Commun. Math. Phys. 279, 251–283 (2008)
Digital Object Identifier (DOI) 10.1007/s00220-008-0417-5

Communications in
Mathematical Physics

Asymptotic Error Rates in Quantum Hypothesis Testing

K. M. R. Audenaert[1,2], **M. Nussbaum**[3], **A. Szkoła**[4], **F. Verstraete**[5]

[1] Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, London SW7 2PG, UK
[2] Dept. of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.
E-mail: Koenraad.Audenaert@rhul.ac.uk
[3] Department of Mathematics, Cornell University, Ithaca, NY 14853, USA.
E-mail: nussbaum@math.cornell.edu
[4] Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany.
E-mail: szkola@mis.mpg.de
[5] Fakultät für Physik, Universität Wien, Boltzmanngasse 5, 1090 Wien, Austria.
E-mail: frank.verstraete@univie.ac.at

Received: 5 September 2007 / Accepted: 27 November 2007
Published online: 8 February 2008 –

Abstract: We consider the problem of discriminating between two different states of a finite quantum system in the setting of large numbers of copies, and find a closed form expression for the asymptotic exponential rate at which the error probability tends to zero. This leads to the identification of the quantum generalisation of the classical Chernoff distance, which is the corresponding quantity in classical symmetric hypothesis testing.

The proof relies on two new techniques introduced by the authors, which are also well suited to tackle the corresponding problem in asymmetric hypothesis testing, yielding the quantum generalisation of the classical Hoeffding bound. This has been done by Hayashi and Nagaoka for the special case where the states have full support.

The goal of this paper is to present the proofs of these results in a unified way and in full generality, allowing hypothesis states with different supports. From the quantum Hoeffding bound, we then easily derive quantum Stein's Lemma and quantum Sanov's theorem. We give an in-depth treatment of the properties of the quantum Chernoff distance, and argue that it is a natural distance measure on the set of density operators, with a clear operational meaning.

1. Introduction

One of the basic tasks in information theory is discriminating between two different information sources, modelled by (time-discrete) stochastic processes. Given a source that generates independent, identically distributed (i.i.d.) random variables, according to one out of two possible probability distributions, the task is to determine which distribution is the true one, and to do so with minimal error, whatever error criterion one chooses.

This basic decision problem has an equally basic quantum-informational incarnation. Given an information source that emits quantum systems (particles) independently and

identically prepared in one out of two possible quantum states, figure out which state is the true one, with minimal error probability.

In both settings, we're dealing with two hypotheses, each one pertaining to one law represented by a probability distribution or a quantum state, respectively, and the discrimination problem is thus a particular instance of a hypothesis testing problem.

In hypothesis testing, one considers a null hypothesis and an alternative hypothesis. The alternative hypothesis is the one of interest and states that "something significant is happening", for example, a cell culture under investigation is coming from a malignant tumor, or some case of flu is the avian one, or an e-mail attachment is a computer virus. In contrast, the null hypothesis corresponds to this not being the case; the cells are normal ones, the flu can be treated with an aspirin, and the attachment is just a nice picture. This is inherently an asymmetric situation, and Neyman and Pearson introduced the idea of similarly making a distinction between type I and type II errors.

- The type I error or "false positive", denoted by α, is the error of accepting the alternative hypothesis when in reality the null hypothesis holds and the results can be attributed merely to chance.
- The type II error or "false negative", denoted by β, is the error of accepting the null hypothesis when the alternative hypothesis is the true state of nature.

The costs associated to the two types of error can be widely different, or even incommensurate. For example, in medical diagnosis, the type I error corresponds to diagnosing a healthy patient with a certain affliction, which can be an expensive mistake, causing a lot of grievance. On the other hand, the type II error may correspond to declaring a patient healthy while in reality (s)he has a life-threatening condition, which can be a fatal mistake.

To treat the state discrimination problem as a hypothesis test, we assign the null hypothesis to one of the two states and the alternative hypothesis to the other one. If all we want to know is which one of the two possible states we are observing, the mathematical treatment is completely symmetric under the interchange of these two states. It therefore fits most naturally in the setting of *symmetric hypothesis testing*, where no essential distinction is made between the two kinds of errors. To wit, in symmetric hypothesis testing, one considers the average, or Bayesian, error probability P_e, defined as the average of α and β weighted by the prior probabilities of the null and the alternative hypothesis, respectively.

This paper will be concerned with symmetric as well as with asymmetric quantum hypothesis testing. Since we have developed the main techniques in the symmetric setting we will start with this case and address the asymmetric setting at the end.

The optimal solution to the symmetric classical hypothesis test is given by the maximum-likelihood (ML) test. Starting from the outcomes of an experiment involving n independent draws from the unknown distribution, one calculates the conditional probabilities (likelihoods) that these outcomes can be obtained when the distribution is the one of the null hypothesis and the one of the alternative hypothesis, respectively. One decides then on the hypothesis for which the conditional probability is the highest. I.e. if the *likelihood ratio* is higher than 1, the null hypothesis is rejected, otherwise it is accepted.

In the quantum setting, the experiment consists of preparing n independent copies of a quantum system in an unknown state, which is either ρ or σ, and performing an optimal measurement on them. We assume that the quantum systems are finite, implying that the states are associated to density operators on a finite-dimensional complex Hilbert space. Under the null hypothesis, the combined n copies correspond to an n-fold tensor

product density operator $\rho^{\otimes n}$, while under the alternative hypothesis, the associated density operator is $\sigma^{\otimes n}$. The null hypothesis is then accepted or rejected according to the outcome of the measurement and the specified decision rule. The task of finding this optimal measurement is so fundamental that it was one of the first problems considered in the field of quantum information theory; it was solved in the one-copy case more than 30 years ago by Helstrom and Holevo [14,17]. We refer to the generalised ML-tests as Holevo-Helstrom tests. In the special case of equal priors, the associated minimal probability of error achieved by the optimal measurement can be calculated from the trace norm distance between the two states:

$$P^*_{e,n}(\rho,\sigma) = \frac{1}{2}(1 - \|\rho^{\otimes n} - \sigma^{\otimes n}\|_1/2), \tag{1}$$

where $\|A\|_1 := \mathrm{Tr}\,|A|$ denotes the trace norm.

Going back to the classical case again, in a seminal paper, H. Chernoff [8] investigated the so-called *asymptotical efficiency* of a class of statistical tests, which includes the likelihood ratio test mentioned before. The probability of error $P_{e,n}$ in discriminating two probability distributions decreases exponentially in n, the number of draws from the distribution: $P_{e,n} \sim \exp(-\xi n)$. For finite n this is a rather crude approximation. However, as n grows larger one finds better and better agreement, and the exponent ξ becomes meaningful in the asymptotic limit. The asymptotical efficiency is exactly the asymptotic limit of this exponent.

Chernoff was able to derive an (almost) closed expression for this asymptotic efficiency, which was later named eponymously in his honour. For two discrete probability distributions p and q, this expression is given by

$$\xi_{CB}(p,q) := -\log\left(\inf_{0\le s\le 1}\sum_i p(i)^{1-s}q(i)^s\right), \tag{2}$$

which is of closed form but for a single variable minimisation. This quantity goes under the alternative names of Chernoff distance, Chernoff divergence and Chernoff information.

While Chernoff's main purpose was to use this asymptotic efficiency measure to compare the power of different tests – the mathematically optimal test need not always be the most practical one – it can also be used as a distinguishability measure between the distributions (states) of the two hypotheses. Indeed, fixing the test, its efficiency for a particular pair of distributions gives a meaningful indication of how well these two distributions can be distinguished by that test. This is especially meaningful if the applied test is the optimal one.

A quantum generalisation of Chernoff's result is highly desirable. Given the large amount of experimental effort in the context of quantum information processing to prepare and measure quantum states, it is of fundamental importance to have a theory that allows to discriminate different quantum states in a meaningful way. Despite considerable effort, however, the quantum generalisation of the Chernoff distance has until recently remained unsolved.

In the previous papers, [21] and [1], this issue was finaly settled and the asymptotic error exponent was identified, when the optimal Holevo-Helstrom strategy for discriminating between the two states is used, by proving that the following version of the Chernoff distance

$$\xi_{QCB}(\rho,\sigma) := -\log\left(\inf_{0\le s\le 1}\mathrm{Tr}[\rho^{1-s}\sigma^s]\right), \tag{3}$$

has the same operational meaning as its classical counterpart: It specifies the asymptotic rate exponent of the minimal error probability $P^*_{e,n}$ (recall definition (1)). Remarkably, it looks like an almost naïve generalisation of the classical expression (2).

We remark that in the literature different extensions of the classical expression have been considered. Indeed, when insisting only on the compatibility with the classical Chernoff distance, there is in principle an infinitude of possiblities. Among those, three especially promising candidate expressions had been put forward by Ogawa and Hayashi [23], who studied their relations and found that there exists an increasing ordering between them. Incidentally, the second candidate coincides with (3) and thus turns out to be the correct one.

Kargin [18] gave lower and upper bounds on the optimal error exponent ξ in terms of the fidelity between the two density operators and found that Ogawa and Hayashi's third candidate (in their increasing arrangement) is a lower bound on the optimal error exponent for faithful states, i.e. it is an achievable rate. Hayashi [11] made progress regarding (3), by showing that for $s = 1/2$, $-\log \mathrm{Tr}[\rho^{1-s}\sigma^s]$ is also an achievable error exponent.

The proof of our main result consists of two parts. In the optimality part, which was first presented in [21], we show that for any test the (Bayesian) error rate $-\frac{1}{n}\log P_{e,n}$ cannot be made arbitrary large but is asymptotically bounded above by ξ_{QCB}. In the achievability part, first put forward in [1], we prove that under the Holevo-Helstrom strategy the bound is actually attained in the asymptotic limit, i.e.

$$\limsup_{n\to\infty}\left(-\frac{1}{n}\log P^*_{e,n}\right) \geq \xi_{QCB}.$$

It is the purpose of this paper to give a complete, detailed, and unified account of these results. We will present the complete proof in Sect. 3. Moreover, we give an in-depth treatment of the properties of the quantum Chernoff distance in Sect. 4. More precisely, we show that it defines a distance measure between quantum states.

Distinguishability measures between quantum states have been used in a wide variety of applications in quantum information theory. The most popular of such measures seems to be Uhlmann's fidelity [28], which happens to coincide with the quantum Chernoff distance when one of the states is pure. The trace norm distance $\|\rho-\sigma\|_1 = \mathrm{Tr}\,|\rho-\sigma|$ has a more natural operational meaning than the fidelity, but lacks monotonicity under taking tensor powers of its arguments. The problem is that one can easily find states $\rho, \sigma, \rho', \sigma'$ such that $\|\rho-\sigma\|_1 < \|\rho'-\sigma'\|_1$ but $\|\rho^{\otimes 2}-\sigma^{\otimes 2}\|_1 > \|\rho'^{\otimes 2}-\sigma'^{\otimes 2}\|_1$. This already happens in the classical setting: take the following 2-dimensional diagonal states

$$\rho = \begin{pmatrix} 1/4 & 0 \\ 0 & 3/4 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 3/4 & 0 \\ 0 & 1/4 \end{pmatrix}, \quad \rho' = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma' = \begin{pmatrix} b & 0 \\ 0 & 1-b \end{pmatrix},$$

where $1 - 1/\sqrt{2} < b < 1/2$. Then $\|\rho-\sigma\|_1 = 1 > 2b = \|\rho'-\sigma'\|_1$, while $\|\rho^{\otimes 2}-\sigma^{\otimes 2}\|_1 = 1 < 2b(2-b) = \|\rho'^{\otimes 2}-\sigma'^{\otimes 2}\|_1$. The quantum Chernoff distance characterises the exponent arising in the asymptotic behaviour of the trace norm distance, in the case of many identical copies, and therefore by construction does not suffer from this problem. As such, the quantum Chernoff distance can be considered as a kind of regularisation of the trace norm distance. For the above-mentioned states, $\xi_{QCB}(\rho,\sigma) = -\log(\sqrt{3}/2)$ (optimal $s = 1/2$) and $\xi_{QCB}(\rho',\sigma') = -\log(1-b)$ (optimal $s = 1$).

A related problem that attracted a lot of attention in the field of quantum information theory was to identify the relative entropy between two quantum states. An information-theoretical way of looking at the classical relative entropy between two probability distributions, or Kullback-Leibler distance, is that it characterises the inefficiency of compressing messages from a source p using an algorithm that is optimal for a source p' (i.e. yields the Shannon information bound for that source). Phrased differently, it quantifies the way one could cheat by telling that the given probability distribution is p while the real one is p'. By proving a quantum version of Stein's lemma [15,24], it has been shown that the quantum relative entropy, as introduced by Umegaki, has exactly the same operational meaning.

When using the relative entropy to distinguish between states, one faces the problem that it is not continuous and is asymmetric under exchange of its arguments, and therefore it does not represent a distance measure in a mathematically strict manner. Furthermore, for pure states, the quantum relative entropy is not very useful, since it is either 0 (when the two states are identical) or infinite (when they are not). In contrast, the quantum Chernoff distance seems to be much more natural in many situations.

On the other hand, (quantum) relative entropy is a crucial notion in *asymmetric* hypothesis testing. There it obtains an operational meaning as the best achievable asymptotic rate of type II errors. Its properties, which are problematic for a candidate for a distance measure, reflect the asymmetry between the null and alternative hypothesis arising from treating the type-I and type-II errors in a different way. As exemplified by the medical diagnosis case mentioned above, the type II error is the one that should be avoided at all costs. Hence, one puts a constraint $\alpha < \epsilon$ on the type I error, and minimises the β-rate. One obtains that the optimal β-rate is the relative entropy of the null hypothesis w.r.t. the alternative, independent of the constrained ϵ. The mathematical derivation of this statement goes under the name of Stein's Lemma. When the constraint consists of a lower bound on the asymptotic exponential rate of the type II error, one obtains what is called the Hoeffding bound.

Asymmetric hypothesis testing has been subject to a quantum theoretical treatment much earlier, although it is a much less natural setting for the basic state discrimination problem. The quantum generalisation of Stein's Lemma was first obtained by Hiai and Petz [15]. Its optimality part was then strengthened by Ogawa and Nagaoka in [24]. In the last few years there has been a lot of progress extending the statement of the lemma in different directions. In [4] the minimal relative entropy distance from a set of quantum states, the null hypothesis, w.r.t. a reference quantum state, the alternative, has been fixed as the best achievable asymptotic rate of the type II errors, see also [13]. This may be seen as a quantum generalisation of Sanov's theorem. In a recent paper [5] an extension of this result to the case where the hypotheses correspond to sources emitting correlated (not necessarily i.i.d.) classical or quantum data has been given. Additionally, an equivalence relation between the achievability part in (quantum) Stein's Lemma and (quantum) Sanov's Theorem has been derived.

Just a few months after the appearance of [21,1], the techniques pioneered in those two papers were used to find a quantum generalisation of the Hoeffding bound under the implicit assumption of equivalent hypotheses, i.e. for states with coinciding supports, thereby (partially) solving another long-standing open problem in quantum hypothesis testing. Just as in the case of the Chernoff distance, the Hoeffding bound contains $\sum_i p(i)^{1-s} q(i)^s$ as a sub-expression, and the quantum generalisation of the Hoeffding bound is obtained by replacing this sub-expression by $\mathrm{Tr}[\rho^{1-s}\sigma^s]$. The optimality of the bound (also called the "converse part") was proven by Nagaoka [20], while its

achievability (the "direct part") was found by Hayashi [12]. Using the same techniques, Hayashi also gave a simple proof of the achievability part of the quantum Stein's Lemma, in that same paper. In Sect. 5 we first formulate and prove an extended version of the classical Hoeffding bound, which allows nonequivalent hypotheses. Secondly, we present a complete proof of the quantum Hoeffding bound in a unified way. Moreover, we derive quantum Stein's Lemma as well as quantum Sanov's Theorem from the quantum Hoeffding bound combined with the mentioned equivalence relation proved in [5].

2. Mathematical Setting and Problem Formulation

We consider the two hypotheses H_0 (null) and H_1 (alternative) that a device prepares finite quantum systems either in the state ρ or in the state σ, respectively. Everywhere in this paper, we identify a state with a density operator, i.e. a positive trace 1 linear operator on a finite-dimensional Hilbert space $\mathcal{H}$ associated to the type of the finite quantum system in question. Since the (quantum) Chernoff distance arises naturally in a Bayesian setting, we supply the prior probabilities π_0 and π_1, which are positive quantities summing up to 1; we exclude the degenerate cases $\pi_0 = 0$ and $\pi_1 = 0$ because these are trivial.

Physically discriminating between the two hypotheses corresponds to performing a generalised (POVM) measurement on the quantum system. In analogy to the classical proceeding one accepts H_0 or H_1 according to a decision rule based on the outcome of the measurement. There is no loss of generality assuming that the POVM consists of only two elements, which we denote by $\{\mathbb{1} - \Pi, \Pi\}$, where Π may be any linear operator on $\mathcal{H}$ with $0 \leq \Pi \leq \mathbb{1}$. We will mostly make reference to this POVM by its Π element, the one corresponding to the alternative hypothesis. The type-I and type-II error probabilities α and β are the probabilities of mistaking σ for ρ, and vice-versa, and are given by

$$\begin{aligned} \alpha &:= \mathrm{Tr}[\Pi\rho], \\ \beta &:= \mathrm{Tr}[(\mathbb{1} - \Pi)\sigma]. \end{aligned}$$

The average error probability P_e is given by

$$P_e = \pi_0\alpha + \pi_1\beta = \pi_0 \,\mathrm{Tr}[\Pi\rho] + \pi_1 \,\mathrm{Tr}[(\mathbb{1} - \Pi)\sigma]. \tag{4}$$

The Bayesian distinguishability problem consists in finding the Π that minimises P_e. A special case is the symmetric one where the prior probabilities π_0, π_1 are equal.

Before we proceed, let us first introduce some basic notations. Abusing terminology, we will use the term 'positive' for 'positive semi-definite' (denoted $A \geq 0$). We employ the positive semi-definite ordering on the linear operators on $\mathcal{H}$ throughout, i.e. $A \geq B$ iff $A - B \geq 0$. For each linear operator $A \in \mathcal{B}(\mathcal{H})$ the *absolute value* $|A|$ is defined as $|A| := (A^*A)^{1/2}$. The Jordan decomposition of a self-adjoint operator A is given by $A = A_+ - A_-$, where

$$A_+ := (|A| + A)/2, \qquad A_- := (|A| - A)/2 \tag{5}$$

are the *positive part* and *negative part* of A, respectively. Both parts are positive by definition, and $A_+A_- = 0$.

There is a very useful variational characterisation of the trace of the positive part of a self-adjoint operator A:

$$\mathrm{Tr}[A_+] = \max_X \{\mathrm{Tr}[AX] : 0 \leq X \leq \mathbb{1}\}. \tag{6}$$

In other words, the maximum is taken over all positive contractive operators. Since the extremal points of the set of positive contractive operators are exactly the orthogonal projectors, we also have

$$\mathrm{Tr}[A_+] = \max_P \{\mathrm{Tr}[AP] : P \geq 0,\, P = P^2\}. \tag{7}$$

The maximiser on the right-hand side is the orthogonal projector onto the range of A_+.

We can now easily prove the quantum version of the Neyman-Pearson Lemma.

Lemma 1 (Quantum Neyman-Pearson). *Let ρ and σ be density operators associated to hypotheses H_0 and H_1, respectively. Let T be a fixed positive number. Consider the POVM with elements $\{\mathbb{1} - \Pi^*, \Pi^*\}$, where Π^* is the projector onto the range of $(T\sigma - \rho)_+$, and let $\alpha^* = \mathrm{Tr}[\Pi^*\rho]$ and $\beta^* = \mathrm{Tr}[(\mathbb{1} - \Pi^*)\sigma]$ be the associated errors. For any other POVM $\{\mathbb{1} - \Pi, \Pi\}$, with associated errors $\alpha = \mathrm{Tr}[\Pi\rho]$ and $\beta = \mathrm{Tr}[(\mathbb{1} - \Pi)\sigma]$, we have*

$$\alpha + T\beta \geq \alpha^* + T\beta^* = T - \mathrm{Tr}[(T\sigma - \rho)_+].$$

Thus if $\alpha \leq \alpha^*$, then $\beta \geq \beta^*$.

Proof. By formulae (6) and (7), for all $0 \leq \Pi \leq \mathbb{1}$ we have $\mathrm{Tr}[\Pi(T\sigma - \rho)] \leq \mathrm{Tr}(T\sigma - \rho)_+ = \mathrm{Tr}[\Pi^*(T\sigma - \rho)]$. In terms of $\alpha, \beta, \alpha^*, \beta^*$, this reads $T(1-\beta) - \alpha \leq T(1-\beta^*) - \alpha^*$, which is equivalent to the statement of the lemma. □

The upshot of this lemma is that the POVM $\{\mathbb{1} - \Pi^*, \Pi^*\}$, where Π^* is the projector on the range of $(T\sigma - \rho)_+$, is the optimal one when the goal is to minimise the quantity $\alpha + T\beta$. In symmetric hypothesis testing the positive number T is taken to be the ratio π_1/π_0 of the prior probabilities.

We emphasize that we have started with the assumption that the physical systems in question are finite systems with an algebra of observables $\mathcal{B}(\mathcal{H})$, i.e. the algebra of linear operators on a finite-dimensional Hilbert space $\mathcal{H}$. This is a purely quantum situation. In the general setting (of statistical mechanics) one associates to a finite physical system, classical or quantum, a finite-dimensional $*$-algebra $\mathcal{A}$. Such an algebra has a block representation $\bigoplus_{i=1}^k \mathcal{B}(\mathcal{H}_i)$, i.e. it is a subalgebra of $\mathcal{B}(\mathcal{H})$, where $\mathcal{H} := \bigoplus_{i=1}^k \mathcal{H}_i$. If the Hilbert spaces $\mathcal{H}_i$ are one-dimensional for all $i = 1, \ldots, k$, then $\mathcal{A}$ is *-isomorphic to the commutative algebra of diagonal $(k \times k)$-matrices. This covers the classical case. Now, in view of Lemma 1 it becomes clear that in the context of hypothesis testing there is no restriction assuming that the algebra of observables of the systems in question is $\mathcal{B}(\mathcal{H})$; indeed, the optimally discriminating projectors Π^* are always in the $*$-subalgebra generated by the two involved density operators ρ and σ. This implies that they are automatically elements of the algebra $\mathcal{A}$ characterising the physical systems. In particular, if the hypotheses correspond to mutually commuting density operators then the problem reduces to a classical one in the sense that the best test Π^* commutes with the density operators as well. Hence it coincides with the classical ML-test, although there are many more possible tests in $\mathcal{B}(\mathcal{H})$ than in the commutative subalgebra of observables of the classical subsystem.

The basic problem we focus on in this paper is to identify how the error probability P_e behaves in the asymptotic limit, i.e. when one has to discriminate between the hypotheses H_0 and H_1 on the basis of a large number n of copies of the quantum systems. This means that we have to distinguish between the n-fold tensor product density operators $\rho^{\otimes n}$ and $\sigma^{\otimes n}$ by means of POVMs $\{\mathbb{1} - \Pi_n, \Pi_n\}$ on $\mathcal{H}^{\otimes n}$.

We define the rate limit s_R for any positive sequence (s_n) as

$$s_R := \lim_{n\to\infty}\left(-\frac{1}{n}\log s_n\right),$$

if the limit exists. Otherwise we have to deal with the lower and upper rate limits $\underline{s}_R$ and $\overline{s}_R$, which are the limit inferior and the limit superior of the sequence $(-\frac{1}{n}\log s_n)$, respectively. In particular, we define the *type-I error rate limit* and the *type-II error rate limit* for a sequence $\Pi := (\Pi_n)$ of quantum measurements (where, as mentioned, each orthogonal projection Π_n corresponds to the alternative hypothesis) as

$$\alpha_R(\Pi) := \lim_{n\to\infty}\left(-\frac{1}{n}\log\alpha_n\right) = \lim_{n\to\infty}\left(-\frac{1}{n}\log \mathrm{Tr}[\rho^{\otimes n}\Pi_n]\right), \tag{8}$$

$$\beta_R(\Pi) := \lim_{n\to\infty}\left(-\frac{1}{n}\log\beta_n\right) = \lim_{n\to\infty}\left(-\frac{1}{n}\log \mathrm{Tr}[\sigma^{\otimes n}(\mathbb{1}-\Pi_n)]\right), \tag{9}$$

if the limits exist. Otherwise we consider the limit inferior and the limit superior $\underline{\alpha}_R(\Pi)$ and $\overline{\alpha}_R(\Pi)$, respectively. Similar definitions hold in the classical case.

3. Bayesian Quantum Hypothesis Testing: Quantum Chernoff Bound

In this section we consider the Bayesian distinguishability problem. This means the goal is to minimise the average error probability P_e, which is defined in (4) and can be rewritten as $P_e = \pi_1 - \mathrm{Tr}[\Pi(\pi_1\sigma - \pi_0\rho)]$. By the Neyman-Pearson Lemma, the optimal test is given by the projector Π^* onto the range of $(\pi_1\sigma - \pi_0\rho)_+$, and the obtained minimal error probability is given by

$$\begin{aligned} P_e^* &= \pi_1 - \mathrm{Tr}[(\pi_1\sigma - \pi_0\rho)_+] \\ &= \pi_1 - (\pi_1 - \pi_0)/2 - \mathrm{Tr}[|\pi_1\sigma - \pi_0\rho|/2] \\ &= \frac{1}{2}\left(1 - \|\pi_1\sigma - \pi_0\rho\|_1\right), \end{aligned}$$

where $\|A\|_1 = \mathrm{Tr}\,|A|$ is the trace norm. We will call Π^* the Holevo-Helstrom projector.

Next, note that the optimal test to discriminate ρ and σ in the case of n copies enforces the use of joint measurements. However, the particular permutational symmetry of n-copy states guarantees that the optimal collective measurement can be implemented efficiently (with a polynomial-size circuit) [2], and hence that the minimum probability of error is achievable with a reasonable amount of resources.

We need to consider the quantity

$$P_{e,n}^* := (1 - \|\pi_1\sigma^{\otimes n} - \pi_0\rho^{\otimes n}\|_1)/2. \tag{10}$$

It turns out that $P_{e,n}^*$ vanishes exponentially fast as n tends to infinity. The theorem below provides the asymptotic value of the exponent $-\frac{1}{n}\log P_{e,n}^*$, i.e. the rate limit of $P_{e,n}^*$, which turns out to be given by the *quantum Chernoff distance*. This is our main result.

Theorem 1. *For any two states ρ and σ on a finite-dimensional Hilbert space, occurring with prior probabilities π_0 and π_1, respectively, the rate limit of $P_{e,n}^*$, as defined by (10), exists and is equal to the quantum Chernoff distance ξ_{QCB},*

$$\lim_{n\to\infty}\left(-\frac{1}{n}\log P_{e,n}^*\right) = \xi_{QCB} := -\log\left(\inf_{0\le s\le 1}\mathrm{Tr}\left(\rho^{1-s}\sigma^s\right)\right). \tag{11}$$

Because the product of two positive operators always has positive spectrum, the quantity $\mathrm{Tr}[\rho^{1-s}\sigma^s]$ is well defined (in the mathematical sense) and guaranteed to be real and non-negative for every $0 \le s \le 1$. As should be, the expression for ξ_{QCB} reduces to the classical Chernoff distance ξ_{CB} defined by (2) when ρ and σ commute.

3.1. Proof of Theorem 1: Optimality Part. In this section, we will show that the best discrimination is specified by the quantum Chernoff distance; that is, ξ_{QCB} is an upper bound on

$$\limsup_{n\to\infty}\left(-\frac{1}{n}\log P_{e,n}\right)$$

for any sequence of tests (Π_n) and $P_{e,n} := \pi_1 - \mathrm{Tr}\,[\pi_1\sigma^{\otimes n} - \pi_0\rho^{\otimes n}]$.

The proof, which first appeared in [21], is essentially based on relating the quantum to the classical case by using a special mapping from a pair of $d \times d$ density matrices (ρ, σ) to a pair of probability distributions (p, q) on a set of cardinality d^2.

Let the spectral decompositions of ρ and σ be given by

$$\rho = \sum_{i=1}^{d} \lambda_i |x_i\rangle\langle x_i|, \quad \sigma = \sum_{j=1}^{d} \mu_j |y_j\rangle\langle y_j|,$$

where $(|x_i\rangle)$ and $(|y_j\rangle)$ are two orthonormal bases of eigenvectors and (λ_i) and (μ_j) are the corresponding sets of eigenvalues of ρ and σ, respectively. Then we map these density operators to the d^2-dimensional vectors

$$p_{i,j} = \lambda_i |\langle x_i|y_j\rangle|^2, \quad q_{i,j} = \mu_j |\langle x_i|y_j\rangle|^2, \tag{12}$$

with $1 \le i, j \le d$. This mapping preserves a number of important properties:

Proposition 1. *With $p_{i,j}$ and $q_{i,j}$ as defined in (12), and $s \in \mathbb{R}$,*

$$\mathrm{Tr}[\rho^{1-s}\sigma^s] = \sum_{i,j} p_{i,j}^{1-s} q_{i,j}^{s}, \tag{13}$$

$$S(\rho\|\sigma) = H(p\|q). \tag{14}$$

Here, $S(\rho\|\sigma)$ is the quantum relative entropy defined as

$$S(\rho\|\sigma) := \begin{cases} \mathrm{Tr}[\rho(\log\rho - \log\sigma)], & \text{if } \mathrm{Supp}\,\rho \le \mathrm{Supp}\,\sigma \\ +\infty, & \text{otherwise,} \end{cases} \tag{15}$$

where $\mathrm{Supp}\,\rho$ denotes the support projection of an operator ρ, and $H(p\|q)$ is the classical relative entropy, or Kullback-Leibler distance,

$$H(p\|q) := \begin{cases} \sum_{i,j} p_{i,j}(\log p_{i,j} - \log q_{i,j}), & \text{if } p \ll q \\ +\infty, & \text{otherwise.} \end{cases} \tag{16}$$

Proof. The proof proceeds by direct calculation. For example:

$$\begin{aligned}\mathrm{Tr}[\rho^{1-s}\sigma^s] &= \sum_{i,j} \lambda_i^{1-s}\mu_j^s |\langle x_i|y_j\rangle|^2 \\ &= \sum_{i,j} \lambda_i^{1-s}\mu_j^s |\langle x_i|y_j\rangle|^{2(1-s)} |\langle x_i|y_j\rangle|^{2s} \\ &= \sum_{i,j} p_{i,j}^{1-s} q_{i,j}^s.\end{aligned}$$

□

A direct consequence of identity (13) is that p and q are normalised if ρ and σ are. Furthermore, tensor powers are preserved by the mapping; that is, if ρ and σ are mapped to p and q, then $\rho^{\otimes n}$ is mapped to $p^{\otimes n}$ and $\sigma^{\otimes n}$ to $q^{\otimes n}$.

Now define the classical and quantum average (Bayesian) error probabilities $P_{e,c}$ and $P_{e,q}$ as

$$P_{e,c}(\phi, p, \pi_0, q, \pi_1) := \sum_i [\pi_0 \phi(i) p_i + \pi_1 (1-\phi(i)) q_i], \tag{17}$$

$$P_{e,q}(\Pi, \rho, \pi_0, \sigma, \pi_1) := \mathrm{Tr}[\pi_0 \Pi \rho + \pi_1 (\mathbb{1} - \Pi)\sigma], \tag{18}$$

where p, q are probability distributions, ρ, σ are density matrices, and π_0, π_1 are the respective prior probabilities of the two hypotheses. Furthermore, ϕ is a non-negative test function $0 \le \phi \le 1$, and Π is a positive semi-definite contraction, $0 \le \Pi \le \mathbb{1}$, so that $\{\mathbb{1} - \Pi, \Pi\}$ forms a POVM.

The main property of the mapping that allows to establish optimality of the quantum Chernoff distance is presented in the following proposition.

Proposition 2. *For all orthogonal projectors Π and all positive scalars η_0, η_1 (not necessarily adding up to 1), and for p and q associated to ρ and σ by the mapping* (12)*,*

$$P_{e,q}(\Pi, \rho, \eta_0, \sigma, \eta_1) \ge \frac{1}{2} \inf_{\phi} P_{e,c}(\phi, p, \eta_0, q, \eta_1),$$

where the infimum is taken over all test functions $0 \le \phi \le 1$.

Note that we have replaced the priors by general positive scalars; this will be useful later on, in proving the optimality of the Hoeffding bound.

Proof. Since Π is a projector, one has $\Pi = \Pi\Pi = \sum_j \Pi|y_j\rangle\langle y_j|\Pi$, where the second equality is obtained by inserting a resolution of the identity $\mathbb{1} = \sum_j |y_j\rangle\langle y_j|$. Likewise, $\mathbb{1} - \Pi$ is also a projector, and using another resolution of the identity, $\mathbb{1} = \sum_i |x_i\rangle\langle x_i|$, we similarly get $\mathbb{1} - \Pi = \sum_i (\mathbb{1} - \Pi)|x_i\rangle\langle x_i|(\mathbb{1} - \Pi)$. This yields

$$\begin{aligned}\mathrm{Tr}[\Pi\rho] &= \sum_i \lambda_i \,\mathrm{Tr}[\Pi|x_i\rangle\langle x_i|] \\ &= \sum_{i,j} \lambda_i \,\mathrm{Tr}[\Pi|y_j\rangle\langle y_j|\Pi|x_i\rangle\langle x_i|] \\ &= \sum_{i,j} \lambda_i |\langle x_i|\Pi|y_j\rangle|^2,\end{aligned}$$

and, similarly,

$$\mathrm{Tr}[(\mathbb{1} - \Pi)\sigma] = \sum_{i,j} \mu_j |\langle x_i|\mathbb{1} - \Pi|y_j\rangle|^2.$$

Then the quantum error probability is given by

$$\begin{aligned} P_{e,q} &= \eta_0 \,\mathrm{Tr}[\Pi\rho] + \eta_1 \,\mathrm{Tr}[(\mathbb{1} - \Pi)\sigma] \\ &= \sum_{i,j} \eta_0 \lambda_i |\langle x_i|\Pi|y_j\rangle|^2 + \eta_1 \mu_j |\langle x_i|\mathbb{1} - \Pi|y_j\rangle|^2. \end{aligned}$$

The infimum of the classical error probability $P_{e,c}$ is obtained when the test function ϕ equals the indicator function $\phi = \chi_{\{\eta_1 q > \eta_0 p\}}$ (corresponding to the maximum likelihood decision rule); hence, the value of this infimum is given by

$$\begin{aligned} \inf_{\phi} P_{e,c} &= \sum_{i,j} \min(\eta_0 p_{i,j}, \eta_1 q_{i,j}) \\ &= \sum_{i,j} \min(\eta_0 \lambda_i, \eta_1 \mu_j) |\langle x_i|y_j\rangle|^2. \end{aligned}$$

For a fixed choice of i, j, let a be the 2×2 non-negative diagonal matrix

$$a := \begin{pmatrix} \eta_0 \lambda_i & 0 \\ 0 & \eta_1 \mu_j \end{pmatrix},$$

and let b be the 2-vector

$$b := (\langle x_i|\Pi|y_j\rangle, \langle x_i|\mathbb{1} - \Pi|y_j\rangle).$$

The i, j-term in the sum for $P_{e,q}$ can then be written as the inner product $\langle b|a|b\rangle$. Similarly, the factor $|\langle x_i|y_j\rangle|^2$ occurring in the i, j-term in the sum for $P_{e,c}$ can then be written as $|b_1 + b_2|^2$.

Now we note that $\langle b|b\rangle = \|b\|_2^2$, while $|b_1 + b_2|^2 \leq \|b\|_1^2$. For d-dimensional vectors, the inequality $\|b\|_2 \geq \|b\|_1/\sqrt{d}$ holds; in our case, $d = 2$. Together with the inequality $a \geq \min(\eta_0 \lambda_i, \eta_1 \mu_j)\mathbb{1}_2$ this yields

$$\langle b|a|b\rangle \geq \min(\eta_0 \lambda_i, \eta_1 \mu_j)\langle b|b\rangle \geq \min(\eta_0 \lambda_i, \eta_1 \mu_j)\frac{1}{2}|b_1 + b_2|^2. \tag{19}$$

Therefore, we obtain, for any i, j,

$$\eta_0 \lambda_i |\langle x_i|\Pi|y_j\rangle|^2 + \eta_1 \mu_j |\langle x_i|\mathbb{1} - \Pi|y_j\rangle|^2 \geq \frac{1}{2} \min(\eta_0 \lambda_i, \eta_1 \mu_j) |\langle x_i|y_j\rangle|^2.$$

As this holds for any i, j, it holds for the sum over i, j, so that a lower bound for the quantum error probability is given by

$$P_{e,q} \geq \frac{1}{2} \sum_{i,j} \min(\eta_0 p_{i,j}, \eta_1 q_{i,j}) = \frac{1}{2} \inf_{\phi} P_{e,c},$$

which proves the proposition. □

Using these properties of the mapping, the proof of optimality of the quantum Chernoff bound is easy.

Proof of optimality of the quantum Chernoff bound. Let hypotheses H_0 and H_1, with priors π_0 and π_1, correspond to the product states $\rho^{\otimes n}$ and $\sigma^{\otimes n}$. Using the mapping (12), these states are mapped to the probability distributions $p^{\otimes n}$ and $q^{\otimes n}$. By Proposition 2, the quantum error probability is bounded from below as

$$P_{e,q}(\Pi_n, \rho^{\otimes n}, \pi_0, \sigma^{\otimes n}, \pi_1) \geq \frac{1}{2} \inf_{\phi_n} P_{e,c}(\phi_n, p^{\otimes n}, \pi_0, q^{\otimes n}, \pi_1). \tag{20}$$

By the classical Chernoff bound, the rate limit of the right-hand side is given by

$$-\log \inf_{0\leq s\leq 1} \sum_{i,j} p_{i,j}^{1-s} q_{i,j}^{s}$$

(provided the priors π_0, π_1 are non-zero) and this is, therefore, an upper bound on the rate limit of the optimal quantum error probability. By Proposition 1 the latter expression is equal to $-\log \inf_{0\leq s\leq 1} \mathrm{Tr}[\rho^{1-s}\sigma^s]$, which is what we set out to prove. □

In a similar way one can prove the converse part of the quantum Hoeffding bound by relating it to the classical problem in the sense of (12), as already noted by Nagaoka in [20]. This will be discussed in Sect. 5.4.

3.2. Proof of Theorem 1: Achievability Part. In this section, we prove the achievability of the quantum Chernoff bound, which is the statement that the error rate limit $\lim_{n\to\infty} \left(-\frac{1}{n} \log P_{e,n}^*\right)$ is not only bounded above by, but is actually equal to the quantum Chernoff distance ξ_{QCB}. This can directly be inferred from the following matrix inequality, which first made its appearance in [1]:

Theorem 2. *Let a and b be positive semi-definite operators, then for all $0 \leq s \leq 1$,*

$$\mathrm{Tr}[a^s b^{1-s}] \geq \mathrm{Tr}[a + b - |a - b|]/2. \tag{21}$$

Note that inequality (21) is also interesting from a purely matrix analytic point of view, as it relates the trace norm to a multiplicative quantity in a highly nontrivial and very useful way.

If we specialise this theorem to states, $a = \sigma$ and $b = \rho$, with $\mathrm{Tr}\,\rho = \mathrm{Tr}\,\sigma = 1$, we obtain

$$Q_s + T \geq 1, \qquad 0 \leq s \leq 1,$$

where $Q_s := Q_s(\rho, \sigma) := \mathrm{Tr}[\rho^{1-s}\sigma^s]$ and $T := T(\rho, \sigma) := \|\rho - \sigma\|_1/2$ is the trace norm distance.

Remark 1. Inequality (21) can be written in the form $\langle b^{1/2} | f_s(\Delta_{a,b}) b^{1/2}\rangle \leq \|a - b\|_1$, where $\Delta_{a,b}$ is the relative modular operator acting on the matrix space endowed with the Hilbert-Schmidt inner product, and f_s is the operator convex function $f_s(t) := 1+t-2t^s$, see [25]. The expression on the left-hand side is a quasi-entropy. This also implies some of the properties of Q_s. For $s = 1/2$ inequality (21) becomes $\|a^{1/2} - b^{1/2}\|^2 = \mathrm{Tr}[(a^{1/2} - b^{1/2})^2] \leq \|a - b\|_1$, which is known to hold also in infinite dimensions.

Remark 2. The inequality $Q_s + T \geq 1$ is *strongly sharp*, which means that for any allowed value of T one can find ρ and σ that achieve equality. Indeed, take the commuting density operators $\rho = |0\rangle\langle 0|$ and $\sigma = (1-T)|0\rangle\langle 0| + T|1\rangle\langle 1|$, then their trace norm distance is T, and $Q_s = 1 - T$.

Proof of achievability of the quantum Chernoff bound from Theorem 2. We will prove the inequality

$$\liminf_{n\to\infty} \left(-\frac{1}{n} \log P^*_{e,n}\right) \geq \xi_{QCB}. \tag{22}$$

Put $a = \pi_1 \sigma^{\otimes n}$ and $b = \pi_0 \rho^{\otimes n}$, so that the right-hand side of (21) turns into

$$(1 - \|\pi_1 \sigma^{\otimes n} - \pi_0 \rho^{\otimes n}\|_1)/2 = P^*_{e,n}.$$

The logarithm of the left-hand side of inequality (21) simplifies to

$$\log(\pi_0^{1-s} \pi_1^s) + n \log \left(\mathrm{Tr}[\rho^{1-s} \sigma^s]\right).$$

Upon dividing by n and taking the limit $n \to \infty$, we obtain $\log Q_s$, independently of the priors π_0, π_1 (as long as the priors are not degenerate, i.e. are different from 0 or 1). Then (22) follows from the fact that the inequality

$$\liminf_{n\to\infty} \left(-\frac{1}{n} \log P^*_{e,n}\right) \geq -\log Q_s$$

holds for all $s \in [0, 1]$ and we can replace the right-hand side by ξ_{QCB}. □

Proof of Theorem 2. The left-hand and right-hand sides of (21) look very disparate, but they can nevertheless be brought closer together by expressing $a + b - |a - b|$ in terms of the positive part $(a - b)_+$. The inequality (21) is indeed equivalent to

$$\begin{aligned} \mathrm{Tr}[a - a^s b^{1-s}] &\leq \mathrm{Tr}[a - (a + b - |a - b|)/2] \\ &= \mathrm{Tr}[(a - b + |a - b|)/2] \\ &= \mathrm{Tr}[(a - b)_+]. \end{aligned} \tag{23}$$

At this point we mention another equivalent formulation of this inequality, which will be used later in the proof of the achievability of the quantum Hoeffding bound. With Π the projector on the range of $(a - b)_+$, we can write:

$$\mathrm{Tr}[a^s b^{1-s}] \geq \mathrm{Tr}[\Pi b + (\mathbb{1} - \Pi) a]. \tag{24}$$

What we do next is strengthening the inequality (23) by replacing its left-hand side by an upper bound, and its right-hand side by a lower bound. Since, for any self-adjoint operator H, we have $H \leq H_+$, we can write

$$\begin{aligned} \mathrm{Tr}[a - a^s b^{1-s}] = \mathrm{Tr}[a^s (a^{1-s} - b^{1-s})] &\leq \mathrm{Tr}[a^s (a^{1-s} - b^{1-s})_+] \\ &= \mathrm{Tr}[a^s \Pi^{(s)} (a^{1-s} - b^{1-s})] \\ &= \mathrm{Tr}[\Pi^{(s)} (a - b^{1-s} a^s)], \end{aligned}$$

where $\Pi^{(s)}$ is the projector on the range of $(a^{1-s} - b^{1-s})_+$. Likewise,

$$\mathrm{Tr}[\Pi^{(s)} (a - b)] \leq \mathrm{Tr}[(a - b)_+],$$

because $\mathrm{Tr}[(a-b)_+]$ is the maximum of $\mathrm{Tr}[\Pi(a-b)]$ over all orthogonal projections Π. Inequality (21) would thus follow if, for that particular $\Pi^{(s)}$,

$$\mathrm{Tr}[\Pi^{(s)}(a-b^{1-s}a^s)] \leq \mathrm{Tr}\,\Pi^{(s)}(a-b).$$

The benefit of this reduction is obvious, as after simplification we get the much nicer statement

$$\mathrm{Tr}[\Pi^{(s)}b^{1-s}(a^s-b^s)] \geq 0.$$

Equally obvious, though, is the risk of this strengthening; it could very well be a false statement. Nevertheless, we show its correctness below.

It is interesting to note the meaning here of this strengthening in the context of the optimal hypothesis test, i.e. when $a=\sigma^{\otimes n}$ and $b=\rho^{\otimes n}$. While the Holevo-Helstrom projectors Π_n^* are optimal for every finite value of n, we can use other projectors that are suboptimal but reach optimality in the asymptotic sense. Here we are indeed using $\Pi^{(s^*)}$, the projector on the range of $(a^{1-s^*}-b^{1-s^*})_+$, where s^* is the minimiser of $\mathrm{Tr}[\rho^{1-s}\sigma^s]$ over $[0,1]$, if it exists. Otherwise we have to use the Holevo-Helstrom projector.

In the next few steps we will further reduce the statement by reformulating the matrix powers in terms of simpler expressions. One can immediately absorb one of them into a and b via appropriate substitutions. As we certainly don't want a power appearing in the definition of the projector $\Pi^{(s)}$, we are led to apply the substitutions

$$A=a^{1-s}, \quad B=b^{1-s}, \quad t=s/(1-s).$$

This yields a value of t between 0 and 1 only when $0\leq s\leq 1/2$. However, this is no restriction since the case $1/2\leq s\leq 1$ can be treated in a completely similar way after applying an additional substitution $s\to 1-s$.

Inequality (21) is thus implied by the lemma below, which ends the proof of Theorem 2. □

Lemma 2. *For matrices $A, B\geq 0$, a scalar $0\leq t\leq 1$, and denoting by P the projector on the range of $(A-B)_+$, the following inequality holds:*

$$\mathrm{Tr}[PB(A^t-B^t)] \geq 0. \tag{25}$$

Proof. To deal with the t^{th} matrix power, we use an integral representation (see, for example [3] (V.56)). For scalars $a\geq 0$ and $0\leq t\leq 1$,

$$a^t = \frac{\sin(t\pi)}{\pi}\int_0^{+\infty} dx\; x^{t-1}\;\frac{a}{a+x}.$$

For other values of t this integral does not converge. This integral can be extended to positive operators in the usual way:

$$A^t = \frac{\sin(t\pi)}{\pi}\int_0^{+\infty} dx\; x^{t-1}\; A(A+x\mathbb{1})^{-1}.$$

To deal with non-invertible A (arising when the states ρ and σ are not faithful), we define $\lim_{x\to 0} A(A+x\mathbb{1})^{-1}=\mathbb{1}$.

The potential benefit of this integral representation is that statements about the integral might follow from statements about the integrand, which is a simpler quantity.

Applying the integral representation to A^t and B^t, we get

$$\mathrm{Tr}[PB(A^t - B^t)] = \frac{\sin(t\pi)}{\pi} \int_0^{+\infty} dx\ x^{t-1}\, \mathrm{Tr}[PB(A(A+x)^{-1} - B(B+x)^{-1})].$$

If the integrand is positive for all $x > 0$ (it is zero for $x = 0$), then the whole integral is positive. The lemma follows if indeed we have

$$\mathrm{Tr}[PB(A(A+x)^{-1} - B(B+x)^{-1})] \geq 0.$$

As a further reduction, we note that a difference can be expressed as an integral of a derivative:

$$f(a) - f(b) = f(b + (a-b)) - f(b) = \int_0^1 dt\ \frac{d}{dt} f(b + (a-b)t).$$

Here, we will apply this to the expression $A(A+x)^{-1} - B(B+x)^{-1}$. Let $\Delta = A - B$. Then

$$A(A+x)^{-1} - B(B+x)^{-1} = \int_0^1 dt\ \frac{d}{dt}(B + t\Delta)(B + t\Delta + x)^{-1}.$$

The potential benefit is again that the required statement might follow from a statement about the integrand, which is a simpler quantity provided one is able to calculate the derivative explicitly. In this case we are not dealing with a stronger statement, because the statement has to hold for the derivative anyway (when A is close to B).

In the present case, we can indeed calculate the derivative:

$$\frac{d}{dt}(B + t\Delta)(B + t\Delta + x)^{-1} = x\ (B + t\Delta + x)^{-1}\ \Delta\ (B + t\Delta + x)^{-1}.$$

Therefore,

$$\begin{aligned}&\mathrm{Tr}[PB(A(A+x)^{-1} - B(B+x)^{-1})] \\ &\quad = x \int_0^1 dt\ \ \mathrm{Tr}[PB(B + t\Delta + x)^{-1}\Delta(B + t\Delta + x)^{-1}].\end{aligned}$$

Again, if the integrand is positive for $0 \leq t \leq 1$, the whole integral is positive. Absorbing t in Δ we need to show, with P the projector on Δ_+:

$$\mathrm{Tr}[PB\,V\,\Delta\,V] \geq 0, \qquad \text{where } V := (B + \Delta + x)^{-1} \geq 0.$$

After all these reductions, the statement is now in sufficiently simple form to allow the final attack. Since $B = V^{-1} - x - \Delta$, we have $BV\Delta V = \Delta(V - V\Delta V) - xV\Delta V$. Positivity of B implies $VBV = V - V\Delta V - xV^2 \geq 0$, thus $V - V\Delta V \geq xV^2$. Furthermore, since $P\Delta = \Delta_+ \geq 0$,

$$\begin{aligned}\mathrm{Tr}[PBV\Delta V] &= \mathrm{Tr}[P(\Delta(V - V\Delta V) - xV\Delta V)] \\ &= \mathrm{Tr}[\Delta_+(V - V\Delta V)] - x\,\mathrm{Tr}[PV\Delta V] \\ &\geq x(\mathrm{Tr}[\Delta_+ V^2] - \mathrm{Tr}[PV\Delta V]).\end{aligned}$$

Because $\mathbb{1} \geq P \geq 0$, $\Delta_+ \geq 0$, and $\Delta_+ \geq \Delta$,

$$\mathrm{Tr}[\Delta_+ V^2] = \mathrm{Tr}[V\Delta_+ V] \geq \mathrm{Tr}[P(V\Delta_+ V)] \geq \mathrm{Tr}[P(V\Delta V)].$$

The conclusion is that, indeed, $\mathrm{Tr}[PBV\Delta V] \geq 0$, which proves the lemma. □

4. Properties of the Quantum Chernoff Distance

In this section, we study the non-logarithmic variety Q of the quantum Chernoff distance ξ_{QCB}, i.e.

$$Q(\rho,\sigma) := \inf_{0\le s\le 1} \mathrm{Tr}[\rho^{1-s}\sigma^s], \tag{26}$$

where ρ, σ are density operators on a fixed finite-dimensional Hilbert space $\mathcal{H}$. All properties of $\xi_{QCB} = -\log Q$ can readily be derived from Q. It will turn out that ξ_{QCB} is not a metric, since it violates the triangle inequality, but it has a lot of properties required of a distance measure on the set of density operators.

4.1. Relation to Fidelity and Trace Distance. The Uhlmann fidelity F between two states is defined as

$$F(\rho,\sigma) := \|\rho^{1/2}\sigma^{1/2}\|_1 = \mathrm{Tr}[(\rho^{1/2}\sigma\rho^{1/2})^{1/2}]. \tag{27}$$

Here, the latter formula is best known, but the first one is easier and makes the symmetry under interchanging arguments readily apparent. The Uhlmann fidelity can be regarded as the quantum generalisation of the so-called Hellinger affinity [29] defined as $B(p_0, p_1) := \sum_i \sqrt{p_0(i)p_1(i)}$, where p_0 and p_1 are classical distributions. It is an upper bound on Q, which can be shown as follows. By definition, for any fixed value of $s \in [0, 1]$, $Q_s = \mathrm{Tr}[\rho^{1-s}\sigma^s]$ is an upper bound on Q. In particular, this is true for $s = 1/2$. Furthermore, by replacing the trace with the trace norm $\|\cdot\|_1$, we get an even higher upper bound. Indeed,

$$Q \le \mathrm{Tr}[\rho^{1/2}\sigma^{1/2}] = \|\rho^{1/4}\sigma^{1/2}\rho^{1/4}\|_1 \le \|\rho^{1/2}\sigma^{1/2}\|_1 = F. \tag{28}$$

In the last inequality we have used the fact ([3], Prop. IX.1.1) that for any unitarily invariant norm $|||AB||| \le |||BA|||$ if AB is normal. In particular, consider the trace norm, with $A = \rho^{1/4}\sigma^{1/2}$ and $B = \rho^{1/4}$.

For a pair of density operators the trace distance T is defined by

$$T(\rho,\sigma) := \frac{1}{2}\|\rho - \sigma\|_1.$$

Fuchs and van de Graaf [10] proved the following relation between F and T:

$$(1-F)^2 \le T^2 \le 1 - F^2. \tag{29}$$

Combining this with inequality (28) yields the upper bound

$$Q^2 + T^2 \le 1. \tag{30}$$

Recall the relation $1 - T \le Q$, following from Theorem 2. Then combining everything yields the chain of inequalities

$$1 - \sqrt{1-F^2} \le 1 - T \le Q \le F \le \sqrt{1-T^2}. \tag{31}$$

There is a sharper lower bound on Q in terms of F, namely

$$F^2 \le Q. \tag{32}$$

This bound is strongly sharp, as it becomes an equality when one of the states is pure [18]. Indeed, for $\rho = |\psi\rangle\langle\psi|$, the minimum of the expression $\mathrm{Tr}[\rho^{1-s}\sigma^s]$ is obtained for $s = 1$ and reduces to $\langle\psi|\sigma|\psi\rangle$, while F is given by the square root of this expression.

We prove (32) in Appendix A, where we also give an alternative proof of the upper bound $Q \le \sqrt{1-T^2}$. Both proofs go through in countably infinite dimensions.

4.2. Range of Q. The maximum value Q can attain is 1, and this happens if and only if $\rho = \sigma$. This follows, for example, from the upper bound $Q^2 + T^2 \leq 1$. The minimal value is 0, and this is only attained for pairs of orthogonal states, i.e. states such that $\mathrm{Tr}\,\rho\sigma = 0$. Consequently the range of the Chernoff distance is $[0, \infty]$ and the infinite value is attained on orthogonal states; this has to be contrasted with the relative entropy, where infinite values are obtained whenever the states have a different support.

4.3. Triangle inequality. As already mentioned, on the set of pure states we have the identity $Q = F^2$. The Uhlmann fidelity F does not obey the triangle inequality; however it can be transformed into a metric by going over to arccos F, while the Chernoff distance on pairs of pure states is equal to $\xi_{QCB} = -\log Q = -2\log F$.

When considering the triangle inequality for ξ_{QCB}, one should note first that in the classical case, the classical expression ξ_{CB} should be expected to behave like a *squared* metric, similarly to the relative entropy or Kullback-Leibler distance. Indeed consider two laws from the normal shift family $N(\mu, 1)$, $\mu \in \mathbb{R}$; then it is easy to see that $\xi_{CB} = (\mu_1 - \mu_2)^2/8$. Thus ξ_{CB} defines a squared metric on the normal shift family, which will not satisfy the triangle inequality due to the square, but $\sqrt{\xi_{CB}}$ will. However $\sqrt{\xi_{CB}}$ does not satisfy the triangle inequality in the general case. To see this, let $Be(\varepsilon)$ be the Bernoulli law with parameter $\varepsilon \in [0, 1]$. Some computations show that $\xi_{CB}(Be(1/2), Be(\varepsilon)) \to \log 2$ and $\xi_{CB}(Be(\varepsilon), Be(1-\varepsilon)) \to \infty$ as $\varepsilon \to 0$. As a consequence we have, for ε small enough,

$$\xi_{CB}^{1/2}(Be(\varepsilon), Be(1-\varepsilon)) > \xi_{CB}^{1/2}(Be(\varepsilon), Be(1/2)) + \xi_{CB}^{1/2}(Be(1/2), Be(1-\varepsilon))$$

contradicting the triangle inequality.

4.4. Convexity of Q_s as a function of s. The target function $s \mapsto Q_s = \mathrm{Tr}[\rho^{1-s}\sigma^s]$ in the variational formula defining Q has the useful property to be convex in $s \in [0, 1]$ in the sense of Jensen's inequality: $Q_{ts_1+(1-t)s_2} \leq t\,Q_{s_1} + (1-t)\,Q_{s_2}$ for all $t \in [0, 1]$. This implies that a local minimum is automatically the global one, which is an important benefit in actual calculations.

Indeed, the function $s \mapsto x^{1-s}y^s$ is analytic for positive scalars x and y, and in this case its convexity may be easily confirmed by calculating the second derivative $x^{1-s}y^s(\log y - \log x)^2$, which is non-negative. If one of the parameters, say x, happens to be 0, then $s \mapsto x^{1-s}y^s$ is a constant function equal to 0 for $s \in [0, 1)$ and equal to 1 at $s = 1$. Hence, it is still convex, albeit discontinuous. Consider then a basis with respect to which the matrix representation of ρ is diagonal

$$\rho = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots).$$

Let the matrix representation of σ (in that basis) be given by

$$\sigma = U\,\mathrm{Diag}(\mu_1, \mu_2, \ldots)U^*,$$

where U is a unitary matrix. Then

$$\mathrm{Tr}[\rho^{1-s}\sigma^s] = \sum_{i,j} \lambda_i^{1-s}\mu_j^s |U_{ij}|^2.$$

As this is a sum with positive weights of convex terms $\lambda_i^{1-s}\mu_j^s$, the sum itself is also convex in s.

4.5. Joint concavity of Q in (ρ, σ). By Lieb's theorem [19], $\mathrm{Tr}[\rho^{1-s}\sigma^s]$ is jointly concave on pairs of density operators (ρ, σ) for each fixed $s \in \mathbb{R}$. Since Q is the point-wise minimum of $\mathrm{Tr}[\rho^{1-s}\sigma^s]$ over $s \in [0, 1]$, it is itself jointly concave as well. Hence the related quantum Chernoff distance is jointly convex, just like the relative entropy.

4.6. Monotonicity under CPT maps. From the joint concavity one easily derives the following monotonicity property: for any completely positive trace preserving (CPT) map Φ on the C^*-algebra $\mathcal{B}(\mathcal{H})$ of linear operators, one has

$$Q(\Phi(\rho), \Phi(\sigma)) \geq Q(\rho, \sigma). \tag{33}$$

We remark that this has been shown as a more general result in the framework of relative modular operators in [25]. Moreover, another proof appeared in [26]. We give an alternative proof omitting the notion of relative modular operators.

First, we note that Q is invariant under unitary conjugations, i.e.

$$Q(U\rho U^*, U\sigma U^*) = Q(\rho, \sigma).$$

Secondly, Q is invariant under addition of an ancilla system: for any density operator τ on a finite-dimensional ancillary Hilbert space we have the identity

$$Q(\rho \otimes \tau, \sigma \otimes \tau) = Q(\rho, \sigma).$$

This is because $\mathrm{Tr}[(\rho \otimes \tau)^{1-s}(\sigma \otimes \tau)^s] = \mathrm{Tr}[\rho^{1-s}\sigma^s]\,\mathrm{Tr}[\tau]$. Exploiting the unitary representation of a CPT map, which is a special case of the Stinespring form, the monotonicity statement follows for general CPT maps if we can prove it for the partial trace map. As noted by Uhlmann [27,7], the partial trace map can be written as a convex combination of certain unitary conjugations. Monotonicity of Q under the partial trace then follows directly from its concavity and its unitary invariance.

4.7. Continuity. By the lower bound $Q + T \geq 1$, the distance measures $1 - Q$ and ξ_{QCB} are continuous in the sense that states that are close in trace distance are also close w.r.t. $1 - Q$ and w.r.t. ξ_{QCB}. Indeed, we have $0 \leq 1 - Q \leq T$ and $\xi_{QCB} = -\log Q \leq -\log(1 - T) = T + O(T^2)$.

4.8. Relation of the Chernoff distance to the relative entropy. In the classical case there is a striking relation between the Chernoff distance ξ_{CB} and the relative entropy $H(\cdot\|\cdot)$. It takes its simplest version if the two involved discrete probability distributions p and q have coinciding supports since then $s \mapsto \log \sum_x p^{1-s}(x)q^s(x) = \log Q_s$ is analytic over $[0, 1]$ and its infimum, which defines the Chernoff distance, may be obtained simply by setting

$$0 = (\log Q_s)' = H(p_s\|p) - H(p_s\|q)$$

(the prime denotes derivation w.r.t. s). Here

$$p_s := \frac{p^{1-s}q^s}{\sum_x p^{1-s}(x)q^s(x)}$$

defines a parametric family of probability distributions interpolating between p and q as the parameter s varies between 0 and 1. In the literature, this family is called the Hellinger arc. It follows that the minimiser $s^* \in [0, 1]$ is uniquely determined by the identity

$$H(p_{s^*}\|q) = H(p_{s^*}\|p). \tag{34}$$

Furthermore, for any $s \in [0, 1]$ we have:

$$H(p_s\|p) = s(\log Q_s)' - \log Q_s, \tag{35}$$

and similarly

$$H(p_s\|q) = -(1-s)(\log Q_s)' - \log Q_s, \tag{36}$$

This may be verified by direct calculation using essentially the identity $\log p^{1-s}q^s = \log p^{1-s} + \log q^s$. For the minimiser s^* the formulas (35) and (36) reduce to

$$H(p_{s^*}\|p) = H(p_{s^*}\|q) = \xi_{CB}(p, q). \tag{37}$$

In the generic case of possibly different supports of p and q one has to modify (34) and (37) slightly, see [22].

It turns out that in the quantum setting the minimiser $s^* \in [0, 1]$ of $\inf_{s\in[0,1]} \log Q_s$ can be characterised by a generalized version of (34). However, the surely more remarkable relation (37) between the Chernoff distance ξ_{CB} and the relative entropy seems to have no quantum counterpart.

We assume again that the involved density operators ρ and σ both have full support, i.e. are invertible. Then $Q_s = \mathrm{Tr}(\rho^{1-s}\sigma^s)$ is an analytic function over $[0, 1]$ and its local infimum over $[0, 1]$, which is a global minimum due to convexity, can be found by differentiating Q_s w.r.t. s:

$$\begin{aligned}\frac{\partial}{\partial s}\mathrm{Tr}[\rho^{1-s}\sigma^s] &= -\mathrm{Tr}[(\log\rho)\,\rho^{1-s}\,\sigma^s] + \mathrm{Tr}[\rho^{1-s}\,\sigma^s\,\log\sigma] \\ &= -\mathrm{Tr}[\rho^{1-s}\,\sigma^s\,\log\rho] + \mathrm{Tr}[\rho^{1-s}\,\sigma^s\,\log\sigma].\end{aligned} \tag{38}$$

The infimum is therefore obtained for an $s \in [0, 1]$ such that

$$\mathrm{Tr}[\rho^{1-s}\sigma^s \log\rho] = \mathrm{Tr}[\rho^{1-s}\sigma^s \log\sigma].$$

This is equivalent to the condition

$$S(\rho_s||\rho) = S(\rho_s||\sigma), \tag{39}$$

where $S(\rho||\sigma)$ denotes the quantum relative entropy defined by (15) and ρ_s is defined as

$$\rho_s = \frac{\rho^{1-s}\sigma^s}{\mathrm{Tr}[\rho^{1-s}\sigma^s]}. \tag{40}$$

Note that ρ_s, with $s \in (0, 1)$, is not a density operator, because it is not even self-adjoint (except in the case of commuting ρ and σ). Nevertheless, as it is basically the product of two positive operators, it has positive spectrum, and its entropy and the relative entropies used in (39) are well-defined. The value of s for which both relative entropies coincide is the minimiser in the variational expression (26) for Q.

The family ρ_s, $s \in [0,1]$, can be considered as a quantum generalisation of the Hellinger arc interpolating between the quantum states ρ and σ, albeit out of the state space, in contrast to the classical case.

When attempting to generalise relation (37) to the quantum setting one has to verify (35) or (36) with density operators ρ, σ replacing the probability distributions p, q. This would require the identity $\mathrm{Tr}\,\rho_s \log \rho^{1-s}\sigma^s = \mathrm{Tr}\,\rho_s(\log \rho^{1-s} + \log \sigma^s)$ to be satisfied. However, this is not the case for arbitrary non-commutative density operators ρ, σ. Thus the second identity in (37) seems to be a classical special case only.

5. Asymmetric Quantum Hypothesis Testing: Quantum Hoeffding Bound

In this section, we consider the applications of our techniques presented in Sect. 3 to the case of asymmetric quantum hypothesis testing. More precisely, we consider a quantum generalisation of the Hoeffding bound and of Stein's Lemma.

5.1. The Classical Hoeffding Bound. The classical Hoeffding bound in information theory is due to Blahut [6] and Csiszár and Longo [9]. The corresponding ideas in statistics were first put forward in the paper [16] by W. Hoeffding, from which the bound got its name. Some authors prefer the more complete name of Hoeffding-Blahut-Csiszár-Longo bound. In the following paragraph we review the basic results in Blahut's terminology; at this point we have to mention that many different notational conventions are in use throughout the literature.

Let p be the distribution associated with the null hypothesis, and q the one associated with the alternative hypothesis.[1] Following [6], and for the purposes of this discussion, we initially assume that p and q are equivalent (mutually absolutely continuous) on a finite sample space. The Hoeffding bound gives the best exponential convergence rate of the type-I error under the constraint that the rate limit of the type-II error is bounded from below by a constant r, i.e. when the type-II error tends to 0 sufficiently fast.

Blahut defines the *error-exponent function* $e(r), r \geq 0$, with respect to two probability densities p and q with coinciding supports, as a minimisation over probability densities x:

$$e(r) = \inf_x \{H(x\|p) : H(x\|q) \leq r\}, \tag{41}$$

where $H(\cdot\|\cdot)$ is again the classical relative entropy defined in (16). This minimisation is a convex minimisation, since the target function is convex in x, and the feasible set, defined by the constraint $H(x\|q) \leq r$, is a convex set. Pictorially speaking, the optimal x is the point in the feasible set that is closest (as measured by the relative entropy) to p. If p itself is in the feasible set (i.e. if $H(p\|q) \leq r$), then the optimal x is p, and $e(r) = 0$. Otherwise, the optimal x is on the boundary of the feasible set, in the sense that $H(x\|q) = r$, and $e(r) > 0$. Obviously, if $r = 0$, the feasible set is the singleton $\{q\}$, and $e(r) = H(q\|p)$.

The error-exponent function is thus a non-increasing, convex function of $r \geq 0$, with the properties that $e(0) = H(q\|p)$ and $e(H(p\|q)) = 0$. It can be expressed in a

[1] In [6], the null hypothesis corresponds to H_2, with distribution q_2, and the alternative hypothesis to H_1, with distribution q_1.

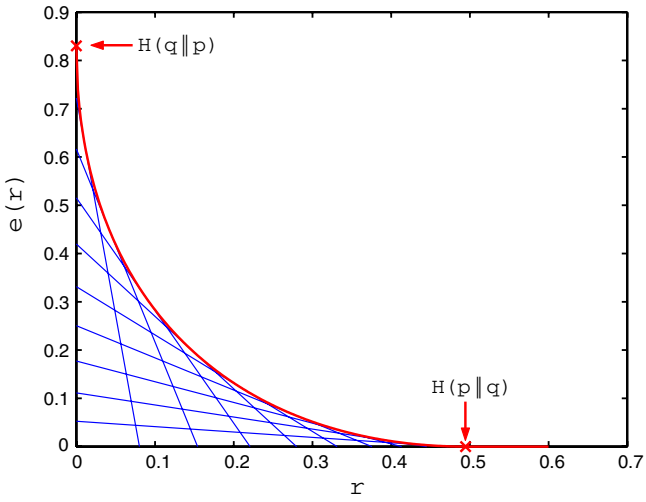

Fig. 1. (Color online) Example plot of the error-exponent function $e(r)$, Eq. (42), for the distributions $p = (0.95, 0.05)$ and $q = (0.5, 0.5)$. The thick (red) line is the graph of $e(r)$, while the thin (blue) lines are instances of the linear function $(-rs - \log \sum_k q_k^s p_k^{1-s})/(1-s)$ for various values of s, of which $e(r)$ is the point-wise maximum. For the chosen p and q, the value of $H(p\|q) = 0.49463$ and the value of $H(q\|p) = 0.83037$

computationally more convenient format as

$$e(r) = \sup_{0 \le s < 1} \frac{-rs - \log \sum_k q_k^s p_k^{1-s}}{1-s}. \tag{42}$$

An example is shown in Fig. 1.

Let $\phi = (\phi_n)$ be a sequence of test functions. Recall the notations $\alpha_R(\phi)$ and $\beta_R(\phi)$ introduced in Sect. 2 for the rate limits (if they exist) of the corresponding type-I and type-II errors, respectively:

$$\alpha_R(\phi) = \lim_{n\to\infty} -\frac{1}{n} \log \alpha_n(\phi), \quad \beta_R(\phi) = \lim_{n\to\infty} -\frac{1}{n} \log \beta_n(\phi).$$

Then the classical HBCL Theorem can be stated as follows.

Theorem 3 *(HBCL). Assume that p, q are mutually absolutely continuous. Then for each $r > 0$ there exists a sequence ϕ of test functions ϕ_n such that the rate limits of the type-II and type-I errors behave like $\beta_R(\phi) \ge r$ and $\alpha_R(\phi) = e(r)$. Moreover, for any sequence ϕ such that $\alpha_R(\phi)$ and $\beta_R(\phi)$ both exist, the relation $\beta_R(\phi) > r$ implies $\alpha_R(\phi) \le e(r)$.*

We remark that for sequences ϕ of test functions ϕ_n for which the rate limits $\alpha_R(\phi)$ or $\beta_R(\phi)$ do not exist, the result still applies to subsequences (ϕ_{n_k}) along which both error rate limits exist. The second part of the HBCL theorem is thus a statement about all accumulation points of $\left(-\frac{1}{n}\log\alpha_n(\phi), -\frac{1}{n}\log\beta_n(\phi)\right)$ for an arbitrary test sequence ϕ.

Referring to Fig. 1, the claim of this theorem is that for any sequence of test functions ϕ the point $(\beta_R(\phi), \alpha_R(\phi))$ cannot be above the graph of $e(r)$ over $r > 0$ and for any point on the graph over $r \ge 0$ one can find a sequence ϕ. Since $\beta_R(\phi) = 0$ may correspond to

the case where $\beta(\phi_n)$ vanishes subexponentially slowly as well as converges to a positive value, a rate limit of type-I error $\alpha_R(\phi)$ larger than $e(0) = H(q\|p)$ is achievable.

The case $\beta_R(\phi) > r \geq H(p\|q)$, where $e(r) = 0$, can be shown to correspond to $\alpha(\phi_n)$ converging to 1, rather than to 0. (This is basically the content of the so-called 'Strong Converse'.) In the case $\beta_R(\phi) = H(p\|q)$ a convergence of $\alpha(\phi_n)$ to 0 is achievable, albeit only subexponantially slowly (this is due to Stein's Lemma.)

Note that in order to obtain a bound on β_R under a constrained α_R one just has to interchange p and q in the theorem.

5.2. Nonequivalent hypotheses. The Chernoff and Hoeffding bounds have typically been treated in the literature under a restrictive assumption that hypotheses p, q are mutually absolutely continuous (equivalent), cf., e.g., Blahut [6]. As a prerequisite for a quantum generalisation, unless one wants to limit oneself to faithful states, one has to understand the classical Hoeffding bound for nonequivalent hypotheses. For the Chernoff bound, a corresponding discussion can be found in [22] without restrictions on the underlying sample space. Here we limit ourselves to finite sample spaces, thereby excluding infinite relative entropies for equivalent measures p, q.

For probability measures p, q on a finite sample space Ω, let D_0 be the support of p, D_1 be the support of q and $B = D_0 \cap D_1$. Let $\psi_0 = p(B)$, $\psi_1 = q(B)$ and note that $\psi_0 > 0, \psi_1 > 0$ unless the measures p, q are orthogonal (which we exclude for triviality). Define conditional measures given the set B: $\tilde{p}(\cdot) = p(\cdot|B), \tilde{q}(\cdot) = q(\cdot|B)$. Note that $\tilde{p}, \tilde{q}$ are equivalent measures; we may have $\tilde{p} = \tilde{q}$. We consider hypothesis testing for a pair of product measures $p^{\otimes n}, q^{\otimes n}$.

Recall that a (nonrandomised) test is a mapping $\phi_n : \Omega^n \mapsto \{0, 1\}$. In our setting, only observations in either D_0^n or D_1^n can occur, so we will modify the sample space to be $D_0^n \cup D_1^n$. We will then establish the relation of tests ϕ_n in the original problem $p^{\otimes n}$ vs. $q^{\otimes n}$ to tests in the 'conditional' problem $\tilde{p}^{\otimes n}$ vs. $\tilde{q}^{\otimes n}$, i.e. to tests $\tilde{\phi}_n : B^n \mapsto \{0, 1\}$. Call a test ϕ_n null admissible if it takes value 0 on $D_0^n \backslash B^n$ and value 1 on $D_1^n \backslash B^n$. These tests correspond to the notion that if a point in the sample space Ω^n is not in B^n, then it identifies the hypothesis errorfree (either p or q). We need only consider null admissible tests; for any test there is a null admissible test with equal or smaller error probabilities α_n, β_n. The restriction $\phi_n|B^n$ gives a test on B^n, i.e. in the conditional problem.

Lemma 3. *There is a one-to-one correspondence between null admissible tests ϕ_n in the original problem $p^{\otimes n}$ vs. $q^{\otimes n}$ and tests $\tilde{\phi}_n$ in the conditional problem $\tilde{p}^{\otimes n}$ vs. $\tilde{q}^{\otimes n}$, given by $\tilde{\phi}_n = \phi_n|B^n$. The errror probabilities satisfy*

$$\alpha_n(\phi) = \psi_0^n \alpha_n\left(\tilde{\phi}\right), \quad \beta_n(\phi) = \psi_1^n \beta_n\left(\tilde{\phi}\right),$$

where $\psi_0 = p(B)$ and $\psi_1 = q(B)$.

Proof. The first claim is obvious, if one takes into account that we took all tests in the original problem to be mappings $\phi_n : D_0^n \cup D_1^n \mapsto \{0, 1\}$. For the relation of error probabilities, note that $p^{\otimes n}(A) = \psi_0^n \tilde{p}^{\otimes n}(A \cap B^n)$, $A \subset D_0^n \cup D_1^n$ and therefore

$$\begin{aligned}\alpha_n(\phi) &= \int \phi_n dp^{\otimes n} = \int_{B^n} \phi_n dp^{\otimes n} \text{ (by null admissibility)} \\ &= \psi_0^n \int \phi_n d\tilde{p}^{\otimes n} = \psi_0^n \int_{B^n} \tilde{\phi}_n d\tilde{p}^{\otimes n} = \psi_0^n \alpha_n\left(\tilde{\phi}\right)\end{aligned}$$

and analogously for $\beta_n(\phi)$. □

This result already allows to state the general Hoeffding bound in terms of the error-exponent function for the conditional problem

$$\tilde{e}(r) = \sup_{0 \le s < 1} \frac{-rs - \log \sum_k \tilde{q}_k^s \tilde{p}_k^{1-s}}{1-s}.$$

Indeed, rate limits $\alpha_R(\phi)$ and $\beta_R(\phi)$ for a null admissible test sequence ϕ exist if and only if they exist for the corresponding test sequence $\tilde{\phi}$, and

$$\alpha_R(\phi) = -\log \psi_0 + \alpha_R\left(\tilde{\phi}\right), \quad \beta_R(\phi) = -\log \psi_1 + \beta_R\left(\tilde{\phi}\right). \tag{43}$$

Proposition 3. *Let p, q be arbitrary probability measures on a finite sample space.*
(i) (achievability) For each $r \ge -\log \psi_1$ there exists a sequence ϕ of test functions ϕ_n such that the rate limits of the type-II and type-I errors behave like $\beta_R(\phi) \ge r$ and $\alpha_R(\phi) = -\log \psi_0 + \tilde{e}(r + \log \psi_1)$. For the case $0 \le r \le -\log \psi_1$, there is a sequence ϕ of test functions ϕ_n obeying $-n^{-1} \log \beta_n(\phi) = -\log \psi_1$ and $\alpha_n(\phi) = 0$ for every n.
(ii) (optimality) Consider any sequence ϕ such that $\alpha_R(\phi)$ and $\beta_R(\phi)$ both exist. If $r \ge -\log \psi_1$ then the relation $\beta_R(\phi) > r$ implies $\alpha_R(\phi) \le -\log \psi_0 + \tilde{e}(r + \log \psi_1)$.

Note that in (ii) the omission of the case $0 \le r \le -\log \psi_1$ means that there is no upper bound on $\alpha_R(\phi)$, as shown by the achievability part ($\alpha_R(\phi)$ has to be set equal to ∞ for a test of vanishing error probability α_n).

Proof. (i) Assume $r \ge -\log \psi_1$ and take a test sequence $\tilde{\phi}_n$ in the conditional problem $\tilde{p}^{\otimes n}$ vs. $\tilde{q}^{\otimes n}$ such that $\beta_R(\tilde{\phi}) \ge r + \log \psi_1$ and $\alpha_R(\tilde{\phi}) = \tilde{e}(r + \log \psi_1)$, which exists according to the HBCL theorem since $\tilde{p}, \tilde{q}$ are mutually absolutely continuous. According to Lemma 3, the corresponding null admissible test ϕ_n satisfies (43) and hence $\beta_R(\phi) \ge r$ and $\alpha_R(\phi) = -\log \psi_0 + \tilde{e}(r + \log \psi_1)$. Furthermore, consider the test $\tilde{\phi}_n \equiv 0$ in $\tilde{p}^{\otimes n}$ vs. $\tilde{q}^{\otimes n}$. This has $\alpha_n(\tilde{\phi}_n) = 0$ and $\beta_n(\tilde{\phi}_n) = 1$, hence the corresponding null admissible test ϕ_n has $\alpha_n(\phi_n) = 0$ and $\beta_n(\phi_n) = \psi_1^n$.

(ii) Using a reduction to the conditional problem $\tilde{p}^{\otimes n}$ vs. $\tilde{q}^{\otimes n}$ similar to the one above, the optimality part also follows immediately from the HBCL theorem. □

Remark. Consider the dual of the test used in the second part of (i), i.e. the null admissible extension of the test $\tilde{\phi}_n \equiv 1$. This one obviously has $\alpha_n(\phi) = \psi_0^n$ and $\beta_n(\phi) = 0$. It can be used for achievability for large r, i.e. it has $\beta_R(\phi) = \infty$ and $\alpha_R(\phi) = -\log \psi_0$.

It is possible to obtain a closed form expression for the Hoeffding bound, using the error-exponent function defined for $r \ge 0$ exactly as in (42), for the case of nonequivalent p, q. The difference is that we now have to admit a value $+\infty$ for certain arguments.

Lemma 4. *For general p, q, the error-exponent function $e(r)$ satisfies*

$$e(r) = \begin{cases} -\log \psi_0 + \tilde{e}(r + \log \psi_1), & \text{for } r \ge -\log \psi_1 \\ \infty, & \text{for } 0 \le r < -\log \psi_1. \end{cases}$$

Remark. For two distinct p, q it is possible that $\tilde{p} = \tilde{q}$. In that case $\tilde{e}(r) = 0$ for $r \ge 0$. It follows that $e(r) = \infty$ for $r < -log \psi_1$ and $e(r) = -\log \psi_0$ for $r \ge -\log \psi_1$. This case will be relevant in the quantum setting when the hypotheses will be represented by two non-orthogonal pure quantum states.

Proof. Assume $r \geq -\log \psi_1$ and set

$$e_s(r) = \frac{-rs - \log Q_s}{1-s},$$

where $Q_s = \sum_k p_k^{1-s} q_k^s$. Let $\tilde{Q}_s = \sum_k \tilde{p}_k^{1-s} \tilde{q}_k^s$ and note $Q_s = \psi_0^{1-s} \psi_1^s \tilde{Q}_s$. Hence

$$\begin{aligned} e_s(r) &= \frac{-rs - (1-s)\log\psi_0 - s\log\psi_1 - \log\tilde{Q}_s}{1-s} \\ &= -\log\psi_0 + \frac{-(r+\log\psi_1)s - \log\tilde{Q}_s}{1-s} = -\log\psi_0 + \tilde{e}_s(r+\log\psi_1), \end{aligned}$$

where $\tilde{e}_s$ is the analogue of the function $e_s(r)$ with Q_s replaced by $\tilde{Q}_s$. Since $e(r) = \sup_{0\leq s<1} e_s(r)$ and the analogue is true for $\tilde{e}_s$ and $\tilde{e}$, the claim follows in the case $r \geq -\log\psi_1$.

Assume now $0 \leq r < -\log\psi_1$ and $\psi_1 < 1$, i.e. $-\log\psi_1 > 0$. Clearly we have $Q_s \to \psi_1$ as $s \nearrow 1$, hence $-rs - \log Q_s \to -r - \log\psi_1 > 0$ as $s \nearrow 1$. Hence $\lim_{s\nearrow 1} e_s(r) = \infty$, and since $e(r) = \sup_{0\leq s<1} e_s(r)$, we also have $e(r) = \infty$. □

In conjunction with Proposition 3 we obtain a closed form description of the Hoeffding bound for possibly nonequivalent measures p, q, in terms of the original error-exponent function $e(r)$.

Theorem 4. *Let p, q be arbitrary probability measures on a finite sample space. Then the statement of the HBCL Theorem (Theorem 3) is true, where the error-exponent function defined in* (42) *obeys $e(r) = \infty$ for $0 \leq r < -\log\psi_1$ if $\psi_1 < 1$.*

We noted already that for $e(r) = \infty$, the bound on $\alpha_R(\phi)$ is achievable in the sense that a test exists having exactly $\alpha_n(\phi) = 0$ for all n.

Using the properties of the rate function $\tilde{e}$ pertaining to equivalent measures $\tilde{p}, \tilde{q}$, as illustrated in Fig. 1, and the representation of Lemma 4 we obtain the following description of the general rate exponent function. In the interval $[0, -\log\psi_1)$ it is infinity. At $r = -\log\psi_1$ it takes value $e(r) = -\log\psi_0 + H(\tilde{q}\|\tilde{p}) = H(\tilde{q}\|p)$. For $r \geq -\log\psi_1$ it is convex and non-increasing. More precisely, over the interval $[-\log\psi_1, -\log\psi_1 + H(\tilde{p}\|\tilde{q}) = H(\tilde{p}\|q)]$ $e(r)$ is convex (even strictly convex) and monotone decreasing. Over the interval $[H(\tilde{p}\|q), \infty)$ it is constant with value $-\log\psi_0$. A visual impression can be obtained by imagining the origin in Fig. 1 shifted to the point $(-\log\psi_1, -\log\psi_0)$. This picture will explicitly appear in Fig. 2 below, in a situation further generalized to two quantum states with different supports.

5.3. Quantum Hoeffding Bound. In the quantum setting the error-exponent function $e(r)$ has to be replaced by a function $e_Q : \mathbb{R}_0^+ \longrightarrow [0, \infty]$ given by

$$e_Q(r) := \sup_{0\leq s<1} \frac{-rs - \log \mathrm{Tr}\,\sigma^s \rho^{1-s}}{1-s}. \tag{44}$$

In view of Proposition 1, $e_Q(r)$ coincides with the error-exponent function $e(r)$ for the pair of probability distributions (p, q) associated with (ρ, σ) via relation (12). Therefore, we can use Lemma 4 to describe properties of the function $e_Q(r)$, or the remarks after Theorem 4.

Recall that for a pair (p, q), we defined a related pair of probability distributions $(\tilde{p}, \tilde{q})$ by conditioning p and q, respectively, on the intersection $B = D_0 \cap D_1$ of the two support sets D_0 and D_1, and also $\psi_0 = p(B)$, $\psi_1 = q(B)$. In the present context, in accordance with (12) we have

$$D_0 = \{(i, j) : 1 \le i, j \le d,\ \lambda_i > 0\},\ D_1 = \left\{(i, j) : 1 \le i, j \le d,\ \mu_j > 0\right\}.$$

Let, as before, $\tilde{e}(r)$ be the error-exponent function pertaining to the pair $(\tilde{p}, \tilde{q})$ according to (42). Then the quantum error-exponent function $e_Q(r)$ for the hypotheses ρ, σ may be represented simply by

$$e_Q(r) = e(r) = \begin{cases} -\log\psi_0 + \tilde{e}(r + \log\psi_1), & \text{for } r \ge -\log\psi_1 \\ \infty, & \text{for } 0 \le r < -\log\psi_1. \end{cases} \tag{45}$$

It obtains its characteristic properties from the classical function being convex and monotone decreasing in the interval $\left[-\log\psi_1, H(\tilde{p}\|q)\right]$ with $e(-\log\psi_1) = H(\tilde{q}\|p)$, and constant with value $-\log\psi_0$ in the interval $[H(\tilde{p}\|q), \infty)$.

Lemma 5. *Let supp ρ, supp σ be the support projections associated with ρ, σ. Then the critical points and extremal values of $e_Q(r)$ may be expressed in a more direct way in terms of the density operators:*

$$\psi_0 = \mathrm{Tr}\left[\rho\ \mathrm{supp}\,\sigma\right],\quad \psi_1 = \mathrm{Tr}\left[\sigma\ \mathrm{supp}\,\rho\right]$$

and

$$H(\tilde{p}\|q) = S_\sigma(\rho\|\sigma)\quad H(\tilde{q}\|p) = S_\rho(\sigma\|\rho),$$

where the entropy type quantities on the right-hand side are defined as

$$S_\sigma(\rho\|\sigma) := \mathrm{Tr}\left[\frac{\rho}{\psi_0}\ \left(\log\frac{\rho}{\psi_0} - \log\sigma\right)\mathrm{supp}\,\sigma\right],$$
$$S_\rho(\sigma\|\rho) := \mathrm{Tr}\left[\frac{\sigma}{\psi_1}\ \left(\log\frac{\sigma}{\psi_1} - \log\rho\right)\mathrm{supp}\,\rho\right].$$

Proof. Note that for $B = D_0 \cap D_1$ we have

$$\begin{aligned}
\psi_0 &= \sum_{(i,j)\in B} \lambda_i\ \left|\langle x_i|y_j\rangle\right|^2 = \sum_{i,j} \lambda_i\ \mathrm{sgn}(\mu_j)\ \left|\langle x_i|y_j\rangle\right|^2 \\
&= \sum_{i,j} \lambda_i\ \left|\langle x_i|\mathrm{sgn}(\mu_j)y_j\rangle\right|^2 = \sum_{i,j} \lambda_i\ \left|\langle x_i|\,(\mathrm{supp}\,\sigma)\,y_j\rangle\right|^2 \\
&= \sum_{i,j} \lambda_i\ \left|\langle (\mathrm{supp}\,\sigma)\,x_i|y_j\rangle\right|^2 = \sum_i \lambda_i\ \|(\mathrm{supp}\,\sigma)\,x_i\|^2 \\
&= \mathrm{Tr}\left[\sum_i \lambda_i\,|(\mathrm{supp}\,\sigma)\,x_i\rangle\,\langle(\mathrm{supp}\,\sigma)\,x_i|\right] = \mathrm{Tr}\left[\rho\ \mathrm{supp}\,\sigma\right]
\end{aligned}$$

and analogously for ψ_1. Furthermore

$$\begin{aligned} H(\tilde{p}\|q) &= \sum_{(i,j)\in B} \tilde{p}_{i,j} \log \frac{\tilde{p}_{i,j}}{q_{i,j}} = \sum_{(i,j)\in B} \lambda_i \left|\langle x_i|y_j\rangle\right|^2 \frac{1}{\psi_0} \log \frac{\lambda_i}{\mu_j \psi_0} \\ &= \sum_{i,j} \operatorname{sgn}(\mu_j) \left|\langle x_i|y_j\rangle\right|^2 \frac{\lambda_i}{\psi_0} \log \frac{\lambda_i}{\psi_0} - \sum_{i,j} \operatorname{sgn}(\mu_j) \left|\langle x_i|y_j\rangle\right|^2 \frac{\lambda_i}{\psi_0} \log \mu_j \\ &= \operatorname{Tr}\left[\frac{\rho}{\psi_0} \left(\log \frac{\rho}{\psi_0}\right) \operatorname{supp}\sigma\right] - \operatorname{Tr}\left[\frac{\rho}{\psi_0} (\log\sigma) \operatorname{supp}\sigma\right], \end{aligned}$$

where the third equality is analogous to the calculation in the proof of Proposition 1. □

To shed some light on the entropy type quantity $S_\sigma(\rho\|\sigma)$, note that it may be rewritten as a difference of usual (Umegaki's) relative entropies:

$$S_\sigma(\rho\|\sigma) = S\left(\frac{\rho}{\psi_0}\operatorname{supp}\sigma\,\|\sigma\right) - S\left(\frac{\rho}{\psi_0}\operatorname{supp}\sigma\,\|\,\frac{\rho}{\psi_0}\right).$$

This may be verified by direct calculations similar to those in the proof of Lemma 5.

The linear operator $\frac{\rho}{\psi_0}\operatorname{supp}\sigma$ is a kind of conditional expectation of ρ. While it is not self-adjoint, the relative entropies on the right-hand side are well defined (in a mathematical sense) and real: first, the entropy of $\frac{\rho}{\psi_0}\operatorname{supp}\sigma$ is defined in terms of its spectrum, which is positive and normalised to 1, hence giving a real, positive entropy, and second, $\operatorname{Tr}[\rho\ \operatorname{supp}\sigma\ \log(\rho)]$ can be written as $\operatorname{Tr}[\operatorname{supp}\sigma\rho\log\rho\ \operatorname{supp}\sigma]$, from which it is evident that this term is also real.

It is easily seen from the above formula that $S_\sigma(\rho\|\sigma)$ coincides with $S(\rho\|\sigma)$ if σ is a faithful state, or more generally if $\operatorname{supp}\rho \leq \operatorname{supp}\sigma$. Otherwise $S(\rho\|\sigma) = \infty$, while $S_\sigma(\rho\|\sigma)$ is finite.

Note also that $S_\rho(\sigma\|\rho) \geq -\log\psi_0$ and equality holds if and only if it holds in $S_\sigma(\rho\|\sigma) \geq -\log\psi_1$. This immediately follows from $S_\rho(\sigma\|\rho) + \log\psi_0 = H(\tilde{p}\|\tilde{q})$, which is seen from Lemma 5. This happens in particular if both ρ and σ are pure states. In this case there is only one pair (i, j) where both $\lambda_i > 0$ and $\mu_j > 0$, hence the set B consists of one element only. In this case we must have $\tilde{p} = \tilde{q}$, hence $H(\tilde{p}\|\tilde{q}) = H(\tilde{q}\|\tilde{p}) = 0$.

The general shape of the quantum error-exponent function $e_Q(r)$ is represented in Fig. 2. If both ρ and σ are pure states then the shape degenerates to 'rectangular' form ($e_Q(r) = \infty$ or $e_Q(r) = -\log\psi_1$).

A quantum generalisation of the HBCL Theorem then reads as follows.

Theorem 5 *(Quantum HBCL). For each $r > 0$ there exists a sequence Π of test projections Π_n on $\mathcal{H}^{\otimes n}$ for which the rate limits of type-I and type-II errors behave like $\alpha_R(\Pi) = e_Q(r)$ and $\beta_R(\Pi) \geq r$, respectively. Moreover, for any sequence Π such that $\alpha_R(\Pi)$ and $\beta_R(\Pi)$ both exist, the relation $\beta_R(\Pi) > r$ implies $\alpha_R(\Pi) \leq e_Q(r)$.*

The statement of the quantum HBCL Theorem is that for every sequence Π (for which both error rate limits exist) the point $(\beta_R(\Pi), \alpha_R(\Pi))$ lies on or below the curve $e_Q(r)$ over $(0, \infty]$, and for every point on the curve over the closed interval $[0, \infty]$ there is a sequence Π achieving it.

We remark that, just like (37), the relationship (41) seems to have no general quantum counterpart, even when both states are faithful. In other words, there is no known subset of linear operators τ with positive spectrum such that $e_Q(r) = \inf_\tau\{S(\tau\|\rho) : S(\tau\|\sigma) \leq r\}$.

To prove the quantum Hoeffding bound, the following lemmas are needed.

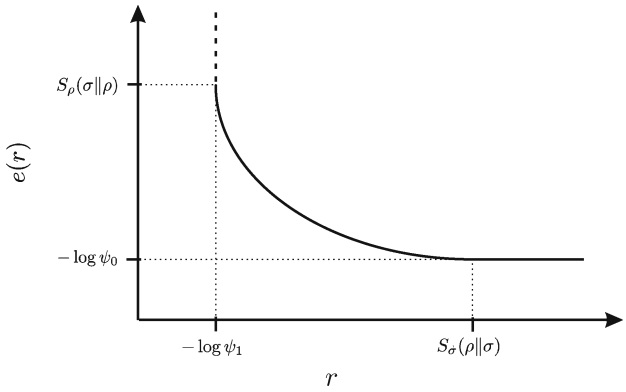

Fig. 2. Example plot of the quantum error-exponent function $e_Q(r)$ in the general case

Lemma 6. *For scalars $x, y > 0$, bounds on $\log(x + y)$ are given by*

$$\max(\log x, \log y) \leq \log(x + y) \leq \max(\log x, \log y) + \log 2. \tag{46}$$

Proof. For the first inequality, put $x = e^a$ and $y = e^b$, and note

$$\begin{aligned} \log(e^a + e^b) &= a + \log(1 + e^{b-a}) \\ &\geq a + \max(0, b - a) \\ &= \max(a, b). \end{aligned}$$

The second inequality follows directly from the fact that the logarithm increases monotonically, so that $\log((x + y)/2) \leq \log \max(x, y)$. □

A direct consequence of this lemma is

Lemma 7. *For two scalar sequences $x_n, y_n > 0$ with rate limits x_R and y_R, the rate limit of $x_n + y_n$ is given by*

$$\lim_{n\to\infty} -\frac{1}{n} \log(x_n + y_n) = \min(x_R, y_R). \tag{47}$$

5.4. Proof of Optimality of the Quantum Hoeffding Bound. Again we use the mapping from the pair (ρ, σ) to the pair (p, q), so that, by Proposition 1, $e(r) = e_Q(r)$. From Proposition 2 we have that for any sequence Π of orthogonal projections Π_n and for any real value of the scalar x, for all $n \in \mathbb{N}$ one as

$$\alpha(\Pi_n) + e^{-nx}\beta(\Pi_n) \geq \frac{1}{2}\left(\alpha(\phi_n) + e^{-nx}\beta(\phi_n)\right),$$

where ϕ_n are classical test functions corresponding to the maximum likelihood decision rule, cf. the proof of Proposition 2. Recall that the type-I and type-II errors are defined as $\alpha(\phi_n) = \sum_i p_i^n \phi_n(i)$ and $\beta(\phi_n) = \sum_i q_i^n(1 - \phi_n(i))$.

On taking the rate limit on the left side, this gives

$$\lim_{n\to\infty} -\frac{1}{n} \log\left(\alpha(\Pi_n) + e^{-nx}\beta(\Pi_n)\right) \leq \liminf_{n\to\infty} -\frac{1}{n} \log\left(\alpha(\phi_n) + e^{-nx}\beta(\phi_n)\right).$$

By possibly taking a subsequence, we can ensure that the rate limits $\alpha_R(\phi)$, $\beta(\phi_n)$ also exist. By Lemma 7, the above simplifies to

$$\min(\alpha_R(\Pi), x + \beta_R(\Pi)) \leq \min(\alpha_R(\phi), x + \beta_R(\phi)). \tag{48}$$

Assume now that $\beta_R(\phi) \leq -\log \psi_1$. Then, by selecting $x < 0$ and $|x|$ sufficiently large, we obtain $x + \beta_R(\Pi) \leq x + \beta_R(\phi)$, and hence $\beta_R(\Pi) \leq -\log \psi_1$. Since $e_Q(r) = \infty$ for $r < \beta_R(\Pi) \leq -\log \psi_1$ according to the discussion above Lemma 5, the claim $\alpha_R(\Pi) \leq e_Q(r)$ holds trivially. Henceforth we assume that $\beta_R(\phi) > -\log \psi_1$.

From the classical HBCL Theorem (more precisely, from Theorem 4), the right-hand side of (48) is bounded above by $\min(e(r), x + \beta_R(\phi))$, for any r with $-\log \psi_1 \leq r < \beta_R(\phi)$. Note that $e(r)$ is continuous for $r \geq -\log \psi_1$ (since it is monotonely nonincreasing and convex). By letting $r \nearrow \beta_R(\phi)$ we obtain an upper bound $\min(e(r), x + r)$ with $r \geq -\log \psi_1$.

We can now prove the optimality part of the quantum HBCL Theorem, using only this upper bound plus the fact that $e(r)$ is monotonously decreasing.

The upper bound $\min(e(r), x + r)$ holds for some particular value r. We will find a further upper bound by maximizing over $r \geq -\log \psi_1$. For this we have to distinguish two cases, depending on the value of x.

a) At $r = -\log \psi_1$ we have $e(r) > x + r$. Since $e(r)$ is decreasing in r and continuous, and $x + r$ is increasing, the maximum of $\min(e(r), x + r)$ is obtained when $e(r) = x + r$. Let $r^*(x) > -\log \psi_1$ be the solution of $x + r = e(r)$. We now have that for any sequence of quantum measurements Π and for any real value of the scalar x,

$$\min(\alpha_R(\Pi), x + \beta_R(\Pi)) \leq x + r^*(x) = e(r^*(x)).$$

b) At $r = -\log \psi_1$ we have $e(r) \leq x + r$. Again by the properties of $e(r)$ and $x + r$, the maximum of $\min(e(r), x + r)$ is $e(r^*)$, attained for $r^*(x) = -\log \psi_1$. We then obtain the upper bound

$$\min(\alpha_R(\Pi), x + \beta_R(\Pi)) \leq e(r^*(x)).$$

Now set $x = \alpha_R(\Pi) - \beta_R(\Pi)$, then both inequalities above yield $\alpha_R(\Pi) \leq e(r^*)$. Assume $r < \beta_R(\Pi)$; we intend to show that this implies $\alpha_R(\Pi) \leq e(r)$. Indeed, in both cases a) and b) r^* is such that

$$e(r^*) \leq x + r^* = \alpha_R(\Pi) - \beta_R(\Pi) + r^* < \alpha_R(\Pi) - r + r^*,$$

hence $r^* - r \geq e(r^*) - \alpha_R(\Pi) \geq 0$. Therefore, from the monotonicity of the error-exponent function $e(r^*) \leq e(r)$ follows and we finally obtain $\alpha_R(\Pi) \leq e(r) = e_Q(r)$. □

5.5. Proof of achievability of the quantum Hoeffding bound. The proof of achievability is mainly due to Hayashi [12], who used inequality (24), which is obtained as a byproduct of the proof of Theorem 2. However, we modify it avoiding any implicit assumption that the involved quantum states are faithful; hence we prove Theorem 5 in full generality, which includes for example the case of two non-orthogonal pure states.

Let us fix an arbitrary $s \in (0, 1)$, and set

$$a = e^{-nx}\sigma^{\otimes n}, \tag{49}$$

$$b = \rho^{\otimes n}, \tag{50}$$

where the value of x will be chosen in due course. Consider the sequence of POVMs $\{(\mathbb{1} - \Pi_n, \Pi_n)\}$ with Π_n the projector on the range of $(a - b)_+$; again element $\mathbb{1} - \Pi_n$ is assigned to the null hypothesis $\rho^{\otimes n}$, and element Π_n is assigned to the alternative hypothesis $\sigma^{\otimes n}$. We will show that this POVM asymptotically attains the Hoeffding bound.

Recall that inequality (24) states

$$\mathrm{Tr}[a^s b^{1-s}] \geq \mathrm{Tr}[\Pi b + (\mathbb{1} - \Pi)a].$$

By positivity of $\mathrm{Tr}[\Pi b]$ and $\mathrm{Tr}[(\mathbb{1} - \Pi)a]$, this implies the two inequalities

$$\mathrm{Tr}[\Pi b], \mathrm{Tr}[(\mathbb{1} - \Pi)a] \leq \mathrm{Tr}[a^s b^{1-s}].$$

These yield the following upper bounds on the α and β errors of the chosen POVM (recall $Q_s = \mathrm{Tr}[\rho^{1-s}\sigma^s]$):

$$\begin{aligned}
\beta_n(\Pi_n) &= \mathrm{Tr}[(\mathbb{1} - \Pi_n)\sigma^{\otimes n}] \\
&= e^{nx}\,\mathrm{Tr}[(\mathbb{1} - \Pi_n)a] \\
&\leq e^{nx}\,\mathrm{Tr}[a^s b^{1-s}] \\
&= e^{nx(1-s)} Q_s^n \\
&= \exp[n(x(1-s) + \log Q_s)], \qquad (51) \\
\alpha_n(\Pi_n) &= \mathrm{Tr}[\Pi_n \rho^{\otimes n}] \\
&= \mathrm{Tr}[\Pi_n b] \\
&\leq \mathrm{Tr}[a^s b^{1-s}] \\
&= e^{-nxs} Q_s^n \\
&= \exp[n(-xs + \log Q_s)]. \qquad (52)
\end{aligned}$$

Choosing x such that $x(1-s) + \log Q_s = -r$ then yields, from (51),

$$\beta_n(\Pi_n) \leq \exp(-nr),$$

and from (52),

$$\begin{aligned}
\alpha_n(\Pi_n) &\leq \exp\left(-n\left(-s\frac{r + \log Q_s}{1-s} - \log Q_s\right)\right) \\
&= \exp\left(-n\frac{-rs - \log Q_s}{1-s}\right) \\
&\leq \exp\left(-n e_Q(r)\right),
\end{aligned}$$

where in the last inequality we have used the fact that the parameter s was arbitrarily chosen from $(0, 1)$.

Thus, for the rate limits we get

$$\beta_R \geq r, \quad \alpha_R \geq e_Q(r).$$

The optimality, proven in the previous subsection, states that $\alpha_R \leq e_Q(r)$ if $\beta_R = r$. Furthermore, since $e_Q(r)$ is a non-increasing function, $\alpha_R \leq e_Q(r)$ if $\beta_R > r$. This implies that for the chosen sequence of POVMs

$$\beta_R = r, \quad \alpha_R = e_Q(r)$$

must hold, which proves that the Hoeffding bound is indeed attained. □

5.6. Quantum Stein's Lemma and quantum version of Sanov's Theorem. The quantum generalisation of Stein's Lemma deals with the asymptotics of the error quantity

$$\beta_n^*(\epsilon) := \inf_{\Pi_n}\{\beta_n(\Pi_n) : \alpha_n(\Pi_n) \leq \epsilon\}, \tag{53}$$

for fixed $0 < \epsilon < 1$. Here, the infimum is taken over all positive semi-definite contractions Π_n on $\mathcal{H}^{\otimes n}$.

Quantum Stein's Lemma states that the rate limit $\beta_R^*(\varepsilon)$ of the sequence $(\beta_n^*(\epsilon))$ exists and is equal to $S(\rho\|\sigma)$, independently of ϵ. It was first obtained by Hiai and Petz [15]. Its optimality part was then strengthened by Ogawa and Nagaoka in [24].

Here we use the quantum HBCL Theorem to prove that the relative entropy $S(\rho\|\sigma)$ is an achievable error rate limit and deduce optimality of this bound from Proposition 1 in [5].

Proof of the quantum Stein's Lemma. We need to show that there is a sequence Π with $\alpha(\Pi_n) \leq \epsilon$ achieving $\beta_R(\Pi) = S(\rho\|\sigma)$. Let $\eta > 0$ be small and set $r = S(\rho||\sigma) - \eta$. Achievability of the quantum Hoeffding bound means that a sequence Π exists for which $\beta_R \geq r$ and $\alpha_R = e_Q(r)$. Since $e_Q(r) > 0$ for all $r < S(\rho\|\sigma)$ and $\eta > 0$, the sequence α_n converges to 0. Thus, from a certain value of n onwards, α_n will get lower than any value $\epsilon > 0$ chosen beforehand. This means that Π is a feasible sequence in (53) for n large enough, exhibiting $\beta_R(\epsilon) \geq r = S(\rho\|\sigma) - \eta$. As this holds for any $\eta > 0$, we find that $\beta_R^*(\epsilon) \geq S(\rho\|\sigma)$.

With $\beta_R^*(\epsilon) \geq S(\rho\|\sigma)$ the two hypotheses associated to the pair of density operators (ρ, σ) satisfy the HP-condition in the terminology of the paper [5]. Thus Proposition 1 in [5] implies $\beta_R^*(\epsilon) = S(\rho\|\sigma)$. □

We remark that in [5] the HP-condition was introduced for (ordered) pairs (Ψ, Φ) of arbitrary correlated states on quantum spin chains, while in the present paper only density operators of the tensor-product form $\rho^{\otimes n}$ have been considered. These correspond to the special case of shift-invariant product states on the infinite spin chain (quantum i.i.d. states). A pair (Ψ, Φ) is said to satisfy the HP-condition if the relative entropy rate $s(\Psi\|\Phi)$ exists and is a lower bound on the lower rate limit $\underline{\beta}_R^*(\varepsilon)$ for all $\varepsilon \in (0, 1)$.

Specifically to our setting (the i.i.d. case), Theorem 1 in [5] states that the achievability part in quantum Stein's Lemma (the HP-condition) is equivalent to a quantum version of Sanov's Theorem, which has been presented in [4] and which is a priori a result extending quantum Stein's Lemma in the following way: Let the null hypothesis H_0 correspond to a family Γ of density operators on $\mathcal{H}$ instead of a single density operator ρ. Let the alternative hypothesis H_1 be still represented by a fixed density operator σ. Then there exists a sequence Π of orthogonal projections Π_n on $\mathcal{H}^{\otimes n}$, respectively, such that for all $\rho \in \Gamma$ the corresponding type-I error vanishes asymptotically, i.e.

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n}\Pi_n] = 0, \tag{54}$$

while the type-II error rate limit $\beta_R(\Pi)$ is equal to the relative entropy distance from Γ to σ:

$$S(\Gamma\|\sigma) := \inf_{\rho\in\Gamma} S(\rho\|\sigma).$$

Moreover $S(\Gamma\|\sigma)$ is the upper bound on type-II error (upper) rate limit, for any sequence Π of POVMs satisfying the constraint (54).

With the above reasoning we obtain the statement of quantum Sanov's Theorem from the quantum HBCL Theorem as well.

Acknowledgements. We thank various institutions for their hospitality: the Max Planck Institute for Quantum Optics (FV, KA), the Erwin Schrödinger Institute in Vienna (FV, KA, AS), and the Physics Department of the National University of Singapore (KA). KA was supported by The Leverhulme Trust (grant F/07 058/U), by the QIP-IRC (www.qipirc.org) supported by EPSRC (GR/S82176/0), by EU Integrated Project QAP, and by the Institute of Mathematical Sciences, Imperial College London. MN has been supported by NSF under grant DMS-03-06497.

The authors are also grateful to the anonymous referee regarding Remark 1 after Theorem 2.

A Proofs of Bounds on Q

Inequality (32) stated in terms of general positive operators is

Theorem 6. *For positive operators A and B, and $0 \leq s \leq 1$,*

$$\|A^{1/2}B^{1/2}\|_1 \leq (\mathrm{Tr}[A^s B^{(1-s)}])^{1/2} (\mathrm{Tr}[A])^{(1-s)/2} (\mathrm{Tr}[B])^{s/2}. \tag{55}$$

Specialising to states, $A = \sigma$ and $B = \rho$, the left-hand side is just $F(\rho, \sigma)$, while the right-hand side is equal to $Q_s(\rho, \sigma)^{1/2}$.

Proof. We rewrite $A^{1/2}B^{1/2}$ as a product of three factors

$$A^{1/2}B^{1/2} = A^{(1-s)/2}(A^{s/2}B^{(1-s)/2})B^{s/2},$$

apply Hölder's inequality on the 1-norm of this product, and exploit the relation

$$\|X^p\|_q = \|X\|_{pq}^p$$

(for $X \geq 0$) a number of times:

$$\begin{aligned}
\|A^{1/2}B^{1/2}\|_1 &= \|A^{(1-s)/2}(A^{s/2}B^{(1-s)/2})B^{s/2}\|_1 \\
&\leq \|A^{(1-s)/2}\|_{2/(1-s)} \, \|A^{s/2}B^{(1-s)/2}\|_2 \, \|B^{s/2}\|_{2/s} \\
&= (\mathrm{Tr}[A])^{(1-s)/2} \, \|A^{s/2}B^{(1-s)/2}\|_2 \, (\mathrm{Tr}[B])^{s/2} \\
&= (\mathrm{Tr}[A^s B^{(1-s)}])^{1/2} \, (\mathrm{Tr}[A])^{(1-s)/2} (\mathrm{Tr}[B])^{s/2}.
\end{aligned}$$

□

We now give a direct proof of inequality (30) that circumvents the proof of (29) and goes through in infinite dimensions. We state it in terms of general positive operators:

Theorem 7. *For positive operators A and B,*

$$\|A - B\|_1^2 + 4(\mathrm{Tr}[A^{1/2}B^{1/2}])^2 \leq (\mathrm{Tr}(A + B))^2. \tag{56}$$

Proof. Consider two general operators P and Q, and define their sum and difference as $S = P + Q$ and $D = P - Q$. We thus have $P = (S + D)/2$ and $Q = (S - D)/2$. Consider the quantity

$$\begin{aligned}
PP^* - QQ^* &= \frac{1}{4}\left((S + D)(S + D)^* - (S - D)(S - D)^*\right) \\
&= \frac{1}{2}(SD^* + DS^*).
\end{aligned}$$

Its trace norm is bounded above as

$$\begin{aligned}\|SD^* + DS^*\|_1/2 &\leq (\|SD^*\|_1 + \|DS^*\|_1)/2 \\ &= \|SD^*\|_1 \\ &\leq \|S\|_2\|D\|_2.\end{aligned}$$

In the last line we have used a specific instance of Hölder's inequality for the trace norm ([3] Cor. IV.2.6). Now put $P = A^{1/2}$ and $Q = B^{1/2}$, which exist by positivity of A and B, and which are themselves positive operators. We get $S, D = A^{1/2} \pm B^{1/2}$, hence

$$\|A - B\|_1 \leq \|A^{1/2} + B^{1/2}\|_2 \ \|A^{1/2} - B^{1/2}\|_2,$$

which upon squaring becomes

$$\begin{aligned}\|A - B\|_1^2 &\leq \mathrm{Tr}(A^{1/2} + B^{1/2})^2 \ \ \mathrm{Tr}(A^{1/2} - B^{1/2})^2 \\ &= \mathrm{Tr}(A + B + A^{1/2}B^{1/2} + B^{1/2}A^{1/2}) \\ &\quad \times \mathrm{Tr}(A + B - A^{1/2}B^{1/2} - B^{1/2}A^{1/2}) \\ &= (\mathrm{Tr}(A + B) + 2\,\mathrm{Tr}(A^{1/2}B^{1/2})) \\ &\quad \times (\mathrm{Tr}(A + B) - 2\,\mathrm{Tr}(A^{1/2}B^{1/2})) \\ &= (\mathrm{Tr}(A + B))^2 - 4(\mathrm{Tr}(A^{1/2}B^{1/2}))^2.\end{aligned}$$

□

References

1. Audenaert, K.M.R., Calsamiglia, J., Munoz-Tapia, R., Bagan, E., Masanes, Ll., Acin, A., Verstraete, F.: Discriminating States: The Quantum Chernoff Bound. Phys. Rev. Lett. **98**, 160501 (2007)
2. Bacon, D., Chuang, I., Harrow, A.: Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms. Phys. Rev. Lett. **97**, 170502 (2006)
3. Bhatia, R.: *Matrix Analysis*. Heidelberg: Springer, 1997
4. Bjelaković, I., Deuschel, J.D., Krüger, T., Seiler, R., Siegmund-Schultze, Ra., Szkoła, A.: A quantum version of Sanov's theorem. Commun. Math. Phys. **260**, 659–671 (2005)
5. Bjelaković, I., Deuschel, J.D., Krüger, T., Seiler, R., Siegmund-Schultze, Ra., Szkoła, A.: *Typical support and Sanov large deviation of correlated states*. http://arxiv.org/list/math/0703772, 2007
6. Blahut, R.E.: Hypothesis Testing and Information Theory. IEEE Trans. Inf. Theory **20**, 405–417 (1974)
7. Carlen, E.A., Lieb, E.H.: Advances in Math. Sciences, AMS Transl. (2) **189**, 59–62 (1999)
8. Chernoff, H.: A Measure of Asymptotic Efficiency for Tests of a Hypothesis based on the Sum of Observations. Ann. Math. Stat. **23**, 493–507 (1952)
9. Csiszár, I., Longo, G.: Studia Sci. Math. Hungarica **6**, 181–191 (1971)
10. Fuchs, C.A., van de Graaf, J.: Cryptographic distinguishability measures for quantum-mechanical states. IEEE Trans. Inf. Theory **45**, 1216 (1999)
11. Hayashi, M.: *Quantum Information, An Introduction*. Berlin: Springer, 2006
12. Hayashi, M.: *Error Exponent in Asymmetric Quantum Hypothesis Testing and Its Application to Classical-Quantum Channel coding*. http://arxiv.org/list/quant-ph/0611013, 2006
13. Hayashi, M.: Asymptotics of quantum relative entropy from a representation theoretical viewpoint. J. Phys. A: Math. Gen. **34**, 3413–3419 (2001)
14. Helstrom, C.W.: *Quantum Detection and Estimation Theory*. New York: Academic Press, 1976
15. Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. Commun. Math. Phys. **143**, 99–114 (1991)
16. Hoeffding, W.: Asymptotically Optimal Tests for Multinomial Distributions. Ann. Math. Statist. **36**, 369–401 (1965)
17. Holevo, A.S.: On Asymptotically Optimal Hypothesis Testing in Quantum Statistics. Theor. Prob. Appl. **23**, 411–415 (1978)

18. Kargin, V.: On the Chernoff distance for efficiency of quantum hypothesis testing. Ann. Statist. **33**, 959–976 (2005)
19. Lieb, E.H.: Convex trace functions and the Wigner-Yanase-Dyson conjecture. Adv. Math. **11**, 267–288 (1973)
20. Nagaoka, H.: *The Converse Part of The Theorem for Quantum Hoeffding Bound*. http://arxiv.org/list/quant-ph/0611289, 2006
21. Nussbaum, M., Szkoła, A.: *A lower bound of Chernoff type in quantum hypothesis testing*. http://arxiv.org/list/quant-ph/0607216, 2006
22. Nussbaum, M., Szkoła, A.: *The Chernoff lower bound in quantum hypothesis testing*. Preprint No. 69/2006, MPI MiS Leipzig
23. Ogawa, T., Hayashi, M.: On error exponents in quantum hypothesis testing. IEEE Trans. Inf. Theory **50**, 1368–1372 (2004)
24. Ogawa, T., Nagaoka, H.: Strong converse and Stein's lemma in quantum hypothesis testing. IEEE Trans. Inf. Theory **46**, 2428 (2000)
25. Petz, D.: Quasi-entropies for finite quantum states. Rep. Math. Phys. **23**, 57–65 (1986)
26. Ruskai, M.B., Lesniewski, A.: Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. J. Math. Phys. **40**, 5702–5742 (1999)
27. Uhlmann, A.: Sätze über Dichtematrizen. Wiss. Z. Karl-Marx Univ. Leipzig **20**, 633–653 (1971)
28. Uhlmann, A.: The 'transition probability' in the state space of a *-algebra. Rep. Math. Phys. **9**, 273 (1976)
29. van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge: University Press, 1998

Communicated by M.B. Ruskai