Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination

Koenraad M.R. Audenaert[1,2,*] and Milán Mosonyi[3,4,†]

[1] *Department of Mathematics, Royal Holloway University of London, Egham TW20 0EX, U.K.*

[2] *Department of Physics and Astronomy, University of Ghent, S9, Krijgslaan 281, B-9000 Ghent, Belgium*

[3] *Física Teòrica: Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain*

[4] *Mathematical Institute, Budapest University of Technology and Economics Egry József u 1., Budapest, 1111 Hungary*

Abstract

We consider the multiple hypothesis testing problem for symmetric quantum state discrimination between r given states $\sigma_1, \ldots, \sigma_r$. By splitting up the overall test into multiple binary tests in various ways we obtain a number of upper bounds on the optimal error probability in terms of the binary error probabilities. These upper bounds allow us to deduce various bounds on the asymptotic error rate, for which it has been hypothesized that it is given by the multi-hypothesis quantum Chernoff bound (or Chernoff divergence) $C(\sigma_1, \ldots, \sigma_r)$, as recently introduced by Nussbaum and Szkoła in analogy with Salikhov's classical multi-hypothesis Chernoff bound. This quantity is defined as the minimum of the pairwise binary Chernoff divergences $\min_{j<k} C(\sigma_j, \sigma_k)$. It was known already that the optimal asymptotic rate must lie between $C/3$ and C, and that for certain classes of sets of states the bound is actually achieved. It was known to be achieved, in particular, when the state pair that is closest together in Chernoff divergence is more than 6 times closer than the next closest pair. Our results improve on this in two ways. Firstly, we show that the optimal asymptotic rate must lie between $C/2$ and C. Secondly, we show that the Chernoff bound is already achieved when the closest state pair is more than 2 times closer than the next closest pair. We also show that the Chernoff bound is achieved when at least $r-2$ of the states are pure, improving on a previous result by Nussbaum and Szkoła. Finally, we indicate a number of potential pathways along which a proof (or disproof) may eventually be found that the multi-hypothesis quantum Chernoff bound is always achieved.

*Electronic address: koenraad.audenaert@rhul.ac.uk
†Electronic address: milan.mosonyi@gmail.com

1 Introduction

Consider a communication scenario where a sender (say, Alice) wishes to send one of r possible messages to a receiver (Bob). To achieve this goal, Alice has a device at her disposal that can prepare r quantum states $\rho_1, \dots, \rho_r$ from some state space $\mathcal{S}(\mathcal{H})$, one for each possible message, which she can then send through a quantum channel Φ, resulting in the states $\sigma_i = \Phi(\rho_i)$ at Bob's side. Bob then has to make a quantum measurement to identify which message was sent. His measurement is described by a set of positive semidefinite operators $E_1, \dots, E_r$, one corresponding to each possible message, that form an incomplete POVM (positive operator-valued measure), i.e., they satisfy $E_1 + \dots + E_r \le I$. The operator $E_0 := I - (E_1 + \dots + E_r)$ corresponds to not making a decision on the identity of the received state. The probability of making an erroneous decision when the message i was sent is then given by $\operatorname{Tr} \sigma_i(I - E_i)$. If we also assume that Alice sends each message i with a certain probability p_i then the best Bob can do is choose the POVM that minimizes the Bayesian error probability

$$P_e(\{E_1, \dots, E_r\}) := \sum_{i=1}^{r} p_i \operatorname{Tr} \sigma_i(I - E_i) = \sum_{i=1}^{r} \operatorname{Tr} A_i(I - E_i),$$

where $A_i := p_i \sigma_i$.

In the classical case, i.e., when the σ_i are mutually commuting, the optimal success probability is known to be reached by the so-called maximum likelihood measurement, and the optimal success probability is given by $\operatorname{Tr} \max\{p_1\sigma_1, \dots, p_r\sigma_r\}$, where the maximum is taken entrywise in some basis that simultaneously diagonalizes all the σ_i. In the general quantum case, no explicit expression is known for the optimal error probability, or for the measurement achieving it, unless $r = 2$, in which case these optimal quantities are easy to find [20, 25]. Moreover, it turns out to be impossible to extend the notion of maximum of a set of real numbers to maximum of a set of positive semidefinite operators on a Hilbert space while keeping all the properties of the former – technically speaking, the positive semidefinite ordering does not induce a lattice structure – and because of this a straightforward generalization of the classical results is not possible. In Section 2.3, we define a generalized notion of maximum for a set of self-adjoint operators, which we call the least upper bound (LUB). This notion reduces to the usual maximum in the classical case, and the optimal success probability can be expressed as $\operatorname{Tr} \operatorname{LUB}(p_1\sigma_1, \dots, p_r\sigma_r)$ [53], giving a direct generalization of the classical expression. We explore further properties of the least upper bound, and its dual, the greatest lower bound (GLB), in Appendix A.

An obvious way to reduce the error probability is to send the same message multiple times. For n repetitions, the optimal error probability is given by

$$P_e^*(A_{1,n}, \dots, A_{r,n}) := \min \left\{ \sum_{i=1}^{r} \operatorname{Tr} A_{i,n}(I - E_i) \,:\, \{E_1, \dots, E_r\} \text{ POVM on } \mathcal{S}(\mathcal{H}^{\otimes n}) \right\}, \quad (1)$$

where $A_{i,n} := p_i \sigma_i^{\otimes n}$. These error probabilities are known to decay exponentially fast in the number of repetitions [4, 38, 39], and hence we are interested in the exponents (which are negative numbers)

$$\underline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right) := \liminf_{n\to\infty}\frac{1}{n}\log P_e^*\left(A_{1,n},\ldots,A_{r,n}\right) \qquad \text{and} \tag{2}$$

$$\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right) := \limsup_{n\to\infty}\frac{1}{n}\log P_e^*\left(A_{1,n},\ldots,A_{r,n}\right), \tag{3}$$

where $\vec{A}_i := \{A_{i,n}\}_{n\in\mathbb{N}}$.

In the case of two possible messages, the theorem for the quantum Chernoff bound [4, 5, 38] states that

$$\underline{p}_e\left(\vec{A}_1,\vec{A}_2\right) = \overline{p}_e\left(\vec{A}_1,\vec{A}_2\right) = \min_{0\le t\le 1}\log\operatorname{Tr}\sigma_1^t\sigma_2^{1-t} =: -C(\sigma_1,\sigma_2), \tag{4}$$

where $C(\sigma_1,\sigma_2)$ is a positive quantity known as the *Chernoff divergence* of σ_1 and σ_2. According to a long-standing conjecture, it is hypothesized that

$$\underline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right) = \overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right) = -\min_{(i,j):\,i\neq j} C(\sigma_i,\sigma_j), \tag{5}$$

i.e., the multi-hypothesis exponent is equal to the worst-case pairwise exponent. Following [39], we call $C(\sigma_1,\ldots,\sigma_r) := \min_{(i,j):\,i\neq j} C(\sigma_i,\sigma_j)$ the *multi-Chernoff bound*. In fact, the lower bound $\underline{p}_e(\vec{A}_1,\ldots,\vec{A}_r) \ge -C(\sigma_1,\ldots,\sigma_r)$ (optimality) follows trivially from the binary case [38], as was pointed out e.g., in [40]. The upper bound $\overline{p}_e(\vec{A}_1,\ldots,\vec{A}_r) \le -C(\sigma_1,\ldots,\sigma_r)$ (achievability) is known to be true for commuting states [48] and when the states σ_i have pairwise disjoint supports [41]. A special case of the latter is when all the states σ_i are pure [40].

Our aim here is to establish decoupling bounds on the *single-shot* error probability by decomposing a multi-hypothesis test into multiple binary tests. These bounds in turn yield bounds on the exponents (2)–(3) in terms of the corresponding pairwise exponents. We remark that the existing asymptotic results mentioned in the previous paragraph also rely implicitly on single-shot decoupling bounds. Regarding lower bounds, it has been shown in [45] that, for any choice of $\sigma_1,\ldots,\sigma_r$, and priors $p_1,\ldots,p_r$, we have

$$P_e^*\left(A_1,\ldots,A_r\right) \ge \frac{1}{r-1}\sum_{(i,j):\,i<j} P_e^*(A_i,A_j), \tag{6}$$

where $A_i := p_i\sigma_i$. Taking then r sequences of states $\sigma_{i,n}$, $n\in\mathbb{N}$, and $\vec{A}_i := \{p_i\sigma_{i,n}\}_{n\in\mathbb{N}}$, we get

$$\underline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right) \ge \max_{(i,j):\,i\neq j}\underline{p}_e\left(\vec{A}_i,\vec{A}_j\right). \tag{7}$$

Note that this is true for arbitrary sequences of states $\{\sigma_{i,n}\}_{n\in\mathbb{N}}$, with no special assumption on the correlations. In the i.i.d. case the right-hand side (RHS) of (7) is exactly $-\min_{(i,j):\,i\neq j} C(\sigma_i,\sigma_j)$, and we recover the optimality part of (5).

Hence, in this paper we will focus on upper decoupling bounds. Upper bounds on the optimal error in terms of the pairwise fidelities can easily be obtained from some results in

[8]:

$$P_e^*(A_1,\dots,A_r) \le \frac{1}{2}\sum_{(i,j):\,i\neq j}\sqrt{p_ip_j}F(\sigma_i,\sigma_j), \tag{8}$$

where $F(\sigma_i,\sigma_j) := \|\sigma_i^{1/2}\sigma_j^{1/2}\|_1$ is the fidelity. We provide a short proof of this bound in Appendix E. When all the σ_i are of rank one, the above bound can be improved as [19]

$$P_e^*(A_1,\dots,A_r) \le \frac{1}{2}\sum_{(i,j):\,i\neq j}\frac{p_i^2+p_j^2}{p_ip_j}F^2(\sigma_i,\sigma_j). \tag{9}$$

Using the Fuchs–van de Graaf inequalities [15], these bounds can easily be translated into bounds in terms of the pairwise error probabilities, and we obtain the following converses to (6) and (7):

Single-shot upper decoupling bounds: For $A_i := p_i\sigma_i$, $i=1,\dots,r$,

$$P_e^*(A_1,\dots,A_r) \le \sum_{(i,j):\,i\neq j}\sqrt{p_i+p_j}\sqrt{P_e^*(A_i,A_j)}. \tag{10}$$

If A_i is rank one for all i then the square root can be removed from the pairwise errors; more precisely,

$$P_e^*(A_1,\dots,A_r) \le \sum_{(i,j):\,i\neq j}(p_i+p_j)\frac{p_i^2+p_j^2}{p_i^2p_j^2}P_e^*\left(A_i,A_j\right). \tag{11}$$

These single-shot bounds immediately yield the following

Asymptotic upper decoupling bounds: For $\vec{A}_i := \{p_i\sigma_{i,n}\}_{n\in\mathbb{N}}$, $i=1,\dots,r$,

$$\overline{p}_e\left(\vec{A}_1,\dots,\vec{A}_r\right) \le \frac{1}{2}\max_{(i,j):\,i\neq j}\overline{p}_e(\vec{A}_i,\vec{A}_j). \tag{12}$$

If $A_{i,n}$ is rank one for all i and n then

$$\overline{p}_e\left(\vec{A}_1,\dots,\vec{A}_r\right) \le \max_{(i,j):\,i\neq j}\overline{p}_e(\vec{A}_i,\vec{A}_j). \tag{13}$$

Note that again (12) and (13) are true for *arbitrary* sequences of states, and in the i.i.d. case we have $\max_{(i,j):\,i\neq j}\overline{p}_e(\vec{A}_i,\vec{A}_j) = -\min_{(i,j):\,i\neq j}C(\sigma_i,\sigma_j) = -C(\sigma_1,\dots,\sigma_r)$. In particular, by (7) and (13) we recover the result of [40], i.e., that (5) is true for pure states. For mixed states, Theorem 3 in [41] gives that $\overline{p}_e(\vec{A}_1,\dots,\vec{A}_r)$ is between $-C(\sigma_1,\dots,\sigma_r)$ and $-\frac{1}{3}C(\sigma_1,\dots,\sigma_r)$. Our bound (12) improves the factor 1/3 in this upper bound to 1/2, which is the best general result known so far.

Analytical proofs for various special cases and extensive numerical simulations for the general case suggest that the square root in (10) – and, consequently, the factor $\frac{1}{2}$ in (12) – can be removed. The fidelity bounds in (8) and (9) were obtained by bounding from above the error probability of the pretty good measurement, using matrix analytic techniques. Here we explore a completely different approach to obtaining upper decoupling bounds. Namely, we show that the optimal error P_e^* can be bounded from above by the sum of the optimal error probabilities of r binary state discrimination problems, where in each of these problems, the goal is to discriminate one of the original hypotheses from all the rest. Thus, finding the optimal error exponent of the symmetric multiple state discrimination problem with i.i.d. hypotheses can be reduced to finding the optimal error exponent of the correlated binary state discrimination problem where one of the hypotheses is i.i.d., while the other is a convex mixture of i.i.d. states. This latter problem is interesting in its own right, as it is arguably the simplest non-i.i.d. state discrimination problem and yet its solution is not yet known, despite considerable effort towards establishing non-i.i.d. analogs of the binary Chernoff bound theorem [21, 22, 23, 31, 32]. Here we make some progress towards the solution of this problem, and provide a complete solution when the i.i.d. state is pure.

The structure of the paper is as follows. In Section 2 we summarize the necessary preliminaries and review the known results that are relevant for the rest of the paper. In particular, we give a short proof of (6), and summarize the known results for the binary case. We also introduce the notion of the least upper bound for self-adjoint operators, and show how the optimal error probability can be expressed in this formalism.

In Section 3 we first review the fidelity bounds of [8] and [19], which are based on the performance of the suboptimal pretty good measurement. Then we follow a similar approach to obtain bounds in terms of pairwise fidelity-like quantities. From these bounds we can recover (8)–(9) up to a constant factor, and for some configurations they are strictly better than (8)–(9). This approach is based on Tyson's bound [50] on the performance of an other suboptimal family of measurements, the square measurements.

In Section 4 we study a special binary problem where one of the hypotheses is i.i.d., while the other one is averaged i.i.d., i.e., a convex mixture of i.i.d. states. In the setting of Stein's lemma (with the averaged state being the null-hypothesis) the corresponding error exponent is known to be the worst-case pairwise exponent [9] (see also [35] for a simple proof), and we conjecture the same to hold in the symmetric setting of the Chernoff bound. Similarly to the case of multiple hypotheses, it is easy to show that the worst-case pairwise exponent cannot be exceeded (optimality). In Theorem 4.3 we present upper decoupling bounds on the error probability, analogous to (10) and (11), which in the asymptotics yield that $1/2$ times the conjectured exponent is achievable. Moreover, when the i.i.d. state is pure then the factor $1/2$ can be removed and we get both optimality and achievability.

In Section 5 we show that the exponential decay rate of the optimal error probability (1) is the same as that of another quantity, which we call the *dichotomic error*. This is defined as the sum of the error probabilities of the binary state discrimination problems where we only want to decide whether hypothesis i is true or not, for every $i = 1, \ldots, r$. In the i.i.d. case these binary problems are exactly of the type discussed in Section 4, and we can directly apply the bounds obtained there to get upper decoupling bounds on both the single-shot and the asymptotic error probabilities, which give the bounds (10)–(13).

In Section 6 we follow Nussbaum's approach [42] to obtain a different kind of decoupling of the optimal error probability. When applied recursively and combined with the bounds of Section 4, this approach provides an alternative way to obtain bounds of the type (10)–(11), which again yield (12)–(13) in the asymptotics.

In Section 7 we show how the various single-shot bounds of the above described approaches translate into bounds for the error rates, i.e., we derive (12)–(13) for the most general scenario, and its variants for more specific settings, where the pairwise rates can be replaced by pairwise Chernoff divergences. In particular, we improve on the result of [40] by showing that (5) holds if at least $r-2$ of the states σ_i are pure. We also give an improvement of Nussbaum's asymptotic result [42], which says that (5) is true if there is a pair of states (σ_i, σ_j) such that $C(\sigma_i, \sigma_j) < \frac{1}{6} C(\sigma_k, \sigma_l)$ for any $(k, l) \neq (i, j)$. Here we show that the constant $\frac{1}{6}$ can be replaced with $\frac{1}{2}$.

Supplementary material is provided in a number of Appendices. In Appendix A, we explore some properties of the least upper bound and the greatest lower bound for self-adjoint operators, which further extend their analogy to the classical notions of minimum and maximum. In Appendix B we show how our approaches work in the classical case (when all operators commute), thus providing various alternative proofs for (5) in the classical case. In Appendix C we review the pure state case; we show an elementary way to derive the Chernoff bound theorem (4) for two pure states, and show how the combination of the single-shot bounds of [19] and [45] yield (5) for an arbitrary number of pure states. In Appendix D we review the dual formulation of the optimal error probability due to [53]. For readers' convenience, we provide a proof for Tyson's and Barnum and Knill's error bounds in Appendix E.

2 Preliminaries

2.1 Notations

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the set of linear operators on $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ denote the set of self-adjoint (Hermitian) operators, $\mathcal{B}(\mathcal{H})_+$ the set of positive semidefinite (PSD) operators, and $\mathcal{S}(\mathcal{H})$ the set of density operators (states), i.e., the set of PSD operators with unit trace.

For X a Hermitian operator, let $|X|$ denote its absolute value (or modulus), $|X| := \sqrt{X^2}$. The Jordan decomposition of X into its positive and negative parts is given by $X = X_+ - X_-$, with $X_\pm = (|X| \pm X)/2$, and $|X| = X_+ + X_-$. It is clear that $X_+ X_- = 0$. As the eigenvalues of $|X|$ are the absolute values of the eigenvalues of X, the eigenvalues of X_+ (X_-) are the positive (negative) eigenvalues of X. We denote the projections onto the support of X_+ and X_- by $\{X > 0\}$ and $\{X < 0\}$, respectively.

We will follow the convention that powers of a positive semidefinite (PSD) operator are only taken on its support. That is, if $a_1, \ldots, a_r$ are the strictly positive eigenvalues of $A \geq 0$, with corresponding spectral projections $P_1, \ldots, P_r$, then $A^s := \sum_{i=1}^r a_i^s P_i$ for every $s \in \mathbb{R}$. In particular, A^0 denotes the projection onto the support of A.

By a POVM we will mean a set of PSD operators $E_1, \ldots, E_r$ such that $E_1 + \ldots + E_r \leq I$. On occasion we will also consider the underlying measurement operators $\{X_k\}_{k=1}^r$, which are sets of operators such that the products $X_k^* X_k$ constitute a POVM.

We will normally not indicate the base of the logarithm, but we will always assume that it is larger than 1, and hence log is a strictly increasing function. We will use the conventions

$\log 0 := -\infty$ and $\log +\infty := +\infty$.

2.2 The problem setting

We will consider a generalized state discrimination problem, where the hypotheses are represented by arbitrary non-zero PSD operators (i.e., not necessarily states). We consider such a generalized setting partly to absorb the priors into the states to make the formalism simpler, and partly because the formalism supports it, and all our results can be formulated and proved in this more general setting. More importantly, however, we need to treat such generalized setups even if we restrict our original hypotheses to be states; see, e.g., Lemma 6.2.

More in detail, in the single-shot case our hypotheses are represented by non-zero PSD operators $A_1, \ldots, A_r$. Occasionally, we will use the notations

$$p_k := \operatorname{Tr} A_k, \qquad \sigma_k := A_k/p_k, \qquad k = 1, \ldots, r.$$

If $p_1 + \ldots + p_r = 1$ then we say that $\{A_k\}$ forms a set of *weighted states.*

For any POVM $E_1, \ldots, E_r$, we define the corresponding *success- and error probabilities* as

$$P_s(\{E_i\}) := \sum_{i=1}^{r} \operatorname{Tr} A_i E_i, \qquad\qquad P_e(\{E_i\}) := \sum_{i=1}^{r} \operatorname{Tr} A_i (I - E_i).$$

These can indeed be interpreted as probabilities in the case of weighted states, whereas in the general case they might take values above 1. Since it will always be obvious what the hypotheses are, we don't indicate them in the above notations. The optimal values of these quantities over all possible choices of POVMs are the optimal success- and error probability

$$P_s^*(A_1, \ldots, A_r) := \max\left\{\sum\nolimits_{i=1}^{r} \operatorname{Tr} A_i E_i \,:\, \{E_1, \ldots, E_r\} \text{ POVM}\right\}, \tag{14}$$
$$P_e^*(A_1, \ldots, A_r) := \min\left\{\sum\nolimits_{i=1}^{r} \operatorname{Tr} A_i (I - E_i) \,:\, \{E_1, \ldots, E_r\} \text{ POVM}\right\}.$$

The maximum and the minimum above exist because the domain of optimization is compact and the functions to optimize are continuous with respect to any natural topology on the set of POVMs on a fixed set of outcomes. We will use the shorthand notations P_s^* and P_e^* when it is clear what the hypotheses are. Note that

$$P_s^*(A_1, \ldots, A_r) + P_e^*(A_1, \ldots, A_r) = \operatorname{Tr} A_0, \qquad \text{where} \qquad A_0 := \sum_{i=1}^{r} A_i.$$

Again, these can be interpreted as probabilities if $1 = \operatorname{Tr} A_0 = \sum_i p_i$, i.e., in the case of weighted states. Note that any POVM that is optimal for P_s^* is also optimal for P_e^* and vice versa. Moreover, there always exists an optimal POVM $\{E_i\}_{i=1}^r$ such that $E_1 + \ldots + E_r = I$.

In the asymptotic setting, our hypotheses are going to be represented by sequences of PSD operators, $\vec{A}_i := \{A_{i,n}\}_{n\in\mathbb{N}}$, $i = 1, \ldots, r$, and we will be interested in the exponents $\underline{p}_e$ and $\overline{p}_e$, defined in (2) and (3), respectively. We say that the i-th hypothesis is *i.i.d.* (for the classical analogy of independent and identically distributed) if $p_{i,n} = \operatorname{Tr} A_{i,n}$ is independent of n (and hence we can define $p_i := p_{i,n}$, $n \in \mathbb{N}$), and $A_{i,n}/p_i = \sigma_i^{\otimes n}$ for every i, where $\sigma_i := \sigma_{i,1}$. We say that the asymptotic state discrimination problem is i.i.d. if all the hypotheses are i.i.d.

When passing from single-shot error bounds to asymptotic error bounds, we will use the following standard lemma without further notice.

Lemma 2.1. Let $a_{i,n}$, $n \in \mathbb{N}$, $i = 1, \ldots, r$, be sequences of positive numbers. Then

$$\max_i \liminf_{n\to\infty} \frac{1}{n} \log a_{i,n} \le \liminf_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^r a_{i,n} \le \limsup_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^r a_{i,n} \le \max_i \limsup_{n\to\infty} \frac{1}{n} \log a_{i,n}.$$

Proof. The first inequality is straightforward from $a_{i,n} \le \sum_i a_{i,n}$, $\forall i$, and the second inequality is obvious. To prove the last inequality, let $M := \max_i \limsup_{n\to\infty} \frac{1}{n} \log a_{i,n}$. If $M = +\infty$ then the assertion is trivial, and hence we assume that $M < +\infty$. By the definition of the limit superior, for every $M' > M$, there exists an $N_{M'}$ such that for all $n \ge N_{M'}$, $a_{i,n} < \exp(M')$, $i = 1, \ldots, r$, and hence $\frac{1}{n} \log \sum_{i=1}^r a_{i,n} < \frac{1}{n} \log r + M'$. Thus $\limsup_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^r a_{i,n} \le M'$. Since this is true for all $M' > M$, the assertion follows. □

2.3 The generalized maximum likelihood error

In the classical state discrimination problem, where the hypotheses are represented by non-negative functions $A_i : \mathcal{X} \to \mathbb{R}_+$ on some finite set $\mathcal{X}$, the optimal success probability is known to be $\sum_x \max\{A_1, \ldots, A_r\}$, and it is achieved by the maximum likelihood measurement (see Appendix B for details). If we consider the A_i as diagonal operators in some fixed basis, then P_s^* can be rewritten as

$$P_s^*(A_1, \ldots, A_r) = \operatorname{Tr} \max\{A_1, \ldots, A_r\}, \tag{15}$$

where $\max\{A_1, \ldots, A_r\}$ is the operator with $\max_i A_i(x)$ in its diagonals. Note that this is not a maximum in the usual sense of PSD ordering; indeed, it is well-known that the PSD ordering does not induce a lattice structure [3], so in general the set of upper bounds to r given self-adjoint operators $A_1, \ldots, A_r$, which is defined as $\mathcal{A} := \{Y : Y \ge A_k,\ k = 1, \ldots, r\}$, has no minimal element, not even when the A_k mutually commute; see, e.g. Example A.1 in Appendix A. However, there is a unique minimal element within $\mathcal{A}$ in terms of the *trace ordering*. We can therefore define a least upper bound in this more restrictive sense as

$$\operatorname{LUB}(A_1, \ldots, A_r) := \arg\min_Y \{\operatorname{Tr} Y : Y \ge A_k,\ k = 1, \ldots, r\}. \tag{16}$$

For the proof of uniqueness, see Appendix A. In a similar vein we can define the *greatest lower bound* (GLB) as

$$\operatorname{GLB}(A_1, \ldots, A_r) := \arg\max_Y \{\operatorname{Tr} Y : Y \le A_k,\ k = 1, \ldots, r\}. \tag{17}$$

Clearly, we have

$$\operatorname{GLB}(A_1, \ldots, A_r) = -\operatorname{LUB}(-A_1, \ldots, -A_r). \tag{18}$$

For further properties of the above notions, see Appendix A.

Note that the set of k-outcome POVMs forms a convex set, and the optimal success probability in (14) is given as the maximum of a linear functional over this convex set. It was shown in [53] that the duality of convex optimization yields

$$P_s^*(A_1,\dots,A_r) = \min\{\operatorname{Tr} Y : Y \geq A_k,\ k=1,\dots,r\} \tag{19}$$

(see also [28] for a different formulation of the same result). Using the definition of the LUB above, this can be rewritten as

$$P_s^*(A_1,\dots,A_r) = \operatorname{Tr}\operatorname{LUB}(A_1,\dots,A_r), \tag{20}$$

in complete analogy with the classical case (15). For readers' convenience, we provide a detailed derivation of (19) in Appendix D.

For an ensemble of PSD operators $\{A_i\}_{i=1}^r$, we define the *complementary operator* of A_i as the operator given by the sum of all other operators in the ensemble:

$$\bar{A}_i := \sum_{j\neq i} A_j = A_0 - A_i,$$

where $A_0 = \sum_i A_i$. The optimal error probability can be expressed in terms of the GLB of the complementary density operators:

$$P_e^*(A_1,\dots,A_r) = \operatorname{Tr}\operatorname{GLB}(\bar{A}_1,\dots,\bar{A}_r). \tag{21}$$

This is easy to show:

$$\begin{aligned} P_e^* &= \min_{\{E_k\}} \sum_k \operatorname{Tr} A_k \sum_{l:\, l\neq k} E_l = \min_{\{E_k\}} \sum_l \operatorname{Tr} E_l \sum_{k:\, k\neq l} A_k = \min_{\{E_k\}} \sum_l \operatorname{Tr} E_l \bar{A}_l = -\max_{\{E_k\}} \sum_l \operatorname{Tr} E_l(-\bar{A}_l) \\ &= -\operatorname{Tr}\operatorname{LUB}(-\bar{A}_1,\dots,-\bar{A}_r) = \operatorname{Tr}\operatorname{GLB}(\bar{A}_1,\dots,\bar{A}_r), \end{aligned}$$

where we used (20), (18), and that an optimal POVM can be chosen so that $E_1+\ldots+E_r = I$. Note that this is in general different from $\operatorname{Tr}\operatorname{GLB}(A_1,\dots,A_r)$, which is the minimal i.e. worst-case success probability $P_{s,\min} = \min_{\{E_k\}} \sum_k \operatorname{Tr} A_k E_k$.

In the binary case, i.e., when $r=2$, we have

$$\begin{aligned} P_s^*(A_1,A_2) &= \max\{\operatorname{Tr} A_1 E + \operatorname{Tr} A_2(I-E) :\ 0 \leq E \leq I\} = \operatorname{Tr} A_2 + \max_{0\leq E\leq I} \operatorname{Tr}(A_1-A_2)E \\ &= \operatorname{Tr} A_2 + \operatorname{Tr}(A_1-A_2)_+ = \frac{1}{2}\operatorname{Tr}(A_1+A_2) + \frac{1}{2}\left\|A_1-A_2\right\|_1, \end{aligned}$$

and the maximum is attained at $E = \{A_1 - A_2 > 0\}$; this is the so-called Holevo-Helström measurement [25, 20]. Consequently, we have

$$\begin{aligned} P_e^*(A_1,A_2) &= \min\{\operatorname{Tr} A_1(I-E) + \operatorname{Tr} A_2 E :\ 0 \leq E \leq I\} \qquad (22) \\ &= \operatorname{Tr}(A_1+A_2) - P_s^*(A,A_2) = \frac{1}{2}\operatorname{Tr}(A_1+A_2) - \frac{1}{2}\left\|A_1-A_2\right\|. \qquad (23) \end{aligned}$$

Comparing these with (20) and (21), and noting that in the binary case $\bar{A}_1 = A_2$, $\bar{A}_2 = A_1$, we obtain

$$\begin{aligned}\operatorname{Tr}\operatorname{LUB}(A_1, A_2) &= \frac{1}{2}\operatorname{Tr}(A_1 + A_2) + \frac{1}{2}\left\|A_1 - A_2\right\|_1, \\ \operatorname{Tr}\operatorname{GLB}(A_1, A_2) &= \frac{1}{2}\operatorname{Tr}(A_1 + A_2) - \frac{1}{2}\left\|A_1 - A_2\right\|_1.\end{aligned}$$

For a more straightforward way to derive these identities, see Appendix A.

In the rest of the paper, we will use the notations $P_e^*(A,B)$, $\operatorname{Tr}\operatorname{GLB}(A,B)$ and $\frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\left\|A-B\right\|_1$ interchangeably for PSD operators A, B.

2.4 Chernoff bound for binary state discrimination

For PSD operators A, B on the same Hilbert space, define

$$\begin{aligned}Q_s(A\|B) &:= \operatorname{Tr} A^s B^{1-s}, \qquad s \in \mathbb{R}, \\ Q_{\min}(A,B) &:= \min_{0\le s\le 1} Q_s(A\|B), \\ C(A,B) &:= -\log Q_{\min}(A,B).\end{aligned} \tag{24}$$

The last quantity, $C(A,B)$ is the *Chernoff divergence* of A and B. As it was shown in Theorem 1 in [4] (see also [5, 7]),

$$\frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\left\|A-B\right\|_1 \le Q_s(A\|B), \qquad s \in [0,1]. \tag{25}$$

Consider now the generalized asymptotic binary hypothesis testing problem with hypotheses $\vec{A}_1, \vec{A}_2$. By (25), we have

$$P_e^*(A_{1,n}, A_{2,n}) = \frac{1}{2}\operatorname{Tr}(A_{1,n} + A_{2,n}) - \frac{1}{2}\left\|A_{1,n} - A_{2,n}\right\|_1 \le Q_s(A_{1,n}\|A_{2,n}) \tag{26}$$

for every $n \in \mathbb{N}$ and $s \in [0,1]$, and hence,

$$\overline{p}_e(\vec{A}_1, \vec{A}_2) = \limsup_{n\to\infty} \frac{1}{n}\log P_e^*(A_{1,n}, A_{2,n}) \le -C(\vec{A}_1, \vec{A}_2), \tag{27}$$

where

$$C(\vec{A}_1, \vec{A}_2) := -\limsup_{n\to\infty}\frac{1}{n}\log Q_{\min}(A_{1,n}, A_{2,n}) = \liminf_{n\to\infty}\frac{1}{n}C(A_{1,n}, A_{2,n}) \tag{28}$$

is the *regularized Chernoff divergence.*

In the i.i.d. case, i.e., when $A_{1,n} = p_1\sigma_1^{\otimes n}$ and $A_{2,n} = p_2\sigma_2^{\otimes n}$ for every $n \in \mathbb{N}$, we have $Q_s(A_{1,n}\|A_{2,n}) = p_1^s p_2^{1-s} Q_s(\sigma_1\|\sigma_2)^n \le \max\{p_1, p_2\} Q_s(\sigma_1\|\sigma_2)^n$, and (26) yields

$$P_e^*(p_1\sigma_1^{\otimes n}, p_2\sigma_2^{\otimes n}) \le \max\{p_1,p_2\} Q_{\min}(\sigma_1,\sigma_2)^n = \max\{p_1,p_2\}\exp\left(-nC(\sigma_1,\sigma_2)\right)$$

for every $n \in \mathbb{N}$. In particular,

$$\limsup_{n\to\infty}\frac{1}{n}\log P_e^*(p_1\sigma_1^{\otimes n}, p_2\sigma_2^{\otimes n}) \le -C(\sigma_1,\sigma_2).$$

(Note that in this case $C(\vec{A}_1, \vec{A}_2) = C(\sigma_1, \sigma_2)$).

The above argument shows that the asymptotic Chernoff divergence (which is equal to the single-shot Chernoff divergence in the i.i.d. case) is an achievable error rate. Optimality means that no faster exponential decay of the optimal error is possible, i.e., that

$$\underline{p}_e(\vec{A}_1, \vec{A}_2) = \liminf_{n\to\infty} \frac{1}{n} \log P_e^*(A_{1,n}, A_{2,n}) \geq -C(\vec{A}_1, \vec{A}_2).$$

This was shown to be true in the i.i.d. case in [38]. Optimality for various correlated scenarios was obtained in [21, 22, 23, 31, 32]; the classes of states covered include Gibbs states of finite-range translation-invariant interactions on a spin chain, and thermal states of non-interacting bosonic and fermionic lattice systems.

2.5 Pairwise discrimination and lower decoupling bounds

Consider the generalized state discrimination problem with hypotheses $A_1, \ldots, A_r$. In this section we review a lower bound on the optimal error probability for discriminating between r given states in terms of the optimal *pairwise* error probabilities, originally given in [45]. Let us thereto define the following quantities:

$$\begin{aligned} P_{s,2}^* \;:=\; P_{s,2}^*(A_1, \ldots, A_r) \;&:=\; \frac{1}{r-1} \sum_{(k,l)\colon k<l} P_s^*(A_k, A_l) = \frac{1}{r-1} \sum_{(k,l)\colon k<l} \operatorname{Tr} \mathrm{LUB}(A_k, A_l) \\ &=\; \frac{1}{r-1} \sum_{(k,l)\colon k<l} \frac{1}{2} \left(\operatorname{Tr}(A_k + A_l) + \|A_k - A_l\|_1\right). \end{aligned} \tag{29}$$

and

$$\begin{aligned} P_{e,2}^* \;:=\; P_{e,2}^*(A_1, \ldots, A_r) \;&:=\; \frac{1}{r-1} \sum_{(k,l)\colon k<l} P_e^*(A_k, A_l) = \frac{1}{r-1} \sum_{(k,l)\colon k<l} \operatorname{Tr} \mathrm{GLB}(A_k, A_l) \\ &=\; \frac{1}{r-1} \sum_{(k,l)\colon k<l} \frac{1}{2} \left(\operatorname{Tr}(A_k + A_l) - \|A_k - A_l\|_1\right). \end{aligned} \tag{30}$$

Note that

$$P_{s,2}^*(A_1, \ldots, A_r) + P_{e,2}^*(A_1, \ldots, A_r) = \sum_i \operatorname{Tr} A_i = P_s^*(A_1, \ldots, A_r) + P_e^*(A_1, \ldots, A_r),$$

explaining the choice $1/(r-1)$ for the normalization.

In the case of weighted states, i.e., when $\operatorname{Tr} A_0 = 1$, we can interpret these quantities as optimal success and error probabilities in a very special setting, whereby the receiver can make use of a particular kind of side information. We shall assume that this side information has been provided by an oracle. The oracle knows the correct value of each symbol sent out by the source, but in the best of oracular traditions, does not quite reveal this information to the receiver. Rather, the oracle provides the receiver with a choice of two symbols, one of which is the correct one and the other is chosen from the remaining values at random, with uniform probability $1/(r-1)$. It is intuitively plausible that the receiver should only try to discriminate between the two options provided.

The optimal success probability in this setup can easily be calculated. From the receiver's viewpoint, the probability that the values of the symbols provided by the oracle are k and l (with $k \neq l$) is $p_k(r-1)^{-1} + (r-1)^{-1}p_l$, and the conditional probability that k is the correct one is $p_k/(p_k+p_l)$. Hence, the receiver's optimal success probability will be

$$P_{s,2}^* = \sum_{(k,l):\, k<l} \frac{p_k + p_l}{r-1} \;\; \frac{1}{2}\left(1 + \left\| \frac{p_k}{p_k+p_l}\sigma_k - \frac{p_l}{p_k+p_l}\sigma_l \right\|_1 \right),$$

which simplifies to (29).

It is intuitively clear that this oracle-assisted success probability should never be smaller than the unassisted optimal success probability, whereby the receiver needs to discriminate between all r possible symbols. The following Theorem, first given in [45], shows that this is indeed the case. Here we give a detailed and slightly simplified proof for readers' convenience. We also provide a different proof and a strengthening of (31) in Theorem 5.2.

Theorem 2.2. For any $A_1, \dots, A_r \in \mathcal{B}(\mathcal{H})_+$,

$$P_s^*(A_1,\dots,A_r) \le P_{s,2}^*(A_1,\dots,A_r), \qquad \text{and} \qquad P_e^*(A_1,\dots,A_r) \ge P_{e,2}^*(A_1,\dots,A_r). \tag{31}$$

Proof. First notice that

$$\sum_{(k,l):\, k\neq l} (\operatorname{Tr} A_k E_k + \operatorname{Tr} A_l E_l) = 2(r-1) \sum_{k=1}^{r} \operatorname{Tr} A_k E_k$$

and

$$\begin{aligned}
\operatorname{Tr} A_k E_k + \operatorname{Tr} A_l E_l &\le \operatorname{Tr} A_k E_k + \operatorname{Tr} A_l (I - E_k) = \operatorname{Tr} A_l + \operatorname{Tr}(A_k - A_l) E_k \\
&\le \operatorname{Tr} A_l + \operatorname{Tr}(A_k - A_l)_+ = \frac{1}{2}(\operatorname{Tr}(A_k + A_l) + \|A_k - A_l\|_1).
\end{aligned}$$

Hence,

$$\begin{aligned}
\sum_{k=1}^{r} \operatorname{Tr} A_k E_k &= \frac{1}{2(r-1)} \sum_{(k,l):\, k\neq l} (\operatorname{Tr} A_k E_k + \operatorname{Tr} A_l E_l) \\
&\le \frac{1}{4(r-1)} \sum_{(k,l):\, k\neq l} (\operatorname{Tr}(A_k + A_l) + \|A_k - A_l\|_1),
\end{aligned}$$

which yields the first assertion. The second assertion, $P_e^* \ge P_{e,2}^*$, is now obvious. □

We conjecture that for any choice of signal states and source probabilities the oracle-assisted error probability can not be arbitrarily smaller than the unassisted one. In particular, we believe:

Conjecture 2.3. There exists a constant c, only depending on the number of hypotheses r, such that for all $A_1, \dots, A_r \ge 0$

$$P_e^*(A_1,\dots,A_r) \le c P_{e,2}^*(A_1,\dots,A_r).$$

We have ample numerical evidence for this conjecture, and this evidence suggests that $c = 4(r-1)$. Several approaches towards a proof will be provided in the next sections.

2.6 Inequalities for various operator distinguishability measures

We will often benefit from inequalities between various operator distinguishability measures. In particular, we will use inequalities between the optimal binary error, the Chernoff divergence, and the fidelity. For positive semidefinite operators $A, B \in \mathcal{B}(\mathcal{H})_+$, their *fidelity* is defined as

$$F(A,B) := \left\| A^{1/2} B^{1/2} \right\|_1 = \operatorname{Tr}(A^{1/2} B A^{1/2})^{1/2}. \tag{32}$$

The following bounds between the fidelity and the trace-norm were shown in [15] for states, and extended to weighted states in [7], where also the sharpness of the inequalities $\frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\|A-B\|_1 \le F(A,B) \le \sqrt{\frac{1}{4}\left(\operatorname{Tr}(A+B)\right)^2 - \frac{1}{4}\|A-B\|_1^2}$ has been shown. The proof for the general case can be obtained exactly the same way as in the above cases.

Lemma 2.4. For any $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$\begin{aligned} P_e^*(A,B) &= \operatorname{Tr}\operatorname{GLB}(A,B) = \frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\|A-B\|_1 \le F(A,B) \\ &\le \sqrt{\frac{1}{4}\left(\operatorname{Tr}(A+B)\right)^2 - \frac{1}{4}\|A-B\|_1^2} = \sqrt{\operatorname{Tr}\operatorname{LUB}(A,B)}\sqrt{\operatorname{Tr}\operatorname{GLB}(A,B)} \\ &\le \sqrt{\operatorname{Tr}(A+B)}\sqrt{\operatorname{Tr}\operatorname{GLB}(A,B)} = \sqrt{\operatorname{Tr}(A+B)}\sqrt{P_e^*(A,B)}. \end{aligned} \tag{33}$$

When A is rank one, we also have the following inequality. This has been stated as an exercise in [37] for states; we provide a proof here for readers' convenience.

Lemma 2.5. Let $A, B \in \mathcal{B}(\mathcal{H})_+$ and assume that A has rank one. Then

$$P_e^*(A,B) = \operatorname{Tr}\operatorname{GLB}(A,B) = \frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\|A-B\|_1 \le \frac{1}{\operatorname{Tr} A} F(A,B)^2. \tag{34}$$

Proof. Let $\tilde{A} := A/\operatorname{Tr} A$. The assumption that A is rank one yields that $\operatorname{Tr} A\tilde{A} = \operatorname{Tr} A$ and $F(A,B) = \sqrt{\operatorname{Tr} AB}$. Using the representation (23), we get

$$\begin{aligned} \frac{1}{2}\operatorname{Tr}(A+B) - \frac{1}{2}\|A-B\|_1 &= \min\{A(I-E) + \operatorname{Tr} BE : 0 \le E \le I\} \\ &\le \operatorname{Tr} A(I-\tilde{A}) + \operatorname{Tr} B\tilde{A} = \frac{1}{\operatorname{Tr} A}\operatorname{Tr} AB = \frac{1}{\operatorname{Tr} A} F(A,B)^2. \quad \square \end{aligned}$$

Remark 2.6. Monotonicity of the fidelity under the trace yields that $F(A,B) \le (\operatorname{Tr} A)^{1/2}(\operatorname{Tr} B)^{1/2}$. If $\operatorname{Tr} B \le \operatorname{Tr} A$ then $F(A,B) \le \operatorname{Tr} A$, or equivalently, $\frac{1}{\operatorname{Tr} A}F(A,B)^2 \le F(A,B)$, and hence the upper bound in (34) is stronger than the inequality in (33). This is the case, for instance, for states. In general, however, the two bounds are not comparable.

According to Theorem 6 in [5], for any $A, B \in \mathcal{B}(\mathcal{H})_+$,

$$F(A,B)^2 \le \operatorname{Tr} A^t B^{1-t} (\operatorname{Tr} A)^{1-t} (\operatorname{Tr} B)^t, \qquad t \in [0,1].$$

In particular, for states ρ, σ,

$$F(\rho,\sigma)^2 \le Q_{\min}(\rho,\sigma) = \min_{0 \le t \le 1} \operatorname{Tr} \rho^t \sigma^{1-t}. \tag{35}$$

3 Upper bounds from suboptimal measurements

Consider the generalized state discrimination problem with hypotheses $A_i \in \mathcal{B}(\mathcal{H})_+$, $i = 1, \ldots, r$. As before, we write $A_i = p_i \sigma_i$, with $\operatorname{Tr} \sigma_i = 1$. When the number of hypotheses r is larger than 2, there is no explicit expression known for the optimal error probability $P_e^*(A_1, \ldots, A_r)$ in general. Obviously, any measurement yields an upper bound on the optimal error probability, some of which are known to have the same asymptotics in the limit of infinitely many copies as the optimal error probability. Here we first review the pretty good measurement (PGM), and the bounds (8)–(9) from [8, 19]. Next, we consider the square measurement (SM), and derive upper bounds on its optimal error probability. These upper bounds sometimes outperform those of (8)–(9).

For every $\alpha \in \mathbb{R}$, define the α-weighted POVM $\mathcal{E}^{(\alpha)}$ by

$$E_k^{(\alpha)} := S_\alpha^{-1/2} A_k^\alpha S_\alpha^{-1/2}, \qquad S_\alpha := \sum_k A_k^\alpha.$$

Note that if $A_k = |\psi_k\rangle\langle\psi_k|$ for some vectors ψ_k then $E_k^{(\alpha)} = |\psi_k^{(\alpha)}\rangle\langle\psi_k^{(\alpha)}|$ with $\psi_k^{(\alpha)} := S_\alpha^{-1/2} \|\psi_k\|^{\alpha-1} \psi_k$, and $\sum_k |\psi_k^{(\alpha)}\rangle\langle\psi_k^{(\alpha)}| = (\sum_k A_k)^0$. In particular, if $\psi_1, \ldots, \psi_r$ are linearly independent then $\psi_1^{(\alpha)}, \ldots, \psi_r^{(\alpha)}$ is an orthonormal system, spanning the same subspace as the original ψ vectors. That is, the above procedure yields an orthogonalization of the original set of vectors, which is different from the Gram-Schmidt orthogonalization in general.

The case $\alpha = 1$ yields the so-called *pretty good measurement* (PG) [16]. Barnum and Knill [8] have shown that in the case of weighted states, the success probability of the PG measurement is bounded below by the square of the optimal success probability: $(P_s^*)^2 \le P_s^{PG} \le P_s^*$, which in turn yields that $\frac{1}{2} P_e^{PG} \le P_e^* \le P_e^{PG}$. In particular, P_e^* and P_e^{PG} have the same exponential decay rate in the asymptotic setting.

Theorem 3.1 (Barnum and Knill)**.**

$$P_e^{PG} \le \frac{1}{2} \sum_{(i,j):\, i \neq j} F(A_i, A_j) = \frac{1}{2} \sum_{(i,j):\, i \neq j} \sqrt{p_i p_j} F(\sigma_i, \sigma_j). \tag{36}$$

Actually, Theorem 4 in [8] gives the upper bound in (36) without the 1/2 pre-factor. We give a short proof of the improved bound in Appendix E. This theorem immediately yields (8). It was shown in [19] that when all the A_i are rank one then

$$P_e^{PG} \le \frac{1}{2} \sum_{(i,j):\, i \neq j} \frac{p_i^2 + p_j^2}{p_i^2 p_j^2} F(A_i, A_j)^2 = \frac{1}{2} \sum_{(i,j):\, i \neq j} \frac{p_i^2 + p_j^2}{p_i p_j} F(\sigma_i, \sigma_j)^2, \tag{37}$$

which yields (9).

The case $\alpha = 2$ yields the *square measurement* (SQ), with POVM elements

$$E_{SQ;k} = X_{SQ;k}^* X_{SQ;k}, \qquad X_{SQ;k} = A_k \left(\sum_k A_k^2 \right)^{-1/2}.$$

This type of measurement has been used by various authors [10, 12, 27, 25], and it features in Tyson's bounds on the error probability [50], which we briefly review below. For a comprehensive overview of the use of the pretty good and the square measurements for state discrimination, see [51].

For any set $\{X_k\}$ of measurement operators (i.e., $\sum_k X_k^* X_k \leq I$), let

$$\Gamma(\{X_k\}) := \operatorname{Tr} A_0 - \sum_{k=1}^{r} \|X_k A_k\|_1 . \tag{38}$$

Minimizing over all possible choices of $\{X_k\}$ yields the optimal value Γ^*:

$$\Gamma^* := \inf_{\{X_k\}: \sum_k X_k^* X_k = I} \Gamma(\{X_k\}). \tag{39}$$

The importance of this quantity Γ comes from a combination of two facts. First, it differs from the error probability P_e only by a factor between 1 and at most 2. Hence, Γ^* is a good approximation of P_e^*, especially in the asymptotic regime. Moreover, unlike the optimal error probability, Γ^* can be calculated explicitly by a closed-form expression. This is the content of the following two theorems, first proven by Tyson [50]. For completeness, we provide short proofs in Appendix E.

Theorem 3.2 (Tyson)**.** Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and $\{E_k = X_k^* X_k\}$ be a POVM. Then

$$\Gamma(\{X_k\}) \leq P_e(\{E_k\}) \leq 2\Gamma(\{X_k\}).$$

In particular, for the optimal POVM and optimal X_k that achieve the minimum:

$$\Gamma^* \leq P_e^* \leq 2\Gamma^*. \tag{40}$$

Theorem 3.3 (Tyson)**.** Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and $A_0 := \sum_i A_i$. Then

$$\Gamma^* = \operatorname{Tr} A_0 - \operatorname{Tr} \left(\sum_{i=1}^{r} A_i^2 \right)^{1/2}, \tag{41}$$

with the optimal measurement operators being those of the SQ measurement.

From (41) it follows that Γ^* can take values between 0 (when all A_i are mutually orthogonal) and $\operatorname{Tr} A_0 - \operatorname{Tr} A_0^2/\sqrt{r}$ (when all A_i are equal), whereas P_e^* lies between 0 and $(\operatorname{Tr} A_0)(1 - 1/r)$.

Tyson's theorems yield that

$$P_e^* \leq P_e^{SQ} \leq 2 \left[\operatorname{Tr} A_0 - \operatorname{Tr} \left(\sum_{i=1}^{r} A_i^2 \right)^{1/2} \right]. \tag{42}$$

Thus any decoupling bound on the RHS of (42) yields a decoupling bound on P_e^*. Here we show the following:

Proposition 3.4. Let $A_1, \ldots, A_r \in B(\mathcal{H})_+$ and $A_0 := \sum_i A_i$. Then

$$\begin{aligned}\operatorname{Tr} A_0 - \operatorname{Tr}\left(\sum_j A_j^2\right)^{1/2} &\le \operatorname{Tr} A_0 - (\operatorname{Tr} A_0)^{\frac{3}{2}}\left((\operatorname{Tr} A_0) + 2\sum_{(i,j):\, i<j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}\right)^{-1/2} \\ &\le \sum_{(i,j):\, i<j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}.\end{aligned}$$

Proof. According to Lieb's theorem, the functional $(B, C) \mapsto \operatorname{Tr}(B^t C^{1-t})$ is jointly concave for $0 < t \le 1$. That is, for PSD operators B_j and C_j,

$$\operatorname{Tr} \sum\nolimits_j B_j^t C_j^{1-t} \le \operatorname{Tr}\left(\sum\nolimits_j B_j\right)^t \left(\sum\nolimits_j C_j\right)^{1-t}.$$

Then, using the fact $|\operatorname{Tr} X| \le ||X||_1$ and Hölder's inequality,

$$\operatorname{Tr} \sum\nolimits_j B_j^t C_j^{1-t} \le \left\|\left(\sum\nolimits_j B_j\right)^t \left(\sum\nolimits_j C_j\right)^{1-t}\right\|_1 \le \left\|\left(\sum\nolimits_j B_j\right)^t\right\|_{1/s} \left\|\left(\sum\nolimits_j C_j\right)^{1-t}\right\|_{1/(1-s)}$$

for every $0 < s < 1$. Now take $t = 2/3$, $s = 1/3$, $B_j = A_j^{1/2}$ and $C_j = A_j^2$, then

$$\begin{aligned}\operatorname{Tr} \sum\nolimits_j A_j^{1/3} A_j^{2/3} &\le \left\|\left(\sum\nolimits_j A_j^{1/2}\right)^{2/3}\right\|_3 \left\|\left(\sum\nolimits_j A_j^2\right)^{1/3}\right\|_{3/2} \\ &= \left(\operatorname{Tr}\left(\sum\nolimits_j A_j^{1/2}\right)^2\right)^{1/3} \left(\operatorname{Tr}\left(\sum\nolimits_j A_j^2\right)^{1/2}\right)^{2/3}.\end{aligned}$$

Obviously, the LHS equals $\operatorname{Tr} A_0$. Taking the 3/2 power and rearranging then yields

$$\begin{aligned}\operatorname{Tr}\left(\sum\nolimits_j A_j^2\right)^{1/2} &\ge (\operatorname{Tr} A_0)^{\frac{3}{2}} \left(\operatorname{Tr}\left(\sum\nolimits_j A_j^{1/2}\right)^2\right)^{-1/2} = (\operatorname{Tr} A_0)^{\frac{3}{2}} \left(\sum\nolimits_{i,j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}\right)^{-1/2} \\ &= (\operatorname{Tr} A_0)^{\frac{3}{2}} \left(\operatorname{Tr} A_0 + 2\sum_{(i,j):\, i<j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}\right)^{-1/2} \\ &\ge \operatorname{Tr} A_0 - \sum_{(i,j):\, i<j} \operatorname{Tr} A_i^{1/2} A_j^{1/2},\end{aligned}$$

where in the last line we exploited the inequality $(a + x)^{-1/2} \ge a^{-1/2} - \frac{1}{2} a^{-3/2} x$. □

Theorem 3.5. Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and let $p_k := \operatorname{Tr} A_k$, $\sigma_k := A_k / \operatorname{Tr} A_k$. Then

$$P_e^*(A_1, \ldots, A_r) \le \sum_{(i,j):\, i \ne j} \operatorname{Tr} A_i^{1/2} A_j^{1/2} \le \sum_{(i,j):\, i \ne j} F(A_i, A_j) \tag{43}$$

$$\le \begin{cases} \sum_{(i,j):\, i \ne j} \sqrt{p_i + p_j}\ \ P_e^*(A_i, A_j)^{1/2}, \\ \sum_{(i,j):\, i \ne j} \sqrt{p_i p_j}\ \ Q_{\min}(\sigma_i, \sigma_j)^{1/2} \end{cases}. \tag{44}$$

In the special case that all states σ_k are pure, we have the improved bound

$$P_e^*(A_1,\ldots,A_r) \le \sum_{(i,j):\, i\neq j} \operatorname{Tr} A_i^{1/2}A_j^{1/2} = \sum_{(i,j):\, i\neq j} \sqrt{p_ip_j}\;\; Q_{\min}(\sigma_i,\sigma_j) \tag{45}$$

$$= \sum_{(i,j):\, i\neq j} \frac{1}{\sqrt{p_ip_j}}\;\; F(A_i,A_j)^2 \le \sum_{(i,j):\, i\neq j} \frac{p_i+p_j}{\sqrt{p_ip_j}}\;\; P_e^*(A_i,A_j). \tag{46}$$

Proof. The first inequalities in (43) and (45) are due to (42) and Proposition 3.4. The second inequality in (43) is obvious from $\operatorname{Tr} A_i^{1/2}A_j^{1/2} \le \|A_i^{1/2}A_j^{1/2}\|_1 = F(A_i,A_j)$. The first bound in (44) follows from lemma 2.4, while the second bound is due to (35). The identities in (45) and (46) are straightforward to verify, and the inequality in (46) is again due to lemma 2.4. □

Remark 3.6. Since the inequalities used to prove (44) can be saturated, the square roots in (44) cannot be removed from $P_e^*(A_i,A_j)^{1/2}$ and $Q_{\min}(\sigma_i,\sigma_j)^{1/2}$ in general.

Remark 3.7. When all states are pure and the prior is uniform (i.e., $p_k = \operatorname{Tr} A_k = 1/r\ \forall k$), we can use another argument. By the inequality $\operatorname{Tr}\rho\sigma \le 1-(||\rho-\sigma||_1/2)^2$ [7], we get

$$\operatorname{Tr}\sigma_j\sigma_k \le 1 - \|\sigma_j-\sigma_k\|_1^2/4 \le 2 - \|\sigma_j-\sigma_k\|_1 .$$

Hence,

$$\operatorname{Tr} A_j^{1/2}A_k^{1/2} = \frac{1}{r}\operatorname{Tr}\sigma_j\sigma_k \le \frac{1}{r}(2-\|\sigma_j-\sigma_k\|_1),$$

so that

$$1-\Gamma^* \ge \left(1+\frac{2}{r}\sum_{(j,k):\, j<k}(2-\|\sigma_j-\sigma_k\|_1)\right)^{-1/2}. \tag{47}$$

Based on extensive numerical simulations, we conjecture that the latter bound also holds for mixed states and for non-uniform priors:

Conjecture 3.8. For any $A_i \ge 0$ with $\sum_i \operatorname{Tr} A_i = 1$,

$$\Gamma^* \le 1-\big(1+4(r-1)P_{e,2}^*\big)^{-1/2} \le 2(r-1)P_{e,2}^*. \tag{48}$$

By Theorem 3.2, this would imply the inequality of Conjecture 2.3 for weighted states:

$$P_e^* \le 4(r-1)P_{e,2}^*.$$

Remark 3.9. Note that for any p_i, p_j, $(p_ip_j)^{3/2} \le p_ip_j \le (p_i^2+p_j)^2/2$, from which it follows that the constants in the bound

$$P_e^* \le \sum_{(i,j):\, i\neq j} \frac{1}{\sqrt{p_ip_j}}\;\; F(A_i,A_j)^2,$$

given in (46), are better than in (37), i.e., (46) gives a tighter upper bound on the optimal error than (37).

To compare the bounds in (36) and (43), first choose all the σ_j to be pure, i.e., $\sigma_j = |\psi_j\rangle\langle\psi_j|$ for some unit vectors ψ_j. Then $\operatorname{Tr} A_i^{1/2} A_j^{1/2} = \sqrt{p_i p_j} |\langle\psi_i, \psi_j\rangle|^2$, while $F(A_i, A_j) = \sqrt{p_i p_j} |\langle\psi_i, \psi_j\rangle|$. Choosing thus the ψ_j close to orthogonal, but not orthogonal, we see that the ratio

$$\frac{\sum_{(i,j):\, i\neq j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}}{\sum_{(i,j):\, i\neq j} F(A_i, A_j)}$$

can be arbitrarily small. By continuity, we can also add a small perturbation to obtain PSD operators A_j of full support with the same property. In this sense, the upper bound $P_e^* \leq \sum_{(i,j):\, i\neq j} \operatorname{Tr} A_i^{1/2} A_j^{1/2}$ in (43) can be arbitrarily better than the bound in (36), for any fixed r. On the other hand, there are configurations for which the bound in (36) outperforms the one in (43), due to the $1/2$ pre-factor in the former.

4 Binary state discrimination: i.i.d. vs. averaged i.i.d

Consider the binary state discrimination problem where one of the hypotheses is i.i.d., i.e., for n copies it is represented by $\rho^{\otimes n}$ for some state ρ, while the other hypothesis is averaged i.i.d., i.e., for n copies it is of the form $\sum_{i=1}^r q_i \sigma_i^{\otimes n}$ for some states $\sigma_1, \ldots, \sigma_r$, and a probability distribution $q_1, \ldots, q_r$. This represents a situation where we have a further uncertainty about the identity of the true state when the second hypothesis is true. Alternatively, this can be considered as a state discrimination problem with $r+1$ i.i.d. hypotheses, where we only want to know whether one of the hypotheses is true or not. If the state ρ has prior probability $0 < p < 1$ then the optimal error probability for n copies is

$$P_e^*\left(p\rho^{\otimes n}, (1-p)\sum_i q_i \sigma_i^{\otimes n}\right) = \frac{1}{2}\left(1 - \left\|p\rho^{\otimes n} - (1-p)\sum_i q_i \sigma_i^{\otimes n}\right\|_1\right).$$

Convexity of the trace-norm implies that

$$\begin{aligned} P_e^*\left(p\rho^{\otimes n}, (1-p)\sum_i q_i \sigma_i^{\otimes n}\right) &\geq \frac{1}{2}\sum_{i=1}^r q_i \left(1 - \left\|p\rho^{\otimes n} - (1-p)\sigma_i^{\otimes n}\right\|_1\right) \\ &= \sum_{i=1}^r q_i P_e^*\left(p\rho^{\otimes n}, (1-p)\sigma_i^{\otimes n}\right), \end{aligned} \tag{49}$$

and hence,

$$\underline{p}_e\left(\{p\rho^{\otimes n}\}_n, \left\{(1-p)\sum_i q_i \sigma_i^{\otimes n}\right\}_n\right) \geq \max_{1\leq i\leq r} \underline{p}_e\left(\{p\rho^{\otimes n}\}_n, \{(1-p)\sigma_i^{\otimes n}\}_n\right) = -\min_i C(\rho, \sigma_i).$$

Based on analytical proofs for various special cases as well extensive numerical search, we conjecture that the following converse decoupling inequality is also true:

Conjecture 4.1.

$$\overline{p}_e\left(\{p\rho^{\otimes n}\}_n, \left\{(1-p)\sum_i q_i \sigma_i^{\otimes n}\right\}_n\right) \leq \max_{1\leq i\leq r} \overline{p}_e\left(\{p\rho^{\otimes n}\}_n, \{(1-p)\sigma_i^{\otimes n}\}_n\right) = -\min_i C(\rho, \sigma_i).$$

This conjecture would immediately yield

Conjecture 4.2.

$$\lim_{n\to\infty} \frac{1}{n} \log P_e^* \left(p\rho^{\otimes n}, (1-p) \sum\nolimits_i q_i \sigma_i^{\otimes n} \right) = -\min_i C(\rho, \sigma_i).$$

Below we will prove Conjecture 4.1 in the case where ρ is a pure state, and prove a weaker version in the general case. These will follow from the following single-shot decoupling bounds, which are the main results of this section:

Theorem 4.3. Let $A, B_1, \dots, B_r \in \mathcal{B}(\mathcal{H})_+$. Then

$$P_e^* \left(A, \sum\nolimits_j B_j \right) \le \sum_j F(A, B_j) \le \sum_j \sqrt{\operatorname{Tr}(A + B_j)} \sqrt{P_e^* (A, B_j)}. \tag{50}$$

If A is rank one then we also have

$$P_e^* \left(A, \sum\nolimits_j B_j \right) \le \begin{cases} \left(\sum_j \operatorname{Tr} B_j \right) \sum_j F\left(\frac{A}{\operatorname{Tr} A}, \frac{B_j}{\operatorname{Tr} B_j} \right)^2 \le \left(\sum_j \operatorname{Tr} B_j \right) \sum_j \frac{\operatorname{Tr} A + \operatorname{Tr} B_j}{(\operatorname{Tr} A)(\operatorname{Tr} B_j)} P_e^*(A, B_j) \\ \left(\sum_j \sqrt{1 + \frac{\operatorname{Tr} B_j}{\operatorname{Tr} A}} \sqrt{P_e^* (A, B_j)} \right)^2 \le \left(\sum_j \left(1 + \frac{\operatorname{Tr} B_j}{\operatorname{Tr} A} \right) \right) \sum_j P_e^* (A, B_j). \end{cases} \tag{51}$$

Before proving Theorem 4.3, we first explore some of its implications. We start with the following:

Corollary 4.4. For every $n \in \mathbb{N}$, let $A_n, B_{1,n}, \dots, B_{r,n} \in \mathcal{B}(\mathcal{H}_n)_+$, where $\mathcal{H}_n$ is some finite-dimensional Hilbert space. If $\limsup_n \operatorname{Tr}(A_n + \sum_j B_{j,n}) < +\infty$ then

$$\overline{p}_e \left(\vec{A}, \sum\nolimits_j \vec{B}_j \right) \le \frac{1}{2} \max_{1 \le j \le r} \overline{p}_e \left(\vec{A}, \vec{B}_j \right).$$

If A_n is rank one for every large enough n and $\limsup_n \operatorname{Tr}(A_n + \sum_j B_{j,n}) / \operatorname{Tr} A_n < +\infty$ then

$$\overline{p}_e \left(\vec{A}, \sum\nolimits_j \vec{B}_j \right) \le \max_{1 \le j \le r} \overline{p}_e \left(\vec{A}, \vec{B}_j \right).$$

Remark 4.5. Note that $P_e^* \left(A, \sum_j B_j \right) \ge P_e^* (A, B_j)$ for every j, and hence

$$\max_{1 \le j \le r} P_e^* (A, B_j) \le P_e^* \left(A, \sum\nolimits_j B_j \right).$$

In the asymptotic setting this yields

$$\max_{1 \le j \le r} \underline{p}_e \left(\vec{A}, \vec{B}_j \right) \le \underline{p}_e \left(\vec{A}, \sum\nolimits_j \vec{B}_j \right),$$

complementing the inequalities of Corollary 4.4.

Applying Corollary 4.4 to the problem of i.i.d. vs. averaged i.i.d. state discrimination, we finally get the following:

Theorem 4.6. In the i.i.d. vs. averaged i.i.d. case described at the beginning of the section,

$$
\begin{aligned}
-\min_i C(\rho,\sigma_i) &\le \liminf_{n\to\infty}\frac{1}{n}\log P_e^*\left(p\rho^{\otimes n},(1-p)\sum\nolimits_i q_i\sigma_i^{\otimes n}\right)\\
&\le \limsup_{n\to\infty}\frac{1}{n}\log P_e^*\left(p\rho^{\otimes n},(1-p)\sum\nolimits_i q_i\sigma_i^{\otimes n}\right)\le -\frac{1}{2}\min_i C(\rho,\sigma_i).
\end{aligned}
$$

If ρ is pure then we have

$$
\lim_{n\to\infty}\frac{1}{n}\log P_e^*\left(p\rho^{\otimes n},(1-p)\sum\nolimits_i q_i\sigma_i^{\otimes n}\right)=-\min_i C(\rho,\sigma_i).
$$

In realistic scenarios it is more natural to assume that the hypotheses are represented by sets of states with many elements (composite hypothesis) rather than one single state (simple hypothesis). Here we briefly consider the simplest such scenario, where we have two hypotheses, of which one is simple, represented by some PSD operator A, and the other one is composite, represented by a finite set of PSD operators $\{B_1,\dots,B_r\}$. For a given POVM $\{E, I-E\}$, the worst-case error probability is given by $\operatorname{Tr} A(I-E)+\max_{1\le i\le r}\operatorname{Tr} B_iE$, and we define

$$
P_e^*\left(A,\{B_i\}_{i=1}^r\right):=\inf\left\{\operatorname{Tr} A(I-E)+\max_{1\le i\le r}\operatorname{Tr} B_iE:\ 0\le E\le I\right\}.
$$

For every i and every E, we have

$$
\operatorname{Tr} A(I-E)+\operatorname{Tr} B_iE\le \operatorname{Tr} A(I-E)+\max_{1\le i\le r}\operatorname{Tr} B_iE\le \operatorname{Tr} A(I-E)+\operatorname{Tr}\sum_{i=1}^r B_iE,
$$

and taking the infimum in E yields

$$
\max_{1\le i\le r}P_e^*(A,B_i)\le P_e^*\left(A,\{B_i\}_{i=1}^r\right)\le P_e^*\left(A,\sum\nolimits_i B_i\right).
$$

Corollary 4.4 then immediately yields the following:

Corollary 4.7. For every $n\in\mathbb{N}$, let $A_n,B_{1,n},\dots,B_{r,n}\in\mathcal{B}(\mathcal{H}_n)_+$, where $\mathcal{H}_n$ is some finite-dimensional Hilbert space, and let

$$
\begin{aligned}
\underline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)&:=\liminf_{n\to+\infty}\frac{1}{n}\log P_e^*\left(A,\{B_i\}_{i=1}^r\right)\\
\overline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)&:=\limsup_{n\to+\infty}\frac{1}{n}\log P_e^*\left(A,\{B_i\}_{i=1}^r\right).
\end{aligned}
$$

If $\limsup_n \operatorname{Tr}(A_n+\sum_j B_{j,n})<+\infty$ then

$$
\max_{1\le i\le r}\underline{p}_e\left(\vec{A},\vec{B}_i\right)\le \underline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)\le \overline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)\le \frac{1}{2}\max_{1\le i\le r}\overline{p}_e\left(\vec{A},\vec{B}_i\right).
$$

If A_n is rank one for every large enough n and $\limsup_n \operatorname{Tr}(A_n+\sum_j B_{j,n})/\operatorname{Tr} A_n<+\infty$ then

$$
\max_{1\le i\le r}\underline{p}_e\left(\vec{A},\vec{B}_i\right)\le \underline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)\le \overline{p}_e\left(\vec{A},\{\vec{B}_i\}_{i=1}^r\right)\le \max_{1\le i\le r}\overline{p}_e\left(\vec{A},\vec{B}_i\right).
$$

Taking now $A_n := \rho^{\otimes n}$, $B_{i,n} := \sigma_i^{\otimes n}$, where $\rho, \sigma_1, \ldots, \sigma_r$ are density operators on some finite-dimensional Hilbert space, we get the following analogous statement to Theorem 4.6:

Theorem 4.8. Let $\rho, \sigma_1, \ldots, \sigma_r$ be density operators on some finite-dimensional Hilbert space. Then

$$
\begin{aligned}
-\min_i C(\rho, \sigma_i) &\le \liminf_{n\to\infty} \frac{1}{n} \log P_e^*\left(\rho^{\otimes n}, \{\sigma_i^{\otimes n}\}_{i=1}^r\right) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log P_e^*\left(\rho^{\otimes n}, \{\sigma_i^{\otimes n}\}_{i=1}^r\right) \le -\frac{1}{2} \min_i C(\rho, \sigma_i).
\end{aligned}
$$

If ρ is pure then we have

$$
\lim_{n\to\infty} \frac{1}{n} \log P_e^*\left(\rho^{\otimes n}, \{\sigma_i^{\otimes n}\}_{i=1}^r\right) = -\min_i C(\rho, \sigma_i).
$$

Now we turn to the proof of Theorem 4.3. For this we will need the following subadditivity property of the fidelity:

Lemma 4.9. Let $A, B_1, \ldots, B_r \in \mathcal{B}(\mathcal{H})_+$. Then

$$
F\left(A, \sum\nolimits_i B_i\right) \le \sum\nolimits_i F(A, B_i). \tag{52}
$$

Proof. The function $X \mapsto \operatorname{Tr}\sqrt{X}$ is subadditive on PSD operators, i.e., if $X, Y \in \mathcal{B}(\mathcal{H})_+$ then $\operatorname{Tr}\sqrt{X+Y} \le \operatorname{Tr}\sqrt{X} + \operatorname{Tr}\sqrt{Y}$. Indeed, assume first that $X, Y > 0$. Then

$$
\begin{aligned}
\operatorname{Tr}\sqrt{X+Y} - \operatorname{Tr}\sqrt{X} &= \int_0^1 \frac{d}{dt} \operatorname{Tr}\sqrt{X+tY}\, dt = \int_0^1 \operatorname{Tr} Y \frac{1}{2}(X+tY)^{-1/2}\, dt \\
&\le \operatorname{Tr}\sqrt{Y} \int_0^1 \frac{t^{-1/2}}{2}\, dt = \operatorname{Tr}\sqrt{Y},
\end{aligned}
$$

where we used the identity $\frac{d}{dt}\operatorname{Tr} f(X+tY) = \operatorname{Tr} Y f'(X+tY)$, and that the function $x \mapsto x^{-1/2}$ is operator monotone decreasing. The assertion for general PSD X and Y then follows by continuity. Thus,

$$
F\left(A, \sum\nolimits_i B_i\right) = \operatorname{Tr}\sqrt{\sum\nolimits_i A^{1/2} B_i A^{1/2}} \le \sum\nolimits_i \operatorname{Tr}\sqrt{A^{1/2} B_i A^{1/2}} = \sum\nolimits_i F(A, B_i). \qquad \square
$$

After this preparation, we are ready to prove Theorem 4.3.

Proof of Theorem 4.3:

$$
P_e^*\left(A, \sum\nolimits_j B_j\right) \le F\left(A, \sum\nolimits_j B_j\right) \le \sum_j F(A, B_j) \le \sum_j \sqrt{\operatorname{Tr}(A+B_j)}\sqrt{P_e^*(A, B_j)},
$$

where we used Lemma 2.4 in the first inequality, the second inequality is due to Lemma 4.9, and the third inequality is again due to Lemma 2.4. This proves (50).

Assume now that A is rank one. Then

$$\begin{aligned}
P_e^*\left(A,\sum\nolimits_j B_j\right) &\le \frac{1}{\operatorname{Tr} A} F\left(A,\sum\nolimits_j B_j\right)^2 \le \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j F(A,B_j)\right)^2 \\
&= \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j (\operatorname{Tr} A)^{\frac{1}{2}}(\operatorname{Tr} B_j)^{\frac{1}{2}} F\left(\frac{A}{\operatorname{Tr} A},\frac{B_j}{\operatorname{Tr} B_j}\right)\right)^2 \\
&\le \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j (\operatorname{Tr} A)(\operatorname{Tr} B_j)\right)\left(\sum\nolimits_j F\left(\frac{A}{\operatorname{Tr} A},\frac{B_j}{\operatorname{Tr} B_j}\right)^2\right) \\
&= \left(\sum\nolimits_j \operatorname{Tr} B_j\right)\sum\nolimits_j \frac{1}{(\operatorname{Tr} A)(\operatorname{Tr} B_j)} F(A,B_j)^2 \\
&\le \left(\sum\nolimits_j \operatorname{Tr} B_j\right)\sum\nolimits_j \frac{\operatorname{Tr} A+\operatorname{Tr} B_j}{(\operatorname{Tr} A)(\operatorname{Tr} B_j)} P_e^*(A,B_j),
\end{aligned}$$

where the first inequality is due to Lemma 2.5, the second inequality is due to Lemma 4.9, in the third inequality we used the Cauchy-Schwarz inequality, and the last inequality follows from Lemma 2.4. This proves the first bound in (51). Alternatively, we may proceed as

$$\begin{aligned}
P_e^*\left(A,\sum\nolimits_j B_j\right) &\le \frac{1}{\operatorname{Tr} A} F\left(A,\sum\nolimits_j B_j\right)^2 \le \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j F(A,B_j)\right)^2 \\
&\le \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j \sqrt{\operatorname{Tr}(A+B_j)}\sqrt{P_e^*(A,B_j)}\right)^2 \\
&\le \frac{1}{\operatorname{Tr} A}\left(\sum\nolimits_j \operatorname{Tr}(A+B_j)\right)\sum\nolimits_j P_e^*(A,B_j),
\end{aligned}$$

where the third inequality is due to Lemma 2.5, and in the last line we used the Cauchy-Schwarz inequality. This proves the second bound in (51). □

We close this section with some discussion of the above results.

Let $\rho,\sigma_1,\ldots,\sigma_r$ be states and $q_1,\ldots,q_r$ be a probability distribution. Then we have

$$\sum_i q_i F(\rho,\sigma_i) \le F\left(\rho,\sum\nolimits_i q_i\sigma_i\right) \le \sum_i \sqrt{q_i} F(\rho,\sigma_i),$$

where the first inequality is a special case of the joint concavity of the fidelity [37, Theorem 9.7], and the second inequality is due to Lemma 4.9 with the choice $A=\rho$ and $B_i = q_i\sigma_i$. Hence, Lemma 4.9 yields a complement to the concavity inequality $\sum_i q_i F(\rho,\sigma_i) \le F(\rho,\sum_i q_i\sigma_i)$. It is natural to ask whether the joint concavity inequality $\sum_i q_i F(\rho_i,\sigma_i) \le F(\sum_i q_i\rho_i,\sum_i q_i\sigma_i)$ can be complemented in the same way, but it is easy to see that the answer is no. Indeed, let $\rho_1=\sigma_2=|x\rangle\langle x|$ and $\rho_2=\sigma_1=|y\rangle\langle y|$ with x,y being orthogonal unit vectors in $\mathbb{C}^2$, and let $q_1=q_2=1/2$. Then $\sum_i q_i\rho_i = \frac{1}{2}I = \sum_i q_i\sigma_i$, and hence $F(\sum_i q_i\rho_i,\sum_i q_i\sigma_i)=1$, while $F(\rho_1,\sigma_1)=F(\rho_2,\sigma_2)=0$, and hence no inequality of the form $F(\sum_i q_i\rho_i,\sum_i q_i\sigma_i)\le c\sum_i F(\rho_i,\sigma_i)$ can hold with some $c>0$.

One can ask the same questions about the quantity $P_e^*(.,.)=\operatorname{Tr}\operatorname{GLB}(.,.)$, which has very similar properties to the fidelity. Indeed, convexity of the trace-norm yields joint concavity of this quantity, i.e., $\operatorname{Tr}\operatorname{GLB}(\sum_i q_i\rho_i,\sum_i q_i\sigma_i) = \frac{1}{2}\left(1-\|\sum_i q_i\rho_i - \sum_i q_i\sigma_i\|_1\right) \ge \sum_i q_i\frac{1}{2}\left(1-\|\rho_i-\sigma_i\|_1\right) = \sum_i q_i \operatorname{Tr}\operatorname{GLB}(\rho_i,\sigma_i)$, and the same example as above shows that this inequality cannot be complemented in general. On the other hand, one may hope that the weaker concavity inequality, where the first argument is a fixed ρ, can be complemented the same way as for the fidelity, i.e., that there exists a constant $c > 0$, depending at most on r, such that

$$\operatorname{Tr}\operatorname{GLB}\left(\rho, \sum\nolimits_i q_i\sigma_i\right) = \frac{1}{2}\left(1 - \left\|\rho - \sum\nolimits_i q_i\sigma_i\right\|_1\right) \le \frac{c}{2}\sum_i (1 - \|\rho - \sigma_i\|_1) = c\sum_i \operatorname{Tr}\operatorname{GLB}(\rho, \sigma_i).$$

More generally, one could ask whether an analogy of the subadditivity inequality (52) holds for $\operatorname{Tr}\operatorname{GLB}(.,.)$, i.e., if there exists a $c > 0$, depending at most on r, such that

$$P_e^*\left(A, \sum\nolimits_j B_j\right) = \operatorname{Tr}\operatorname{GLB}\left(A, \sum\nolimits_j B_j\right) \le c\sum_j \operatorname{Tr}\operatorname{GLB}(A, B_j) = c\sum_j P_e^*(A, B_j) \qquad (53)$$

holds for any PSD $A, B_1, \ldots, B_r$, where $c > 0$ depends only on r. This would give an improvement over Theorem 4.3, and prove Conjecture 4.1. Note that (53) is true when A is of rank one, according to Theorem 4.3, and also when all the operators are commuting, as we show in Appendix B. However, as it turns out, no such c exists in the general case.

Counterexamples are as follows: for $r = 2$, take $A = \varepsilon|\psi_1\rangle\langle\psi_1|$, $B_1 = |\psi_2\rangle\langle\psi_2|$ and $B_2 = |\psi_3\rangle\langle\psi_3|$ with ε small and ψ_2 and ψ_3 very close and almost orthogonal to ψ_1. For example, consider

$$\psi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \psi_2 = \begin{pmatrix} \sin\alpha \\ \cos\alpha \end{pmatrix}, \qquad \psi_3 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix},$$

with $\sin^2\alpha = \varepsilon/2$. Then $B_1 + B_2 = \operatorname{Diag}(\varepsilon, 2-\varepsilon)$ and $\operatorname{Tr}\operatorname{GLB}(A, B_1 + B_2) = \frac{1}{2}(\varepsilon + 2 - |2 - \varepsilon|) = \varepsilon$. However, one can check that $\operatorname{Tr}\operatorname{GLB}(A, B_1) = \operatorname{Tr}\operatorname{GLB}(A, B_2) \approx \varepsilon^2/2$ for very small ε. Thus, the LHS of (53) is linear in ε, whereas its RHS is quadratic, meaning that the RHS can be arbitrarily smaller than the LHS in the sense that the RHS/LHS ratio can be arbitrarily small.

One might get the impression that this failure is due to the fact that A_1 has very small trace. Thus one could try to amend inequality (53) by dividing the RHS by that trace (making both sides linear in ε):

$$\operatorname{Tr}\operatorname{GLB}\left(A, \sum\nolimits_j B_j\right) \le \frac{1}{\operatorname{Tr} A} \sum_j \operatorname{Tr}\operatorname{GLB}(A, B_j). \qquad (54)$$

This is a sensible amendment as it resonates with the appearance of the factor $\frac{1}{\operatorname{Tr} A}$ in (34) in our treatment of the pure state case, and furthermore, initial numerical simulations seemed to bolster the claim. However, this inequality is false too. We can use the direct sum trick based on Lemma A.9, and replace A by $A \oplus (1 - \operatorname{Tr} A)|x\rangle\langle x|$ and B_i by $B_i \oplus 0$ in the counterexample of the previous paragraph, where x is a unit vector in some auxiliary Hilbert space. This does not change the $\operatorname{Tr}\operatorname{GLB}$ terms but changes $\operatorname{Tr} A$ to 1, thereby eliminating its supposedly compensating effect. Thus inequality (54) is violated to arbitrarily high extent. Moreover, the same argument excludes the possibility to fix inequality (54) by replacing $1/\operatorname{Tr} A$ with $f(\operatorname{Tr} A)$ where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\lim_{x\to 0^+} f(x) = +\infty$.

The problems presented by the above example could be eliminated if we allowed the cross-term $\operatorname{Tr}\operatorname{GLB}(B_1, B_2)$ to appear (with some positive constant) on the RHS of (53), since the term $\operatorname{Tr} \operatorname{GLB}(B_1, B_2)$ is close to 1 and swamps the distinction between ε and ε^2. Although such a bound is too weak for proving Conjecture 4.1, it would be just the right tool to prove Conjecture 2.3, as we will see in the next section.

5 Dichotomic discrimination

Consider the generalized state discrimination problem with hypotheses $A_1, \ldots, A_r$. In this section we show an intermediate step towards proving Conjecture 2.3 in the form of a partial decoupling bound. Namely, we prove (in Theorem 5.2) that the multiple state discrimination error is bounded from above by the *dichotomic error*, which is the sum of the error probabilities of discriminating one A_i from the rest of the hypotheses. Using then the bounds obtained in Section 4, we get full decoupling bounds (Theorem 5.5).

Define the complementary operators as $\bar{A}_i := \sum_{j \neq i} A_j = A_0 - A_i$, where $A_0 := \sum_i A_i$. If we only want to decide whether the true hypothesis is A_i or not, i.e., we want to discriminate between A_i and $\bar{A}_i$, then the corresponding optimal error is given by

$$P^*_{e,dich,i} := P^*_e(A_i, \bar{A}_i) = \operatorname{Tr} \operatorname{GLB}(A_i, \bar{A}_i) = \frac{1}{2}(\operatorname{Tr} A_0 - \left\| A_i - \bar{A}_i \right\|_1). \tag{55}$$

We will call this a *dichotomic* discrimination, and $P^*_{e,dich,i}$ the i-th optimal *dichotomic error*. Let us define $P^*_{e,dich}$ as the sum of the r optimal dichotomic errors corresponding to each of the A_i:

$$P^*_{e,dich}(A_1, \ldots, A_r) := P^*_{e,dich} := \sum_{i=1}^{r} P^*_{e,dich,i} = \sum_{i=1}^{r} P^*_e(A_i, \bar{A}_i). \tag{56}$$

We show in Theorem 5.2 that the optimal multi-hypothesis discrimination error P^*_e is well-approximated by $P^*_{e,dich}$; more precisely,

$$\frac{1}{2} P^*_{e,dich} \leq P^*_e \leq P^*_{e,dich}. \tag{57}$$

In particular, these bounds, together with the fact that $P^*_{e,dich}$ is a number between 0 and $\operatorname{Tr} A_0$, show that there exists a POVM $\{E_k\}_{k=1}^r$ for which $P_e\left(\{E_k\}_{k=1}^r\right) = P^*_{e,dich}$. Therefore, we can rightly call $P^*_{e,dich}$ the *dichotomic error*. Moreover, these inequalities show that P^*_e and $P^*_{e,dich}$ have the same exponential behavior in the limit of many i.i.d. copies of the hypotheses.

We need some preparation to prove the bounds in (57). First, we give a number of useful expressions for $P^*_{e,dich}$. Since $A_i - \bar{A}_i = 2A_i - A_0$, we have $||A_i - \bar{A}_i||_1 = 2 \operatorname{Tr}(2A_i - A_0)_+ +$

$\operatorname{Tr} A_0 - 2 \operatorname{Tr} A_i$. Then

$$
\begin{aligned}
P^*_{e,dich} &= \frac{1}{2}\sum_{i=1}^{r}(\operatorname{Tr} A_0 - ||A_i - \bar{A}_i||_1) \\
&= \frac{1}{2}\sum_{i=1}^{r} 2\operatorname{Tr} A_i - 2\operatorname{Tr}(2A_i - A_0)_+ \\
&= \operatorname{Tr} A_0 - \sum_{i=1}^{r} \operatorname{Tr}(2A_i - A_0)_+ \\
&= \operatorname{Tr} A_0 - \sum_{i=1}^{r} \operatorname{Tr}(A_i - \bar{A}_i)_+ . \qquad (58)
\end{aligned}
$$

These expressions show that the quantity $P^*_{e,dich}$ is a number between 0 and $\operatorname{Tr} A_0$.

Next, we prove Lemma 5.1 below, which we will use for the proof of the upper bound in (57). Note that the map $f : X \mapsto X^*X$ is operator convex on $\mathcal{B}(\mathcal{H})$, as it was pointed out in [43, Lemma 5]. Indeed, for any $X_1, X_2 \in \mathcal{B}(\mathcal{H})$ and any $t \in [0, 1]$, we have

$$
tf(X_1) + (1-t)f(X_2) - f(tX_1 + (1-t)X_2) = t(1-t)(X_1 - X_2)^*(X_1 - X_2) \geq 0.
$$

In particular, for $X_1, \ldots, X_r \in \mathcal{B}(\mathcal{H})$, we have $\left(\sum_i X_i\right)^* \left(\sum_i X_i\right) \leq r \sum_i X_i^* X_i$, and operator monotony of the square root yields

$$
\left|\sum\nolimits_i X_i\right| \leq \sqrt{r}\left(\sum\nolimits_i |X_i|^2\right)^{1/2}. \qquad (59)
$$

Lemma 5.1. Let $\{P_i\}_{i=1}^r$ be a set of r projectors. Define $P_0 = \sum_{i=1}^r P_i$. Then

$$
0 \leq \sum_i (2P_i - P_0)_+ \leq \mathrm{I},
$$

i.e. the set of operators $\{(2P_i - P_0)_+\}_{i=1}^r$ forms an (incomplete) POVM.

Proof. Let $X_i := |2P_i - P_0|$, with P_i and P_0 as defined in the statement of the lemma. By (59),

$$
\sum\nolimits_i |2P_i - P_0| \leq \sqrt{r}\left(\sum\nolimits_i (2P_i - P_0)^2\right)^{1/2}.
$$

Considering the facts that the P_i are projectors, i.e. $P_i^2 = P_i$, and that P_0 is equal to their sum, the expression $\sum_i (2P_i - P_0)^2$ simplifies to

$$
\sum_i (2P_i - P_0)^2 = \sum_i (4P_i + P_0^2 - 2P_iP_0 - 2P_0P_i) = 4P_0 + rP_0^2 - 4P_0^2 = 4P_0 + (r-4)P_0^2.
$$

Now note the following:

$$
\begin{aligned}
r(4P_0 + (r-4)P_0^2) &\leq r(4P_0 + (r-4)P_0^2) + 4(I - P_0)^2 \\
&= 4rP_0 + (r^2 - 4r)P_0^2 + 4I - 8P_0 + 4P_0^2 \\
&= 4I + 4(r-2)P_0 + (r-2)^2 P_0^2 \\
&= (2I + (r-2)P_0)^2.
\end{aligned}
$$

Thus, we get

$$\sum_i |2P_i - P_0| \le 2I + (r-2)P_0.$$

To rewrite this in terms of the positive parts, we use the relation $|X| = 2X_+ - X$. This gives

$$\sum_i |2P_i - P_0| = 2\sum_i (2P_i - P_0)_+ - \sum_i (2P_i - P_0) = 2\sum_i (2P_i - P_0)_+ - (2-r)P_0.$$

Hence, we finally obtain

$$\sum_i (2P_i - P_0)_+ = \frac{1}{2}\left(\sum_i |2P_i - P_0| + (2-r)P_0\right) \le I,$$

as we set out to prove. □

Now we are ready to prove (57).

Theorem 5.2. For any $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$,

$$P^*_{e,2} \le \frac{1}{2} P^*_{e,dich} \le P^*_e \le P^*_{e,dich}. \tag{60}$$

Proof. The first inequality follows by a straightforward computation:

$$\begin{aligned} P^*_{e,2} &= \frac{1}{r-1} \sum_{(k,l):\, k<l} P^*_e(A_k, A_l) = \frac{1}{2(r-1)} \sum_{(k,l):\, k\neq l} P^*_e(A_k, A_l) \\ &= \frac{1}{2}\sum_k \frac{1}{r-1} \sum_{l:\, l\neq k} P^*_e(A_k, A_l) \le \frac{1}{2}\sum_k \frac{1}{r-1} \sum_{l: l\neq k} P^*_e(A_k, \bar{A}_k) = \frac{1}{2}\sum_k P^*_e(A_k, \bar{A}_k) = \frac{1}{2} P^*_{e,dich}. \end{aligned}$$

The inequality is due to the fact that $A_l \le \bar{A}_k$ for $l \neq k$, and hence $P^*_e(A_k, A_l) \le P^*_e(A_k, \bar{A}_k)$.

Next we prove the second inequality. Let $\{E_i\}_{i=1}^r$ be the optimal POVM for P^*_e. Clearly,

$$\operatorname{Tr}(2A_i - A_0)_+ \ge \operatorname{Tr}(2A_i - A_0)E_i = 2\operatorname{Tr} A_i E_i - \operatorname{Tr} A_0 E_i.$$

Summing over i yields

$$\sum_{i=1}^r \operatorname{Tr}(2A_i - A_0)_+ \ge 2\sum_{i=1}^r \operatorname{Tr} A_i E_i - \operatorname{Tr} A_0 \sum_{i=1}^r E_i \ge 2\sum_{i=1}^r \operatorname{Tr} A_i E_i - \operatorname{Tr} A_0 = 2P^*_s - \operatorname{Tr} A_0.$$

Hence, by (58),

$$P^*_{e,dich} = \operatorname{Tr} A_0 - \sum_{i=1}^r \operatorname{Tr}(2A_i - A_0)_+ \le \operatorname{Tr} A_0 - (2P^*_s - \operatorname{Tr} A_0) = 2P^*_e.$$

We will use Lemma 5.1 to prove the last inequality in (60). The trace of the positive part X_+ of a Hermitian operator X can be expressed as $\operatorname{Tr} XP$ with P the projector on the support of X_+. In particular, if P_i is the projector on the support of $(A_i - \bar{A}_i)_+$, we have

$$\sum_{i=1}^r \operatorname{Tr}(A_i - \bar{A}_i)_+ = \sum_i \operatorname{Tr}(A_i - \bar{A}_i)P_i.$$

Defining $P_0 := \sum_i P_i$, the summation on the right-hand side can be rewritten in the following way:

$$\sum_i \operatorname{Tr}(A_i - \bar{A}_i)P_i = \sum_i \operatorname{Tr}(2A_i - A_0)P_i = 2\sum_i \operatorname{Tr} A_iP_i - \operatorname{Tr} A_0P_0 = \sum_i \operatorname{Tr}(2P_i - P_0)A_i$$
$$\leq \sum_i \operatorname{Tr}(2P_i - P_0)_+A_i \leq \max_{\{E_i\}\text{ POVM}} \sum_i \operatorname{Tr} E_iA_i = P_s^*,$$

where the last inequality follows from the fact that the set of operators $\{(2P_i - P_0)_+\}_{i=1}^r$ forms an (incomplete) POVM by Lemma 5.1. Hence

$$P_{e,dich}^* = \operatorname{Tr} A_0 - \sum_i \operatorname{Tr}(A_i - \bar{A}_i)P_i \geq \operatorname{Tr} A_0 - P_s^* = P_e^*. \qquad \square$$

Remark 5.3. Validity of the last inequality in (60) in the classical case is a simple consequence of the fact that in a list of positive numbers only the largest one can be bigger than half their sum. Hence, for diagonal states

$$\sum_i (2A_i - A_0)_+ = \operatorname{LUB}(\{(2A_i - A_0)_+\}) = (2\operatorname{LUB}(\{A_i\}) - A_0)_+ \leq \frac{r\operatorname{LUB}(\{A_i\}) - A_0}{r-1}.$$

Taking the trace then yields

$$\operatorname{Tr} A_0 - P_{e,dich}^* = \sum_i \operatorname{Tr}(2A_i - A_0)_+ \leq \frac{r\operatorname{Tr}\operatorname{LUB}(\{A_i\}) - \operatorname{Tr} A_0}{r-1} = \frac{rP_s^* - \operatorname{Tr} A_0}{r-1} = \operatorname{Tr} A_0 - \frac{r}{r-1}P_e^*,$$

which is slightly stronger than what we needed to prove.

Note that the proof presented above for the first two inequalities in (60) gives an alternative proof of the inequality $P_{e,2}^* \leq P_e^*$ from Theorem 2.2. Moreover, we have obtained a strengthening of this inequality, by including $\frac{1}{2}P_{e,dich}^*$ in between $P_{e,2}^*$ and P_e^*.

Theorem 5.2 shows that the pairwise error does not exceed one half of the dichotomic error, and we conjecture that it can not be less than the dichotomic error up to another constant factor (depending only on the number of hypotheses). More precisely, we have the following:

Conjecture 5.4. There exists a constant c, at most depending on the number of hypotheses r, such that for all $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$,

$$P_{e,dich}^*(A_1, \ldots, A_r) \leq cP_{e,2}^*(A_1, \ldots, A_r).$$

Explicitly,

$$P_{e,dich}^* = \sum_{i=1}^r \operatorname{Tr}\operatorname{GLB}(A_i, \bar{A}_i) \leq c(r)\frac{1}{r-1}\sum_{(i,j):\, i\neq j} \operatorname{Tr}\operatorname{GLB}(A_i, A_j)$$
$$= \tilde{c}(r) \sum_{(i,j):\, i\neq j} P_e^*(A_i, A_j). \tag{61}$$

Numerical simulations suggest that $c(r) = 4(r-1)$ is best possible. Clearly, validity of this conjecture would prove validity of Conjecture 2.3. We can prove this conjecture for pure states, and for commuting states (see Appendix B), whereas for mixed states we are able to prove a weaker inequality:

Theorem 5.5. Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and $p_i := \operatorname{Tr} A_i$. Then

$$P_e^*(A_1, \ldots, A_r) \le P_{e,dich}^*(A_1, \ldots, A_r) \le \sum_{(i,j):\, i \ne j} F(A_i, A_j) \le \sum_{(i,j):\, i \ne j} \sqrt{p_i + p_j} \sqrt{P_e^*(A_i, A_j)}. \tag{62}$$

If A_i is rank one for all i then

$$\begin{aligned} P_e^*(A_1, \ldots, A_r) &\le P_{e,dich}^*(A_1, \ldots, A_r) \\ &\le \begin{cases} (\operatorname{Tr} A_0) \sum_{(i,j):\, i \ne j} \frac{1}{p_i p_j} F(A_i, A_j)^2 \le (\operatorname{Tr} A_0) \sum_{(i,j):\, i \ne j} \frac{p_i + p_j}{p_i p_j} P_e^*(A_i, A_j), \\ \frac{\operatorname{Tr} A_0}{\min_i \operatorname{Tr} A_i} \sum_{(i,j):\, i \ne j} P_e^*(A_i, A_j). \end{cases} \end{aligned} \tag{63}$$

Proof. The inequality $P_e^*(A_1, \ldots, A_r) \le P_{e,dich}^*(A_1, \ldots, A_r)$ is due to Theorem 5.2, and the rest is immediate from the definition of $P_{e,dich}^*$ and Theorem 4.3. □

Remark 5.6. Note that the bound $P_e^*(A_1, \ldots, A_r) \le \sum_{(i,j):\, i \ne j} F(A_i, A_j)$ in (62) is the same as in (43), but weaker than the bound in (36), due to the $1/2$ prefactor in the latter.

To compare the bounds in (63) to the other bounds obtained previously, we consider the most relevant case where $\operatorname{Tr} A_0 = p_1 + \ldots + p_r = 1$. Then (63) tells that $P_e^*(A_1, \ldots, A_r) \le \sum_{(i,j):\, i \ne j} \frac{1}{p_i p_j} F(A_i, A_j)^2$. Since $\frac{1}{p_i p_j} \le \frac{1}{2} \frac{p_i^2 + p_j^2}{p_i^2 p_j^2}$, this bound is better than the one in (37), and the two coincide if and only if $p_1 = \ldots = p_r$. On the other hand, $\frac{1}{p_i p_j} > \frac{1}{\sqrt{p_i p_j}}$ (we assume that all $p_i > 0$), and hence the fidelity bound in (63) is strictly worse than the one in (46).

To close this section, we formulate two further conjectures that would imply Conjecture 5.4. We have seen in the previous section that no bound of the form $\operatorname{Tr} \operatorname{GLB}(A_1, \overline{A_1}) \le c \sum_{l=2}^r \operatorname{Tr} \operatorname{GLB}(A_1, A_l)$ may hold in general, but amending the RHS with cross terms, i.e., error probabilities between A_k, A_l, $k, l \ne 1$ may yield a valid upper bound. Although such a bound would not have been useful for the purposes of Section 4, it would be sufficient for Conjecture 5.4, and numerical simulations suggest that it is indeed true. Hence, we have the following

Conjecture 5.7. There exist constants c_1 and c_2, at most depending on the number of hypotheses r, such that for all $A_i \ge 0$,

$$\operatorname{Tr} \operatorname{GLB}(A_1, \bar{A}_1) \le c_1 \sum_{l=2}^r \operatorname{Tr} \operatorname{GLB}(A_1, A_l) + c_2 \sum_{k,l=2:k \ne l}^r \operatorname{Tr} \operatorname{GLB}(A_k, A_l). \tag{64}$$

An equivalent conjecture in terms of POVM elements (using the primal SDP characterization of error probabilities) is:

Conjecture 5.8. There exist constants c_1 and c_2, at most depending on the number of hypotheses r, such that for any $0 \le F_i \le \mathrm{I}$ (for $2 \le i \le r$) and $0 \le G_{j,k} \le \mathrm{I}$ (for $2 \le j < k \le r$) there exists an E in the intersection of operator intervals

$$\left.\begin{array}{r} 0 \\ \mathrm{I} - c_1 \sum_{i=2}^r (\mathrm{I} - F_i) \end{array}\right\} \le E \le \left\{\begin{array}{l} \mathrm{I} \\ c_1 F_j + c_2 \left(\sum_{k=j+1}^r G_{j,k} + \sum_{k=2}^{j-1} (\mathrm{I} - G_{k,j})\right), \quad j = 2, \ldots, r. \end{array}\right. \tag{65}$$

Note, however, that operator intervals behave very differently than ordinary intervals of real numbers and are not very well understood. See for example the papers by Ando on this subject (e.g. [2]).

Proof of equivalence of Claims 5.7 and 5.8. The correspondence between the two claims is based on the following equivalent characterizations of the error probabilities:

$$\begin{aligned}
\mathrm{Tr}\,\mathrm{GLB}(A_1, \overline{A_1}) &= \min_E \mathrm{Tr}((\mathrm{I} - E)A_1 + E \sum_{j=2}^r A_j) \\
\mathrm{Tr}\,\mathrm{GLB}(A_1, A_i) &= \min_{F_i} \mathrm{Tr}((\mathrm{I} - F_i)A_1 + F_i A_i) \\
\mathrm{Tr}\,\mathrm{GLB}(A_j, A_k) &= \min_{G_{j,k}} \mathrm{Tr}(G_{j,k} A_j + (\mathrm{I} - G_{j,k})A_k),
\end{aligned}$$

where $E, F_i, G_{j,k}$ are POVM elements and satisfy $0 \le E, F_i, G_{j,k} \le \mathrm{I}$. Hence (64) holds if and only if

$$\begin{aligned}
0 \le\ & c_1 \sum_{i=2}^r \min_{F_i} \mathrm{Tr}((\mathrm{I} - F_i)A_1 + F_i A_i) \\
& + c_2 \sum_{j=2}^r \sum_{k=2:k\ne j}^r \min_{G_{j,k}} \mathrm{Tr}(G_{j,k} A_j + (\mathrm{I} - G_{j,k})A_k) \\
& - \min_E \mathrm{Tr}((\mathrm{I} - E)A_1 + E \sum_{j=2}^r A_j) \\
=\ & \min_{F_i} c_1 \sum_{i=2}^r \mathrm{Tr}((\mathrm{I} - F_i)A_1 + F_i A_i) \\
& + \min_{G_{j,k}} c_2 \sum_{j=2}^r \sum_{k=2:k\ne j}^r \mathrm{Tr}(G_{j,k} A_j + (\mathrm{I} - G_{j,k})A_k) \\
& + \max_E - \mathrm{Tr}((\mathrm{I} - E)A_1 + E \sum_{j=2}^r A_j) \\
=\ & \min_{F_i} \min_{G_{j,k}} \max_E \mathrm{Tr}\, A_1 \left(E - \mathrm{I} + c_1 \sum_{i=2}^r (\mathrm{I} - F_i)\right) \\
& + \sum_{j=2}^r \mathrm{Tr}\, A_j \left(-E + c_1 F_j + c_2 \left(\sum_{k=j+1}^r G_{j,k} + \sum_{k=2}^{j-1} (\mathrm{I} - G_{k,j})\right)\right)
\end{aligned}$$

holds for all $A_i \geq 0$. This quantification can be rephrased as the requirement that the minimization of the RHS over all $A_i \geq 0$ is non-negative. By von Neumann's minimax theorem, the order between this minimization and the minimization over E can be interchanged:

$$\begin{aligned} 0 \quad \leq \quad & \min_{F_i} \min_{G_{j,k}} \max_{E} \min_{A_1 \geq 0} \operatorname{Tr} A_1 \left(E - \mathrm{I} + c_1 \sum_{i=2}^{r} (\mathrm{I} - F_i) \right) \\ & + \sum_{j=2}^{r} \min_{A_j \geq 0} \operatorname{Tr} A_j \left(-E + c_1 F_j + c_2 \left(\sum_{k=j+1}^{r} G_{j,k} + \sum_{k=2}^{j-1} (\mathrm{I} - G_{k,j}) \right) \right). \end{aligned}$$

The minimizations over F_i and $G_{j,k}$ and the maximization over E can now be replaced by quantifications: for all POVM elements F_i and $G_{j,k}$ there should exist a POVM element E such that

$$\begin{aligned} 0 \quad \leq \quad & \min_{A_1 \geq 0} \operatorname{Tr} A_1 \left(E - \mathrm{I} + c_1 \sum_{i=2}^{r} (\mathrm{I} - F_i) \right) \\ & + \sum_{j=2}^{r} \min_{A_j \geq 0} \operatorname{Tr} A_j \left(-E + c_1 F_j + c_2 \left(\sum_{k=j+1}^{r} G_{j,k} + \sum_{k=2}^{j-1} (\mathrm{I} - G_{k,j}) \right) \right). \end{aligned}$$

Since $\operatorname{Tr} AB \geq 0$ for all $A \geq 0$ if and only if $B \geq 0$, this is so if and only if

$$E - \mathrm{I} + c_1 \sum_{i=2}^{r} (\mathrm{I} - F_i) \geq 0$$

and, for all $j \geq 2$,

$$-E + c_1 F_j + c_2 \left(\sum_{k=j+1}^{r} G_{j,k} + \sum_{k=2}^{j-1} (\mathrm{I} - G_{k,j}) \right) \geq 0.$$

Combining this with the requirement $0 \leq E \leq \mathrm{I}$ yields the inequalities of Claim 5.8. □

6 Nussbaum's mixed exponents approach

In [42] Nussbaum presented a different approach towards splitting up the multi-hypothesis testing problem into pairwise tests, in which one pair of hypotheses is treated in a preferential way. This leads to an upper bound on the total error probability in which different pairwise error probabilities appear with different exponents. Here we generalize his approach and by combining it with our results we improve his bounds on the total error probability.

First we need a lemma about POVM elements, the content of which is implicit in [42]:

Lemma 6.1. For any E and Q satisfying $0 \leq E, Q \leq \mathrm{I}$,

$$\frac{1}{2} Q^{1/2} E Q^{1/2} \leq \mathrm{I} - Q + E. \tag{66}$$

Proof. For any operator X, we have $0 \leq (X-2)^*(X-2)$, which can be rewritten as $X^*X/2 \leq (\mathrm{I}-X)^*(\mathrm{I}-X)+\mathrm{I}$. In particular, let E be positive definite and Q positive semidefinite and let $X = E^{1/2}Q^{1/2}E^{-1/2}$. Then we obtain

$$\frac{1}{2}E^{-1/2}Q^{1/2}EQ^{1/2}E^{-1/2} \leq E^{-1/2}(\mathrm{I}-Q^{1/2})E(\mathrm{I}-Q^{1/2})E^{-1/2}+\mathrm{I},$$

which yields, after multiplying with $E^{1/2}$ on the left and on the right,

$$\frac{1}{2}Q^{1/2}EQ^{1/2} \leq (\mathrm{I}-Q^{1/2})E(\mathrm{I}-Q^{1/2})+E.$$

By continuity, this inequality also holds for positive semidefinite E. If we now impose $E, Q \leq \mathrm{I}$ then the RHS can be bounded above by a simplified expression:

$$(\mathrm{I}-Q^{1/2})E(\mathrm{I}-Q^{1/2})+E \leq (\mathrm{I}-Q^{1/2})^2+E \leq \mathrm{I}-Q+E. \qquad \square$$

Nussbaum's result relies on the following decomposition lemma, proven by him for the case of uniform priors and for $K = 2$. We provide the lemma in full generality, and with a somewhat shorter proof, but still based on Nussbaum's main idea to decompose the POVM in a clever way into two parts.

Lemma 6.2. Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$. For all $1 \leq K \leq r$,

$$P_e^*(A_1, \ldots, A_r) \leq 2P_e^*(A_1, \ldots, A_K) + P_e^*(3A^{(K)}, A_{K+1}, \ldots, A_r), \tag{67}$$

where $A^{(K)} := \sum_{i=1}^K A_i$.

Proof. Let $\mathcal{F} = \{F_1, \ldots, F_K\}$ be the optimal POVM for discriminating between $A_1, \ldots, A_K$, and let $\mathcal{E}^{(K)} = \{Q, E_{K+1}, \ldots, E_r\}$ be the optimal POVM for discriminating between $3A^{(K)}$, $A_{K+1}, \ldots, A_r$. Define $E_i := Q^{1/2}F_iQ^{1/2}$ for $i = 1, \ldots, K$. Then $\mathcal{E} = \{E_1, \ldots, E_r\}$ is a POVM.

In terms of the POVM $\mathcal{E}$ we have, for $i = 1, \ldots, K$,

$$\mathrm{Tr}\, A_iE_i = \mathrm{Tr}\, A_i(Q-(Q-E_i)) = \mathrm{Tr}\, A_iQ - \mathrm{Tr}\, A_iQ^{1/2}(\mathrm{I}-F_i)Q^{1/2}.$$

The total error probability for this POVM (an upper bound on P_e^*) is given by

$$\begin{aligned} P_e(\mathcal{E}) &= \sum_{i=1}^K \mathrm{Tr}\, A_i(\mathrm{I}-E_i) + \sum_{i=K+1}^r \mathrm{Tr}\, A_i(\mathrm{I}-E_i) \\ &= \mathrm{Tr}\, A^{(K)}(\mathrm{I}-Q) + \sum_{i=1}^K \mathrm{Tr}\, A_iQ^{1/2}(\mathrm{I}-F_i)Q^{1/2} + \sum_{i=K+1}^r \mathrm{Tr}\, A_i(\mathrm{I}-E_i). \end{aligned}$$

Using (66) of Lemma 6.1, the second sum can be bounded above by

$$2\left(\mathrm{Tr}\sum_{i=1}^K A_i(\mathrm{I}-F_i) + \mathrm{Tr}\, A^{(K)}(\mathrm{I}-Q)\right) = 2P_e^*(A_1, \ldots, A_K) + 2\,\mathrm{Tr}\, A^{(K)}(\mathrm{I}-Q).$$

Then

$$P_e^* \le P_e(\mathcal{E}) \le 2P_e^*(A_1,\dots,A_K) + 3\operatorname{Tr} A^{(K)}(\mathrm{I}-Q) + \sum_{i=K+1}^r \operatorname{Tr} A_i(\mathrm{I}-E_i)$$
$$= 2P_e^*(A_1,\dots,A_K) + P_e^*\left(3A^{(K)}, A_{K+1},\dots,A_r\right),$$

proving (67). □

The above lemma yields immediately the following:

Theorem 6.3. Let $A_1,\dots,A_r \in \mathcal{B}(\mathcal{H})_+$. Then

$$P_e^*(A_1,\dots,A_r) \le 2^{r-2}P_e^*(A_1,A_2) + 3\sum_{k=2}^{r-1} 2^{r-1-k} P_e^*\left(\sum_{i=1}^k A_i, A_{k+1}\right). \tag{68}$$

Proof. Applying Lemma 6.2 recursively, we get

$$\begin{aligned} P_e^*(A_1,\dots,A_r) &\le 2P_e^*(A_1,\dots,A_{r-1}) + P_e^*(3A^{(r-1)}, A_r) \\ &\le 4P_e^*(A_1,\dots,A_{r-2}) + 2P_e^*(3A^{(r-2)}, A_{r-1}) + P_e^*(3A^{(r-1)}, A_r) \\ &\le \dots \\ &\le 2^{r-2}P_e^*(A_1,A_2) + \sum_{k=2}^{r-1} 2^{r-1-k}P_e^*(3A^{(k)}, A_{k+1}). \end{aligned}$$

Note that $P_e^*(3A^{(k)}, A_{k+1}) \le 3P_e^*(A^{(k)}, A_{k+1})$, and thus we obtain (68). □

Remark 6.4. Note that the upper bound in (68) is similar to the bound $P_e^* \le P_{e,dich}^*$ in Theorem 5.2, but the two are not directly comparable regarding their tightness.

Combining now Theorem 4.3 with the above theorem, we finally get the following decoupling bound in terms of the optimal pairwise error probabilities:

Theorem 6.5. Let $A_1,\dots,A_r \in \mathcal{B}(\mathcal{H})_+$ and $\kappa := 3\max_{1\le i<j\le r}\sqrt{\operatorname{Tr} A_i + \operatorname{Tr} A_j}$. Then

$$P_e^*(A_1,\dots,A_r) \le 2^{r-2}P_e^*(A_1,A_2) + \kappa\sum_{k=2}^{r-1} 2^{r-1-k} \sum_{l=1}^k \sqrt{P_e^*(A_l, A_{k+1})}. \tag{69}$$

If A_k is of rank one for $k = 3,\dots,r$, then

$$P_e^*(A_1,\dots,A_r) \le 2^{r-2}P_e^*(A_1,A_2) + \kappa'\sum_{k=2}^{r-1} 2^{r-1-k} \sum_{l=1}^k P_e^*(A_l, A_{k+1}), \tag{70}$$

where $\kappa' := 3\operatorname{Tr} A_0 / \left(\min_{3\le i\le r}\operatorname{Tr} A_i\right)$.

Proof. Applying Theorem 4.3 to each term in the summand in (68) yields the inequalities of the theorem. □

Remark 6.6. The constants in (69) and (70) are in general worse than the ones in Theorems 3.5 and 5.5. On the other hand, (69) outperforms all the previous bounds in the sense that for one pair of states, it contains the optimal binary error probability instead of its square root. We will explore the consequences of this in the next section.

7 Asymptotics: the Chernoff bound

The various single-shot decoupling bounds, that we obtained in the previous sections for the multiple state discrimination problem, can be summarized as follows:

Lemma 7.1. For every $r \in \mathbb{N}$, there exist $\kappa_r, \kappa_r' > 0$ such that for all $A_1, \dots, A_r \in \mathcal{B}(\mathcal{H})_+$,

$$P_e^*(A_1, \dots, A_r) \le \kappa_r (\operatorname{Tr} A_0)^{1/2} \sum_{(i,j):\, i \ne j} P_e^*(A_i, A_j)^{1/2}. \tag{71}$$

If all but at most two of the A_i are of rank 1 then we also have

$$P_e^*(A_1, \dots, A_r) \le \kappa_r' \frac{\operatorname{Tr} A_0}{\min_i \operatorname{Tr} A_i} \sum_{(i,j):\, i \ne j} P_e^*(A_i, A_j). \tag{72}$$

Proof. The bound in (71) can be obtained from either of the following: the bound (36) of [8] using the Fuchs–van de Graaf inequalities; from the bound (44) of Theorem 3.5; from the bound (62) of Theorem 5.5; and from (69) of Theorem 6.5.

The bound (72) follows from (70) of Theorem 6.5. (We can assume without loss of generality that at most hypotheses 1 and 2 are not represented by rank one operators.) However, when all the A_i are pure, (72) also follows from any of the following: from the bound (37) of [19] using the Fuchs–van de Graaf inequalities; from the bound (46) of Theorem 3.5; and from the bound (63) of Theorem 5.5. □

Armed with these upper bounds, we now turn to the study of its asymptotic behavior. Let our hypotheses be represented by the sequences $\vec{A}_i := \{A_{i,n}\}_{n \in \mathbb{N}}$, $i = 1, \dots, r$, and define $A_{0,n} := \sum_{i=1}^r A_{i,n}$. Recall the definitions of $\underline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right)$ and $\overline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right)$ from (2)–(3). Due to Theorem 2.2, we have

$$\underline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right) \ge \max_{(i,j):\, i \ne j} \underline{p}_e\left(\vec{A}_i, \vec{A}_j\right). \tag{73}$$

Our aim here is to complement the above inequality by giving upper bounds on $\overline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right)$ in terms of the pairwise exponents. Recall the definition of the asymptotic Chernoff divergence from (28),

$$C(\vec{A}_1, \vec{A}_2) = \liminf_{n \to \infty} \frac{1}{n} C(A_{1,n}, A_{2,n}) = -\limsup_{n \to \infty} \frac{1}{n} \log \min_{0 \le s \le 1} \operatorname{Tr} A_{1,n}^s A_{2,n}^{1-s}.$$

We conjecture that the following converse to (73) holds under very mild conditions:

$$\overline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right) \le \max_{(i,j):\, i \ne j} \overline{p}_e\left(\vec{A}_i, \vec{A}_j\right) \le -\min_{(i,j):\, i \ne j} C(\vec{A}_1, \vec{A}_2). \tag{74}$$

Note that the second inequality is always true, due to (27). Below we show that the weaker inequality

$$\overline{p}_e\left(\vec{A}_1, \dots, \vec{A}_r\right) \le \frac{1}{2} \max_{(i,j):\, i \ne j} \overline{p}_e\left(\vec{A}_i, \vec{A}_j\right) \le -\frac{1}{2} \min_{(i,j):\, i \ne j} C(\vec{A}_1, \vec{A}_2) \tag{75}$$

is always true as long as $\limsup_{n\to\infty}\frac{1}{n}\log\operatorname{Tr}A_{0,n}=0$, which is trivially satisfied in the case of weighted states. We also show (74) in a number of special cases.

We have the following general result:

Theorem 7.2. Assume that $\limsup_{n\to\infty}\frac{1}{n}\log\operatorname{Tr}A_{0,n}=0$. Then

$$\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right)\le\frac{1}{2}\max_{(i,j):\,i\neq j}\overline{p}_e\left(\vec{A}_i,\vec{A}_j\right)\le-\frac{1}{2}\min_{(i,j):\,i\neq j}C(\vec{A}_1,\vec{A}_2). \tag{76}$$

Assume, moreover, that $A_{i,n}$ is of rank one for every $n\in\mathbb{N}$ for at least $r-2$ of the hypotheses. If $\limsup_{n\to\infty}\frac{1}{n}\log\frac{\operatorname{Tr}A_{0,n}}{\min_i\operatorname{Tr}A_{i,n}}=0$, then we have the stronger inequality

$$\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right)\le\max_{(i,j):\,i\neq j}\overline{p}_e\left(\vec{A}_i,\vec{A}_j\right)\le-\min_{(i,j):\,i\neq j}C(\vec{A}_i,\vec{A}_j). \tag{77}$$

Proof. Immediate from Lemma 7.1, Lemma 2.1, and (27). □

We can also prove (74) in the following special cases, by using Theorem 6.5.

Theorem 7.3. Assume that (74) holds for hypotheses $\vec{A}_i$, $i=1,\ldots,r$, and that $A_{r+1,n}$ is rank one for every n. If $\limsup_{n\to\infty}\frac{1}{n}\log\frac{\operatorname{Tr}\sum_{i=1}^r A_i}{\operatorname{Tr}A_{r+1,n}}=0$ then

$$\begin{aligned}\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r,\vec{A}_{r+1}\right)&\le\max\left\{\overline{p}_e\left(\vec{A}_i,\vec{A}_j\right):\,1\le i<j\le r+1\right\}\\&\le-\min\left\{C(\vec{A}_i,\vec{A}_j):\,1\le i<j\le r+1\right\}.\end{aligned}$$

Proof. By (67),

$$P_e^*(A_{1,n},\ldots,A_{r,n})\le 2P_e^*(A_1,\ldots,A_r)+P_e^*(3A^{(r)},A_{r+1}).$$

Applying then (51) to the second term yields the assertion. □

Remark 7.4. Note that the binary case (27), combined with a recursive application of Theorem 7.3, gives an alternative proof of the second part of Theorem 7.2.

Inequality (74) has been proved in [42] for the i.i.d. case under the assumption that there exists a pair of states σ_k,σ_l, $k\neq l$, such that $C(\sigma_k,\sigma_l)\le\frac{1}{6}C(\sigma_i,\sigma_j)$ for every $(i,j)\neq(k,l)$, $i\neq j$. Theorem 7.5 below shows that the constant 1/6 can be improved to 1/2.

Theorem 7.5. Assume that $\limsup_{n\to\infty}\frac{1}{n}\log\operatorname{Tr}A_{0,n}=0$. For any pair (k,l), $k\neq l$,

$$\begin{aligned}\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right)&\le\max\left\{\overline{p}_e(\vec{A}_k,\vec{A}_l),\,\frac{1}{2}\overline{p}_e(\vec{A}_i,\vec{A}_j),\,i\neq j,\,(i,j)\neq(k,l)\right\}\\&\le-\min\left\{C(\vec{A}_k,\vec{A}_l),\,\frac{1}{2}C(\vec{A}_i,\vec{A}_j),\,i\neq j,\,(i,j)\neq(k,l)\right\}.\end{aligned}$$

In particular, if there exists a pair (k,l), $k\neq l$, such that $\overline{p}_e(\vec{A}_k,\vec{A}_l)\ge\frac{1}{2}\overline{p}_e(\vec{A}_i,\vec{A}_j)$ or $C(\vec{A}_k,\vec{A}_l)\le\frac{1}{2}C(\vec{A}_i,\vec{A}_j)$, $i\neq j$, $(i,j)\neq(k,l)$, then

$$\overline{p}_e\left(\vec{A}_1,\ldots,\vec{A}_r\right)\le\max_{(i,j):\,i\neq j}\overline{p}_e\left(\vec{A}_i,\vec{A}_j\right)\le-\min_{(i,j):\,i\neq j}C(\vec{A}_i,\vec{A}_j).$$

Proof. Immediate from Theorem 6.5. □

Finally, we note that in many important cases, we have the optimality relation

$$\underline{p}_e\left(\vec{A}_i, \vec{A}_j\right) \geq -C\left(\vec{A}_i, \vec{A}_j\right). \tag{78}$$

For instance, this happens in the standard state discrimination problem if the hypotheses i, j are i.i.d. [38], or Gibbs states of a finite-range, translation-invariant Hamiltonian on a spin chain [22], or Gibbs states of interaction-free fermionic or bosonic chains [31, 32]. In these cases, if $\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{Tr} A_{0,n} = 0$ then we have

$$-\min_{(i,j):\, i\neq j} C\left(\vec{A}_i, \vec{A}_j\right) \leq \underline{p}_e\left(\vec{A}_1, \ldots, \vec{A}_r\right) \leq \overline{p}_e\left(\vec{A}_1, \ldots, \vec{A}_r\right) \leq -\frac{1}{2} \min_{(i,j):\, i\neq j} C\left(\vec{A}_i, \vec{A}_j\right).$$

If, moreover, (74) is satisfied then we get the stronger statement

$$\lim_{n\to+\infty} \frac{1}{n} \log P_e^*\left(A_{1,n}, \ldots, A_{r,n}\right) = -\min_{(i,j):\, i\neq j} C\left(\vec{A}_i, \vec{A}_j\right).$$

Appendix

A Least upper bound and greatest lower bound for operators

As mentioned already in Section 2.3, for a set $A_1, \ldots, A_r$ of self-adjoint operators on the same Hilbert space, the set of upper bounds $\mathcal{A} := \{Y : Y \geq A_k,\, k = 1, \ldots, r\}$ has no minimal element in general. The following example shows that a minimal element may not exist even if all the A_k commute with each other.

Example A.1. Let $\mathcal{H} = \mathbb{C}^2$, and let the operators $A_1, A_2, Y_{\alpha,\beta,\delta}$ be given by their matrices in the standard basis of $\mathbb{C}^2$ as

$$A_1 := \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad A_2 := \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad Y_{\alpha,\beta,\delta} := \begin{bmatrix} 2+\alpha & \delta \\ \overline{\delta} & 2+\beta \end{bmatrix}.$$

Let $\mathcal{I} := \{(\alpha, \beta, \delta) \in \mathbb{R}^3 : \alpha, \beta \geq 0,\, \min\{\alpha, \beta\} + \alpha\beta \geq |\delta|^2\}$. It is easy to see that $\mathcal{U}(A_1, A_2) = \{Y : Y \geq A_1, A_2\} = \{Y_{\alpha,\beta,\delta} : (\alpha, \beta, \delta) \in \mathcal{I}\}$. Assume that $\mathcal{U}(A_1, A_2)$ has a minimal element Y. The assumption $Y \geq A_1, A_2$ yields that $Y_{11} \geq 2$ and $Y_{22} \geq 2$, while the assumption that $Y \leq Y_{\alpha,\beta,\delta}$ for all $(\alpha, \beta, \delta) \in \mathcal{I}$ yields that $Y_{11} \leq 2$ and $Y_{22} \leq 2$. Hence, $0 \leq Y - A_1 = \begin{bmatrix} 1 & Y_{12} \\ \overline{Y_{12}} & 0 \end{bmatrix}$, which yields $Y_{12} = 0$, i.e., $Y = 2I$. Now, $Y_{\alpha,\beta,\delta} - Y \geq 0$ if and only if $\alpha, \beta \geq 0$ and $\alpha\beta \geq |\delta|^2$, which defines a strictly smaller set than $\mathcal{I}$, contradicting our initial assumption that Y is a lower bound to $\mathcal{U}(A_1, A_2)$.

In general, the set $\mathcal{A} := \{Y : Y \geq A_k,\, k = 1, \ldots, r\}$ is the intersection of r cones, and the intersection of two cones is not itself a cone, unless one is completely contained in the other. Thus, $\mathcal{A}$ has no unique minimal element in general, in the sense that there would be an element Y_0 such that $Y_0 \leq Y$ for all $Y \in \mathcal{A}$. Rather, there is an infinity of minimal elements, in the sense that there is an infinity of operators $Y \in \mathcal{A}$ for which no other $Y' \in \mathcal{A}$ exists such that $Y' \leq Y$, and these minima constitute the boundary of $\mathcal{A}$ [2]. The upshot is that one can not define a least upper bound on the basis of the PSD ordering alone.

However, there is a unique minimal element within $\mathcal{A}$ in terms of the *trace ordering*. We can therefore define a least upper bound in this more restrictive sense as

$$\mathrm{LUB}(A_1, \ldots, A_r) := \arg\min_Y \{\operatorname{Tr} Y : Y \geq A_k,\ k = 1, \ldots, r\}. \tag{79}$$

To make sense of the definition, we have to prove the uniqueness of the minimizer. For the proof, we will need the following simple fact, which has been stated, e.g., in [2] without a proof. Here we provide a proof for readers' convenience.

Lemma A.2. Let $D, T \in \mathcal{B}(\mathcal{H})$ be self-adjoint operators such that $D \geq \pm T$. Then D is positive semidefinite, and its support dominates the support of T.

Proof. First, $D \geq \pm T$ implies $D \geq (T + (-T))/2 = 0$, proving that D is PSD. Let $\mathcal{H}_1$ denote the support of D, and decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$. Then D and T can be written in the corresponding block forms as $D = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{12}^* & T22 \end{bmatrix}$, and positive semidefiniteness of $D \pm T$ implies $0 \geq T_{22} \geq 0$. Using again that $D + T \geq 0$, we finally obtain that $T_{12} = 0$, too, from which the assertion follows. □

Theorem A.3. Let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_{\mathrm{sa}}$ be a finite number of self-adjoint operators. Then in the set $\mathcal{A} := \{Y : Y \geq A_1, \ldots, A_r\}$ there is a unique element with minimal trace.

Proof. Let us assume that there are two distinct elements Y_1 and Y_2 in $\mathcal{A}$ with minimal trace $\operatorname{Tr} Y_1 = \operatorname{Tr} Y_2$. Let $Y_m = (Y_1 + Y_2)/2$ and $\Delta = (Y_1 - Y_2)/2$. Then $Y_1 = Y_m + \Delta$ and $Y_2 = Y_m - \Delta$, and $Y_1, Y_2 \geq A_i$ implies $Y_m - A_i \geq \pm\Delta$. Hence, by Lemma A.2, there exists a constant $c_i > 0$ such that $Y_m - A_i \geq c_i|\Delta|$ for every $i = 1, \ldots, r$. Taking $c := \min_i c_i$, we have $Y_m - c|\Delta| \geq A_i$, $i = 1, \ldots, r$. Thus, $Y_m - c|\Delta| \in \mathcal{A}$, but $\operatorname{Tr}(Y_m - c|\Delta|) = \operatorname{Tr} Y_m - c \operatorname{Tr}|\Delta| < \operatorname{Tr} Y_m = \operatorname{Tr} Y_i$, $i = 1, 2$, contradicting our original assumption. □

Next, we explore some properties of the LUB. It is easy to see from (79) that the LUB satisfies the *translation property*:

$$\mathrm{LUB}(A_1 + B, \ldots, A_r + B) = \mathrm{LUB}(A_1, \ldots, A_r) + B. \tag{80}$$

This is because the addition $X \mapsto X + B$, with a fixed self-adjoint operator B, is an order-preserving operation. Furthermore, the LUB is jointly homogeneous: for any $c \geq 0$,

$$\mathrm{LUB}(cA_1, \ldots, cA_r) = c\,\mathrm{LUB}(A_1, \ldots, A_r). \tag{81}$$

The positive part and modulus can be expressed in terms of the LUB.

Lemma A.4. For all Hermitian operators A,

$$A_+ = \mathrm{LUB}(A, 0), \qquad \text{and} \qquad |A| = \mathrm{LUB}(A, -A). \tag{82}$$

Proof. Consider the set $\mathcal{A} = \{Y : Y \geq A, Y \geq 0\}$. Clearly, $A_+ \in \mathcal{A}$. By Weyl's monotonicity principle, the eigenvalues of any $Y \in \mathcal{A}$ are non-negative and not smaller than those of A; that is, $\lambda_j(Y) \geq \lambda_j(A)$, where λ_j denotes the j^{th} largest eigenvalue. Hence, $\lambda_j(Y) \geq \lambda_j(A_+)$, since the spectrum of A_+ consists of the positive eigenvalues of A and zero. As the sum of all eigenvalues is the trace, A_+ is an element (and therefore *the* element) in $\mathcal{A}$ with minimal trace.

Using (80) and (81), the modulus $|A| = 2A_+ - A$ can be similarly expressed as $|A| = 2\,\mathrm{LUB}(A, 0) - A = \mathrm{LUB}(A, -A)$. □

Remark A.5. We emphasize again that the LUB is a minimum with respect to the trace ordering and not the PSD ordering. In particular, $X \geq A$ and $X \geq -A$ doesn't imply $X \geq |A|$. A counterexample can be easily given by taking $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $X = \begin{bmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 2 \end{bmatrix}$. However, as lemma A.2 shows, there always exists a positive constant c, depending on A and X, such that $X \geq c|A|$.

By lemma A.4 and (80), $(A - B)_+ = \mathrm{LUB}(A, B) - B$. This immediately leads to a closed form expression for the LUB of two Hermitian operators:

Lemma A.6. For all Hermitian operators A, B,

$$\mathrm{LUB}(A, B) = B + (A - B)_+ = \frac{1}{2}(A + B + |A - B|) = A + (A - B)_-. \tag{83}$$

From these expressions it is clear that for $A, B \geq 0$, the LUB is PSD as well.

In a similar vein we can define the *greatest lower bound* (GLB) as

$$\mathrm{GLB}(A_1, \ldots, A_r) := \arg\max_Y \{\mathrm{Tr}\, Y : Y \leq A_k,\ k = 1, \ldots, r\}. \tag{84}$$

Clearly, we have

$$\mathrm{GLB}(A_1, \ldots, A_r) = -\,\mathrm{LUB}(-A_1, \ldots, -A_r). \tag{85}$$

Hence, for two operators, we get

Lemma A.7. For all Hermitian operators A, B,

$$\mathrm{GLB}(A, B) = \frac{1}{2}(A + B - |A - B|) = A - (A - B)_+ = B - (A - B)_-. \tag{86}$$

A warning is in order about the sign of the GLB. When A and B commute, their GLB is given by the entrywise minimum in the joint eigenbasis. If A and B are also PSD, then clearly their GLB will be PSD. When A and B are PSD but do not commute, however, their GLB need not be PSD; only the trace of their GLB will be guaranteed to be non-negative. The reason is that while the function $x \mapsto x_+ = \max(0, x)$ is monotone increasing, it is also convex and therefore not operator monotone. Thus, for $A, B \geq 0$, $(A - B)_+ \leq A$ need not be true. For a concrete counterexample, take $A = |x\rangle\langle x|$, $B = |y\rangle\langle y|$ with $x = (1, 1)$, $y = (1, i)$; then it is easy to check that $0 \not\leq \mathrm{LUB}(A, B)$. Similarly, the LUB of two negative semidefinite operators need not be negative semidefinite.

Both LUB and GLB are monotonous in their arguments with respect to the PSD ordering.

Lemma A.8. For all Hermitian operators $\{A_i\}$ and $\{B_i\}$, if $A_i \le B_i$ then

$$\begin{aligned} \operatorname{Tr}\operatorname{LUB}(A_1,\ldots,A_r) &\le \operatorname{Tr}\operatorname{LUB}(B_1,\ldots,B_r), && (87)\\ \operatorname{Tr}\operatorname{GLB}(A_1,\ldots,A_r) &\le \operatorname{Tr}\operatorname{GLB}(B_1,\ldots,B_r). && (88)\end{aligned}$$

Proof. By definition, $\operatorname{LUB}(B_1,\ldots,B_r) \ge B_i \ge A_i$, for all i, so that $\operatorname{LUB}(B_1,\ldots,B_r)$ is an upper bound on all A_i. In general it is not the minimal one, hence $\operatorname{Tr}\operatorname{LUB}(B_1,\ldots,B_r) \ge \operatorname{Tr}\operatorname{LUB}(A_1,\ldots,A_r)$. Monotonicity for the GLB follows from this and the correspondence (85). □

The LUB and GLB (and their trace) behave in the expected way with respect to the direct sum:

Lemma A.9. For any pair of sets of $A_i \in \mathcal{B}(\mathcal{H}_1)_{\mathrm{sa}}$ and $B_i \in \mathcal{B}(\mathcal{H}_2)_{\mathrm{sa}}$, $i = 1,\ldots,r$,

$$\begin{aligned} \operatorname{LUB}(\{A_i \oplus B_i\}) &= \operatorname{LUB}(\{A_i\}) \oplus \operatorname{LUB}(\{B_i\}) && (89)\\ \operatorname{GLB}(\{A_i \oplus B_i\}) &= \operatorname{GLB}(\{A_i\}) \oplus \operatorname{GLB}(\{B_i\}). && (90)\end{aligned}$$

Proof. Consider first the LUB. Let $X := \operatorname{LUB}(\{A_i \oplus B_i\})$, and let P_i denote the projection onto $\mathcal{H}_i$ in the direct sum $\mathcal{H}_1 \oplus \mathcal{H}_2$. Then $P_1XP_1 \oplus P_2XP_2 \ge A_i \oplus B_i$ for all i, and $\operatorname{Tr} X = \operatorname{Tr} P_1XP_1 \oplus P_2XP_2$. The uniqueness of the LUB then yields $X = P_1XP_1 \oplus P_2XP_2$.

The proof for the GLB goes exactly the same way. □

This lemma has an important consequence. For every set of subnormalized states $\{A_i\}_{i=1}^r$ there is a set of normalized states $\{\sigma_i\}_{i=1}^r$ such that $\operatorname{Tr}\operatorname{GLB}(\{A_i\}) = \operatorname{Tr}\operatorname{GLB}(\{\sigma_i\})$; namely $\sigma_i = A_i \oplus (1 - \operatorname{Tr} A_i)|i\rangle\langle i|$, where $\{|i\rangle\}_{i=1}^r$ is an orthonormal system. This is because the 'appended' states $B_i = (1 - \operatorname{Tr} A_i)|i\rangle\langle i|$ are mutually orthogonal so that $\operatorname{Tr}\operatorname{GLB}(\{B_i\}) = 0$. Similar statements can be made when the arguments of Tr GLB are linear combinations of states. The upshot of this is two-fold. First, for a large class of statements it allows one to restrict to normalized states to prove them. Secondly, it aids the heuristic processes of coming up with reasonable conjectures and finding counterexamples (see, e.g., at the end of Section 4).

Finally, we give another representation of the least upper bound as the max-relative entropy center in the case where all the operators are positive semidefinite. For PSD operators $A, B \in \mathcal{B}(\mathcal{H})_+$, their max-relative entropy $D_{\max}(A \,\|\, B)$ is defined as [13, 47]

$$D_{\max}(A \,\|\, B) := \inf\{\gamma :\ A \le 2^\gamma B\}.$$

For a set of states $\mathcal{A} \subset \mathcal{S}(\mathcal{H})$, its max-relative entropy radius $R_{\max}(\mathcal{A})$ is defined as $R_{\max}(\mathcal{A}) := \inf_{\omega\in\mathcal{S}(\mathcal{H})} \sup_{\sigma\in\mathcal{A}} D_{\max}(\sigma \,\|\, \omega)$. For the interpretation of this quantity in quantum information theory, see, e.g. [28, 33, 34] and references therein. We extend this definition to general positive semidefinite operators by keeping the reference ω varying only over the set of states. That is, for a set of PSD operators $\mathcal{A} \subset \mathcal{B}(\mathcal{H})_+$, its max-relative entropy radius $R_{\max}(\mathcal{A})$ is defined as

$$R_{\max}(\mathcal{A}) := \inf_{\omega\in\mathcal{S}(\mathcal{H})} \sup_{A\in\mathcal{A}} D_{\max}(A \,\|\, \omega).$$

Any state ω where the infimum above is attained is called a $D_{\max}$-*divergence center* of $\mathcal{A}$.

If $\mathcal{A} = \{0\}$ then $R_{\max}(\mathcal{A}) = -\infty$, and any state is a divergence center. Assume for the rest that $\mathcal{A} = \{A_1, \ldots, A_r\}$ is finite, and it contains a non-zero element, and hence $R := R_{\max}(\mathcal{A})$ is a finite number. By definition, for every $n \in \mathbb{N}$, there exists an $\omega_n \in \mathcal{S}(\mathcal{H})$ such that $A \leq 2^{R+1/n}\omega_n$ for every $A \in \mathcal{A}$. Since $\mathcal{S}(\mathcal{H})$ is compact, there exists a subsequence n_k, $k \in \mathbb{N}$, such that ω_{n_k}, $k \in \mathbb{N}$, is convergent. Let $\omega^* := \lim_{k\to\infty} \omega_{n_k}$; then $A \leq 2^R\omega^*$ for every $A \in \mathcal{A}$, and hence ω^* is a divergence center. Thus, the set of divergence centers is non-empty. Obviously, if ω is a divergence center then $2^R\omega$ is an upper bound to $\mathcal{A}$, and hence $2^R\omega \geq \mathrm{LUB}(\mathcal{A}) =: L$. Let $\tilde{R} := \log_2 \operatorname{Tr} L$ and $\tilde{\omega} := L/\operatorname{Tr} L$. Then $2^R\omega \geq L$ yields $R \geq \tilde{R}$, while $A \leq 2^{\tilde{R}}\tilde{\omega}$ due to the definition of $\mathrm{LUB}(\mathcal{A})$, and hence $R \leq \tilde{R}$. Thus, $R = \tilde{R}$, i.e., $\operatorname{Tr}(2^R\omega) = \operatorname{Tr} \mathrm{LUB}(\mathcal{A})$. Taking into account that $2^R\omega \geq L$, this implies that $2^R\omega = \mathrm{LUB}(\mathcal{A})$. Thus, the $D_{\max}$-divergence center is unique, and is equal to $\mathrm{LUB}(\mathcal{A})/\operatorname{Tr} \mathrm{LUB}(\mathcal{A})$, while $R_{\max}(\mathcal{A}) = \log \operatorname{Tr} \mathrm{LUB}(\mathcal{A})$.

According to [53] (see also Appendix D), this can be rewritten as $\log P_e^*(A_1, \ldots, A_r) = R_{\max}(\mathcal{A})$. A similar expression for the optimal error probability in terms of the max-relative entropy has been given in [28].

B The classical case

In the classical case the hypotheses (in the single-shot setting) are represented by non-negative functions $A_i : \mathcal{X} \to \mathbb{R}_+$, where $\mathcal{X}$ is some finite set, and POVM elements are replaced by non-negative functions $E_i : \mathcal{X} \to \mathbb{R}_+$, satisfying $\sum_i E_i(x) \leq 1$, $\forall x \in \mathcal{X}$, which we may call a classical POVM. The success probability corresponding to a classical POVM $\{E_i\}$ is $P_s(\{E_i\}) = \sum_{i=1}^r \sum_{x\in\mathcal{X}} A_i(x)E_i(x)$. We can assign to each non-negative function $F : \mathcal{X} \to \mathbb{R}_+$ a diagonal operator $\hat{F}$ on $\mathbb{C}^{\mathcal{X}}$ in an obvious way, and under this identification we get $\sum_{i=1}^r \sum_{x\in\mathcal{X}} A_i(x)E_i(x) = \sum_{i=1}^r \operatorname{Tr} \hat{A}_i\hat{E}_i$, which is the success probability corresponding to hypotheses $\hat{A}_i$ and POVM elements $\hat{E}_i$. On the other hand, if $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ are mutually commuting then there exists a basis in $\mathcal{H}$, labeled by the elements of some finite set $\mathcal{X}$, such that $A_i = \sum_{x\in\mathcal{X}} \langle x|A_i|x\rangle \, |x\rangle\langle x|$. Moreover, for any operator $E \in \mathcal{B}(\mathcal{H})$, we have $\operatorname{Tr} A_iE = \sum_{x\in\mathcal{X}} \tilde{A}(x)\tilde{E}_(x)$, where for $F \in \mathcal{B}(\mathcal{H})$, we let $\tilde{F} : \mathcal{X} \to \mathbb{C}$ be defined by $\tilde{F}(x) := \langle x|F|x\rangle$. In particular, if $E_1, \ldots, E_r$ is a POVM then $\tilde{E}_1, \ldots, \tilde{E}_r$ is a classical POVM, and $\sum_{i=1}^r \sum_{x\in\mathcal{X}} \tilde{A}_i(x)\tilde{E}_i(x) = \sum_{i=1}^r \operatorname{Tr} A_iE_i$. Hence, if the operators representing the hypotheses are diagonal in a given basis then it is enough to consider POVM elements that are also diagonal in the same basis, which reduces the problem into a classical one. Thus, the classical case can be represented both by functions and diagonal operators, and we will not make a difference in the notation between the two representations in what follows.

Consider first the classical binary state discrimination problem with hypotheses $A_1 = A$ and $A_2 = \sum_{i=1}^r B_i$. Then we have the following strengthening of Theorem 4.3:

$$P_e^*\left(A, \sum\nolimits_i B_i\right) \leq \sum_i P_e^*(A, B_i). \tag{91}$$

Indeed,

$$
\begin{aligned}
P_e^*\left(A, \sum_i B_i\right) &= \frac{1}{2}\operatorname{Tr}\left(A + \sum_i B_i\right) - \frac{1}{2}\left\|A - \sum_i B_i\right\|_1 \\
&= \frac{1}{2}\sum_x \left(A(x) + \sum_i B_i(x) - \left|A(x) + \sum_i B_i(x)\right|\right) \\
&= \frac{1}{2}\sum_x f_x\left(\sum_i B_i(x)\right),
\end{aligned}
$$

where $f_x(t) := t + x - |t - x| = 2\min\{t, x\}$. It is easy to see that f_x is subadditive for every x, and hence the above can be continued as

$$
\begin{aligned}
P_e^*\left(A, \sum_i B_i\right) &= \frac{1}{2}\sum_x f_x\left(\sum_i B_i(x)\right) \le \frac{1}{2}\sum_x\sum_i f_x(B_i(x)) = \sum_i \frac{1}{2}\sum_x f_x(B_i(x)) \\
&= \sum_i \left(\frac{1}{2}\operatorname{Tr}(A + B_i) - \frac{1}{2}\|A - B_i\|_1\right) = \sum_i P_e^*(A, B_i).
\end{aligned}
$$

Combining this with Theorem 5.2, we get

$$
P_e^*(A_1, \dots, A_r) \le P_e^* = \sum_{k=1}^r P_e^*\left(A_k, \sum_{l\ne k} A_l\right) \le \sum_{(k,l):\, k\ne l}^r P_e^*(A_k, A_l), \tag{92}
$$

proving Conjecture 2.3 with $c = 2(r-1)$. Below we give a more direct proof of this, without using Theorem 5.2.

Consider now the i.i.d. vs. averaged i.i.d. problem as in Section 4, with hypotheses $A_{1,n} = p\rho^{\otimes n}$ and $B_{i,n} = (1-p)q_i\sigma_i^{\otimes n}$, where $p \in (0,1)$ and q is a probability distribution. Then we have

$$
\begin{aligned}
&\sum_i q_i P_e^*\left(p\rho^{\otimes n}, (1-p)\sigma_i^{\otimes n}\right) \le P_e^*\left(p\rho^{\otimes n}, (1-p)\sum_i q_i\sigma_i^{\otimes n}\right) \\
&\qquad \le \sum_i P_e^*\left(p\rho^{\otimes n}, (1-p)q_i\sigma_i^{\otimes n}\right) \le \sum_i P_e^*\left(p\rho^{\otimes n}, (1-p)\sigma_i^{\otimes n}\right),
\end{aligned}
$$

where the first inequality is due to the convexity of the trace-norm, the second is due to the subadditivity relation (91), and the last inequality is obvious from the definition of the error probability. This yields immediately Conjecture 4.2 in the classical case, i.e.,

$$
\lim_{n\to\infty}\frac{1}{n}\log P_e^*\left(p\rho^{\otimes n}, (1-p)\sum_i q_i\sigma_i^{\otimes n}\right) = -\min_i C(\rho, \sigma_i).
$$

Consider now the classical single-shot state discrimination problem with hypotheses $A_1, \dots, A_r : \mathcal{X} \to \mathbb{R}_+$, and let $m(x) := \max_k A_k(x)$. We say that a POVM $\{E_k\}_{k=1}^r$ is a *maximum likelihood* POVM if $E_k(x) = 0$ when $A_k(x) < m(x)$, and for every $x \in \mathcal{X}$, $\sum_k E_k(x) = 1$. For any POVM $\{E_k\}_{k=1}^r$, we have

$$
P_s(E_1, \dots, E_r) = \sum_k\sum_x A_k(x)E_k(x) \le \sum_k\sum_x m(x)E_k(x) \le \sum_x m(x) = \operatorname{Tr}\max\{A_1, \dots, A_r\},
$$

where $\max\{A_1, \ldots, A_r\} := \sum_x m(x)|x\rangle\langle x|$. The above inequality holds with equality if and only if $\{E_k\}_{k=1}^r$ is a maximum likelihood POVM, and hence we have

$$P_s^*(A_1, \ldots, A_r) = \operatorname{Tr}\max\{A_1, \ldots, A_r\}.$$

Now let $E_1, \ldots, E_r$ be a maximum likelihood measurement. Then the individual error probabilities are, for each k,

$$P_{e,k} = \sum_x A_k(x)\left(1 - E_k(x)\right) = \sum_{x:\, A_k(x) < m(x)} A_k(x) + \sum_{x:\, A_k(x) = m(x)} A_k(x)\left(1 - E_k(x)\right).$$

Obviously, if $A_k(x) < m(x)$ then there exists an $l \neq k$ such that $A_k(x) < A_l(x)$, and if $A_k(x) = m(x)$ and $A_k(x)\left(1 - E_k(x)\right) > 0$ then there exists and $l \neq k$ such that $A_k(x) = A_l(x)$. Hence,

$$P_{e,k} \leq \sum_{l \neq k} \sum_{x:\, A_k(x) < A_l(x)} A_k(x) + \sum_{l \neq k} \sum_{x:\, A_k(x) = A_l(x)} A_k(x) = \sum_{l \neq k} \sum_{x:\, A_k(x) \leq A_l(x)} A_k(x).$$

Thus,

$$\begin{aligned}
P_e^*(A_1, \ldots, A_r) = \sum_{k=1}^r P_{e,k} &\leq \sum_{k=1}^r \sum_{l \neq k} \sum_{x:\, A_k(x) \leq A_l(x)} A_k(x) \\
&\leq \sum_{k=1}^r \sum_{l \neq k} \left[\sum_{x:\, A_k(x) \leq A_l(x)} A_k(x) + \sum_{x:\, A_k(x) > A_l(x)} A_l(x) \right] \\
&= \sum_{k=1}^r \sum_{l \neq k} \left[\frac{1}{2} \operatorname{Tr}(A_k + A_l) - \frac{1}{2} \|A_k - A_l\|_1 \right] \\
&= \sum_{(k,l):\, k \neq l} P_e^*(A_k, A_l),
\end{aligned}$$

and we recover (92).

C The pure state case

Let $A_1, A_2 \in \mathcal{B}(\mathcal{H})_+$ be rank one operators; then we can write them as $A_i = |x_i\rangle\langle x_i| = p_i|\psi_i\rangle\langle\psi_i| = p_i\sigma_i$, where $p_i := \operatorname{Tr} A_i$. Many of the divergence measures coincide in this case; indeed, it is easy to see that

$$|\langle\psi_1, \psi_2\rangle|^2 = F(\sigma_1, \sigma_2)^2 = Q_s(\sigma_1\|\sigma_2) = Q_{\min}(\sigma_1, \sigma_2) = \exp(-C(\sigma_1, \sigma_2)), \qquad s \in [0, 1].$$

A straightforward computation gives that $\|A_1 - A_2\|_1 = \sqrt{(p_1 + p_2)^2 - 4p_1p_2|\langle\psi_1, \psi_2\rangle|^2}$, and hence

$$P_e^*(A_1, A_2) = \frac{1}{2}\operatorname{Tr}(A_1 + A_2) - \frac{1}{2}\|A_1 - A_2\|_1 = \frac{2p_1p_2|\langle\psi_1, \psi_2\rangle|^2}{p_1 + p_2 + \|A_1 - A_2\|_1}$$

Noting that $0 \leq \|A_1 - A_2\|_1 \leq p_1 + p_2$, we get

$$\frac{p_1 p_2}{p_1 + p_2} |\langle \psi_1, \psi_2 \rangle|^2 \leq P_e^*(A_1, A_2) \leq \frac{2p_1 p_2}{p_1 + p_2} |\langle \psi_1, \psi_2 \rangle|^2 \tag{93}$$

Consider now two sequences of rank one operators $\vec{A_i} = \{A_{i,n}\}_{n \in \mathbb{N}}$, $i = 1, 2$, and let $p_{i,n} := \operatorname{Tr} A_{i,n}$, and $A_{i,n} = p_{i,n} |\psi_{i,n}\rangle\langle\psi_{i,n}| = p_{i,n}\sigma_{i,n}$. Applying (93) to each n, we get

$$\frac{p_{1,n} p_{2,n}}{p_{1,n} + p_{2,n}} \exp\left(-C(\sigma_{1,n}, \sigma_{2,n})\right) \leq P_e^*(A_{1,n}, A_{2,n}) \leq \frac{2p_{1,n} p_{2,n}}{p_{1,n} + p_{2,n}} \exp\left(-C(\sigma_{1,n}, \sigma_{2,n})\right).$$

If we assume now that $0 < \liminf_n p_{i,n} \leq \limsup_n p_{i,n} < +\infty$, $i = 1, 2$, then taking the limit $n \to \infty$ in the above formula yields

$$\lim_{n \to \infty} \frac{1}{n} \log P_e^*(A_{1,n}, A_{2,n}) = -C(\vec{\sigma_1}, \vec{\sigma_2}) = -C\left(\vec{A}_1, \vec{A}_2\right),$$

where the last identity is straightforward to verify. Thus in the pure state case we can get the Chernoff bound theorem from the above elementary argument, without using the trace inequality of [4] or the reduction to classical states from [38].

Consider now the case $r > 2$, and let $A_1, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ be rank one operators. Let $E_i := A_0^{-1/2} A_i A_0^{-1/2}$ be the POVM elements of the pretty good measurement, where $A_0 := \sum_{i=1}^r A_i$. It was shown in Appendix A of [19] that for every i,

$$\operatorname{Tr} A_i(I - E_i) \leq \frac{1}{p_i} \sum_{j:\, j \neq i} |\langle x_i, x_j \rangle|^2.$$

Summing it over i, we get

$$P_e^*(A_1, \ldots, A_r) \leq P_e(\{E_1, \ldots, E_r\}) \leq \sum_{i=1}^r \frac{1}{p_i} \sum_{j:\, j \neq i} |\langle x_i, x_j \rangle|^2 \leq \frac{1}{\min_i p_i} \sum_{(i,j):\, i \neq j} \exp(-C(A_i, A_j)), \tag{94}$$

while Theorem 2.2 yields

$$P_e^*(A_1, \ldots, A_r) \geq \frac{1}{r-1} \sum_{(k,l):\, k<l} P_e^*(A_k, A_l) \geq \frac{1}{r-1} \sum_{(k,l):\, k<l} \frac{1}{p_k + p_l} \exp(-C(A_k, A_l)), \tag{95}$$

where the last inequality is due to (93). Note that (94) also yields a decoupling bound for the error probabilities, as $|\langle x_i, x_j \rangle|^2 / p_i \leq (1 + p_j/p_i) P_e^*(A_i, A_j)$ by (93), and hence

$$P_e^*(A_1, \ldots, A_r) \leq \sum_{i=1}^r \frac{1}{p_i} \sum_{j:\, j \neq i} |\langle x_i, x_j \rangle|^2 \leq \frac{\operatorname{Tr} A_0}{\min_i p_i} \sum_{(i,j):\, i \neq j} P_e^*(A_i, A_j).$$

Consider now the asymptotic case, with hypotheses $\vec{A_i}$, $i = 1, \ldots, r$, and assume as before that $0 < \liminf_n p_{i,n} \leq \limsup_n p_{i,n} < +\infty$, $\forall i$. Applying (94) and (95) to every n, and taking the limit $n \to \infty$, we get

$$\lim_{n \to \infty} \frac{1}{n} \log P_e^*(A_{1,n}, \ldots, A_{r,n}) = -\max_{(i,j):\, i \neq j} C\left(\vec{A}_i, \vec{A}_j\right).$$

D Semidefinite program representations of success and error probabilities

The average success probability of a POVM $\{E_k\}$ for discriminating between r PSD operators $\{A_k\}_{k=1}^r$ is given by

$$P_s(\{E_k\}) = \sum_{k=1}^{r} \mathrm{Tr}(A_k E_k), \tag{96}$$

and the optimal success probability P_s^* is the maximum over all POVMs:

$$P_s^* = \max\left\{P_s(\{E_k\}) : \{E_k\}_{k=1}^r \ \text{ POVM}\right\}. \tag{97}$$

In this section we consider the consequences of the following simple observation [53]: in (97) the maximum of a linear functional is taken over the set of POVMs, which is a convex set. This optimization problem is therefore a so-called *semidefinite program* (SDP) [52]. One consequence is that P_s^* can be efficiently calculated numerically by SDP solvers even when no closed form analytical solution exists. Another, theoretically important consequence is that the duality theory of SDPs allows to express the value of P_s^* in a dual way as a minimization problem [14, 27].

The Lagrangian of problem (97) is

$$\begin{aligned}\mathcal{L} &= \sum_k \mathrm{Tr}(A_k E_k) + \sum_k \mathrm{Tr}(Z_k E_k) + \mathrm{Tr}\, Y\left(I - \sum_k E_k\right) \\ &= \mathrm{Tr}\, Y + \sum_k \mathrm{Tr}\, E_k(A_k + Z_k - Y),\end{aligned}$$

where the operators Z_k and Y are the Lagrange multipliers of the problem. If the Z_k are taken to be PSD, we see that always $P_s(\{E_k\}) \le \mathcal{L}$. This does not change when maximizing over all POVMs, and certainly not when in the maximization of $\mathcal{L}$ over the E_k the POVM constraints are dropped. Hence $P_s^* \le \max_{E_k} \mathcal{L}$. This unconstrained maximization is easy to do; when $Y = A_k + Z_k$ for all k, $\max_{E_k} \mathcal{L} = \mathrm{Tr}\, Y$, otherwise it is positive infinity. Minimizing this upper bound over all PSD Z_k and all Y yields the best upper bound on P_s^*. The positivity condition on the Z_k can be replaced by requiring that for all k, $Y \ge A_k$. Minimizing over such Y then gives

$$P_s^* \le \min_Y \{\mathrm{Tr}\, Y : Y \ge A_k,\ k = 1\ldots, r\}, \tag{98}$$

which is again an SDP, called the *dual* of the original (*primal*) SDP (see, e.g., [53] or [14], equations (15) and (16)).

Therefore, the optimal success probability is bounded above by the trace of the LUB of all weighted density operators:

$$P_s^* \le \mathrm{Tr}\, \mathrm{LUB}(A_1, \ldots, A_r). \tag{99}$$

Note that in the classical case (all A_i are diagonal, with diagonal elements $A_i(j) = p_i q_i(j)$) the LUB is the entrywise maximum, so that the dual SDP reproduces the maximum-likelihood formula $P_s^* = \sum_j \max_i(A_i(j))$.

The difference between the maximum of the primal SDP (P_s^*) and the minimum of the dual SDP is called the *duality gap*. One can show that the duality gap is zero, provided some mild technical conditions are satisfied (e.g. Slater's conditions), in which case equality holds:

$$P_s^* = \operatorname{Tr} \operatorname{LUB}(A_1, \ldots, A_r). \tag{100}$$

If the duality gap is zero, then the optimal Z_k (denoted by Z_k^*) and the optimal POVM $\{E_k^*\}$ must necessarily satisfy a simple relation, called the *complementary slackness* condition. Indeed, Let Y^* be the operator where the minimum on the RHS of (98) is attained. As $\operatorname{Tr} Y^* = \operatorname{Tr}(Y^* \sum_k E_k^*)$ and $Y^* - A_k = Z_k^*$, the equality $\sum_k \operatorname{Tr} A_k E_k^* = \operatorname{Tr} Y^*$ implies $\sum_k \operatorname{Tr}(Z_k^* E_k^*) = 0$. Since all Z_k^* and E_k^* are required to be PSD, this actually means that

$$Z_k^* E_k^* = 0, \quad \forall k. \tag{101}$$

A simple consequence of these complementary slackness conditions is obtained by summing over k: $\sum_k Z_k^* E_k^* = 0$. Noting that $Z_k = Y - A_k$, this yields

$$Y^* = \sum_k A_k E_k^*. \tag{102}$$

Combined with the conditions $Y^* \geq A_k$ for all k, these are the optimality conditions first obtained by Yuen, Kennedy and Lax [53].

E Short proofs of Barnum and Knill's and Tyson's bounds

Proof of Theorem 3.1. The main ingredient of the proof is the following lemma (a slight improvement over Lemma 5 in [8], which lacked the factor $\frac{1}{2}$). Let M be a positive semidefinite $n \times n$ matrix, symmetrically partitioned as the 2×2 block matrix $M = \begin{pmatrix} X & Y \\ Y^* & Z \end{pmatrix}$, where X is $n_1 \times n_1$, Y is $n_1 \times n_2$ and Z is $n_2 \times n_2$ (with $n = n_1 + n_2$). Let M^2 be partitioned conformally. Then the off-diagonal blocks of M and M^2 satisfy

$$||M_{1,2}||_2^2 \leq \frac{1}{2} ||(M^2)_{1,2}||_1.$$

Note that the validity of this lemma does not extend to general $m \times m$ partitions.

> *Proof of lemma.* We have $M_{1,2} = Y$ and $(M^2)_{1,2} = XY + YZ$. Let us, without loss of generality, assume that $n_1 \leq n_2$. From the singular value decomposition of Y we can obtain a basis for representing M in which Y is pseudo-diagonal with non-negative diagonal elements. Let (for $i = 1, \ldots, n_1$) x_i and y_i be the diagonal elements of X and Y, and z_i the first n_1 diagonal elements of Z, all of which are non-negative. As M is PSD, any of its principal submatrices is PSD too, and we have $y_i \leq \sqrt{x_i z_i} \leq (x_i + z_i)/2$. Thus
>
> $$||Y||_2^2 = \sum_{i=1}^{n_1} y_i^2 \leq \frac{1}{2} \sum_{i=1}^{n_1} x_i y_i + y_i z_i = \frac{1}{2} \sum_{i=1}^{n_1} (XY + YZ)_{i,i} \leq \frac{1}{2} ||XY + YZ||_1,$$

as required. The last inequality follows from the inequality $|\operatorname{Tr} A| \leq ||A||_1$ applied to the square matrix obtained by padding $XY + YZ$ with extra rows containing zero (an operation that does not affect the trace norm). □

To prove Theorem 3.1, let X be the $r \times 1$ column matrix $X := (A_j^{1/2} A_0^{-1/4})_{j=1}^r$. Then $X^*X = \sum_{j=1}^r A_0^{-1/4} A_j A_0^{-1/4} = A_0^{1/2}$. Furthermore, let $N = XX^*$. Then $N_{i,j} = A_i^{1/2} A_0^{-1/2} A_j^{1/2}$ and $(N^2)_{i,j} = (XX^*XX^*)_{i,j} = A_i^{1/2} A_j^{1/2}$.

For each value of $i = 1, \ldots, r$ we now apply the lemma to the 2×2 block matrix $M = \begin{pmatrix} X & Y \\ Y^* & Z \end{pmatrix}$ where $X = N_{i,i}$, Y is the i-th row of N, but with the i-th column removed, and Z is the submatrix of N with the i-th row and i-th column removed. Thus, $M_{1,2} = Y$ is itself a row block matrix consisting of the $r-1$ blocks $A_i^{1/2} A_0^{-1/2} A_j^{1/2}$ for fixed i and $j \neq i$. Likewise, $(M^2)_{1,2}$ is a row block matrix consisting of the $r-1$ blocks $A_i^{1/2} A_j^{1/2}$. The lemma then implies, for all i,

$$\begin{aligned} \sum_{j:j\neq i} \operatorname{Tr} A_i A_0^{-1/2} A_j A_0^{-1/2} &= \sum_{j:j\neq i} ||A_i^{1/2} A_0^{-1/2} A_j^{1/2}||_2^2 = ||M_{1,2}||_2^2 \\ &\leq \frac{1}{2} ||(M^2)_{1,2}||_1 = \frac{1}{2} ||(A_i^{1/2} A_j^{1/2})_{j\neq i}||_1 \\ &\leq \frac{1}{2} \sum_{j:j\neq i} ||A_i^{1/2} A_j^{1/2}||_1 = \frac{1}{2} \sum_{j:j\neq i} F(A_i, A_j). \end{aligned}$$

The last inequality is just the triangle inequality for the trace norm. Summing over all i yields the stated bound on the error probability P_e^{PG}. □

Proof of Theorem 3.2. For any operator X with $\|X\| \leq 1$ and any quantum state σ we have

$$1 - \|X\sigma\|_1 \leq 1 - \operatorname{Tr}(X^*X\sigma) \leq 1 - \|X\sigma\|_1^2 \leq 2(1 - \|X\sigma\|_1). \tag{103}$$

The first two inequalities both follow from Hölder's inequality ([11], Cor IV.2.6):

$$\operatorname{Tr}(X^*X\sigma) \leq \|X^*X\sigma\|_1 \leq \|X\sigma\|_1 \; \|X\| \leq \|X\sigma\|_1 ,$$

and

$$\|X\sigma\|_1^2 = \left\|(X\sigma^{1/2})\sigma^{1/2}\right\|_1^2 \leq \left\|X\sigma^{1/2}\right\|_2^2 \; \left\|\sigma^{1/2}\right\|_2^2 = \operatorname{Tr} X^*X\sigma$$

and the last inequality in (103) follows from $1 - x^2 \leq 2(1-x)$, $x \in \mathbb{R}$. Applying (103) for $\sigma_k := A_k / \operatorname{Tr} A_k$ and X_k and summing over k yields

$$\sum_{k=1}^r (\operatorname{Tr} A_k)\,(1 - \|X_k\sigma_k\|_1) \leq \sum_{k=1}^r (\operatorname{Tr} A_k)\,(1 - \operatorname{Tr}(X_k^* X_k \sigma_k)) \leq 2 \sum_{k=1}^r (\operatorname{Tr} A_k)\,(1 - \|X_k\sigma_k\|_1) .$$

Taking the minimum over all X_k then yields the inequalities of the theorem. □

Proof of Theorem 3.3. In the following we will abbreviate the expression $\sum_{k=1}^r A_k^2$ by S.

First note that the operator $\bigoplus_k X_k A_k$ is a pinching of the block operator $(X_j A_k)_{j,k}$. This operator is the product of the column block operator $\mathcal{X} := (X_j)_{j,1}$ and the row block operator $\mathcal{A} := (A_k)_{1,k}$. Because unitarily invariant norms do not increase under pinchings, we get

$$\sum_k \|X_k A_k\|_1 = \left\| \bigoplus_k X_k A_k \right\|_1 \le \|(X_j A_k)_{j,k}\|_1 = \|\mathcal{X}\mathcal{A}\|_1 = \mathrm{Tr}\,(\mathcal{A}^*\mathcal{X}^*\mathcal{X}\mathcal{A})^{1/2}\,.$$

Noting that $\mathcal{X}^*\mathcal{X} = \sum_k X_k^* X_k = I$ yields

$$\mathrm{Tr}\,(\mathcal{A}^*\mathcal{X}^*\mathcal{X}\mathcal{A})^{1/2} = \mathrm{Tr}\,(\mathcal{A}^*\mathcal{A})^{1/2} = \|\mathcal{A}\|_1 = \|\mathcal{A}^*\|_1 = \mathrm{Tr}\,S^{1/2}.$$

Hence,

$$\sum_k \|X_k A_k\|_1 \le \mathrm{Tr}\,S^{1/2}.$$

Equality can be achieved by taking the SQ measurement, $X_k = A_k S^{-1/2}$. Indeed, from $X_k A_k = A_k S^{-1/2} A_k \ge 0$, we get

$$\sum_k \|X_k A_k\|_1 = \sum_k \mathrm{Tr}(X_k A_k) = \mathrm{Tr}\left(S^{-1/2}\sum_k A_k^2\right) = \mathrm{Tr}\,S^{1/2}.$$

This shows that the maximum of $\sum_k \|X_k A_k\|_1$ over any complete set of r measurement operators $\{X_k\}$ is achieved for $X_k = A_k S^{-1/2}$ and is given by $\mathrm{Tr}\,S^{1/2}$. Hence, $\Gamma^* = \mathrm{Tr}\,A_0 - \mathrm{Tr}\,S^{1/2}$, which is (41). □

Acknowledgments

The authors are grateful to Giulio Chiribella for pointing out the decoupling bounds of [8]. This work was partly supported by an Odysseus grant from the Flemish FWO (KA) and by the Marie Curie International Incoming Fellowship "QUANTSTAT" within the 7th European Community Framework Programme (MM). MM also acknowledges support by the European Research Council (Advanced Grant "IRQUAT"). The authors are grateful to the following institutions for their hospitality: the Fields Institute in Toronto (KA and MM), during the Thematic Program on Mathematics in Quantum Information, 2009, where this research has started, and the University of Ulm (KA).

References

[1] T. Ando, *Concavity of certain maps on positive definite operators and applications to Hadamard operators*, Linear Algebra Appl. **26**, 203–241 (1979).

[2] T. Ando, *Parameterization of minimal points of some convex sets of operators*, Acta Sci. Math. (Szeged) **57**, 3–10 (1993).

[3] T. Ando, *Problem of infimum in the positive cone*, in *Analytic and geometric inequalities and applications*, T.M. Rassias and H.M. Srivastava (Eds.), Mathematics and its Applications **478**, 1–12, Kluwer Acad. Publ., Dordrecht (1999).

[4] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan and F. Verstraete, *Discriminating states: the quantum Chernoff bound*, Phys. Rev. Lett. **98** 160501, (2007)

[5] K.M.R. Audenaert, M. Nussbaum, A. Szkoła and F. Verstraete, *Asymptotic error rates in quantum hypothesis testing*, Commun. Math. Phys. **279**, 251–283, (2008)

[6] K. Audenaert M. Mosonyi, F. Verstraete, *Quantum state discrimination bounds for finite sample size*, J. Math. Phys. **53**, 122205, (2012).

[7] K.M.R. Audenaert, *Comparisons between quantum state distinguishability measures*, Quant. Inf. Comp. **14**(1&2) 31–38. In press (2014).

[8] H. Barnum and E. Knill, *Reversing quantum dynamics with near-optimal quantum and classical fidelity*, J. Math. Phys. **43**, 2097–2106 (2002).

[9] I. Bjelaković, R. Siegmund-Schultze: *An ergodic theorem for the quantum relative entropy*; Commun. Math. Phys. **247**, 697–712, (2004)

[10] V.P. Belavkin and V. Maslov, *Design of Optimal Dynamic Analyzer: Mathematical Aspects of Wave Pattern Recognition*, In Mathematical Aspects of Computer Engineering, edited by V. Maslov, pp. 146-237 (Mir, Moscow 1987);

[11] R. Bhatia, *Matrix Analysis*, Springer Verlag, (1997)

[12] P.J. Curlander, *Quantum Limitations on Communication Systems*, Ph.D. Thesis, MIT, Cambridge, MA, (1979).

[13] N. Datta, *Min- and Max-Relative Entropies and a New Entanglement Monotone*, IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2816–2826, (2009).

[14] Y.C. Eldar, A. Megretski and G.C. Verghese, *Designing optimal quantum detectors via semidefinite programming*, IEEE Trans. Inf. Theory **49**, 1007–1012, (2003)

[15] C.A. Fuchs, J. van de Graaf, *Cryptographic distinguishability measures for quantum-mechanical states*, IEEE Transactions on Information Theory, Volume: 45, Issue: 4, pp. 1216 - 1227, (1999).

[16] P. Hausladen and W. Wootters, *A "pretty good" measurement for distinguishing quantum states*, J. Mod. Opt. **41**, 2385 (1994).

[17] M. Hayashi, *Quantum Information: an Introduction*, Springer (2006).

[18] M. Hayashi, A. Kawachi and H. Kobayashi, *Quantum measurements for hidden subgroup problems with optimal sample*, Quantum Information and Computation **8**, 0345–0358, (2008).

[19] P. Hayden, D. Leung, G. Smith, *Multiparty data hiding of quantum information*, Phys. Rev. A **71**, 062339, (2005).

[20] C.W. Helström, *Quantum Detection and Estimation Theory*, Academic Press, New York, (1976)

[21] F. Hiai, M. Mosonyi, T. Ogawa, *Large deviations and Chernoff bound for certain correlated states on the spin chain*, J. Math. Phys. **48**, 123301, (2007)

[22] F. Hiai, M. Mosonyi, T. Ogawa, *Error exponents in hypothesis testing for correlated states on a spin chain*, J. Math. Phys. **49**, 032112, (2008)

[23] F. Hiai, M. Mosonyi, M. Hayashi, *Quantum hypothesis testing with group symmetry*, J. Math. Phys. **50**, 103304, (2009).

[24] A.S. Holevo, *Information-theoretical aspects of quantum measurement*, Problemy Peredachi Informatsii **9**(2), 31–42, (1973) (in Russian); English translation: A. S. Kholevo, Problems of Information Transmission, **9**, 110–118, (1973).

[25] A.S. Holevo, *On Asymptotically Optimal Hypothesis Testing in Quantum Statistics, Theor. Prob. Appl.* **23**, pp. 411–415, (1978)

[26] W.Y. Hwang and J. Bae, *Minimum-error state discrimination constrained by the non-signalling principle*, J. Math. Phys. **51**, 022202, (2010);

[27] M. Jezek, J. Rehacek and J. Fiurasek, *Finding optimal strategies for minimum-error quantum-state discrimination*, Phys. Rev. A **65**, 060301(R) (2002).

[28] R. König, R. Renner, C. Schaffner, *The operational meaning of min- and max-entropy*, IEEE Trans. Inf. Th. **55**(9), 4337–4347, (2009).

[29] A. Montanaro, *On the distinguishability of random quantum states*, Comm. Math. Phys. **273**(3), 619–636, (2007);

[30] A. Montanaro, *A lower bound on the probability of error in quantum state discrimination*, IEEE Inf. Theory Workshop, May 2008, 378–380, (2008);

[31] M. Mosonyi, F. Hiai, T. Ogawa, M. Fannes, *Asymptotic distinguishability measures for shift-invariant quasi-free states of fermionic lattice systems*, J. Math. Phys. **49**, 072104, (2008)

[32] M. Mosonyi, *Hypothesis testing for Gaussian states on bosonic lattices*, J. Math. Phys. **50**, 032104, (2009)

[33] M. Mosonyi and N. Datta, *Generalized relative entropies and the capacity of classical-quantum channels*, J. Math. Phys. **50**, 072104, (2009).

[34] M. Mosonyi, F. Hiai, *On the quantum Renyi relative entropies and related capacity formulas*, IEEE Transactions on Information Theory, vol. 57, pp. 2474–2487, (2011).

[35] M. Mosonyi, *Rényi divergences and the classical capacity of finite compound channels*, arXiv:1310.7525, (2013).

[36] A. Nayak and J. Salzman, *Limits on the ability of quantum states to convey classical messages*, Journal of the ACM, **53**(1), 184–206, (2006)

[37] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, (2000).

[38] M. Nussbaum, A. Szkoła: *The Chernoff lower bound for symmetric quantum hypothesis testing*, Ann. Statist. **37**(2), 1040–1057, (2009).

[39] M. Nussbaum, A. Szkoła, *Exponential error rates in multiple state discrimination on a quantum spin chain*, J. Math. Phys. **51**, 072203, (2010).

[40] M. Nussbaum, A. Szkoła, *Asymptotically optimal discrimination between multiple pure quantum states*, In: Theory of Quantum Computation, Communication and Cryptography. 5th Conference, TQC 2010, Leeds, UK. Revised Selected Papers. Lecture Notes in Computer Science, Vol 6519, van Dam, Wim; Kendon, Vivien M.; Severini, Simone (Eds.), Springer, 1–8, (2011).

[41] M. Nussbaum, A. Szkoła: *An asymptotic error bound for testing multiple quantum hypotheses*, Ann. Statist. **39**, 3211–3233, (2011).

[42] M. Nussbaum, *Attainment of the multiple quantum Chernoff bound for certain ensembles of mixed states*, Proceedings of the First International Workshop on Entangled Coherent States and Its Application to Quantum Information Science, (Usuda, T.S., Kato, K., Eds.), Tamagawa University, Tokyo, Japan, 77–81 (2013).

[43] T. Ogawa, M. Hayashi: *On error exponents in quantum hypothesis testing*; IEEE Trans. Inform. Theory **50**, 1368–1372, (2004).

[44] A. Peres and W.K. Wootters, *Optimal Detection of Quantum Information*, Phys. Rev. Lett. **66**, 1119–1122, (1991).

[45] D.W. Qiu, *Minimum-error discrimination between mixed quantum states*, Phys. Rev. A **77**, 012328, (2008).

[46] D.W. Qiu and L. Li, *Minimum-error discrimination of quantum states: New bounds and comparisons*, Phys. Rev. A **81**, 04232, (2010).

[47] R. Renner, *Security of Quantum Key Distribution*, PhD dissertation, Swiss Federal Institute of Technology Zurich, Diss. ETH No. 16242, (2005).

[48] N.P. Salikhov, *Asymptotic properties of error probabilities of tests for distinguishing between several multinomial testing schemes*, Dokl. Akad. Nauk SSSR **209**, 54–57, (Russian, 1973).

[49] N.P. Salikhov, *On a generalization of Chernoff divergence*, Teor. Veroyatn. Primen. **43**, 294–314, (Russian, 1998). English translation in Theory Probab. Appl. **47**(2), 286–298, (2003).

[50] J. Tyson, *Two-sided estimates of minimum-error distinguishability of mixed quantum states via generalized Holevo-Curlander bounds*, J. Math. Phys. **50**, 032106, (2009).

[51] J. Tyson, *Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates*, J. Math. Phys. **51**, 092204, (2010).

[52] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Review **38**, 49–95, (1996).

[53] H.P. Yuen, R.S. Kennedy and M. Lax, *Optimum testing of Multiple Hypotheses in Quantum Detection Theory*, IEEE Trans. Inform. Theory **21**(2), 125–134, (1975).