j, it follows that $Q_i^{\mathcal{M}} = P_i^{\mathcal{M}_{\mathcal{N}}}$ for $i \in [r]$. This implies

$$\xi'(\mathcal{N}(\rho_1), \ldots, \mathcal{N}(\rho_r)) = \sup_{\mathcal{M}} \xi_{\mathrm{cl}}(Q_1^{\mathcal{M}}, \ldots, Q_r^{\mathcal{M}}) \tag{D4}$$
$$= \sup_{\mathcal{M}} \xi_{\mathrm{cl}}(P_1^{\mathcal{M}_{\mathcal{N}}}, \ldots, P_r^{\mathcal{M}_{\mathcal{N}}}) \tag{D5}$$
$$\leq \xi'(\rho_1, \ldots, \rho_r), \tag{D6}$$

which means that ξ' satisfies the data-processing inequality. In the case when the states $\rho_1, \ldots, \rho_r$ commute, Theorem 6 and Proposition 14 give the following classical data-processing inequality

$$\xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r) \geq \xi_{\mathrm{cl}}(P_1^{\mathcal{M}}, \ldots, P_r^{\mathcal{M}}). \tag{D7}$$

Also, the inequality in (D7) is saturated for the measurement channel corresponding to a common eigenbasis of the commuting states. Therefore, we get

$$\xi'(\rho_1, \ldots, \rho_r) = \xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r). \tag{D8}$$

We thus conclude that ξ' is the minimal multivariate quantum Chernoff divergence.

Appendix E: Proof of Proposition 16

Define a map $\xi'' : \mathcal{D}^r \to [0, \infty]$ by

$$\xi''(\rho_1, \ldots, \rho_r) := \inf_{(\mathcal{P}, \{P_i\}_{i\in[r]})} \{\xi_{\mathrm{cl}}(P_1, \ldots, P_r) : \mathcal{P}(P_i) = \rho_i \quad \text{for all } i \in [r]\}, \tag{E1}$$

as given on the right-hand side of (170). We first show that ξ'' is an upper bound on any multivariate Chernoff divergence. Let $\xi : \mathcal{D}^r \to [0, \infty]$ be any multivariate quantum Chernoff divergence, and let $\rho_1, \ldots, \rho_r$ be arbitrary quantum states. Given a preparation channel $\mathcal{P}$ and probability measures $P_1, \ldots, P_r$ satisfying

$$\mathcal{P}(P_i) = \rho_i, \qquad \text{for } i \in [r], \tag{E2}$$

we have

$$\xi(\rho_1, \ldots, \rho_r) = \xi(\mathcal{P}(P_1), \ldots, \mathcal{P}(P_r)) \leq \xi_{\mathrm{cl}}(P_1, \ldots, P_r). \tag{E3}$$

In (E3), we used the assumptions that ξ satisfies the data-processing inequality and reduces to the multivariate classical Chernoff divergence for commuting states. By taking the infimum in (E3) over preparation channels and probability measures satisfying (E2), we thus get

$$\xi(\rho_1, \ldots, \rho_r) \leq \xi''(\rho_1, \ldots, \rho_r). \tag{E4}$$

We now show that ξ'' is a multivariate quantum Chernoff divergence, i.e., it satisfies the data-processing inequality and reduces to the multivariate classical Chernoff divergence for commuting states. Let $\mathcal{N}$ be any quantum channel. We have

$$\xi''(\mathcal{N}(\rho_1), \ldots, \mathcal{N}(\rho_r)) = \inf_{\substack{(\mathcal{P}, \{P_i\}_{i\in[r]}) \\ \mathcal{P}(P_i) = \mathcal{N}(\rho_i)}} \xi_{\mathrm{cl}}(P_1, \ldots, P_r) \tag{E5}$$

$$\leq \inf_{\substack{(\mathcal{P}, \{P_i\}_{i\in[r]}) \\ \mathcal{P}(P_i) = \rho_i}} \xi_{\mathrm{cl}}(P_1, \ldots, P_r) \tag{E6}$$

$$= \xi''(\rho_1, \ldots, \rho_r), \tag{E7}$$

where the inequality follows because for every preparation channel $\mathcal{P}$ satisfying $\mathcal{P}(P_i) = \rho_i$, its concatenation with $\mathcal{N}$ gives another preparation channel $\mathcal{N} \circ \mathcal{P}$ that satisfies $(\mathcal{N} \circ \mathcal{P})(P_i) = \mathcal{N}(\mathcal{P}(P_i)) = \mathcal{N}(\rho_i)$. If the states $\rho_1, \ldots, \rho_r$ commute, then by the classical data-processing inequality, for any preparation channel $\mathcal{P}$ and probability measures $P_1, \ldots, P_r$ satisfying (E2), we get

$$\xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r) = \xi_{\mathrm{cl}}(\mathcal{P}(P_1), \ldots, \mathcal{P}(P_r)) \leq \xi_{\mathrm{cl}}(P_1, \ldots, P_r). \tag{E8}$$

Also, the last inequality is equality for probability distributions prepared from a spectral decomposition of the commuting states in a common orthonormal basis. Therefore, we get

$$\xi''(\rho_1, \ldots, \rho_r) = \xi_{\mathrm{cl}}(\rho_1, \ldots, \rho_r). \tag{E9}$$

We thus conclude that ξ'' is the maximal multivariate quantum Chernoff divergence.

Appendix F: Additivity of the optimal error exponent

Lemma 25 *Let $\mathcal{E} = \{(\eta_i, \rho_i) : i \in [r]\}$ be an ensemble of states. The following equality holds*

$$\mathrm{E}(\rho_1, \ldots, \rho_r) = \frac{1}{\ell} \mathrm{E}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}) \quad \textit{for all } \ell \in \mathbb{N}, \tag{F1}$$

where $\mathrm{E}(\rho_1, \ldots, \rho_r)$ *is the optimal error exponent defined in* (31).

Proof First, we have that

$$\mathrm{E}(\rho_1, \ldots, \rho_r) \leq \frac{1}{\ell} \mathrm{E}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}) \qquad \text{for all } \ell \in \mathbb{N}, \tag{F2}$$

because $\left\{-\frac{1}{n\ell} \ln \mathrm{Err}(\mathcal{E}^{n\ell})\right\}_{n\in\mathbb{N}}$ is a subsequence of $\left\{-\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^{n})\right\}_{n\in\mathbb{N}}$. We now prove the inequality converse to (F2). Let $\{M_{k,\ell}(1), \ldots, M_{k,\ell}(r)\}$ be a POVM attaining $\mathrm{Err}(\mathcal{E}^{k\ell})$ for all $k, \ell \in \mathbb{N}$. Then for all $n \in \mathbb{N}$ such that $n \geq \ell$, we have

$$\mathrm{Err}(\mathcal{E}^n) \leq \sum_{i\in[r]} \eta_i \operatorname{Tr}\left[\rho_i^{\otimes n} \left(M_{\lfloor \frac{n}{\ell} \rfloor, \ell}(i) \otimes \mathbb{I}^{\otimes (n - \lfloor \frac{n}{\ell} \rfloor)}\right)\right] \tag{F3}$$

$$= \sum_{i\in[r]} \eta_i \operatorname{Tr}\left[\rho_i^{\otimes \lfloor \frac{n}{\ell} \rfloor \ell} M_{\lfloor \frac{n}{\ell} \rfloor, \ell}(i)\right] \tag{F4}$$

$$= \mathrm{Err}(\mathcal{E}^{\lfloor \frac{n}{\ell} \rfloor \ell}). \tag{F5}$$

This implies

$$\mathrm{E}(\rho_1, \ldots, \rho_r) = \liminf_{n\to\infty} -\frac{1}{n} \ln \mathrm{Err}(\mathcal{E}^n) \tag{F6}$$

$$\geq \liminf_{n\to\infty} -\frac{1}{\lfloor \frac{n}{\ell} \rfloor \ell} \ln \mathrm{Err}(\mathcal{E}^{\lfloor \frac{n}{\ell} \rfloor \ell}) \tag{F7}$$

$$= \frac{1}{\ell} \liminf_{k\to\infty} -\frac{1}{k} \ln \mathrm{Err}(\mathcal{E}^{k\ell}) \tag{F8}$$

$$= \frac{1}{\ell} \mathrm{E}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}). \tag{F9}$$

This completes the proof. □

Appenix G: Limit of the regularized maximal multivariate quantum Chernoff divergence

Here we provide a proof of equation (175). We first observe that the multivariate classical Chernoff divergence is subadditive, i.e.,

$$\xi_{\mathrm{cl}}(P_1 \otimes Q_1, \ldots, P_r \otimes Q_r) \leq \xi_{\mathrm{cl}}(P_1, \ldots, P_r) + \xi_{\mathrm{cl}}(Q_1, \ldots, Q_r) \tag{G1}$$

for all sets of probability densities $\{P_1, \ldots, P_r\}$ and $\{Q_1, \ldots, Q_r\}$ on a measureable space $(\Omega, \mathcal{A})$. This follows easily from the definitions of the Hellinger transform (19) and multivariate Chernoff divergence (23). So, from the definition (170), we have for $\ell, m \in \mathbb{N}$ that

$$\xi_{\max}(\rho_1^{\otimes(\ell+m)}, \ldots, \rho_r^{\otimes(\ell+m)})$$

$$= \inf_{\substack{(\mathcal{P}^{(\ell+m)}, \{P_i^{(\ell+m)}\}_{i\in[r]}) \\ \mathcal{P}^{(\ell+m)}(P_i^{(\ell+m)})=\rho_i^{\otimes \ell}\otimes\rho_i^{\otimes m}}} \xi_{\mathrm{cl}}(P_1^{(\ell+m)}, \ldots, P_r^{(\ell+m)}) \tag{G2}$$

$$\leq \inf_{\substack{(\mathcal{P}^{(\ell)}\otimes\mathcal{P}^{(m)}, \{P_i^{(\ell)}\otimes P_i^{(m)}\}_{i\in[r]}) \\ \mathcal{P}^{(\ell)}(P_i^{(\ell)})=\rho_i^{\otimes \ell}, \mathcal{P}^{(m)}(P_i^{(m)})=\rho_i^{\otimes m}}} \xi_{\mathrm{cl}}(P_1^{(\ell)}\otimes P_1^{(m)}, \ldots, P_r^{(\ell)}\otimes P_r^{(m)}) \tag{G3}$$

$$\leq \inf_{\substack{(\mathcal{P}^{(\ell)}\otimes\mathcal{P}^{(m)}, \{P_i^{(\ell)}\otimes P_i^{(m)}\}_{i\in[r]}) \\ \mathcal{P}^{(\ell)}(P_i^{(\ell)})=\rho_i^{\otimes \ell}, \mathcal{P}^{(m)}(P_i^{(m)})=\rho_i^{\otimes m}}} \left(\xi_{\mathrm{cl}}(P_1^{(\ell)}, \ldots, P_r^{(\ell)}) + \xi_{\mathrm{cl}}(P_1^{(m)}, \ldots, P_r^{(m)})\right) \tag{G4}$$

$$= \inf_{\substack{(\mathcal{P}^{(\ell)}, \{P_i^{(\ell)}\}_{i\in[r]}) \\ \mathcal{P}^{(\ell)}(P_i^{(\ell)})=\rho_i^{\otimes \ell}}} \left(\xi_{\mathrm{cl}}(P_1^{(\ell)}, \ldots, P_r^{(\ell)})\right) + \inf_{\substack{(\mathcal{P}^{(m)}, \{P_i^{(m)}\}_{i\in[r]}) \\ \mathcal{P}^{(m)}(P_i^{(m)})=\rho_i^{\otimes m}}} \left(\xi_{\mathrm{cl}}(P_1^{(m)}, \ldots, P_r^{(m)})\right) \tag{G5}$$

$$= \xi_{\max}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}) + \xi_{\max}(\rho_1^{\otimes m}, \ldots, \rho_r^{\otimes m}). \tag{G6}$$

We have thus proved that the sequence $\left(\xi_{\max}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell})\right)_{\ell\in\mathbb{N}}$ is subadditive. It then follows from Fekete's subadditive lemma [15] that the limit $\lim_{\ell\to\infty} \xi_{\max}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell})/\ell$ exists and is given by

$$\lim_{\ell\to\infty} \frac{1}{\ell}\xi_{\max}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}) = \inf_{\ell\in\mathbb{N}} \frac{1}{\ell}\xi_{\max}(\rho_1^{\otimes \ell}, \ldots, \rho_r^{\otimes \ell}). \tag{G7}$$

Appendix H: Properties of the extended max-relative entropy in Equation (201)

Recall the definition of extended max-relative entropy from (201) for a Hermitian operator X and a positive semidefinite operator σ:

$$D_{\max}(X\|\sigma) := \ln \inf_{\lambda\geq 0} \{\lambda : -\lambda\sigma \leq X \leq \lambda\sigma\}. \tag{H1}$$

We illustrate some special cases of extended max-relative entropy as follows. If $X = 0$, then, for all positive semi-definite σ, the choice $\lambda = 0$ satisfies $-\lambda\sigma \leq X \leq \lambda\sigma$. This implies that $D_{\max}(X\|\sigma) = -\infty$ in this case. In the case when X is nonzero and σ is zero, the support of X is not contained in the support of σ. This implies that $D_{\max}(X\|\sigma) = +\infty$ in this case.

We now present several properties of the extended max-relative entropy.

Proposition 26 (Monotonicity). *Let X be a Hermitian operator, and let σ', σ be positive semi-definite operators such that $\sigma' \leq \sigma$. Then*

$$D_{\max}(X\|\sigma) \leq D_{\max}(X\|\sigma'). \tag{H2}$$

Proof Given an arbitrary $\lambda \geq 0$ that satisfies $-\lambda\sigma' \leq X \leq \lambda\sigma'$, this λ also satisfies $-\lambda\sigma \leq X \leq \lambda\sigma$. Consequently,

$$D_{\max}(X\|\sigma) = \ln \inf_{\lambda \geq 0}\{\lambda : -\lambda\sigma \leq X \leq \lambda\sigma\} \tag{H3}$$

$$\leq \ln \inf_{\lambda \geq 0}\{\lambda : -\lambda\sigma' \leq X \leq \lambda\sigma'\} \tag{H4}$$

$$= D_{\max}(X\|\sigma'), \tag{H5}$$

concluding the proof. □

Proposition 27 (Supremum representation). *For a Hermitian operator X and a positive semi-definite operator σ, the following equality holds:*

$$D_{\max}(X\|\sigma) = \sup_{\varepsilon>0} D_{\max}(X\|\sigma + \varepsilon I) = \lim_{\varepsilon \searrow 0} D_{\max}(X\|\sigma + \varepsilon I). \tag{H6}$$

Proof We conclude the second equality in (H6) because $\sigma + \varepsilon I \leq \sigma + \varepsilon' I$ holds for $0 < \varepsilon \leq \varepsilon'$, and applying Proposition 26 allows us to conclude that, for fixed X and σ, the function $\varepsilon \mapsto D_{\max}(X\|\sigma + \varepsilon I)$ is monotone non-increasing.

For all $\varepsilon > 0$, the operator inequality $\sigma \leq \sigma + \varepsilon I$ holds. By applying Proposition 26, we conclude that $D_{\max}(X\|\sigma) \geq D_{\max}(X\|\sigma + \varepsilon I)$. So it remains to prove that this is actually an equality. To see that equality holds, we consider two separate cases. First suppose that the support of X is contained in the support of σ. Then, the following equality holds as a consequence of (202):

$$D_{\max}(X\|\sigma + \varepsilon I) = \ln \left\| (\sigma + \varepsilon I)^{-1/2} X (\sigma + \varepsilon I)^{-1/2} \right\|_\infty . \tag{H7}$$

The equality $D_{\max}(X\|\sigma) = \lim_{\varepsilon \searrow 0} D_{\max}(X\|\sigma + \varepsilon I)$ follows as a consequence of the continuity of the operator norm. Now suppose that the support of X is not contained in the support of σ. Let $|v\rangle \in \mathrm{supp}(X) \setminus \mathrm{supp}(\sigma)$ be a unit vector. Consider that

$$\ln \left\| (\sigma + \varepsilon I)^{-1/2} X (\sigma + \varepsilon I)^{-1/2} \right\|_\infty \geq \ln \left| \langle v | (\sigma + \varepsilon I)^{-1/2} X (\sigma + \varepsilon I)^{-1/2} | v \rangle \right|$$
$$= \ln\Big(|\langle v | X | v \rangle| \, \varepsilon^{-1} \Big) . \tag{H8}$$

Thus, by taking the $\varepsilon \searrow 0$ limit, we see that $\lim_{\varepsilon \searrow 0} D_{\max}(X\|\sigma + \varepsilon I) = +\infty$ in this case, consistent with the definition in (201). □

Proposition 28 (Data-processing inequality). *Let X be a Hermitian operator and σ a positive semi-definite operator. Let $\mathcal{N}$ be a positive map (a special case of which is a quantum channel, i.e., a completely positive and trace-preserving map). Then*

$$D_{\max}(X\|\sigma) \geq D_{\max}(\mathcal{N}(X)\|\mathcal{N}(\sigma)). \tag{H9}$$

Proof A special case of this inequality follows from [62, Lemma 2] by taking the limit $\alpha \to \infty$. Here we prove it for all positive maps, for X an arbitrary Hermitian operator, and σ an arbitrary positive semi-definite operator. Suppose that $\lambda \geq 0$ is such that $-\lambda\sigma \leq X \leq \lambda\sigma$. Then, the following inequality holds $-\lambda\mathcal{N}(\sigma) \leq \mathcal{N}(X) \leq \lambda\mathcal{N}(\sigma)$, from the assumption that $\mathcal{N}$ is a positive map. Consequently, we get

$$D_{\max}(X\|\sigma) = \ln \inf_{\lambda\geq 0}\{\lambda : -\lambda\sigma \leq X \leq \lambda\sigma\} \tag{H10}$$

$$\geq \ln \inf_{\lambda\geq 0}\{\lambda : -\lambda\mathcal{N}(\sigma) \leq \mathcal{N}(X) \leq \lambda\mathcal{N}(\sigma)\} \tag{H11}$$

$$= D_{\max}(\mathcal{N}(X)\|\mathcal{N}(\sigma)), \tag{H12}$$

concluding the proof. □

Proposition 29 (Joint quasi-convexity). *Let $\mathscr{X}$ be a finite alphabet and p a probability distribution on $\mathscr{X}$. Let X^x and σ^x be Hermitian and positive semi-definite operators, respectively, for all $x \in \mathscr{X}$. Then*

$$\max_{x\in\mathscr{X}} D_{\max}(X^x\|\sigma^x) \geq D_{\max}\left(\sum_{x\in\mathscr{X}} p(x)X^x \middle\| \sum_{x\in\mathscr{X}} p(x)\sigma^x\right). \tag{H13}$$

Proof If $\lambda \geq 0$ satisfies $-\lambda\sigma^x \leq X^x \leq \lambda\sigma^x$ for all $x \in \mathscr{X}$, then we also have $-\lambda\sum_{x\in\mathscr{X}} p(x)\sigma^x \leq \sum_{x\in\mathscr{X}} p(x)X^x \leq \lambda\sum_{x\in\mathscr{X}} p(x)\sigma^x$. This gives

$$D_{\max}\left(\sum_{x\in\mathscr{X}} p(x)X^x \middle\| \sum_{x\in\mathscr{X}} p(x)\sigma^x\right) = \ln \inf_{\lambda\geq 0}\left\{\lambda : -\lambda\sum_{x\in\mathscr{X}} p(x)\sigma^x \right.$$

$$\left. \leq \sum_{x\in\mathscr{X}} p(x)X^x \leq \lambda\sum_{x\in\mathscr{X}} p(x)\sigma^x\right\} \tag{H14}$$

$$\leq \ln \inf_{\lambda\geq 0}\left\{\lambda : -\lambda\sigma^x \leq X^x \leq \lambda\sigma^x, \forall x \in \mathscr{X}\right\} \tag{H15}$$

$$= \max_{x\in\mathscr{X}} \ln \inf_{\lambda\geq 0}\left\{\lambda : -\lambda\sigma^x \leq X^x \leq \lambda\sigma^x\right\} \tag{H16}$$

$$= \max_{x\in\mathscr{X}} D_{\max}(X^x\|\sigma^x), \tag{H17}$$

concluding the proof. □

Proposition 30 (Non-negativity and faithfulness). *Let X be a Hermitian operator of unit trace, and let σ be a quantum state. Then $D_{\max}(X\|\sigma) \geq 0$. Also, under the same conditions, $D_{\max}(X\|\sigma) = 0$ if and only if $X = \sigma$.*

Proof For every $\lambda \geq 0$ satisfying $-\lambda\sigma \leq X \leq \lambda\sigma$, we have that $\lambda = \mathrm{Tr}[\lambda\sigma] \geq \mathrm{Tr}\, X = 1$, implying that $\ln\lambda \geq 0$. By definition, we then get $D_{\max}(X\|\sigma) \geq 0$.

If $X = \sigma$, then it trivially follows by definition that $D_{\max}(X\|\sigma) = 0$. Conversely, suppose that $D_{\max}(X\|\sigma) = 0$. This implies $-\sigma \leq X \leq \sigma$, and hence $\sigma - X \geq 0$. By the Helstrom-Holevo Theorem [28, Eq. (5.1.17)], and the fact that $\mathrm{Tr}[\sigma - X] = 0$, we get

$$\frac{1}{2}\|\sigma - X\|_1 = \sup_{M\geq 0}\{\mathrm{Tr}[M(\sigma - X)] : M \leq \mathbb{I}\} \tag{H18}$$

$$\leq \inf_{Y\geq 0}\{\mathrm{Tr}[Y] : Y \geq \sigma - X\}, \tag{H19}$$

where the last inequality follows by the weak duality of the SDP given in (H18). A feasible point in (H19) is given by $Y = \sigma - X$, and we have $\mathrm{Tr}[Y] = \mathrm{Tr}[\sigma - X] = 0$. It thus follows from (H19) that $\|\sigma - X\|_1 \leq 0$, which implies $\|\sigma - X\|_1 = 0$. We have thus shown that $\sigma = X$. □

Proposition 31 (Lower semi-continuity). *The function* $(X, \sigma) \mapsto D_{\max}(X\|\sigma)$*, with domain* $\mathrm{Herm}(\mathcal{H}) \times \mathcal{L}_+(\mathcal{H})$ *and range* $\mathbb{R} \cup \{-\infty, +\infty\}$*, is lower semi-continuous.*

Proof Here we follow arguments similar to those given in [43] (see also [48, Lemma 18], whose short proof we follow verbatim). Recall the supremum representation in Proposition 27. For all $\varepsilon > 0$, the functions defined by $(X, \sigma) \mapsto D_{\max}(X\|\sigma + \varepsilon I)$ are continuous because the second argument has full support. Since the pointwise supremum of continuous functions is lower semi-continuous, it follows that the function $(X, \sigma) \mapsto D_{\max}(X\|\sigma)$ is lower semi-continuous. □

If A, B are Hermitian operators on a Hilbert space $\mathcal{H}$, then it is easy to prove that the kernel of their tensor product is given by $\ker(A\otimes B) = \ker(A)\otimes\mathcal{H}+\mathcal{H}\otimes\ker(B)$. We use this observation in the proof of the next property.

Proposition 32 (Additivity). *Let* X_1, X_2 *be nonzero Hermitian operators, and let* σ_1, σ_2 *be nonzero positive semi-definite operators. Then,*

$$D_{\max}(X_1 \otimes X_2\|\sigma_1 \otimes \sigma_2) = D_{\max}(X_1\|\sigma_1) + D_{\max}(X_2\|\sigma_2). \tag{H20}$$

Proof First, suppose that $\mathrm{supp}(X_1) \nsubseteq \mathrm{supp}(\sigma_1)$. This implies that $\mathrm{supp}(X_1 \otimes X_2) \nsubseteq \mathrm{supp}(\sigma_1 \otimes \sigma_2)$. Indeed, let $|x_1\rangle \in \mathrm{supp}(X_1)\setminus \mathrm{supp}(\sigma_1)$. Also, $X_2 \neq 0$ implies that there exists a nonzero vector $|x_2\rangle \in \mathrm{supp}(X_2)$. We thus have $(X_1\otimes X_2)(|x_1\rangle\otimes|x_2\rangle) \neq 0$ and $(\sigma_1 \otimes \sigma_2)(|x_1\rangle \otimes |x_2\rangle) = 0$, implying that $\mathrm{supp}(X_1 \otimes X_2) \nsubseteq \mathrm{supp}(\sigma_1 \otimes \sigma_2)$. Also, the assumption that X_2 and σ_2 are nonzero implies that $D_{\max}(X_2\|\sigma_2) > -\infty$. Therefore, in this case, both $D_{\max}(X_1 \otimes X_2\|\sigma_1 \otimes \sigma_2)$ and $D_{\max}(X_1\|\sigma_1) + D_{\max}(X_2\|\sigma_2)$ are equal to ∞. We also get by similar arguments for the case $\mathrm{supp}(X_2) \nsubseteq \mathrm{supp}(\sigma_2)$ that both $D_{\max}(X_1 \otimes X_2\|\sigma_1 \otimes \sigma_2)$ and $D_{\max}(X_1\|\sigma_1) + D_{\max}(X_2\|\sigma_2)$ are equal to ∞.

To complete the proof, we now consider the case when $\mathrm{supp}(X_1) \subseteq \mathrm{supp}(\sigma_1)$ and $\mathrm{supp}(X_2) \subseteq \mathrm{supp}(\sigma_2)$. In this case, we have $\mathrm{supp}(X_1 \otimes X_2) \subseteq \mathrm{supp}(\sigma_1 \otimes \sigma_2)$. This is because we have $\ker(\sigma_1) \subseteq \ker(X_1)$ and $\ker(\sigma_2) \subseteq \ker(X_2)$, which gives

$$\ker(\sigma_1 \otimes \sigma_2) = \ker(\sigma_1) \otimes \mathcal{H} + \mathcal{H} \otimes \ker(\sigma_2) \tag{H21}$$

$$\subseteq \ker(X_1) \otimes \mathcal{H} + \mathcal{H} \otimes \ker(X_2) \tag{H22}$$

$$= \ker(X_1 \otimes X_2). \tag{H23}$$

We thus have

$$D_{\max}(X_1 \otimes X_2 \| \sigma_1 \otimes \sigma_2) = \ln \left\| (\sigma_1^{-1/2} \otimes \sigma_2^{-1/2})(X_1 \otimes X_2)(\sigma_1^{-1/2} \otimes \sigma_2^{-1/2}) \right\|_\infty \tag{H24}$$

$$= \ln \left\| \sigma_1^{-1/2} X_1 \sigma_1^{-1/2} \otimes \sigma_2^{-1/2} X_2 \sigma_2^{-1/2} \right\|_\infty \tag{H25}$$

$$= \ln \left(\left\| \sigma_1^{-1/2} X_1 \sigma_1^{-1/2} \right\|_\infty \cdot \left\| \sigma_2^{-1/2} X_2 \sigma_2^{-1/2} \right\|_\infty \right) \tag{H26}$$

$$= \ln \left\| \sigma_1^{-1/2} X_1 \sigma_1^{-1/2} \right\|_\infty + \ln \left\| \sigma_2^{-1/2} X_2 \sigma_2^{-1/2} \right\|_\infty \tag{H27}$$

$$= D_{\max}(X_1 \| \sigma_1) + D_{\max}(X_2 \| \sigma_2), \tag{H28}$$

concluding the proof. □

Appendix I: Proof of Equation (221)

Let $\omega \in \mathcal{D}$ be arbitrary and $(s_1, \ldots, s_r) \in \mathbb{R}^r$ be any probability vector. Since the quantum states $\rho_1, \ldots, \rho_r$ have full support, we have

$$\sum_{i \in [r]} s_i D(\omega \| \rho_i)$$

$$= \sum_{i \in [r]} s_i \operatorname{Tr}[\omega(\ln \omega - \ln \rho_i)] \tag{I1}$$

$$= \operatorname{Tr}[\omega \ln \omega] - \operatorname{Tr}\left[\omega \left(\sum_{i \in [r]} s_i \ln \rho_i \right) \right] \tag{I2}$$

$$= \operatorname{Tr}[\omega \ln \omega] - \operatorname{Tr}\left[\omega \ln \exp \left(\sum_{i \in [r]} s_i \ln \rho_i \right) \right] \tag{I3}$$

$$= \operatorname{Tr}[\omega \ln \omega] - \operatorname{Tr}\left[\omega \ln \left(\frac{\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)}{\operatorname{Tr}\left[\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)\right]} \cdot \operatorname{Tr}\left[\exp \left(\sum_{i \in [r]} s_i \ln \rho_i \right) \right] \right) \right] \tag{I4}$$

$$= \operatorname{Tr}[\omega \ln \omega] - \operatorname{Tr}\left[\omega \ln \left(\frac{\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)}{\operatorname{Tr}\left[\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)\right]} \right) \right] - \ln \operatorname{Tr}\left[\exp \left(\sum_{i \in [r]} s_i \ln \rho_i \right) \right] \tag{I5}$$

$$= D\left(\omega \middle\| \frac{\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)}{\operatorname{Tr}\left[\exp\left(\sum_{i \in [r]} s_i \ln \rho_i\right)\right]} \right) - \ln \operatorname{Tr}\left[\exp \left(\sum_{i \in [r]} s_i \ln \rho_i \right) \right] \tag{I6}$$

$$\geq -\ln \mathrm{Tr}\left[\exp\left(\sum_{i\in[r]} s_i \ln \rho_i\right)\right], \tag{I7}$$

where the inequality follows from the non-negativity of quantum relative entropy for quantum states. The lower bound is achieved by picking $\omega = \frac{\exp(\sum_{i\in[r]} s_i \ln \rho_i)}{\mathrm{Tr}[\exp(\sum_{i\in[r]} s_i \ln \rho_i)]}$, so that

$$\begin{aligned}
&\inf_{\omega\in\mathcal{D}} \sum_{i\in[r]} s_i D(\omega\|\rho_i) \\
&\quad = \inf_{\omega\in\mathcal{D}} D\left(\omega\middle\|\frac{\exp\left(\sum_{i\in[r]} s_i \ln \rho_i\right)}{\mathrm{Tr}\left[\exp\left(\sum_{i\in[r]} s_i \ln \rho_i\right)\right]}\right) - \ln \mathrm{Tr}\left[\exp\left(\sum_{i\in[r]} s_i \ln \rho_i\right)\right] && \text{(I8)} \\
&\quad = -\ln \mathrm{Tr}\left[\exp\left(\sum_{i\in[r]} s_i \ln \rho_i\right)\right]. && \text{(I9)}
\end{aligned}$$

This directly gives (221).

Acknowledgements We are especially grateful to Milán Mosonyi for several clarifying discussions about quantum hypothesis testing, as well as to Kaiyuan Ji, Felix Leditzky, Vishal Singh, and Aaron Wagner for insightful discussions. We also thank the anonymous referee and the editor for many extensive helpful comments that improved the manuscript. HKM and MMW acknowledge support from the National Science Foundation under Grant No. 2304816.

Data availability Data sharing was not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Audenaert, K.M.R., Calsamiglia, J., Munoz-Tapia, R., Bagan, E., Masanes, Ll., Acin, Antonio, Verstraete, Frank: Discriminating states: The quantum Chernoff bound. Phys. Rev. Lett. **98**(16), 160501 (2007). arXiv:quant-ph/0610027
2. Ando, T.: Lebesgue-type decomposition of positive operators. Acta Sci. Math. (Szeged) **38**(3–4), 253–260 (1976)
3. Barnett, S.M., Croke, S.: Quantum state discrimination. Adv. Opt. Photonics **1**(2), 238–278 (2009). arXiv:0810.1970
4. Bacon, D., Childs, A.M., van Dam, W.D.: Optimal measurements for the dihedral hidden subgroup problem. Chic. J. Theor. Comput. Sci. **2006**, 2 (2006). arXiv:quant-ph/0501044
5. Barrett, J., Cavalcanti, E.G., Lal, R., Maroney, O.J.E.: No ψ-epistemic model can fully explain the indistinguishability of quantum states. Phys. Rev. Lett. **112**(25), 250403 (2014). arXiv:1310.8302
6. Billingsley, P.: Probability and Measure. Wiley Series in Probability and Statistics. Wiley, Hoboken (1995)
7. Bandyopadhyay, S., Jain, R., Oppenheim, J., Perry, C.: Conclusive exclusion of quantum states. Phys. Rev. A **89**(2), 022336 (2014). arXiv:1306.4683

8. Bae, J., Kwek, L.-C.: Quantum state discrimination and its applications. J. Phys. A: Math. Theor. **48**(8), 083001 (2015). arXiv:1707.02571
9. Borwein, J., Lewis, A.: Convex Analysis. Springer, New York (2006)
10. Born, M.: Quantenmechanik der Stoßvorgänge. Z. Physik **38**(11), 803–827 (1926)
11. Collins, R.J., Donaldson, R.J., Dunjko, V., Wallden, P., Clarke, P.J., Andersson, E., Jeffers, J., Buller, G.S.: Realization of quantum digital signatures without the requirement of quantum memory. Phys. Rev. Lett. **113**, 040502 (2014). arXiv:1311.5760
12. Caves, C.M., Fuchs, C.A., Schack, R.: Conditions for compatibility of quantum-state assignments. Phys. Rev. A **66**(6), 062111 (2002). arXiv:quant-ph/0206110
13. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Stat. **23**(4), 493–507 (1952)
14. Datta, N., Leditzky, F.: A limit of the quantum Rényi divergence. J. Phys. A: Math. Theor. **47**(4), 045304 (2014)
15. Fekete, M.: über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. Math. Z. **17**(1), 228–249 (1923)
16. Fazekas, I., Liese, F.: Some properties of the Hellinger transform and its application in classification problems. Comp. Math. Appl. **31**(8), 107–116 (1996)
17. Furuya, K., Lashkari, N., Ouseph, S.: Monotonic multi-state quantum f-divergences. J. Math. Phys. **64**(4), 042203 (2023)
18. Grigelionis, B.: On Hellinger Transforms for Solutions of Martingale Problems, pp. 107–116. Springer, New York, UK (1993)
19. Havlíček, V., Barrett, J.: Simple communication complexity separation from quantum state antidistinguishability. Phys. Rev. Res. **2**(1), 013326 (2020). arXiv:1911.01927
20. Helstrom, C.W.: Quantum detection and estimation theory. J. Stat. Phys. **1**, 231–252 (1969)
21. Hiai, F.: Equality cases in matrix norm inequalities of Golden-Thompson type. Lin. Multilin. Algebr. **36**(4), 239–249 (1994)
22. Heinosaari, T., Kerppo, O.: Antidistinguishability of pure quantum states. J. Phys. A: Math. Theor. **51**(36), 365303 (2018). arXiv:1804.10457
23. Hiai, F., Mosonyi, M.: Different quantum f-divergences and the reversibility of quantum operations. Rev. Math. Phys. **29**(07), 1750023 (2017)
24. Holevo, A.S.: An analogue of statistical decision theory and noncommutative probability theory. Trudy Moskovskogo Matematicheskogo Obshchestva **26**, 133–149 (1972)
25. Horodecki, M., Shor, P.W., Ruskai, M.B.: Entanglement breaking channels. Rev. Math. Phys. **15**(06), 629–641 (2003)
26. Hayashi, M., Tomamichel, M.: Correlation detection and an operational interpretation of the Rényi mutual information. J. Math. Phys. **57**(10), 102201 (2016)
27. Jacod, J.: Filtered statistical models and Hellinger processes. Stoch. Process. Appl. **32**(1), 3–45 (1989)
28. Khatri, S., Wilde, M.M.: Principles of quantum communication theory: a modern approach. (2020). arXiv:2011.04672v1
29. Katariya, V., Wilde, M.M.: Geometric distinguishability measures limit quantum channel estimation and discrimination. Quantum Inf. Process. **20**(2), 78 (2021). arXiv:2004.10708
30. Cam, L.M.L., Yang, G.L.: Asymptotics in Statistics: Some Basic Concepts. Springer Science & Business Media (2000)
31. Leifer, M., Duarte, C.: Noncontextuality inequalities from antidistinguishability. Phys. Rev. A **101**(6), 062113 (2020). arXiv:2001.11485
32. LeCam, L.: On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Ann. Math. Stat. **41**(3), 802–828 (1970)
33. Leifer, M.S.: Is the quantum state real? An extended review of ψ-ontology theorems. Quanta **3**, 67–155 (2014). arXiv:1409.1570
34. Li, K.: Discriminating quantum states: The multiple Chernoff distance. Ann. Stat. **44**(4), 1661–1679 (2016). arXiv:1508.06624
35. Leang, C.C., Johnson, D.H.: On the asymptotics of m-hypothesis Bayesian detection. IEEE Trans. Inf. Theory **43**(1), 280–282 (1997)
36. Liese, F., Miescke, Klaus-J..: Statistical Decision Theory. Springer, New York, NY (2010)
37. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum-mechanical entropy. J. Math. Phys. **14**(12), 1938–1941 (1973)

38. Matusita, K.: On the notion of affinity of several distributions and some of its applications. Ann. Inst. Stat. Math. **19**, 181–192 (1967)
39. Matsumoto, K.: A new quantum version of f-divergence (2013). arXiv:1311.4722
40. Matsumoto, K.: On maximization of measured f-divergence between a given pair of quantum states (2014). arXiv:1412.3676
41. Matsumoto, K.: A new quantum version of f-divergence. In: Ozawa, M., Butterfield, J., Halvorson, H., Rédei, M., Kitajima, Y., Buscemi, F. (eds.) Reality and measurement in algebraic quantum theory. Springer Proceedings in Mathematics & Statistics, vol. 261, pp. 229–273. Springer, Singapore (2018)
42. Mosonyi, M., Bunth, G., Vrana, P.: Geometric relative entropies and barycentric Rényi divergences. (2022). arXiv:2207.14282v2
43. Mosonyi, M., Hiai, F.: On the quantum Rényi relative entropies and related capacity formulas. IEEE Trans. Inf. Theory **57**(4), 2474–2487 (2011)
44. Mosonyi, M., Ogawa, T.: Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies. Commun. Math. Phys. **334**, 1617–1648 (2015)
45. Nussbaum, M., Szkoła, A.: The Chernoff lower bound for symmetric quantum hypothesis testing. Ann. Stat. **37**(2), 1040–1057 (2009). arXiv:quant-ph/0607216
46. Pusey, M.F., Barrett, J., Rudolph, T.: On the reality of the quantum state. Nat. Phys. **8**(6), 475–478 (2012). arXiv:1111.3328
47. Russo, V., Sikora, J.: Inner products of pure states and their antidistinguishability. Phys. Rev. A **107**(3), L030202 (2023). arXiv:2206.08313
48. Rubboli, R., Tomamichel, M.: New additivity properties of the relative entropy of entanglement and its generalizations. (2024). arXiv:2211.12804
49. Ruskai, M.B.: Inequalities for quantum entropy: A review with conditions for equality. J. Math. Phys. **43**(9), 4358–4375 (2002)
50. Salikhov, N.P.: Asymptotic properties of error probabilities of tests for distinguishing between several multinomial testing schemes. Dokl. Akad. Nauk SSSR **209**(1), 54–57 (1973)
51. Salikhov, N.P.: On one generalization of Chernov's distance. Theory Prob. Appl. **43**(2), 239–255 (1999)
52. Salikhov, N.P.: Optimal sequences of tests for several polynomial schemes of trials. Theory Prob. Appl. **47**(2), 286–298 (2003)
53. Schervish, M.J.: Theory Stat. Springer Science & Business Media, New York (2012)
54. Shiryaev, A.N.: Probability-1, vol. 95. Springer, New York (2016)
55. Sion, M.: On general minimax theorems. Pac. J. Math. **8**, 171–176 (1958)
56. Strasser, H.: Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory, vol. 7. Walter de Gruyter, Berlin (2011)
57. Torgersen, E.N.: Measures of information based on comparison with total information and with total ignorance. Ann. Stat. **9**(3), 638–657 (1981)
58. Torgersen, E.N.: Comparison of Statistical Experiments, vol. 36. Cambridge University Press, Cambridge (1991)
59. Toussaint, G.T.: Some properties of Matusita's measure of affinity of several distributions. Ann. Inst. Stat. Math. **26**(3), 389–394 (1974)
60. Umegaki, H.: Conditional expectation in an operator algebra, IV (Entropy and information). Kodai Math. Semin. Rep. **14**, 59–85 (1962)
61. Wilde, M.M.: Quantum Information Theory, second Cambridge University Press, Cambridge (2017)
62. Wang, X., Wilde, M.M.: α-Logarithmic negativity. Phys. Rev. A **102**, 032416 (2020). arXiv:1904.10437
63. Yuen, H., Kennedy, R., Lax, M.: Optimum testing of multiple hypotheses in quantum detection theory. IEEE Trans. Inf. Theory **21**(2), 125–134 (1975)