AN ASYMPTOTIC ERROR BOUND FOR TESTING MULTIPLE QUANTUM HYPOTHESES

BY MICHAEL NUSSBAUM\textsuperscript{1} AND ARLETA SZKOŁA\textsuperscript{2}

Cornell University and Max Planck Institute

We consider the problem of detecting the true quantum state among \( r \) possible ones, based on measurements performed on \( n \) copies of a finite-dimensional quantum system. A special case is the problem of discriminating between \( r \) probability measures on a finite sample space, using \( n \) i.i.d. observations. In this classical setting, it is known that the averaged error probability decreases exponentially with exponent given by the worst case binary Chernoff bound between any possible pair of the \( r \) probability measures. Define analogously the multiple quantum Chernoff bound, considering all possible pairs of states. Recently, it has been shown that this asymptotic error bound is attainable in the case of \( r \) pure states, and that it is unimprovable in general. Here we extend the attainability result to a larger class of \( r \)-tuples of states which are possibly mixed, but pairwise linearly independent. We also construct a quantum detector which universally attains the multiple quantum Chernoff bound up to a factor \( 1/3 \).

1. Introduction. Consider a finite set \( \Sigma = \{P_1, \ldots, P_r\} \) of probability distributions on a sample space \( \Omega \), and the problem of discriminating between them on the basis of observed i.i.d. data. It is well known that for the maximum likelihood decision rule, the error probability (Bayesian for uniform prior) decreases exponentially, with a rate given by the worst case among the possible pairwise hypothesis testing problems. Indeed if \( \xi_{\text{CB}}(P_i, P_j) \) represents the rate of exponential decay of the error probability for deciding between \( P_i \) and \( P_j \), given by the classical Chernoff bound

\[
\xi_{\text{CB}}(P_i, P_j) = -\log \inf_{0 \leq s \leq 1} \int (dP_i)^{1-s} (dP_j)^s,
\]

1 Supported by NSF Grants DMS-08-05632 and DMS-11-06460.
2 Supported by German Research Foundation (DFG) via the project “Quantum Statistics: Decision Problems and Entropic Functionals on State Spaces.”
AMS 2000 subject classifications. 62P35, 62G10.
Key words and phrases. Quantum statistics, density operators, Bayesian discrimination, exponential error rate, Holevo–Helstrom tests, quantum Chernoff bound.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in The Annals of Statistics, 2011, Vol. 39, No. 6, 3211–3233. This reprint differs from the original in pagination and typographic detail.
then the multiple Chernoff bound pertaining to the set $\Sigma$ has been defined as

$$\xi_{CB}(\Sigma) := \min \{\xi_{CB}(P_i, P_j) : P_i, P_j \in \Sigma, P_i \neq P_j\}$$

(Salikhov [28–30]). If $\pi_n$ is the maximum likelihood rule for sample size $n$, with values in $\{1, \ldots, r\}$, then, under a uniform prior on $\Sigma$

$$-\frac{1}{n} \log \Pr(\pi_n \neq i) \to \xi_{CB}(\Sigma) \quad \text{as } n \to \infty$$

and since $\pi_n$ is also Bayesian here, the quantity $\xi_{CB}(\Sigma)$ is the best possible asymptotic error exponent for any decision rule under a uniform prior.

On terminology. When large deviation type limits are written in logarithmic form as in (1.2), then the right-hand side $\xi_{CB}(\Sigma)$ is referred to as the rate of exponential decay or, in information theory, as the asymptotic error exponent, to be maximized by decision rules. Throughout the paper, we adhere to this formulation as a convenient equivalent to minimizing asymptotic error.

We consider here the analogous problem in a quantum statistical setting, where $\Sigma = \{\rho_1, \ldots, \rho_r\}$ is a set of density operators on the finite-dimensional complex Hilbert space $\mathbb{C}^d$. Recall that by definition a density operator $\rho$, describing the state of a physical system, is a complex, self-adjoint, positive semidefinite matrix satisfying the normalization condition $\text{tr}[\rho] = 1$. If all operators $\rho_i \in \Sigma$ commute, then the corresponding matrix representations are jointly diagonalizable, and the problem becomes one of discriminating between the associated finite probability distributions appearing on the matrix diagonal.

The starting point for our investigation is the recent extension of the Chernoff binary testing bound to the quantum setting [2,3,22]. In full analogy to the classical case, the quantum Chernoff bound specifies the asymptotic error in the decision problem between $\rho_i$ and $\rho_j$, based on a rule using the outcomes of measurements performed on $n$ copies of the basic quantum system.

The case of multiple hypotheses ($r > 2$) represented by quantum states has received some interest in the literature over the past three decades; cf. [6,12,14,15,25,26,34] and overviews in [7,9,11]. While in the binary case ($r = 2$) the optimal quantum test is described explicitly by the Holevo–Helstrom projections, in the case $r > 2$ only an implicit description in terms of an extremal problem is available (Holevo [14], Yuen, Kennedy, Lax [34]). Parthasarathy [26] has dubbed the quantum Bayes rule “quantum maximum likelihood,” in view of the fact that in the classical case, for a finite number of hypotheses, the Bayes rule for uniform prior is indeed maximum likelihood.

Numerous new contributions to multiple quantum hypothesis testing appeared in the very recent past, for example, [1,4,17–19,21,27,32,33]. The main focus has been on characterizing the Bayes rule of [14,34] and finding
approximations to it. We focus here on the asymptotics of the error probability based on measurements performed on $n$ copies of the basic quantum system. The true state is thus described by the $n$th tensor power $\rho_i \otimes^n$ of one of the original density operators $\rho_i \in \Sigma$. Parthasarathy [26] established consistency of the Bayes rule and also an exponential rate of decay of the error probability, without specifying the error exponent. The first step toward finding the optimal asymptotic error, for which a similar structure as in the classical case (1.2) was conjectured, was made in [24]. It was shown that if all $\rho_i$ are pure states $\text{rank}(\rho_i) = 1$, then the optimal asymptotic error is given by $\xi_{QCB}(\Sigma)$, defined as the worst case error for quantum discrimination between any pair of distinct states involved. Thus, the situation is indeed analogous to the classical case (1.2), and the quantity $\xi_{QCB}(\Sigma)$ describing the asymptotics of the error probability should be termed the multiple quantum Chernoff bound.

The fact that $\xi_{QCB}(\Sigma)$ is valid as a lower error bound is relatively straightforward to prove; for a precise statement of the result from [24]; cf. Theorem 1. Attainability for pure states has been shown in [24] by constructing a measurement based on a Gram–Schmidt orthonormalization of the $r$ unit vectors representing the $\rho_i \in \Sigma$. It should be mentioned that earlier Holevo [16] showed such a measurement to be an approximation to the Bayes rule. In [23], it was shown that without any restriction on the nature of the states, an asymptotic error $\xi_{QCB}(\Sigma)$ is achievable up to a factor which is between $2/r(r-1)$ and 1, for $r$ being the number of hypotheses.

In the present paper, we develop a new decision rule generalizing two known asymptotically optimal ones, in the following sense: if all states commute, the method reduces to classical maximum likelihood (as does the Bayes rule of [14, 34]). If all states are pure, then it coincides with the orthonormalization algorithm of [24]. We establish that this rule attains asymptotic error $\xi_{QCB}(\Sigma)$ for a class of $r$-tuples of states which fulfill Condition (LI) below. The condition allows for mixed states but excludes faithful ones (full rank density matrices). We then show that a modified version of our rule is near optimal, in the sense that it attains at least $\frac{1}{3} \xi_{QCB}(\Sigma)$ universally.

The outline of our paper is as follows. In Section 2, we introduce notation, specify the mathematical framework, and state precisely our main results in Theorems 2 and 3. Some further discussion of the quantum Bayes rule, of results in statistics resembling the multiple Chernoff bound and other topics follows at the end of that section. In Section 3, our new quantum decision rule is developed, along with Lemma 1 providing a basic error bound. Section 4 treats the case of pairwise linearly independent states [Condition (LI) and Theorem 2]. Section 5 shows how our decision rule reduces to maximum likelihood in the commuting case, such that Lemma 1 reproduces the multiple Chernoff bound of [28, 29]. Section 6 concerns the general attainability of the near optimal error bound (Theorem 3).
2. Notation and preliminaries. We will describe here the formalism for the simplest possible nonclassical setup of discrimination between several quantum hypotheses. A density matrix $\rho$ is a complex, self-adjoint, positive, $d \times d$ matrix satisfying the normalization condition $\text{tr}[\rho] = 1$, where $\text{tr}[\cdot]$ is the trace operation. Here, “positive” means nonnegative definite. We identify a $d \times d$ density matrix with a quantum state on $\mathbb{C}^d$; we also use “matrix” and “operator” interchangeably. The $r$ hypotheses are described by states $H_i: \rho = \rho_i, i = 1, \ldots, r$. Physically discriminating between these states corresponds to performing a measurement on the quantum system. Mathematically a quantum decision rule with $r$ possible outcomes is a set of complex self-adjoint positive matrices $d \times d$ matrices $E = \{E_1, \ldots, E_r\}$ satisfying $\sum_{i=1}^r E_i = 1$ where $1$ is the unit matrix. The $r$-tuple $E$ is often called a POVM (positive operator valued measure); we will refer to it as a quantum multiple test or a quantum detector. In the special case where all $E_i$ are projections, the $r$-tuple $E$ is called a PVM (projection valued measure) or von Neumann measurement. The individual success probability, that is, the probability to accept hypothesis $H_i$ when $\rho_i$ is the true state, is given by

$$Succ_i(E) := \text{tr}[\rho_i E_i].$$

The corresponding individual error probability, that is, the probability of rejecting the true state $\rho_i$ according to the decision rule, is

$$Err_i(E) = 1 - Succ_i(E) = \text{tr}[\rho_i (1 - E_i)].$$

The total (averaged) error probability is then

$$Err(E) := \frac{1}{r} \sum_{i=1}^r Err_i(E) = \frac{1}{r} \sum_{i=1}^r \text{tr}[\rho_i (1 - E_i)].$$

The above describes the basic setup where the finite dimension $d$ is arbitrary and the hypotheses are equiprobable. We consider the quantum analog of having $n$ i.i.d. observations. For this, the $r$ hypotheses are assumed to be $\rho_i^{\otimes n}$, $i = 1, \ldots, r$, where $\rho^{\otimes n}$ is the $n$-fold tensor product of $\rho$ with itself (a $d^n \times d^n$ matrix). The detectors $E = \{E_1, \ldots, E_r\}$ now operate on the states $\rho_i^{\otimes n}$, that is, the dimension of the components $E_i$ is $d^n \times d^n$, but $E_i$ need not have tensor product structure. The corresponding total error probability of a detector $E$ is now

$$Err_n(E) = 1 - \frac{1}{r} \sum_{i=1}^r \text{tr}[\rho_i^{\otimes n} E_i].$$

For the case of two hypotheses $r = 2$, the Bayes test for each $n \in \mathbb{N}$ is known to be the Holevo–Helstrom hypothesis test. It is given by the detector $E_{(n)}^* = \frac{1}{n} \sum_{i=1}^n \rho_i^{\otimes n}$.
\{1 - \Pi_n^*, \Pi_n^*\} \text{ where}\\
\Pi_n^* = \text{supp}(\rho_{2}^\otimes n - \rho_{1}^\otimes n)_+,

where \text{supp} \alpha \text{ is the projection onto the space spanned by the columns of } \alpha \text{ and } a_+ \text{ denotes the positive part of a self-adjoint operator } a. \text{ Thus, if } a = \sum_i \lambda_i S_i \text{ is the spectral decomposition using projections } S_i \text{, then } a_+ := \sum_{\lambda_i > 0} \lambda_i S_i \text{ and supp } a_+ = \sum_{\lambda_i > 0} S_i. \text{ The Bayes test is unique up to a possible reassignment of the projections } S_i \text{ corresponding to zero eigenvalues of } a = \rho_{2}^\otimes n - \rho_{1}^\otimes n. \text{ For } r > 2, \text{ the Bayes detector has been described in } [14, 34]. \text{ Explicit expressions for its } r \text{ components are not known in general; for the convenience of the reader, we present the available implicit description below at the end of this section.}

If for a sequence of detectors } E_n \text{ the limit } \lim_{n \to \infty} -\frac{1}{n} \log \text{Err}_n(E_n) \text{ exists, we refer to it as the (asymptotic) error exponent. For two density matrices } \rho_1 \text{ and } \rho_2, \text{ the quantum Chernoff bound is defined by}

\begin{equation}
\xi_{\text{QCB}}(\rho_1, \rho_2) := -\log \inf_{0 \leq s \leq 1} \text{tr}[\rho_1^{1-s} \rho_2^s].
\end{equation}

(2.1)

The basic properties of } \xi_{\text{QCB}}(\rho_1, \rho_2) \text{ have been discussed in } [3]. \text{ Some distance-like properties have been noted by Calsamiglia et al. } [8]. \text{ For the binary discrimination problem, it is known that the Holevo–Helstrom (Bayes) detector } E_n^* \text{ satisfies}

\begin{equation}
\lim_{n \to \infty} -\frac{1}{n} \log \text{Err}_n(E_n^*) = \xi_{\text{QCB}}(\rho_1, \rho_2),
\end{equation}

(2.3)

thus specifying } \xi_{\text{QCB}}(\rho_1, \rho_2) \text{ as the optimal error exponent (cf. } [2, 3, 22]), \text{ and providing the quantum analog of the classical Chernoff bound, that is, (1.2) for } r = 2. \text{ For a set } \Sigma = \{\rho_1, \ldots, \rho_r\} \text{ of density operators on } \mathbb{C}^d, \text{ where } r \geq 2, \text{ we have introduced in } [24] \text{ the multiple quantum Chernoff bound } \xi_{\text{QCB}}(\Sigma)

\begin{equation}
\xi_{\text{QCB}}(\Sigma) := \min\{\xi_{\text{QCB}}(\rho_i, \rho_j) : 1 \leq i < j \leq r\}.
\end{equation}

(2.2)

If all the states are jointly diagonalizable (commuting), then (2.2) reduces to the classical multiple Chernoff bound (1.1), as it was defined in [28, 30] for hypotheses represented by probability distributions. Taking the minimum over different pairs of hypotheses corresponds to the worst case in any of the associated binary hypothesis testing problems. The following well-known result shows that } \xi_{\text{QCB}}(\Sigma) \text{ as a rate exponent cannot be exceeded (cf. } [24], \text{ Theorem 1).}

\textbf{Theorem 1.} \textit{Let } \Sigma = \{\rho_1, \ldots, \rho_r\} \text{ be a finite set of hypothetic states on } \mathbb{C}^d. \text{ Then for any sequence } \{E_n\}_{n \in \mathbb{N}} \text{ of quantum detectors relative to } \Sigma^\otimes n, \text{ respectively, one has}

\begin{equation}
\lim_{n \to \infty} \sup -\frac{1}{n} \log \text{Err}_n(E_n) \leq \xi_{\text{QCB}}(\Sigma).
\end{equation}

(2.3)
The above theorem has been extended in [23] to the case of quantum hypotheses which correspond to identically distributed but not necessary independent observations. The corresponding upper bound in (2.3) is then replaced by a mean generalized Chernoff distance, as introduced in [13] for a stationary observation scheme in binary case. In [23], it was also shown, again in a wider model corresponding to a class of correlated observations, that quantum detectors with an exponential decay of $\text{Err}(E_n)$ can be constructed, with error exponent $\phi \xi_{QCB}(\Sigma)$ where $2/r(r-1) \leq \phi \leq 1$. The method used in [23] yields a factor $\phi$ which may be close to one for special ensembles of states, but the guaranteed factor $2/r(r-1)$ decreases with the number of hypotheses.

The following two theorems represent our main results. The support $\text{supp}(\rho)$ of a state $\rho$ is the subspace of $\mathbb{C}^d$ spanned by its columns. Consider:

**Condition (LI).** $\text{supp}(\rho_i) \cap \text{supp}(\rho_j) = \{0\}$ for all $i \neq j$.

The condition is equivalent to requiring that $\rho_i$ and $\rho_j$ are linearly independent, in the sense that for any two bases of $\text{supp}(\rho_i)$ and $\text{supp}(\rho_j)$, the union set of vectors is linearly independent. This is obviously fulfilled for a set $\Sigma$ of $r$ distinct pure states, but the condition allows for mixed states if $d > 2$. Indeed, Condition (LI) restricts the dimension of the supports $\text{supp}(\rho_i)$ according to the inequality $\text{supp}(\rho_i) + \text{supp}(\rho_j) \leq d$ that is valid for all $i \neq j$. However, as long as none of the density matrices is of full rank, that is, rank equal to $d$, no constraints on the number $r$ of distinct hypothetic states are imposed by Condition (LI).

**Theorem 2.** Let $\Sigma$ be a finite set of states on $\mathbb{C}^d$ fulfilling Condition (LI). Then there exists a sequence $\{E_n\}_{n \in \mathbb{N}}$ of quantum detectors relative to $\Sigma^\otimes n$, respectively, such that

$$\lim_{n \to \infty} -\frac{1}{n} \log \text{Err}(E_n) = \xi_{QCB}(\Sigma).$$

Due to the following theorem in the i.i.d. situation—as considered in the present paper— an error exponent of $\frac{1}{3} \xi_{QCB}(\Sigma)$ can always be achieved, independently of both the (finite) number $r$ of hypotheses and the special configuration of the corresponding states.

**Theorem 3.** Let $\Sigma$ be a finite set of states on $\mathbb{C}^d$. Then there exists a sequence $\{E_n\}_{n \in \mathbb{N}}$ of quantum detectors relative to $\Sigma^\otimes n$, respectively, such that

$$\liminf_{n \to \infty} -\frac{1}{n} \log \text{Err}(E_n) \geq \frac{1}{3} \xi_{QCB}(\Sigma).$$

Our results are constructive in the sense that we provide an explicitly computable quantum detector attaining the bounds. This detector reduces
to classical maximum likelihood in the commuting case (cf. Section 5), as
does the Bayes rule, and hence attains the optimal rate exponent (1.2); cf. [28]. Thus our method can be seen as an alternative to the quantum
Bayes rule. The above error bound is a fortiori true for the latter, and
also for computable approximations to it having at most 2 times its error
probability (Tyson [32, 33]). Our results along with those of [23] allow the
counterexample to consider in Theorem 3, the factor 1/3 can be removed; cf. also the
discussion point 5 beyond.

To further discuss the context of the main results, we note the following
points.

1. The quantum Bayes rule (Holevo [14], Yuen et al. [34]; cf. also Parthasarathy [25, 26] and Hayashi [11]). Let \( \Sigma = \{\rho_1, \ldots, \rho_r\} \) be such that all \( \rho_i \) are
   distinct states on \( \mathbb{C}^d \). Let be
   \( E = \{E_1, \ldots, E_r\} \) where \( E_i \) are positive self-adjoint \( d \times d \) with \( \sum_{i=1}^{r} E_i = I \). Define
   \[
   (2.4) \quad \mu = \max_{E \in \mathcal{E}} \text{Succ}(E) := \max_{E \in \mathcal{E}} \sum_{i=1}^{r} \text{tr}[\rho_i E_i].
   \]
   Then there exists a unique operator \( M \) on \( \mathbb{C}^d \) satisfying
   \[
   \text{tr}[M] = \mu, \quad M \geq \rho_i, \quad i = 1, \ldots, r.
   \]
   Maximizers \( E^* = \{E_1^*, \ldots, E_r^*\} \in \mathcal{E} \) of (2.4) exist by compactness and conti-
   nuity, and any such maximizer (a Bayes rule) satisfies
   \[
   (2.5) \quad (M - \rho_i)E_i^* = E_i^*(M - \rho_i) = 0, \quad i = 1, \ldots, r.
   \]
   A proof using only elementary calculus can be found in [25], Theorem 3.1.

If \( r = 2 \), then the Holevo–Helstrom rule \( \{1 - \Pi, \Pi\} \) for \( \Pi = \text{supp}(\rho_2 - \rho_1) \) is a Bayes rule. If all states \( \rho_i, i = 1, \ldots, r \) commute, hence \( \rho_i \) can be rep-
  reseted as diagonal matrix with diagonal elements \( p_{ij}, j = 1, \ldots, d \), then \( M \) is a diagonal matrix with diagonal elements \( m_j = \max_{i=1,\ldots,r} p_{ij} \). Then any
   Bayes rule \( E^* \) with diagonal matrices \( E_i^* \) is maximum likelihood, assigning 0
   or 1 to the diagonals of \( E_i^* \), such that a 1 is at \( (j, j) \) only if \( p_{ij} = m_j \).

2. Pretty good measurement. Let \( \Sigma = \{\rho_1, \ldots, \rho_r\} \) be a set of pairwise distinct density operators with respective a priori probabilities \( p_i \). Define the positive semi-definite operator \( \rho = \sum_{i=1}^{r} p_i \rho_i \). A possible quantum detector
   relative to \( \Sigma \) is of the form
   \[
   E_i^{\text{PGM}} := \rho^{-1/2} p_i \rho_i \rho^{-1/2}, \quad i = 1, \ldots, r.
   \]
   (The inverse is understood to be taken on the support of \( \rho \) only.) It represents
   the widely investigated POVMs called pretty good measurements (PGM).
These are known to be a good approximation of the quantum Bayes rule: if $\Sigma$ is a set of pure states, then the averaged success probability $\text{Succ}(\text{PGM}) = \sum_{i=1}^{r} p_i \text{Succ}_i(\text{PGM})$ is lower bounded by a result of Barnum and Knill [5],

$$\text{Succ}(\text{PGM}) \geq \left( \max_{E \in \mathcal{E}} \sum_{i=1}^{r} p_i \text{Succ}_i(E) \right)^2,$$

where $\mathcal{E}$ denotes the set of quantum detectors relative to $\Sigma$. For further bounds on $\text{Succ}(\text{PGM})$ refering also to the general case of mixed states see [21] and references therein. To the best of our knowledge, in the literature, the PGM has not been successfully used to study the optimal asymptotic error exponent.

3. Classical results resembling the multiple Chernoff bound. Let $\Sigma$ be a statistical experiment having finite parameter space $\{\theta_1, \ldots, \theta_r\}$, and $\Sigma^n$ be the associated product experiment corresponding to i.i.d. observations. Torgersen [31] considered $\delta(\Sigma_n, \Sigma_a)$, the deficiency (in the Le Cam sense) of $\Sigma_n$ with respect to the fully informative experiment $\Sigma_a$. Here $\Sigma_a$ may be identified, up to equivalence, with the set of $r$ point masses concentrated on $\theta_1, \ldots, \theta_r$. It was shown ([31], Theorem 4.2) that

$$-\frac{1}{n} \log \delta(\Sigma_n, \Sigma_a) \to \xi_{\text{CB}}(\Sigma) \quad \text{as } n \to \infty$$

with $\xi_{\text{CB}}(\Sigma)$ defined in (1.1). Krob and von Weizsäcker [20] considered the Shannon capacity $C(\Sigma^n)$ of $\Sigma^n$ construed as a communication channel, and showed that $C(\Sigma^n)$ approaches its upper bound $\log r$ exponentially quickly, with rate exponent $\xi_{\text{CB}}(\Sigma)$:

$$-\frac{1}{n} \log(\log r - C(\Sigma^n)) \to \xi_{\text{CB}}(\Sigma) \quad \text{as } n \to \infty.$$

4. Linearly independent states. A stronger condition than Condition (LI) would be that all states $\{\rho_1, \ldots, \rho_r\}$ are linearly independent (in the sense that for any selected $r$ bases of the spaces $\text{supp}(\rho_i), i = 1, \ldots, r$, the union set of vectors is linearly independent.) The paper [10] gives examples of such ensembles of states, and shows that under this stronger condition, the Bayes detector $E = \{E_1, \ldots, E_r\}$ consists of projections $E_i$ (is a von Neumann measurement or PVM). Lemma 2 implies that our pairwise Condition (LI) on $\Sigma$ implies the stronger one for $\Sigma^{\otimes n}$, that is, the states $\rho_1^{\otimes n}, \ldots, \rho_r^{\otimes n}$ are linearly independent for sufficiently large $n$.

5. Other special ensembles. It can be shown that there are other situations besides Condition (LI) where the error exponent $\xi_{\text{QCB}}(\Sigma)$ is attainable exactly. One condition, which does not impose any rank restrictions on the states and thus allows for full rank density matrices $\rho_i$, is as follows. For a set $\Sigma = \{\rho_1, \ldots, \rho_r\}$ of density operators where $r > 2$, let $\Sigma_{(i,j)}$ be the set
where a pair $\rho_i, \rho_j$, is removed, that is, $\Sigma(i,j) = \Sigma \setminus \{\rho_i, \rho_j\}$ for $1 \leq i < j \leq r$. Assume there is a pair $(i, j)$ such that
\[
\xi_{QCB}(\Sigma) \leq \frac{1}{6} \xi_{QCB}(\Sigma(i,j)).
\]
This condition can replace Condition (LI) in the statement of Theorem 2, that is, the multiple quantum Chernoff bound is then attainable. The proof, not to be presented here, consists in a combination of the sample splitting method of [23] with Theorem 3. This further supports the conjecture that the result of Theorem 3 is not final and the factor $1/3$ there may be removed.

Throughout the paper, we use the notation $j \in \{1, \ldots, d\}$ and $j \in [1, d]$ interchangeably.

3. The detection algorithm. In this section, we construct a sequence $E^{(n)}$, $n \in \mathbb{N}$, of quantum detectors for $\Sigma \otimes n$. The construction does not rely on the existence of asymptotically optimal quantum tests for the binary case. It is rather a modification of a construction used in [24] which yields asymptotically optimal quantum tests for a set of pure states. At the same time, it represents a quantum extension of the classical ML method, different from the Bayes rule described in (2.5).

Consider again the classical case where a set $\Sigma = \{P_1, \ldots, P_r\}$ of probability distributions is given on a finite sample space $\Omega$ with cardinality $d$. An obvious algorithmic description of a ML decision rule $\varphi: \Omega \to \{1, \ldots, r\}$ is as follows. For each $\omega \in \Omega$, find a maximal element in $\{P_i(\omega)\}_{i=1}^r$, say $P_i^*(\omega)$, and then decide $\varphi(\omega) = i^*$. Alternatively, one may successively find the largest probabilities among all $P_i(\omega)$, identify which $P_i$ and which $\omega$ they are from, and assign a corresponding decision on this $\omega$. This iterative approach can be expressed in a simple algorithm in pseudocode as follows.

**Algorithm 1** (Classical ML rule).

**Initialize.** Let $\Pi_0 = \{P_i(\omega), i = 1, \ldots, r, \omega \in \Omega\}$ be the $r \times d$-matrix of all probabilities.

For $s = 1$ to $d$:

(i) In $\Pi_{s-1}$ find a maximal entry, $P_i^*(\omega^*)$ say. Set $\omega_s = \omega^*$ and decide $\varphi(\omega_s) = i^*$.

(ii) In $\Pi_{s-1}$, all $P_i(\omega_s), i = 1, \ldots, r$ are replaced by $-1$; the resulting $r \times d$-matrix is $\Pi_s$.

After $s = d$ steps, the matrix $\Pi_d$ has entries $-1$ only (a value serving as an indicator, chosen to be smaller than any probability). We also have enumerated the elements of $\Omega$ as $\omega_1, \ldots, \omega_d$; on each of these, a decision $\varphi(\omega_s)$ has been made, which is ML by construction.

In the quantum case, there is no initial sample space $\Omega$; it only appears after defining a measurement, which in our context can be taken to be an
orthonormal basis \{e_s\}_{s=1}^d of \mathbb{C}^d. After this basis is fixed, the sample space \(\Omega = \{\omega_s\}_{s=1}^d\) can be identified with the basis itself, or more precisely with the set of pertaining projectors, such that each \(\omega_s = |e_s\rangle\langle e_s|\), and a classical nonrandomized decision rule \(\varphi: \Omega \to \{1, \ldots, r\}\) has to be found. Then the quantum decision rule \(E = \{E_1, \ldots, E_r\}\) is given by the PVM

\[
E_i = \sum_{s: \varphi(|e_s\rangle\langle e_s|) = i} |e_s\rangle\langle e_s|, \quad i = 1, \ldots, r. \tag{3.1}
\]

The algorithm we will describe constructs the basis elements \(e_j\) and the pertaining decision \(\varphi(\cdot)\) iteratively, combining the ML principle underlying Algorithm 1 with a Gram–Schmidt orthogonalization.

For each \(1 \leq i \leq r\) let

\[
\rho_i = \sum_{j=1}^d \lambda_{ij} |v_{ij}\rangle\langle v_{ij}| \tag{3.2}
\]

be a spectral decomposition of the density matrix \(\rho_i\), where \(\lambda_{ij}, j = 1, \ldots, d\), are the eigenvalues of \(\rho_i\) appearing with their multiplicity, in arbitrary order, and \(|v_{ij}\rangle\) are the corresponding normalized eigenvectors in \(\mathbb{C}^d\). Here \(|v_{ij}\rangle\) denotes the dual vector such that in this notation \(|v_{ij}\rangle\langle v_{ij}|\) describes an orthogonal projector onto the one-dimensional subspace of \(\mathbb{C}^d\) spanned by \(|v_{ij}\rangle\).

We stress that zero eigenvalues are included with their multiplicity since \(d\) in (3.2) is the dimension of \(\rho_i\).

**Algorithm 2** (A quantum decision rule).

**Initialize.** Let \(\Lambda_0 = \{\lambda_{ij}, i = 1, \ldots, r, j = 1, \ldots, d\}\) be the \(r \times d\)-matrix of all eigenvalues. Let \(e_0 = 0\) be the zero vector in \(\mathbb{C}^d\).

For \(s = 1 \to d\):

(i) In \(\Lambda_{s-1}\) find a maximal entry, \(\lambda_{i^*, j^*}\) say. Set \(e_s\) to be a unit vector such that

\[
e_s \in \text{span}(e_1, \ldots, e_{s-1}, v_{i^*, j^*}), \quad e_s \perp \text{span}(e_1, \ldots, e_{s-1}) \tag{3.3}
\]

and decide \(\varphi(|e_s\rangle\langle e_s|) = i^*\).

(ii) In \(\Lambda_{s-1}\), all \(\lambda_{ij}\) such that \(v_{ij} \in \text{span}(e_1, \ldots, e_s)\) are replaced by \(-1\); the resulting \(r \times d\)-matrix is \(\Lambda_s\).

Again, after \(s = d\) steps, the matrix \(\Lambda_s\) has entries \(-1\) only. We also have constructed an orthonormal basis \(e_1, \ldots, e_d\) and on each of these, an associated decision \(\varphi(|e_s\rangle\langle e_s|)\). The crucial step (3.3) is recognized to define a Gram–Schmidt orthogonalization process. The quantum detector now is given by the PVM (3.1).

To bound the error probability of this detector, we need to introduce some further notation. In each step \(s\) of Algorithm 2, in part (i) we have selected
an index pair \((i^*, j^*)\) where \(\lambda_{i^*,j^*}\) is a maximal entry of the matrix \(\Lambda_{s-1}\); set \((i(s), j(s)) = (i^*, j^*)\). The sequence of vectors \(
{v_i(s), j(s)}_{s=1}^d\) is linearly independent by construction. For each \(s \in \n{1, d}\) define a \(d \times s\) matrix \(V_s\)

\[
V_s := (v_{i(1), j(1)}, \ldots, v_{i(s), j(s)}),
\]

that is, the columns of \(V_s\) are the vectors \(v_{i(k), j(k)}, k \in \n{1, s}\). We refer to the \(s \times s\)-matrix

\[
\Gamma_s := V_s^* V_s
\]

as a Gram matrix of \(\n{v_{i(k), j(k)}}_{k=1}^s\). For each \(s \in \n{1, d}\) the matrix \(\Gamma_s\) is nonsingular and the matrix

\[
P_s := V_s (V_s^* V_s)^{-1} V_s^* = V_s \Gamma_s^{-1} V_s^*
\]

represents an orthogonal projection onto \(\text{span}(v_{i(1), j(1)}, \ldots, v_{i(s), j(s)})\), an \(s\)-dimensional subspace of \(\mathbb{C}^d\). Additionally, we set \(P_0 = 0\) and define for \(s \in \n{1, d}\)

\[
P(s) := P_s - P_{s-1}.
\]

Observe that the \(P(s)\) represent one-dimensional orthogonal projectors, which are mutually orthogonal, such that \(P(s) = |e_s\rangle \langle e_s|\) for the unit vectors \(e_s\) defined in \((3.3)\). The latter can be taken to be \(e_s = \|P(s) v_{i(s), j(s)}\|^{-1} P(s) v_{i(s), j(s)}\) (or a sign changed version).

Furthermore, define an index \(N\) as

\[
N = \max\{s \in \n{1, d} : \lambda_{i(s), j(s)} > 0\}.
\]

It can be seen from the proof of Lemma 1 below that if \(N < d\), then \(N\) can serve as an early stopping index for Algorithm 2, in the following sense: the obtained set of orthonormal vectors \(\n{e_s}_{s=1}^N\) can be completed to a basis of \(\mathbb{C}^d\) in an arbitrary way and the decisions \(\varphi(e_s), s > N\), can be taken arbitrarily. This is related to the fact that for all further steps \(s > N\), the remaining eigenvalues \(\lambda_{ij}\) listed in the matrix \(\Lambda_s\) are 0; in Algorithm 1 this corresponds to the case that there exist \(\omega \in \Omega\) which are outside the support of all \(P_i\).

We use the notation \(\lambda_{\min}(\cdot)\) for the minimal eigenvalue of a self-adjoint matrix.

**Lemma 1.** Let \(\Sigma = \n{\rho_i}_{i=1}^r\) be an arbitrary set of density matrices on \(\mathbb{C}^d\). Then the detector \(E = \n{E_i}_{i=1}^r\) constructed in Algorithm 2 fulfills

\[
\text{Err}(E) \leq \lambda_{\min}^{-1}(\Gamma_N) r^{-1} \sum_{1 \leq i, j \leq r, \rho_i \neq \rho_j} \inf_{s \in [0, 1]} \text{tr}[\rho_i^{1-s} \rho_j^s],
\]

where \(\Gamma_N\) is the Gram matrix according to \((3.5)\) for index \(s = N\) defined in \((3.7)\).

**Proof.** Define \(J\) to be the subset of \([1, r] \times [1, d]\) consisting of all pairs \((i(s), j(s)), s \in [1, d]\), and \(J_i := \n{j \in [1, d] : (i, j) \in J}\). For given \(i \in [1, r],\)
consider the corresponding individual success probability of the detector defined by (3.1):

\[
\text{Succ}_s(E) = \text{tr}[\rho_i E_i] = \sum_{j=1}^d \lambda_{ij} \text{tr}[v_{ij}] |v_{ij} \rangle \langle E_i| \geq \sum_{j \in J_i} \lambda_{ij} \text{tr}[v_{ij}] |v_{ij} \rangle \langle E_i|,
\]

where the right-hand side is set 0 if the set \( J_i \) is empty. For any \( j \in J_i \), let \( s(i, j) \) be the unique index \( s \in [1, d] \) such that \( (i, j) = (i(s), j(s)) \). If \( J_i \) is nonempty, then

\[
E_i = \sum_{j \in J_i} |e_{s(i,j)} \rangle \langle e_{s(i,j)}| = \sum_{j \in J_i} P(s(i,j))
\]

with \( P(s) \) defined in (3.6), hence \( E_i \geq P(s(i,j)) \) for all \( j \in J_i \), in the sense of the ordering for self-adjoint matrices. This implies

\[
\text{tr}[v_{ij}] |v_{ij} \rangle \langle E_i| \geq |v_{ij} \rangle \langle v_{ij}| P(s(i,j)) |v_{ij} \rangle = |v_{ij} \rangle \langle v_{ij}| P_s(i,j) |v_{ij} \rangle - |v_{ij} \rangle \langle v_{ij}| P_s(i,j)_{-1} |v_{ij} \rangle.
\]

Recall that the matrices \( P_s \) are constructed as orthogonal projectors onto \( \text{span}(v_{i(1),j(1)}, \ldots, v_{i(s),j(s)}) \), and since for \( j \in J_i \) and \( s = s(i, j) \) we have \( v_{ij} = v_{i(s),j(s)} \), it follows that for \( s = s(i, j) \)

\[
|v_{ij} \rangle \langle v_{ij}| P_s(i,j) |v_{ij} \rangle = |v_{i(s),j(s)} \rangle \langle v_{i(s),j(s)}| P_s |v_{i(s),j(s)} \rangle = 1.
\]

Consequently,

\[
\text{Succ}_s(E) \geq \sum_{j \in J_i} \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| P(s(i,j)) |v_{ij} \rangle = \sum_{j \in J_i} \lambda_{ij} - \sum_{j \in J_i} \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| P_s(i,j)_{-1} |v_{ij} \rangle.
\]

For the individual error probability under state \( \rho_i \) this implies, setting \( J_i^c := [1, d] \setminus J_i \),

\[
\text{Err}_i(E) = 1 - \text{Succ}_i(E) \leq \sum_{j \in J_i} \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| P_s(i,j)_{-1} |v_{ij} \rangle + \sum_{j \in J_i^c} \lambda_{ij} \]

\[= S_1 + S_2,
\]

say.

**Bounding the term \( S_1 \).** Consider only those terms in

\[
S_1 = \sum_{j \in J_i} \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| P_s(i,j)_{-1} |v_{ij} \rangle,
\]

where \( \lambda_{ij} > 0 \). Since for \( j \in J_i \) we have \( \lambda_{ij} = \lambda_{i(s),j(s)} \) for some \( s = s(i, j) \in [1, d] \), the assumption \( \lambda_{ij} > 0 \) implies \( s(i, j) \leq N \). Recall that \( P_s = V_s \Gamma_s^{-1} V_s^* \), \( s = 1, \ldots, d \), and that each \( \Gamma_{s-1} \) is a principal submatrix of \( \Gamma_s \). As a consequence, \( \lambda_{\min}(\Gamma_s) \geq \lambda_{\min}(\Gamma_N) \), \( s \in [1, N] \), and for \( j \in J_i \), if not \( s(i, j) = 1 \),

\[
\lambda_{ij} |v_{ij} \rangle \langle v_{ij}| P_s(i,j)_{-1} |v_{ij} \rangle \leq \lambda_{\min}(\Gamma_s(i,j)_{-1}) \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| V_s(i,j)_{-1} V_s^* \]

\[\leq \lambda_{\min}(\Gamma_N) \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| V_s(i,j)_{-1} V_s^* \]

\[\leq \lambda_{\min}(\Gamma_N) \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| V_s(i,j)_{-1} V_s^* \]

\[\leq \lambda_{\min}(\Gamma_N) \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| V_s(i,j)_{-1} V_s^* \]

\[\leq \lambda_{\min}(\Gamma_N) \lambda_{ij} |v_{ij} \rangle \langle v_{ij}| V_s(i,j)_{-1} V_s^* \]
where $\lambda_{\min}(\Gamma_N) > 0$ by construction. Formally setting $V_0 = 0 \in \mathbb{C}^d$, the above inequality holds also if $s(i,j) = 1$. One obtains the upper bound

$$S_1 = \sum_{j \in J_i} \lambda_{ij} \langle v_{ij} | P_{s(i,j)-1} | v_{ij} \rangle$$

(3.12)

$$\leq \lambda_{\min}^{-1}(\Gamma_N) \sum_{j \in J_i} \lambda_{ij} \langle v_{ij} | V_{s(i,j)-1} V_{s(i,j)-1}^* | v_{ij} \rangle$$

$$= \lambda_{\min}^{-1}(\Gamma_N) \sum_{j \in J_i} \lambda_{ij} \sum_{k=1}^{s(i,j)-1} |\langle v_{i(k),j(k)} | v_{ij} \rangle|^2.$$

The identity above is based on the fact that the columns of $V_{s(i,j)-1}$ are given by the vectors $v_{i(k),j(k)}$, $k \in [1, s(i,j) - 1]$. Note that in (3.12), for every pair of vectors occurring in $\langle v_{i(k),j(k)} | v_{ij} \rangle$ the corresponding eigenvalues satisfy $\lambda_{i(k),j(k)} \geq \lambda_{ij}$ by construction. This implies

$$\lambda_{ij} \leq \lambda_{ij}^{1-s} \lambda_{i(k),j(k)}^s$$

(3.13)

for every $s \in [0,1]$. Recall that every eigenvalue $\lambda_{i(k),j(k)}$ pertains to a state $\rho_{i(k)}$; we may assume $i(k) \neq i$, since otherwise necessarily $j(k) \neq j$ and thus $\langle v_{i(k),j(k)} | v_{i(k),j} \rangle = 0$. Setting now $m = i(k)$ and assuming $m \neq i$, we will apply inequality (3.13) for an exponent $s$ which is allowed to depend on $i$ and $m$. Denote by $s(i,m) = s(m,i) \in [0,1]$ the exponent associated to the pair of indices $(i,m) \in [1,r]^2$. Observe that for any subset $D_m \subset [1,d]$

$$\sum_{j \in J_i, j' \in D_m} \sum_{j' \in D_m} \lambda_{ij}^{1-s(i,m)} \lambda_{m,j'}^{s(i,m)} |\langle v_{m,j'} | v_{ij} \rangle|^2$$

(3.14)

$$\leq \sum_{j \in J_i, j' = 1} |\langle v_{m,j'} | v_{ij} \rangle|^2,$$

where on the right-hand side of the inequality we are just adding positive reals. It now follows from (3.12), (3.13) and (3.14) that

$$S_1 \leq \lambda_{\min}^{-1}(\Gamma_N) \sum_{j \in J_i, 1 \leq m \leq r, m \neq i} \sum_{j' = 1}^d \lambda_{ij}^{1-s(i,m)} \lambda_{m,j'}^{s(i,m)} |\langle v_{m,j'} | v_{ij} \rangle|^2.$$

**Bounding the term $S_2$.** We have

$$S_2 = \sum_{j \in J_i^c} \lambda_{ij} = \sum_{j \in J_i^c} \lambda_{ij} \langle v_{ij} | v_{ij} \rangle.$$

Consider only those terms where $\lambda_{ij} > 0$. By definition of $J_i^c$, there exists $s \in [1,d]$ such that $v_{ij} \in \text{span}(v_{i(1),j(1)}, \ldots, v_{i(s),j(s)})$. Then $\lambda_{i(k),j(k)} \geq \lambda_{ij}$ for
\( k \in [1, s] \), hence \( \lambda_i(s)_j(s) > 0 \) and consequently \( s \leq N \). We also have \( \langle v_{ij} | P_s | v_{ij} \rangle = \langle v_{ij} | P_s | v_{ij} \rangle \), so the same reasoning as for \( S_1 \) leads to

\[
S_2 \leq \lambda_{\min}^{-1} (\Gamma_N) \sum_{j \in J^*} \sum_{1 \leq m \leq r, m \neq i, j}^{d} \lambda_{ij}^{1-s(i,m)} \lambda_{m,j'}^{s(i,m)} |\langle v_{m,j'} | v_{ij} \rangle|^2.
\]

Putting together (3.15) and (3.16), we obtain

\[
\text{Err}_i(E) \leq \lambda_{\min}^{-1} (\Gamma_N) \sum_{j=1}^{d} \sum_{1 \leq m \leq r, m \neq i, j}^{d} \lambda_{ij}^{1-s(i,m)} \lambda_{m,j'}^{s(i,m)} |\langle v_{m,j'} | v_{ij} \rangle|^2.
\]

Since \( s(i,m), m \neq i \), are arbitrary in \([0, 1]\), we obtain

\[
\text{Err}_i(E) \leq \lambda_{\min}^{-1} (\Gamma_N) \sum_{1 \leq m \leq r, m \neq i}^{\text{inf}} \text{tr}[\rho_i^{1-s} \rho_m^s].
\]

By averaging over \( i \in [1, r] \), we obtain (3.8). \( \square \)

4. **Pairwise linearly independent states.** The main difficulty for utilizing Lemma 1 for an asymptotic error bound is the control of the minimal eigenvalue of the Gram matrix \( \Gamma_N \). Imposing Condition (LI) on the set \( \Sigma = \{ \rho_1, \ldots, \rho_r \} \) is one way to achieve that control, resulting in Theorem 2. Observe that this condition is equivalent to requiring that for each pair \( \rho_i, \rho_j, i \neq j \), the joint set of eigenvectors pertaining to a nonzero eigenvalue is linearly independent. Lemma 2 below implies in this case: the Gram matrix \( \Gamma_N \) associated to the tensor product set \( \Sigma \otimes^n = \{ \rho_1 \otimes^n, \ldots, \rho_r \otimes^n \} \) has minimal eigenvalue bounded away from zero as \( n \to \infty \).

For each of the original \( \rho_i \), let \( d_i := \text{rank}(\rho_i) \) the number of nonzero eigenvalues. Condition (LI) implies that for any \( i \neq j \) we have \( d_i + d_j \leq d \), and since \( d_i \geq 1 \) this implies that all \( d_i < d \). In this case \( \text{rank}(\rho_i \otimes^n) = d_i^n < d^n \). Let \( \mathcal{V}_n \) be the set of eigenvectors of \( \rho_1 \otimes^n, \ldots, \rho_r \otimes^n \) pertaining to a nonzero eigenvalue; more precisely, if we assume spectral representations

\[
\rho_i \otimes^n = \sum_{j=1}^{\text{rank}(\rho_i \otimes^n)} \lambda_{ij} |v_{ij}\rangle \langle v_{ij}|
\]

with unit vectors \( v_{ij} \) and eigenvalues \( \lambda_{ij} > 0 \), then \( \mathcal{V}_n \) is the double array

\[
\mathcal{V}_n = \{ v_{ij}, j \in [1, d_i^n], i \in [1, r] \}
\]

so that \( \#\mathcal{V}_n = D_n := \sum_{i=1}^r d_i^n \).

**Lemma 2.** Let \( \Sigma = \{ \rho_1, \ldots, \rho_r \} \) be a set of density matrices in \( \mathbb{C}^d \), fulfilling Condition (LI). Let \( \mathcal{V}_n \) be the set of eigenvectors defined above and let \( \tilde{\Gamma}_n \) its \( D_n \times D_n \) Gram matrix. Then

\[
\lambda_{\min}(\tilde{\Gamma}_n) = 1 + o(1) \quad \text{as} \quad n \to \infty.
\]
PROOF. We will first argue for the generic case $n = 1$, and subsequently impose the tensor product structure on the $\rho_i$. As above, let $\{v_{ij}\}_{j=1}^{d_i}$ be the eigenvectors of $\rho_i$ pertaining to a nonzero eigenvalue. Define a $d \times d_i$ matrix
\[
U_i := (u_{i1}, \ldots, u_{id_i}),
\]
that is, the columns of $U_i$ are the vectors $u_{ij}$, $j \in [1, d_i]$. Furthermore, define a $d \times D$ matrix (where $D = \sum_{i=1}^{r} d_i$)
\[
U := (U_1 | \cdots | U_r)
\]
made up of submatrices $U_i$. Now, for $n > 1$ replace the matrices $U_i$ in (4.2) by their $n$th tensor powers $U_i^\otimes n$. Then for $n \geq 1$ the $d^n \times d^n$ blocks $U_i^\otimes n$ correspond to eigenvectors of $\rho_i^\otimes n$, and $D$ is now of dimension $d^n \times D_n$ where $D_n = \sum_{i=1}^{r} d_i^n$. For the $D_n \times D_n$ Gram matrix $\hat{\Gamma}_n := U^*U$ we show (4.1).

We will again begin with the case $n = 1$ and develop a representation of $U$ which takes account of its block structure in terms of $U_i^*U_j$. To this end, for $i \in [1, r]$ define $d_i \times D$ matrices
\[
E_i = (0_{d_i \times d_1} | \cdots | 0_{d_i \times d_{i-1}} | 1_{d_i} | 0_{d_i \times d_{i+1}} | \cdots | 0_{d_i \times d_r}),
\]
where we denote a $k \times l$ matrix of $0$'s by $0_{k \times l}$ and the $k$-dimensional unit matrix by $1_k$. Then it is easily seen that $U = \sum_{i=1}^{r} U_i E_i$ and consequently
\[
\hat{\Gamma}_1 = U^*U = \sum_{i,j=1}^{r} E_i^*U_i^*U_j E_j.
\]
Here $U_i^*U_i = 1_{d_i}$, $i \in [1, r]$, so that
\[
1_D = \sum_{i=1}^{r} E_i^*U_i^*U_i E_i.
\]
We define
\[
\Delta := \hat{\Gamma}_1 - 1_D
\]
and write $\hat{\Gamma}_1 = 1_D + \Delta$. Moreover, for $j < i$ we define
\[
\Delta_{ij} = E_i^*U_i^*U_j E_j + E_j^*U_j^*U_i E_i.
\]
Clearly $\Delta_{ij}$ is Hermitian, and by construction $\Delta = \sum_{i=2}^{r} \sum_{j=1}^{i-1} \Delta_{ij}$. Now, with $\norm{a} = \lambda_{\max}(a^2)$ being the operator norm of a Hermitian matrix $a$, we have
\[
\lambda_{\min}(\hat{\Gamma}_1) = \min_{\norm{v} = 1} \langle v | 1_D + \Delta | v \rangle = 1 + \min_{\norm{v} = 1} \langle v | \Delta | v \rangle
\]
\[
\geq 1 - \norm{\Delta} \geq 1 - \sum_{i=2}^{r} \sum_{j=1}^{i-1} \norm{\Delta_{ij}} = 1 - \sum_{i=2}^{r} \sum_{j=1}^{i-1} \lambda_{\max}^{1/2}(\Delta_{ij}^2),
\]
where the second inequality is by the triangle inequality for the operator norm.
For the case \( n > 1 \), replacing the matrices \( U_i \) in (4.2) by their \( n \)-th tensor powers \( U_i^\otimes n \) leads to a representation of \( \bar{\Gamma}_n \) analogous to (4.3). Here the matrices \( E_i \) have to be replaced by \( E_{i,n} \), defined analogously to \( E_i \) with \( d_i \) replaced by \( d_i^n \), \( i \in [1, r] \). Furthermore, we define \( \Delta_n \) and \( \Delta_{ij,n} \) analogously to (4.4) and (4.5) with \( U_i, E_i \) replaced by \( U_i^\otimes n \) and \( E_{i,n} \). In order to prove (4.1) we use the analog of (4.6) holding for \( \bar{\Gamma}_n \) and \( \Delta_n \), which is

\[
\lambda_{\min}(\bar{\Gamma}_n) \geq 1 - \sum_{i=2}^{r} \sum_{j=1}^{i-1} \lambda_{\max}^{1/2}(\Delta_{ij,n}).
\]

It now suffices to show that for all \( i \in [2, r] \), \( j \in [1, i - 1] \)

\[
(4.7) \quad \lambda_{\max}^{1/2}(\Delta_{ij,n}) \to 0 \quad \text{as} \quad n \to \infty.
\]

Clearly, we have

\[
\Delta_{ij,n} = E_{i,n}^*(U_i^* U_j)^\otimes n E_{j,n} + E_{j,n}^*(U_j^* U_i)^\otimes n E_{i,n}
\]

and by a computation, since \( E_{i,n} E_{i,n}^* = 1_{d_i^n} \) and \( E_{j,n} E_{i,n}^* = 0_{d_i^n \times d_j^n} \) for \( j < i \),

\[
\Delta_{ij,n}^2 = E_{i,n}^*(U_i^* U_j U_j^* U_i)^\otimes n E_{i,n} + E_{j,n}^*(U_j^* U_i U_i^* U_j)^\otimes n E_{j,n}.
\]

The two hermitian matrices composing \( \Delta_{ij,n}^2 \) are orthogonal, and their non-zero eigenvalues are those of \( (U_i^* U_j U_j^* U_i)^\otimes n \) and \( (U_j^* U_i U_i^* U_j)^\otimes n \), respectively. Hence,

\[
(4.8) \quad \lambda_{\max}(\Delta_{ij,n}^2) = \max\{\lambda_{\max}(U_i^* U_j U_j^* U_i)^\otimes n, \lambda_{\max}(U_j^* U_i U_i^* U_j)^\otimes n\}
\]

\[
= \max\{\lambda^n_{\max}(U_i^* U_j U_j^* U_i), \lambda^n_{\max}(U_j^* U_i U_i^* U_j)\}.
\]

Let \( P_i = U_i U_i^* \) be the projection operator onto the space \( \text{supp}(\rho_i) = \text{span}(U_i) \). Note that \( U_i^* P_j U_i \) and \( P_i P_j P_i \) have the same set of nonzero eigenvalues, hence by Lemma 3 below and Condition (LI) we have \( \lambda_{\max}(U_i^* P_j U_i) \leq 1 \) and \( \lambda_{\max}(U_j^* P_i U_j) \leq 1 \). It follows

\[
\lambda^n_{\max}(U_i^* P_j U_i) \to 0 \quad \text{as} \quad n \to \infty,
\]

\[
\lambda^n_{\max}(U_j^* P_i U_j) \to 0 \quad \text{as} \quad n \to \infty,
\]

hence by (4.8) \( \lambda_{\max}(\Delta_{ij,n}^2) \to 0 \). Thus, (4.7) is established. \( \square \)

**Lemma 3.** Let \( \mathcal{L}_0, \mathcal{L}_1 \) be linear subspaces of \( \mathbb{C}^d \) and \( P_0, P_1 \) be the corresponding projection operators. Then \( \mathcal{L}_0 \cap \mathcal{L}_1 = \{0\} \) if and only if

\[
\lambda_{\max}(P_0 P_1 P_0) < 1.
\]

**Proof.** It is obvious that always \( \lambda_{\max}(P_0 P_1 P_0) \leq 1 \), so it suffices to prove that \( \mathcal{L}_0 \cap \mathcal{L}_1 \neq \{0\} \) is equivalent to \( \lambda_{\max}(P_0 P_1 P_0) = 1 \). Assume there exists \( x \in \mathcal{L}_0 \cap \mathcal{L}_1, \|x\| > 0 \), then \( P_ix = x, i = 0, 1 \) and hence \( P_0 P_1 P_0 x = x \) so
that \( \lambda_{\text{max}}(P_0 P_1 P_0) = 1 \). For the other direction, assume

\[
\lambda_{\text{max}}(P_0 P_1 P_0) = 1.
\]

Then there exist \( v_0 \in \mathbb{C}^d, \|v_0\| = 1 \) such that \( \langle v_0 | P_0 P_1 P_0 | v_0 \rangle = 1 \). Here \( \|P_0 v_0\| \leq 1 \) by the properties of projections. Assume \( \|P_0 v_0\| < 1 \). Then for \( u_0 = P_0 v_0 \) we have

\[
\langle v_0 | P_0 P_1 P_0 | v_0 \rangle = \langle u_0 | P_1 | u_0 \rangle < 1,
\]

which contradicts the assumption (4.9). Hence, we must have \( \|P_0 v_0\| = 1 \) and hence \( v_0 \in \mathcal{L}_0 \) and \( P_0 v_0 = v_0 \). Then

\[
1 = \langle v_0 | P_0 P_1 P_0 | v_0 \rangle = \langle v_0 | P_1 | v_0 \rangle,
\]

which implies \( v_0 \in \mathcal{L}_1 \) by an analogous reasoning. Hence, \( v_0 \in \mathcal{L}_0 \cap \mathcal{L}_1 \) where \( \|v_0\| = 1 \), hence \( \mathcal{L}_0 \cap \mathcal{L}_1 \neq \{0\} \). \( \square \)

**Proof of Theorem 2.** We utilize the detector constructed in Algorithm 2, applied to the tensor product case \( \Sigma = \Sigma \otimes n \); call this detector \( E^{(n)} \). Lemma 2 implies that the set \( \mathcal{V}_n \) is a linearly independent set for sufficiently large \( n \). As a consequence, when Lemma 1 is applied to the tensor product set \( \Sigma \otimes n = \{\rho_1 \otimes n, \ldots, \rho_r \otimes n\} \), the matrix \( \Gamma_N \) occurring there equals \( \hat{\Gamma}_n \) up to a rearrangement and \( \lambda_{\text{min}}(\Gamma_N) = \lambda_{\text{min}}(\hat{\Gamma}_n) \). We find from (3.8) that

\[
\text{Err}(E^{(n)}) \leq \lambda_{\text{min}}^{-1}(\hat{\Gamma}_n) r^{-1} \sum_{1 \leq i, j \leq r, j \neq i} \inf_{s \in [0, 1]} \text{tr}[(\rho_i^{\otimes n})^{1-s} (\rho_j^{\otimes n})^s]
\]

\[
= r^{-1} (1 + o(1)) \sum_{1 \leq i, j \leq r, j \neq i} \left( \inf_{s \in [0, 1]} \text{tr}[\rho_i^{1-s} \rho_j^s] \right)^n.
\]

Recall the definition (2.1) of the pairwise quantum Chernoff bound \( \xi_{\text{QCB}}(\rho_i, \rho_j) \); then

\[
\text{Err}(E^{(n)}) \leq r^{-1} (1 + o(1)) \sum_{1 \leq i, j \leq r, j \neq i} \exp(-n \xi_{\text{QCB}}(\rho_i, \rho_j)).
\]

Taking log of both sides and dividing by \( n \), the limit of the right-hand side above is determined by the smallest of the \( \xi_{\text{QCB}}(\rho_i, \rho_j) \), which according to (2.2) coincides with \( \xi_{\text{QCB}}(\Sigma) \). The theorem follows. \( \square \)

**5. Commuting states.** Suppose all the density matrices \( \rho_i \) are commuting: \( \rho_i \rho_j = \rho_j \rho_i \) for all \( i, j \in [1, r] \). Then the \( \rho_i \) have a common set of eigenvectors \( v_j, j \in [1, d] \). The spectral decompositions (3.2) now are

\[
\rho_i = \sum_{j=1}^d \lambda_{i,j} |v_j \rangle \langle v_j|, \quad i \in [1, r].
\]

Also, w.l.o.g., by applying a unitary transformation, we can assume that all \( \rho_i \) are diagonal matrices and \( v_j \) is a canonical basis vector of \( \mathbb{C}^d \). Then
the set of eigenvalues of $\rho_i$ represents a probability distribution $P_i$ on a finite sample space $\Omega$, $\#\Omega = d$, where each $\omega \in \Omega$ can be identified with one of the projections $|v_j\rangle\langle v_j|$. 

With this identification, Algorithm 2 reduces essentially to Algorithm 1. Indeed, in the orthogonalization step (3.3), the newly appearing unit vector $v_{i*,j}$ in step $s$ is one of the basis vectors $v_j$. By induction, it follows that the constructed basis $e_1, \ldots, e_d$ coincides with $v_1, \ldots, v_d$ up to possible reindexing and change of sign. Thus, the classical decision rule $\varphi$ found in Algorithm 2 on the sample space elements $|e_j\rangle\langle e_j|$ is equivalent to a decision rule on $\Omega$, constructed according to Algorithm 1, and the latter is a maximum likelihood rule. The ML rule is not unique in general; in case of nonuniqueness, any version may result from Algorithm 1, according to the choice of a maximal entry in step (i).

In Lemma 1, $\Gamma_L$ is the Gram matrix pertaining to $\{v_j\}_{j=1}^d$, that is unity. Thus, we obtain

$$\text{Err}(E) \leq r^{-1} \sum_{1 \leq i,j \leq r, j \neq i} \inf_{s \in [0,1]} \text{tr}[\rho_i^{1-s} \rho_j^s]$$

and reasoning further as in (4.10) and (4.11), we have thus reproduced the attainability result for the multiple classical Chernoff bound (cf. (1.2) and [28, 29]).

6. A near optimal rate in the general case. We establish that, as stated in Theorem 3, in the general case of a finite number of quantum hypotheses there exist quantum tests that achieve an error exponent equal to the generalized quantum Chernoff distance up to a factor $1/3$.

To construct the detector attaining the exponential bound in the general case, we will modify Algorithm 2 such that it assumes certain density matrices $\tilde{\rho}_i$, which represent $\varepsilon$-perturbations of embeddings of the original $\rho_i$ into a higher-dimensional space $\mathbb{C}^D$, $D > d$. These states $\tilde{\rho}_i$ are not observable; the detector will be applied to the extensions of $\rho_i$, which are observable.

Set $D = (r+1)d$ and consider the $k$th canonical unit vector $f_k$ in $(r+1)d$-dimensional space $\mathbb{C}^D$. Reindex the basis vectors $f_k$ such that $f_{i,j} = f_{(i-1)d+j}$ for $(i,j) \in [1, r+1] \times [1, d]$ and define subspaces

$$S_i = \text{span}\{f_{i,j}\}_{j=1}^d.$$

Then $\mathbb{C}^D$ is a direct sum $\mathbb{C}^D = \bigoplus_{i=1}^{r+1} S_i$ where all $S_i$ are isomorphic to $\mathbb{C}^d$. Let the operator $F$ represent the canonical embedding $F: \mathbb{C}^d \to S_1$. Recall the spectral representation (3.2) of $\rho_i$ with eigenvectors $v_{i,j} \in \mathbb{C}^d$; setting $u_{i,j} = Fv_{i,j}$, we may equivalently assume that instead of $\rho_i$ we measure
a $D \times D$ density matrix $\rho_{0,i}$ having spectral representation

$$\rho_{0,i} = \sum_{i=1}^{d} \lambda_{i,j} |u_{i,j}\rangle \langle u_{i,j}|.$$  

For $\varepsilon \in (0,1)$ and $\delta_{\varepsilon} = (1 - \varepsilon^2)^{1/2}$, define vectors

$$\tilde{u}_{i,j} := \delta_{\varepsilon} u_{i,j} + \varepsilon f_{i+1,j}$$

for $(i,j) \in J = [1,r] \times [1,d]$. Then, since $(u_{i,j}|f_{i+1,j}) = 0$, the vectors $\tilde{u}_{i,j}$ are unit vectors; define density matrices

$$(6.1) \quad \tilde{\rho}_{i} = \sum_{i=1}^{d} \lambda_{i,j} |\tilde{u}_{i,j}\rangle \langle \tilde{u}_{i,j}|, \quad i \in [1,r].$$

Relative to this set of density matrices on $\mathbb{C}^{D}$, satisfying

$$(6.2) \quad \text{tr}[\tilde{\rho}_{i}^{1-s} \tilde{\rho}_{j}^{s}] = \delta_{\varepsilon} \text{tr}[\rho_{i}^{1-s} \rho_{j}^{s}]$$

construct a detector according to (3.1) and Algorithm 2, and call this $\tilde{E}_{\varepsilon}$. Then each $\tilde{E}_{\varepsilon,i}$ is a projection matrix in $\mathbb{C}^{D}$ and $\sum_{i=1}^{r} \tilde{E}_{\varepsilon,i} = 1_{D}$. Define now $E_{\varepsilon,i}$ as the upper $d \times d$ submatrix of $\tilde{E}_{\varepsilon,i}$. Then $E_{\varepsilon,i}$ is a positive matrix and $\sum_{i=1}^{r} E_{\varepsilon,i} = 1_{d}$, so that

$$(6.3) \quad E_{\varepsilon} := \{E_{\varepsilon,i}\}_{i=1}^{r}$$

constitutes a POVM in $\mathbb{C}^{d}$.

It should be noted that $E_{\varepsilon,i}$ are not projections, that is, $E_{\varepsilon}$ is a general POVM but not a PVM, contrary to the detector constructed in Algorithm 2. However, $E_{\varepsilon}$ results from a PVM $\tilde{E}_{\varepsilon}$ in a higher-dimensional space by taking submatrices. This relationship holds between POVMs and PVMs in general, on the basis of Naimark’s theorem; cf. Parthasarathy [25] for a discussion.

**Lemma 4.** Let $\Sigma = \{\rho_{i}\}_{i=1}^{r}$ be an arbitrary set of density matrices on $\mathbb{C}^{d}$. For sufficiently small $\varepsilon > 0$, the detector $E_{\varepsilon}$ constructed in (6.3) fulfills

$$(6.4) \quad \text{Err}(E_{\varepsilon}) \leq r^{-1} \left(2\varepsilon + \varepsilon^{-2} \sum_{1 \leq i,j \leq r, j \neq i} \inf_{s \in [0,1]} \text{tr}[\rho_{i}^{1-s} \rho_{j}^{s}]\right).$$

**Proof.** Consider the Gram matrix $\tilde{\Gamma}_{J}$ of the set of vectors $\{\tilde{u}_{i,j}, (i,j) \in J\}$. Since for $(i,j) \in J$ and $(k,l) \in J$ we have

$$\langle \tilde{u}_{i,j}|\tilde{u}_{k,l}\rangle = \delta_{\varepsilon}^{2} \langle u_{i,j}|u_{k,l}\rangle + \varepsilon^{2} \langle f_{i+1,j}|f_{k+1,l}\rangle$$

it follows that $\tilde{\Gamma}_{J}$ is a convex combination of two Gram matrices, which implies that

$$\lambda_{\min}(\tilde{\Gamma}_{J}) \geq \varepsilon^{2}.$$
Hence, \( \{ \tilde{u}_{i,j}, (i,j) \in J \} \) is a set of \( rd \) linearly independent vectors in \( \mathbb{C}^D \). Since Algorithm 2 eliminates from \( V_1(\Sigma) \) all eigenvectors pertaining to zero eigenvalues, the sequence \( V_1(\Sigma) \) of length \( L \) contains exactly the vectors \( \{ \tilde{u}_{i,j}, (i,j) \in J \} \) pertaining to nonzero \( \lambda_{i,j} \) in (6.1). Their full Gram matrix \( \Gamma_L \) as given by (3.5) for \( s = L \) is a submatrix of \( \tilde{\Gamma}_J \) (after rearrangement) and hence also fulfills

\[
\lambda_{\min}(\Gamma_L) \geq \varepsilon^2.
\]

Consider the error probability of the POVM \( E_\varepsilon \)

\[
\text{Err}(E_\varepsilon) = 1 - r^{-1} \sum_{i=1}^{r} \text{tr}[\tilde{E}_{\varepsilon,i} \rho_{0,i}]
\]

(6.6)

\[
= 1 - r^{-1} \sum_{i=1}^{r} \text{tr}[\tilde{E}_{\varepsilon,i} \tilde{\rho}_i] + r^{-1} \sum_{i=1}^{r} \text{tr}[\tilde{E}_{\varepsilon,i} (\tilde{\rho}_i - \rho_{0,i})].
\]

Now according to Lemma 1, (6.5), and (6.2) we have

\[
1 - r^{-1} \sum_{i=1}^{r} \text{tr}[\tilde{E}_{\varepsilon,i} \tilde{\rho}_i] \leq \varepsilon^{-2} r^{-1} \sum_{1 \leq i < j \leq s \leq [0, 1]} \inf_{s \in [0, 1]} \text{tr}[\tilde{\rho}_i^{1-s} \tilde{\rho}_j^s]
\]

\[
\leq \varepsilon^{-2} r^{-1} \sum_{1 \leq i < j \leq s \leq [0, 1]} \inf_{s \in [0, 1]} \text{tr}[\rho_i^{1-s} \rho_j^s].
\]

For the second term on the right-hand side of (6.6) note that

\[
\tilde{\rho}_i - \rho_{0,i} = \sum_{j=1}^{d} \lambda_{i,j} (|\tilde{u}_{i,j}\rangle \langle \tilde{u}_{i,j}| - |u_{i,j}\rangle \langle u_{i,j}|).
\]

Here we have

\[
|\tilde{u}_{i,j}\rangle \langle \tilde{u}_{i,j}| - |u_{i,j}\rangle \langle u_{i,j}|
\]

\[
= |\delta \varepsilon u_{i,j} + \varepsilon f_{i+1,j}\rangle \langle \delta \varepsilon u_{i,j} + \varepsilon f_{i+1,j}| - |u_{i,j}\rangle \langle u_{i,j}|
\]

\[
= -\varepsilon^2 |u_{i,j}\rangle \langle u_{i,j}| + \delta \varepsilon |u_{i,j}\rangle \langle f_{i+1,j} + \delta \varepsilon |f_{i+1,j}\rangle \langle u_{i,j}| + \varepsilon^2 |f_{i+1,j}\rangle \langle f_{i+1,j}| + \delta \varepsilon |u_{i,j} + f_{i+1,j}\rangle \langle u_{i,j} + f_{i+1,j}| + (\delta \varepsilon - \varepsilon^2) |u_{i,j}\rangle \langle u_{i,j}| + |f_{i+1,j}\rangle \langle f_{i+1,j}|
\]

\[
- 2\varepsilon^2 |u_{i,j}\rangle \langle u_{i,j}|
\]

Since the matrix

\[
(\delta \varepsilon - \varepsilon^2) |u_{i,j}\rangle \langle u_{i,j}| + |f_{i+1,j}\rangle \langle f_{i+1,j}|
\]

\[
+ 2\varepsilon^2 |u_{i,j}\rangle \langle u_{i,j}|
\]
is positive for sufficiently small $\varepsilon$, we have
\[
|\tilde{u}_{i,j} \langle \tilde{u}_{i,j}| - |u_{i,j} \langle u_{i,j}| \leq \delta \varepsilon |u_{i,j} + f_{i+1,j} \langle u_{i,j} + f_{i+1,j}|
\]
consequently
\[
\text{tr}[\tilde{E}_{\varepsilon,i}(\tilde{\rho}_i - \rho_{0,i})] \leq \sum_{j=1}^{d} \lambda_{i,j} \text{tr}[\tilde{E}_{\varepsilon,i}(\delta \varepsilon |u_{i,j} + f_{i+1,j} \langle u_{i,j} + f_{i+1,j}|)]
\]
\[
\leq \sum_{j=1}^{d} \lambda_{i,j} \text{tr}[(\delta \varepsilon |u_{i,j} + f_{i+1,j} \langle u_{i,j} + f_{i+1,j}|)]
\]
\[
= \delta \varepsilon \sum_{j=1}^{d} \lambda_{i,j} : 2 \leq 2 \varepsilon.
\]

**Proof of Theorem 3.** We denote the factor of $\varepsilon^{-2}$ in (6.4) by $K_1$, and in the $n$-fold tensor product case, where $\rho_i$ is replaced by $\rho_i^\otimes n$, by $K_n$, respectively. To find the best upper bound in (6.4), we minimize the expression $2\varepsilon^2 + \varepsilon^{-2}K_n$ in $\varepsilon$. The solution is $\varepsilon = K_n^{-1/3}$ and the value at the minimum is $3K_n^{-1/3}$. Since $K_n$ tends to zero as $n$ goes to infinity it is ensured that for sufficiently large $n$, the value $K_n^{-1/3}$ is small enough to satisfy the condition of Lemma 4. Thus from (6.4) we obtain
\[
\text{Err}(E_{\varepsilon}^{(n)}) \leq 3n^{-1} \left( \sum_{1 \leq i,j \leq r, j \neq i} \inf_{s \in [0,1]} \text{tr}[\rho_i^\otimes n]^{1-s} (\rho_j^\otimes n)^s \right)^{1/3},
\]
where $E_{\varepsilon}^{(n)}$ denotes the respective detectors in the tensor product case $\Sigma^\otimes n$. It follows
\[
\frac{1}{n} \log \text{Err}(E^{(n)}) \leq \frac{1}{3} \frac{1}{n} \log \left( \sum_{1 \leq i,j \leq r, j \neq i} \inf_{s \in [0,1]} \text{tr}[\rho_i^\otimes n]^{1-s} (\rho_j^\otimes n)^s \right) + o(1)
\]
\[
= \frac{1}{3} \log \xi_{QCB}(\Sigma) + o(1),
\]
which proves our claim. □

**REFERENCES**

[1] Assalini, A., Cariolaro, G. and Pierobon, G. (2010). Efficient optimal minimum error discrimination of symmetric quantum states. *Phys. Rev. A* **81** 012315.

[2] Audenaert, K. M. R., Casamiglia, J., Munoz-Tapia, R., Bagan, E., Masanes, L., Acin, A. and Verstraete, F. (2007). Discriminating states: The quantum chernoff bound. *Phys. Rev. Lett.* **98** 160501.
[3] Audenaert, K. M. R., Nussbaum, M., Szkola, A. and Verstraete, F. (2008). Asymptotic error rates in quantum hypothesis testing. Comm. Math. Phys. 279 251–283. MR2377635
[4] Barnett, S. M. and Croke, S. (2009). On the conditions for discrimination between quantum states with minimum error. J. Phys. A 42 062001, 4. MR2525420
[5] Barnum, H. and Knill, E. (2002). Reversing quantum dynamics with near-optimal quantum and classical fidelity. J. Math. Phys. 43 2097–2106. MR1893662
[6] Belavkin, V. P. (1975). Optimal multiple quantum statistical hypothesis testing. Stochastics 1 315–345. MR0446192
[7] Bergou, J. A., Herzog, U. and Hillery, M. (2004). Discrimination of quantum states. In Quantum State Estimation. Lecture Notes in Physics 649 417–465. Springer, Berlin. MR2180170
[8] Calsamiglia, J., Munoz-Tapia, R., Masanes, L., Acin, A. and Bagan, E. (2008). The quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states. Phys. Rev. A 77 032311.
[9] Chefles, A. (2000). Quantum state discrimination. Contemp. Phys. 41 401.
[10] Eldar, Y. C. (2003). von Neumann measurement is optimal for detecting linearly independent mixed quantum states. Phys. Rev. A (3) 68 052303, 4. MR2026802
[11] Hayashi, M. (2006). Quantum Information. An introduction. Springer, Berlin. MR228302
[12] Helstrom, C. W. (1976). Quantum Detection and Estimation Theory. Academic Press, New York.
[13] Hiai, F., Mosonyi, M. and Ogawa, T. (2007). Large deviations and Chernoff bound for certain correlated states on a spin chain. J. Math. Phys. 48 123301, 19. MR2377831
[14] Holevo, A. S. (1973). Statistical decision theory for quantum systems. J. Multivariate Anal. 3 337–394. MR0337245
[15] Holevo, A. S. (1974). Remarks on optimal quantum measurements. Probl. Inf. Transm. 10 51–55.
[16] Holevo, A. S. (1978). On asymptotically optimal hypothesis testing in quantum statistics. Theory Probab. Appl. 23 411–415.
[17] Hwang, W.-Y. and Bae, J. (2010). Minimum-error state discrimination constrained by the no-signaling principle. J. Math. Phys. 51 022202, 11. MR2605029
[18] Kimura, G., Miyadera, T. and Imai, H. (2008). Optimal state discrimination in general probabilistic theories. Phys. Rev. A 79 062306.
[19] König, R., Renner, R. and Schaffner, C. (2009). The operational meaning of min- and max-entropy. IEEE Trans. Inform. Theory 55 4337–4347. MR2582886
[20] Krob, J. and V. Weizsäcker, H. (1997). On the rate of information gain in experiments with a finite parameter set. Statist. Decisions 15 281–294. MR1484382
[21] Montanaro, A. (2007). On the distinguishability of random quantum states. Comm. Math. Phys. 273 619–636. MR2318859
[22] Nussbaum, M. and Szkola, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. Ann. Statist. 37 1040–1057. MR2502660
[23] Nussbaum, M. and Szkola, A. (2010). Exponential error rates in multiple state discrimination on a quantum spin chain. J. Math. Phys. 51 072203, 11. MR2681076
[24] Nussbaum, M. and Szkola, A. (2011). Asymptotically optimal discrimination between multiple pure quantum states. In Theory of Quantum Computation, Communication and Cryptography. 5th Conference, TQC 2010, Leeds, UK. Revised Selected Papers (W. van Dam, V. M. Kendon and S. Severini, eds.). Lecture Notes in Computer Science 6519 1–8. Springer, Berlin.
[25] Parthasarathy, K. R. (1999). Extremal decision rules in quantum hypothesis testing. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* 2 557–568. MR1810813
[26] Parthasarathy, K. R. (2001). On consistency of the maximum likelihood method in testing multiple quantum hypotheses. In *Stochastics in Finite and Infinite Dimensions* 361–377. Birkhäuser, Boston, MA. MR1797096
[27] Qiu, D. and Li, L. (2010). Minimum-error discrimination of quantum states: New bounds and comparisons. *Phys. Rev. A* 81 042329.
[28] Salihov, N. P. (1973). Asymptotic properties of error probabilities of tests for distinguishing between several multinomial testing schemes. *Dokl. Akad. Nauk SSSR* 209 54–57. MR0356327
[29] Salihov, N. P. (1998). On a generalization of Chernoff distance. *Teor. Veroyatn. Primen.* 43 294–314. Translation in *Theory Probab. Appl.* 43 (1999) 239–255. MR1679004
[30] Salihov, N. P. (2002). Optimal sequences of tests for the discrimination of several multinomial schemes of trials. *Teor. Veroyatn. Primen.* 47 270–285. Translation in *Theory Probab. Appl.* 47 (2003) 286–298. MR2001833
[31] Torgersen, E. N. (1981). Measures of information based on comparison with total information and with total ignorance. *Ann. Statist.* 9 638–657. MR0615440
[32] Tyson, J. (2009). Two-sided estimates for minimum-error distinguishability of mixed quantum states via generalized Holevo–Curlander bounds. *J. Math. Phys.* 50 032106, 10. MR2510897
[33] Tyson, J. (2010). Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates. *J. Math. Phys.* 51 092204, 35. MR2742804
[34] Yuen, H. P., Kennedy, R. S. and Lax, M. (1975). Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inform. Theory* 21 125–134. MR0397511