Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states

J. Calsamiglia and R. Muñoz-Tapia
Grup de Física Teòrica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Ll. Masanes
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

A. Acin
ICREA and ICFO-Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain

E. Bagan
Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA and Grup de Física Teòrica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
(Received 20 September 2007; published 7 March 2008)

Hypothesis testing is a fundamental issue in statistical inference and has been a crucial element in the development of information sciences. The Chernoff bound gives the minimal Bayesian error probability when discriminating two hypotheses given a large number of observations. Recently the combined work of Audenaert *et al.* [Phys. Rev. Lett. **98**, 160501 (2007)] and Nussbaum and Szkola [e-print arXiv:quant-ph/0607216] has proved the quantum analog of this bound, which applies when the hypotheses correspond to two quantum states. Based on this quantum Chernoff bound, we define a physically meaningful distinguishability measure and its corresponding metric in the space of states; the latter is shown to coincide with the Wigner-Yanase metric. Along the same lines, we define a second, more easily implementable, distinguishability measure based on the error probability of discrimination when the same local measurement is performed on every copy. We study some general properties of these measures, including the probability distribution of density matrices, defined via the volume element induced by the metric. It is shown that the Bures and the local-measurement based metrics are always proportional. Finally, we illustrate their use in the paradigmatic cases of qubits and Gaussian infinite-dimensional states.

DOI: 10.1103/PhysRevA.77.032311 PACS number(s): 03.67.Hk, 03.65.Ta

I. INTRODUCTION

About fifty years ago Herman Chernoff proved his famous bound, which characterizes the asymptotic behavior of the minimal probability of error when discriminating two hypothesis given a large number of observations [1]. Its quantum analog was recently conjectured [2] and finally proven by combining the results of two recent publications [3,4]. In this quantum setting one is confronted with the problem of knowing the minimum error probability in identifying one of two possible known states of which N identical copies are given. Hereafter we will refer to this minimum simply as the error probability P_e. This problem is widely known as quantum state discrimination.[1] Its difficulty (but also its appeal) lies in the fact that quantum mechanics only allows for full discrimination of such states when they are orthogonal. This has both fundamental and practical implications that lie at the heart of quantum mechanics and its applications.

For these past fifty years the classical Chernoff bound—as well as hypothesis testing in general—has proved to be extremely useful in all branches of science. Likewise, one would expect its quantum version to be far more than a mere academic issue. The characterization and control of quantum devices is a necessary requirement for quantum computation and communication, and quantum hypothesis testing is specially designed for assessing the performance of these tasks. Particularly important examples for which state discrimination plays an essential role are quantum cryptography [7], classical capacity of quantum channels [8], or even quantum algorithms [9]. Equally important are some new theorems concerning different quantum extensions of hypothesis testing: the quantum Stein's lemma, proved some years ago [10,11], and the quantum Hoeffding bound, recently established in Refs. [12–14].

In this paper we study the classical and the quantum Chernoff bounds in connection to measures of distinguishability for quantum states, putting special emphasis on the qubit and Gaussian cases. We start by reviewing classical and quantum hypothesis testing and the corresponding Chernoff bounds in Secs. II and III, respectively (the latter includes the before mentioned recent results by Nussbaum and Szkola [4] and Audenaert *et al.* [3]). In Sec. IV we discuss the notion of a distinguishability measure for quantum states. We briefly motivate an important instance of such a notion based on classical statistical measures, that is, the quantum fidelity, and move to a fully operational alternative, based on the

[1] See Refs. [5] and [6] for two reviews on the recent and more historical developments of this field, respectively.

asymptotic rate exponent of the error probability in symmetric quantum hypothesis testing: the quantum Chernoff measure.[2] We also discuss a similar distinguishability measure derived from the same rate exponent when the decision is based on N identical single-copy (local) measurements—instead of the collective measurements on the N copies assumed in the derivation of the quantum Chernoff bound. In Sec. V we study the metrics induced by the previously defined measures of distinguishability and give explicit expressions for general d-dimensional systems. We also give the probability distribution of the eigenvalues of a $d\times d$ density matrix based on the quantum Chernoff metric (induced by the corresponding distinguishability measure). We find that the metric based on local measurements is discontinuous and has to be defined piecewise: on the set of pure states, where it agrees with the Fubini-Study metric, and, separately, on the set of strictly mixed states, where it agrees with one-half the Bures-Uhlmann metric. The quantum Chernoff metric, in contrast, is continuous and smoothly interpolates between the Fubini-Study and one-half the Bures-Uhlmann metrics. In Sec. VI we concentrate on the particular case of two-level systems and study in some depth the differences between the quantum Chernoff measure and metric and those based on identical local measurements. In Sec. VII we give explicit expressions of the quantum Chernoff measure and its corresponding induced metric for general Gaussian states. Finally, we state our conclusions in Sec. VIII.

II. CLASSICAL HYPOTHESIS TESTING: CHERNOFF BOUND

One of the most fundamental problems in statistical decision theory is that of choosing between two possible explanations or models, that we will refer to as hypothesis H_0 and H_1, where the decision is based on a set of data collected from measurements or observations. For example, a medical team has to decide whether a patient is healthy (hypothesis H_0) or has certain disease (hypothesis H_1) in view of the results of some clinical test. Often, H_0 is called the working hypothesis or null hypothesis, while H_1 is called the alternative hypothesis. In general these two hypotheses do not have to be treated on equal footing, since wrongly accepting or rejecting one of them might have very different consequences. These two types of errors, i.e., the rejection of a true null hypothesis or the acceptance of a false null hypothesis, are called type-I or -II errors, respectively, and their corresponding probabilities will be denoted by $p(1|H_0)\equiv p_0(1)$ and $p(0|H_1)\equiv p_1(0)$ throughout the paper. In our example, failure to diagnose the disease is a type-II error, whereas it is a type-I error to wrongly conclude that the healthy patient has the disease. Of course, it would be desirable to minimize the two types of errors at the same time. However, this is typically not possible since reducing those of one type entails increasing those of the other type. Hence, a common way to proceed is to minimize the errors of one type, while keeping those of the other type bounded by a constant (which may depend on the number of observations). Another (Bayesian-like) approach consists in minimizing a linear combination of the two error probabilities $P_e=\pi_0 p(1|H_0)+\pi_1 p(0|H_1)$, where π_0 and π_1 can be interpreted as the *a priori* probabilities that we assign to the occurrence of each hypothesis. In this paper we consider this latter approach, which is known as symmetric hypothesis testing.

For the sake of simplicity, we assume to start with that $\pi_0=\pi_1=1/2$, and we deal with tests that have only two possible outcomes $b=0,1$. This is, for example, the situation that corresponds to the identification of a biased coin that can be (with equal probability) of one of two types: 0 or 1 (corresponding to hypothesis H_0 or H_1, respectively). If it is of the type 0 the probabilities of obtaining head and tail are, respectively, $p_0(0)=p$ and $p_0(1)=1-p\equiv\bar{p}$, while if it is of type 1 we write $p_1(0)=q$ and $p_1(1)=1-q\equiv\bar{q}$. The test consists in tossing the coin, which has two possible outcomes: either head ($b=0$) or tail ($b=1$).

If we can toss the coin only once (single observation), it is easy to convince oneself that the minimum (average) probability of error is attained when we accept the hypothesis (decide that the tossed coin is of the type) for which the observed outcome occurs with largest probability. Therefore,[3]

$$P_e=\frac{1}{2}\sum_{b=0}^{1}\min\{p_0(b),p_1(b)\}\leq\frac{1}{2}\min_{s\in[0,1]}\sum_{b=0}^{1}p_0^s(b)p_1^{1-s}(b)\equiv P_{\mathrm{CC}},\tag{1}$$

where we have used the inequality $\min\{p,q\}\leq p^s q^{1-s}$. The subscript "CC" stands for classical Chernoff. This expression also holds for tests with more than two outcomes. We just need to extend the sum over b to the entire range of possible outcomes. In what follows, we leave the range of b unspecified whenever an expression is valid for an arbitrary number of outcomes.

Next, let us assume we can toss the coin N times. The set of possible outcomes (the sample space) is the N-fold Cartesian product of {0, 1} (or {head, tail}). The two probability distributions of these outcomes, $p_0^{(N)}(b^{(N)})$ and $p_1^{(N)}(b^{(N)})$, will be given by the product of the corresponding single-observation distributions $p_i^{(N)}(b^{(N)})=p_i(b_1)p_i(b_2)\cdots p_i(b_N)$, where now $b^{(N)}\equiv(b_1,b_2,\ldots,b_N)\in\{0,1\}^{\times N}$, and one immediately obtains [15]

[2] By "operational" is meant "defined though a specific procedure or task," in contradistinction to "purely mathematical."

[3] In this formula, as well as in most of the formulas involving minimization throughout the paper, one should properly write $\inf_{s\in[0,1]}$ instead of $\min_{s\in[0,1]}$ since the minimum may not exist if p_0 and p_1 (ρ_0 and ρ_1 in the quantum case) are degenerate and have different support. This is so because in this case the continuity of the argument of $\min_{s\in[0,1]}$ in all these equations is guaranteed only in the open interval (0,1) and (end-point) singularities may occur at $s=0,1$. We will overlook this mathematical subtlety in the main text to simplify the exposition.

$$P_e \leq \frac{1}{2}\min_{s\in[0,1]}\left(\sum_b p_0^s(b)p_1^{1-s}(b)\right)^N. \tag{2}$$

This is the Chernoff bound [1]. It is specially important because it can be proved to give the exact asymptotic rate exponent of the error probability, that is,

$$P_e \sim e^{-NC(p_0,p_1)};$$

$$C(p_0,p_1) \equiv -\min_{s\in[0,1]} \log \sum_b p_0^s(b)p_1^{1-s}(b). \tag{3}$$

The so-called "Chernoff information," or Chernoff distance $C(p_0,p_1)$, can also be written in terms of the Kullback-Leibler divergence $K(p_0/p_1)=\Sigma_b p_0(b)\log[p_0(b)/p_1(b)]$ [15]:

$$C(p_0,p_1) = K(p_{s^*}/p_0) = K(p_{s^*}/p_1), \tag{4}$$

where

$$p_s(b) = \frac{p_0^s(b)p_1^{1-s}(b)}{\sum_b p_0^s(b)p_1^{1-s}(b)}; \quad s\in[0,1] \tag{5}$$

is a family of probability distributions known as the Hellinger arc that interpolates between p_0 and p_1, and s^* is the value of s at which the second equality in Eq. (4) holds. In other words, it is the point at which p_s is equidistant to both p_0 and p_1 (in terms of the Kullback-Leibler distance). It can be shown that s^* is also the value of s that minimizes the right-hand side of Eq. (3).

For the case of measurements with two outcomes, such as the example of the coins discussed above, one can give a closed expression for the Chernoff distance, which we denote in this binary case as $C(p,q)$:

$$C(p,q) = \xi \log\frac{\xi}{p} + \bar{\xi}\log\frac{\bar{\xi}}{\bar{p}}, \tag{6}$$

with

$$\xi \equiv \frac{\log(\bar{q}/\bar{p})}{\log(p/\bar{p}) + \log(\bar{q}/q)}; \quad \bar{\xi} \equiv 1-\xi. \tag{7}$$

The parameter ξ has a very straightforward interpretation. If N_0 is the number of heads (of 0's) after N trials, which according to the distribution p_0 occurs with probability

$$P_0(N_0) = \binom{N}{N_0} p^{N_0}\bar{p}^{N-N_0} \tag{8}$$

[according to the distribution p_1 it occurs with probability $P_1(N_0)$, defined the same way but with p replaced by q], then ξ is the fraction of heads above which one must decide in favor of p_0. That is, if $N_0 \geq \xi N$ one accepts hypothesis H_0, while if $N_0 < \xi N$ one accepts H_1. Asymptotically, the contribution to the error probability is dominated by situations where $N_0=\xi N$, i.e., by events that occur with the same probability for both hypotheses (see Fig. 1). The probability of such events is clearly a lower bound to the probability of error. It is straightforward to check that $-\lim_{N\to\infty}\log P_0(\xi N)/N$ [or, equivalently, $-\lim_{N\to\infty}\log P_1(\xi N)/N$] coincides with the upper bound given by the Chernoff distance $C(p,q)$. This proves that the Chernoff bound is indeed attainable.

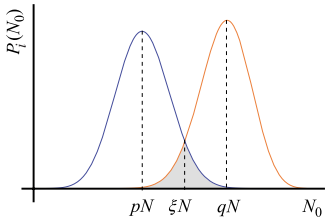

FIG. 1. (Color online) Each curve represents the probability to obtain N_0 heads after N tosses of a bias coin that can be of one of two types, 0 or 1. The probability that the coin of type 0 (1) produces a head at any given toss is p (q). For large N these curves approach Gaussian distributions centered at pN and qN, respectively. The point ξN where they cross defines the decision boundary (see main text). The error probability is given by the shaded area.

III. QUANTUM HYPOTHESIS TESTING: THE QUANTUM CHERNOFF BOUND

We now tackle discrimination (symmetric hypothesis testing) in a quantum scenario. We consider two sources 0 and 1 that produce states described, respectively, by the density matrices ρ_0 and ρ_1 acting on a Hilbert space $\mathcal{H}$. We are given N copies of a state ρ with the promise that they have been produced either by the source 0 (with prior probability π_0) or by the source 1 (with prior probability $\pi_1=1-\pi_0$). Accordingly, we can formulate two hypothesis (H_0 and H_1) about the identity (0 or 1, respectively) of the source that has produced these copies. We wish to find a protocol to determine, with minimal error probability, which hypothesis better explains the nature of the N copies. No matter how complicated this protocol might be, it is clear that the output must be classical: we have to settle for one of the two hypotheses. Therefore the protocol develops in two stages. First, to obtain information about the states we must necessarily make a (quantum) measurement, which in contrast to the classical world is an inherently random and destructive process. Second, one has to provide a classical algorithm that processes the measurement outcomes (classical data) and produces the best answer (H_0 or H_1). Quantum mechanics allows for a convenient description of this two-step process by assigning to each answer H_0 and H_1 a single positive operator valued measure (POVM) element E_0 and E_1, respectively ($E_b \geq 0$ acts on $\mathcal{H}^{\otimes N}$; $E_0+E_1=\mathbb{1}$). The probability that this POVM measurement gives the answer H_b conditioned to $\rho=\rho_i$ is $p_i(b)=\mathrm{tr}(\rho_i^{\otimes N}E_b)$.

The problem thus reduces to finding the set of operators $\{E_b\}_{b=0}^1$ that minimize the mean probability of error, For the simplest case of a single copy ($N=1$) and two equiprobable hypotheses ($\pi_0=\pi_1=1/2$) it is [16]

$$P_e = \frac{1}{2}[p_0(1)+p_1(0)] = \frac{1}{2}[\mathrm{tr}(\rho_0 E_1) + \mathrm{tr}(\rho_1 E_0)]. \tag{9}$$

Since $E_0=\mathbb{1}-E_1$, we can introduce the *Helstrom matrix* $\Gamma \equiv \rho_1-\rho_0$, as is common in quantum state discrimination, and write

$$P_e = \frac{1}{2} - \frac{1}{2}\mathrm{tr}(E_1\Gamma), \tag{10}$$

which only needs to be optimized with respect to E_1. We note that Γ has some negative eigenvalues, as $\mathrm{tr}\,\Gamma=0$. This necessarily implies that the minimum error probability is attained if E_1 is the projector on the subspace of positive eigenvalues of Γ. We will denote this projector by $\{\Gamma>0\}$ and define the positive part of Γ as $\Gamma_+=\{\Gamma>0\}\Gamma$. Taking into account that Γ is traceless, we obtain

$$\mathrm{tr}(E_1\Gamma) = \mathrm{tr}\,\Gamma_+ = \frac{1}{2}\mathrm{tr}|\Gamma|, \tag{11}$$

where the matrix $|A|$ (absolute value of A) is defined to be $|A|=\sqrt{A^\dagger A}$. We arrive at the final result [16]

$$P_e = \frac{1}{2}\left(1 - \frac{1}{2}\mathrm{tr}|\rho_1-\rho_0|\right). \tag{12}$$

The problem of discriminating multiple copies (arbitrary N) is thus formally solved by replacing ρ_i by $\rho_i^{\otimes N}$ in the above equations. Indeed, if we do not have any restrictions on the type of measurements performed on the N copies, $E_1=\{\rho_0^{\otimes N}-\rho_1^{\otimes N}<0\}$, and the mean probability of error is just

$$P_e = \frac{1}{2}\left(1 - \frac{1}{2}\mathrm{tr}|\rho_1^{\otimes N}-\rho_0^{\otimes N}|\right). \tag{13}$$

However, the computation of the trace norm of the Helmstrom matrix in Eq. (13) is tedious and, moreover, this equation provides little information about the large N behavior of the error probability, which is what the Chernoff bound is about.

The quantum version of the Chernoff (upper) bound was presented very recently in Ref. [3]. There it is shown that

$$P_e \le \frac{1}{2}\min_{s\in[0,1]} \mathrm{tr}\,\rho_0^s\rho_1^{1-s} \equiv P_{\mathrm{QC}} \equiv \frac{1}{2}Q \tag{14}$$

(the subscript QC stands for quantum Chernoff), which holds for arbitrary density matrices. Moreover, this bound can be very efficiently computed. The bound (14) is a straightforward application of the following theorem [3].

Theorem 1. Let A and B be two positive operators, then for all $0\le s\le 1$,

$$\mathrm{tr}(A^sB^{1-s}) \ge \frac{1}{2}\mathrm{tr}(A+B-|A-B|). \tag{15}$$

The proof of this theorem involves advanced methods in matrix algebra and we refer the interested reader to Ref. [3]. Instead, here we will give a simple proof of the inequality (14) where instead of minimizing over s, the particular value $s=1/2$ will be chosen.

We first notice that one obtains an upper bound to P_e by picking any particular positive operator E_1 (and, accordingly, E_0) in Eq. (9). A convenient choice is $\tilde{E}_1=\{\rho_0^{1/2}-\rho_1^{1/2}<0\}$ (and thus $\tilde{E}_0=\{\rho_0^{1/2}-\rho_1^{1/2}\ge 0\}$), where, as above, $\{A>0\}$ stands for the projector onto the subspace spanned by the eigenstates of A with positive eigenvalue. After the following series of inequalities we arrive at the desired result [17]:

$$\begin{aligned} 2P_e &\le \mathrm{tr}(\tilde{E}_1\rho_0) + \mathrm{tr}(\tilde{E}_0\rho_1) = \mathrm{tr}(\rho_0^{1/2}\rho_0^{1/2}\{\rho_0^{1/2}-\rho_1^{1/2}<0\}) \\ &\quad + \mathrm{tr}(\rho_1^{1/2}\rho_1^{1/2}\{\rho_0^{1/2}-\rho_1^{1/2}\ge 0\}) \le \mathrm{tr}(\rho_0^{1/2}\rho_1^{1/2}\{\rho_0^{1/2}-\rho_1^{1/2} \\ &< 0\}) + \mathrm{tr}(\rho_0^{1/2}\rho_1^{1/2}\{\rho_0^{1/2}-\rho_1^{1/2}\ge 0\}) = \mathrm{tr}[\rho_0^{1/2}\rho_1^{1/2}(\{\rho_0^{1/2} \\ &\quad - \rho_1^{1/2}<0\} + \{\rho_0^{1/2}-\rho_1^{1/2}\ge 0\})] = \mathrm{tr}(\rho_0^{1/2}\rho_1^{1/2}), \end{aligned} \tag{16}$$

where in the second inequality we have used

$$\begin{aligned} (\rho_1^{1/2}-\rho_0^{1/2})\{\rho_0^{1/2}-\rho_1^{1/2}<0\} &\ge 0, \\ (\rho_0^{1/2}-\rho_1^{1/2})\{\rho_0^{1/2}-\rho_1^{1/2}\ge 0\} &\ge 0. \end{aligned} \tag{17}$$

The general proof (for all s) follows the same steps but taking $\tilde{E}_1=\{\rho_0^{1-s}-\rho_1^{1-s}<0\}$ if $0\le s<1/2$ and $\tilde{E}_1=\{\rho_0^s-\rho_1^s<0\}$ if $1/2<s\le 1$. In this case, the inequality analogous to the second one in Eq. (16) requires the two additional nonobvious relations

$$\begin{aligned} \mathrm{tr}[\rho_1^{1-s}(\rho_0^s-\rho_1^s)\{\rho_0^{1-s}-\rho_1^{1-s}\ge 0\}] &\ge 0, \quad 0\le s<\frac{1}{2}, \\ \mathrm{tr}[\rho_0^s(\rho_1^{1-s}-\rho_0^{1-s})\{\rho_1^s-\rho_0^s\ge 0\}] &\ge 0, \quad \frac{1}{2}\le s<1. \end{aligned} \tag{18}$$

These inequalities follow immediately from the following non-trivial lemma, which constitutes the core of the proof [3]:

Lemma 1. Let A and B be two positive operators, then for all $0\le t\le 1$,

$$\mathrm{tr}[\{A-B\ge 0\}B(A^t-B^t)] \ge 0. \tag{19}$$

Before proceeding with the asymptotic limit, several comments about Eq. (14) are in order. (i) The exponential fall-off of the probability of error when a number N of copies is available follows immediately from $\mathrm{tr}(A\otimes B)=\mathrm{tr}\,A\ \mathrm{tr}\,B$:

$$P_e \le \frac{Q^N}{2} = \frac{1}{2}\exp\{-N[-\min_{s\in[0,1]} \log \mathrm{tr}\,\rho_0^s\rho_1^{1-s}]\}. \tag{20}$$

Remarkably enough, this rate exponent, which we may call quantum Chernoff information because of its analogy with $C(p_0,p_1)$, is asymptotically attainable, as follows from the results of Ref. [4]. This is the quantum extension of the classical result (3) and was first conjectured by Ogawa and Hayashi in Ref. [2]. (ii) If the two matrices ρ_0 and ρ_1 commute the bound reduces to the classical Chernoff bound (1), where the two probability distributions are given by the spectrum of the two density matrices. (iii) The function $Q_s=\mathrm{tr}\,\rho_0^s\rho_1^{1-s}$ (whose minimum gives the best bound) is a convex function of s in [0, 1], which means that a stationary point will automatically be the global minimum (see Ref. [3] for a proof). This is a very useful fact when computing the quantum Chernoff bound (14). (iv) Q is jointly concave in (ρ_0,ρ_1), unitarily invariant, and non-decreasing under trace preserving quantum operations [3]. (v) The quantum Chernoff bound

gives a tighter bound than that given by the quantum fidelity

$$F(\rho_0,\rho_1) \equiv (\mathrm{tr}\sqrt{\sqrt{\rho_0}\rho_1\sqrt{\rho_0}})^2 = (\mathrm{tr}|\sqrt{\rho_0}\sqrt{\rho_1}|)^2, \quad (21)$$

which is the most widely used quantum distinguishability measure (see next section). This follows from the following set of inequalities:

$$P_e \leq P_{\mathrm{QC}} \leq \frac{\mathrm{tr}\,\rho_0^{1/2}\rho_1^{1/2}}{2} \leq \frac{\mathrm{tr}|\sqrt{\rho_0}\sqrt{\rho_1}|}{2} = \frac{\sqrt{F(\rho_0,\rho_1)}}{2}. \quad (22)$$

In fact, the fidelity also provides a lower-bound to the probability of error [18]:

$$\frac{1-\sqrt{1-F(\rho_0,\rho_1)}}{2} \leq P_e. \quad (23)$$

In the case where one of the states (say ρ_0) is pure the upper bound to the error probability can be made tighter [19,20]:

$$P_e \leq P_{\mathrm{QC}} = \frac{Q}{2} = \frac{1}{2}F(\rho_0,\rho_1). \quad (24)$$

(vi) The quantum Chernoff bound can be easily extended to the case where the two states ρ_0 and ρ_1 (sources) are not equiprobable:

$$P_e \leq \min_{s\in[0,1]} \pi_0^s\pi_1^{1-s}\,\mathrm{tr}\,\rho_0^s\rho_1^{1-s}. \quad (25)$$

(vii) The permutation invariance of the N-copy density matrices $\rho_i^{\otimes N}$ guarantees that the optimal collective measurement can be implemented efficiently (with a polynomial-size circuit known as quantum Schur transform) [21], and hence that the minimum probability of error is achievable with reasonable resources.

As stated above, for multiple-copy discrimination the error probability decreases exponentially with the number N of copies $P_e \sim \exp[-ND(\rho_0,\rho_1)]$ as N goes to infinity [15]. The error (rate) exponent $D(\rho_0,\rho_1)$ is defined generically by

$$D(\rho_0,\rho_1) = -\lim_{N\to\infty}\frac{1}{N}\log P_e \quad (26)$$

and characterizes the asymptotic behavior of the error probability. From Eq. (20) we readily see that if the best (joint) measurement is used it coincides with the quantum Chernoff information

$$D_{\mathrm{QC}}(\rho_0,\rho_1) = -\min_{s\in[0,1]} \log\,\mathrm{tr}\,\rho_0^s\rho_1^{1-s}, \quad (27)$$

where the equality holds because of the attainability of Eq. (20) discussed above and we have added the subscript QC. Moreover, this asymptotic value is also attained by the square root (or "pretty-good") measurement (see Refs. [22,23] for the precise definition). This immediately follows from the known bounds [24,25] $P_e \leq P_e^{\mathrm{SRM}} \leq 2P_e$, where P_e^{SRM} is the error probability of discrimination when the square root measurement is used.

Before closing this section, we briefly come back to the fidelity bounds in Eqs. (22)–(24) and simply note that the first two inequalities translate into the following bounds to the rate exponent:

$$-\frac{1}{2}\log F(\rho_0,\rho_1) \leq D_{\mathrm{QC}}(\rho_0,\rho_1) \leq -\log F(\rho_0,\rho_1). \quad (28)$$

If one of the states is pure Eq. (24) implies that the factor 1/2 in Eq. (28) becomes 1 and we have the exact relation

$$D_{\mathrm{QC}}(\rho_0,\rho_1) = -\log F(\rho_0,\rho_1). \quad (29)$$

IV. DISTINGUISHABILITY MEASURES

In this section we aim to define a measure of distinguishability between states using the results reviewed in Sec. III. Before doing so we will briefly outline how classical statistical methods can be used to (partially) accomplish this goal. We will then discuss an operational measure of distinguishability based on the error probability in multiple-copy state discrimination, leading to the quantum Chernoff measure. Finally we will define the analogous quantity for local discrimination protocols.

A. Classical statistical approach

The notion of distance between states is a fundamental issue that has been studied for a long time. A straightforward way to define such a distance is to take any suitable norm in the space of states. However, a more physical approach, kick-started by the pioneering work in Ref. [26], is to relate the inherently probabilistic nature of quantum measurements to classical statistical measures of distinguishability between probability distributions.

In particular, the author in Ref. [26] uses the notion of statistical distance

$$d_S(p_0,p_1) = \arccos\sqrt{\mathcal{F}(p_0,p_1)}, \quad (30)$$

as a measure of distinguishability between the probability distributions p_0 and p_1, where

$$\mathcal{F}(p_0,p_1) = \left(\sum_b \sqrt{p_0(b)p_1(b)}\right)^2 \quad (31)$$

is the statistical fidelity. Accordingly, he defines a distinguishability measure between quantum states ρ_0 and ρ_1 by maximizing $d_S(p_0,p_1)$ [i.e., minimizing $\mathcal{F}(p_0,p_1)$] over all possible POVM measurements, characterized by all possible sets of operators $\{E_b\}_{b=1}^M$ with outcome probabilities given by $p_0(b)=\mathrm{tr}(E_b\rho_0)$ and $p_1(b)=\mathrm{tr}(E_b\rho_1)$. The statistical distance as such makes sense only when the number of samplings of the probability distribution is large. Hence, in the quantum extension of this notion it is implicitly assumed that one performs the same measurement on each of a large number N of copies of the state $\rho\in\{\rho_0,\rho_1\}$. The optimization over such local repeated measurements leads to one of the most widely used distinguishability measures [27]: The (quantum) fidelity $F(\rho_0,\rho_1)$, defined in Eq. (21).

The fidelity, or statistical distance, has many desirable properties: (i) it is easily computable; (ii) for pure states it reduces to the standard distance given by the angle between rays in the Hilbert space $\mathcal{H}$; (iii) as mentioned above, it provides bounds to P_e. Nevertheless, a strict physical inter-

pretation is so far unclear, and its definition is based on repeated local measurements, while quantum mechanics allows for much more general ways to access the information contained in the N copies, via collective measurements on the whole of them.

B. Quantum Chernoff distance

A very natural and also operational distinguishability measure is provided by the error probability of discrimination. As a first candidate, one could take this very error probability P_e for a given fixed number N of copies. However, the choice of a particular N in such a definition would not only be arbitrary but also problematic since one can find examples [15], where $P_e(\rho_0,\rho_1;N)>P_e(\rho_0',\rho_1';N)$, whereas $P_e(\rho_0,\rho_1;M)<P_e(\rho_0',\rho_1';M)$ for a different number M of copies. A straightforward way to go around this problem is to use the asymptotic expressions for $N\to\infty$ and define the distinguishability measure as the largest rate exponent in Eq. (26). We further note that the presence of the logarithm ensures that $D(\rho_0,\rho_1)=0$ if and only if $\rho_0=\rho_1$, while the minus sign makes distinguishability decrease as discrimination becomes more difficult, i.e., as P_e increases.

The quantum Chernoff information $D_{\mathrm{QC}}(\rho_0,\rho_1)$ is therefore a physically meaningful and efficiently computable distinguishability measure. Note that Eq. (27) does not *stricto sensu* define a distance, since it does not fulfill the triangular inequality. It has, however, all of the other properties that one should expect from a reasonable measure. This, in itself, is already a remarkable fact since, as far as measures and metrics are concerned, there is usually a compromise among operational definiteness, computability, and contractivity [28]. For instance, the distance proposed in Ref. [29], although having an operational definition, is not contractive.

We point out that another operational distinguishability measure can be obtained in asymmetric hypothesis testing by minimizing the type-II error rate while keeping the type-I error rate upper bounded by a fixed value. The optimal error rate in this situation is provided by the quantum Stein's lemma [10,11] and leads to the well-known quantum relative entropy. Despite having an operational meaning, the quantum relative entropy has two obvious drawbacks as a distinguishability measure: it is not symmetric in its arguments and it diverges if one of the states is pure.

C. Classical Chernoff distance: Local measurements

In the derivation of the quantum Chernoff bound one optimizes over all possible quantum measurements, in particular over quantum joint measurements on $\mathcal{H}^{\otimes N}$, that act over all the N copies coherently. It is of great interest, both theoretically and in practice, to know whether such joint measurements are strictly necessary to attain the bound or one can make do with separable ones (which include those that can be implemented with local operations and classical communication, simply known as LOCC measurements). As far as we are aware, the answer to this is unknown. This question is also relevant in connection with the operational meaning attached to $D(\rho_0,\rho_1)$. In this section we focus on this operational aspect and compute $D(\rho_0,\rho_1)$ from its definition in Eq. (26) assuming that the discrimination protocol P_e refers to is constrained to make use of the same individual measurements, defined by a local POVM $\{E(b)\}_{b=1}^{M}$, on each of the N available copies. We loosely refer to these protocols as local. Local protocols are relevant from the theoretical point of view since they help to elucidate the role of quantum correlated measurements in asymptotic hypothesis testing. For example, in quantum phase estimation local measurements suffice to achieve the collective bounds [22]. Here, we will show that these protocols do not achieve the quantum Chernoff bound. In addition, from a more practical point of view, local protocols are much simpler to implement experimentally, specially in a situation where the number of subsystems is increasingly large.

In such a local protocol, after the measurements have been performed we have a sample of N elements of the probability distribution $p_i(b)=\mathrm{tr}(E_b\rho_i)$, $i=0,1$, based on which we have to discriminate between the candidate H_0 or H_1. In such a scenario the error probability, which we call P_e^{loc}, can be obtained using the classical Chernoff bound (1) applied to the distributions p_0 and p_1. One can thus define the error exponent (26) and thereby introduce a new operational distinguishability measure based on local discrimination

$$D_{\mathrm{CC}}(\rho_0,\rho_1)=-\min_{\{E_b\}}\min_{s\in[0,1]}\log\sum_b p_0^s(b)p_1^{1-s}(b), \quad (32)$$

where the subscript CC reminds us that we have made use of the classical Chernoff bound.

The measure $D_{\mathrm{CC}}(\rho_0,\rho_1)$ is obtained by maximizing the rate exponent over all possible single-copy generalized measurements $\{E_b\}_{b=1}^{M}$ (just as is done for the fidelity). Unfortunately, there is no simple closed expression for this maximum for general mixed states. However, we do encounter again the relation (22) with the fidelity: since the square root of the statistical fidelity $\mathcal{F}(p_0,p_1)$ upper bounds P_{CC} in (1), it also upper bounds the local error probability P_e^{loc}. That is,

$$P_e^{\mathrm{loc}}\le P_{\mathrm{CC}}\le\min_{\{E_b\}}\frac{\sqrt{\mathcal{F}(p_0,p_1)}}{2}=\frac{\sqrt{F(\rho_0,\rho_1)}}{2} \quad (33)$$

and

$$D_{\mathrm{CC}}(\rho_0,\rho_1)\ge-\frac{1}{2}\log F(\rho_0,\rho_1). \quad (34)$$

Since $D_{\mathrm{QC}}(\rho_0,\rho_1)\ge D_{\mathrm{CC}}(\rho_0,\rho_1)$, we note that whenever $D_{\mathrm{QC}}(\rho_0,\rho_1)=-(1/2)\log F(\rho_0,\rho_1)$ the inequality (34) has to be saturated. This, in turn, means that in this situation one can optimally discriminate between H_0 and H_1 just by performing a fixed local measurement on each of the N copies (no collective measurements are required to attain the quantum Chernoff bound).

There is still another important situation when the quantum Chernoff bound is attainable by local measurements: when one of the states (say ρ_0) is pure. If this is the case, Eq. (24) holds and $D_{\mathrm{QC}}(\rho_0,\rho_1)=-\log F(\rho_0,\rho_1)$. To prove that $D_{\mathrm{CC}}(\rho_0,\rho_1)=D_{\mathrm{QC}}(\rho_0,\rho_1)$, let us consider the two-outcome measurement defined by $E_0=\rho_0$, $E_1=\mathbb{1}-\rho_0$. Note

that $p_0(1)=\mathrm{tr}(E_1\rho_0)=0$ and $p_0(0)=\mathrm{tr}(E_0\rho_0)=1$. After performing this measurement on each of the N copies the protocol proceeds as follows: we accept H_0 if all of the outcomes are 0, otherwise we accept H_1. One may refer to this classical data processing as unanimity vote [30]. The error probability can be easily computed by noticing that no error occurs unless we get N times the outcome 0 [since $p_0(1)=0$]. Therefore,

$$P_e^{\mathrm{loc}} = \pi_1 p_1^N(0) = \pi_1[\mathrm{tr}(\rho_0\rho_1)]^N = \pi_1[F(\rho_0,\rho_1)]^N, \quad (35)$$

where the last equality holds because ρ_0 is assumed to be a pure state. From this equation it follows immediately that $D_{\mathrm{CC}}(\rho_0,\rho_1)=-\log F(\rho_0,\rho_1)=D_{\mathrm{QC}}(\rho_0,\rho_1)$, and the quantum Chernoff bound is attainable by local measurements. It also follows from the first equality in Eq. (35) that this result corresponds to taking the limit $s\to 0$ in Eq. (1).

V. METRICS

The set of states of a quantum system, as that of classical probability distributions on a given sample space,[4] can be endowed with a metric structure [31], and thus thought of as a Riemannian manifold. This enables us to relate geometrical concepts (e.g., distance, volume, curvature, parallel transport) to physical ones (e.g., state discrimination and estimation, geometrical phases). Among the novel applications of metrics in quantum information, they have been recently used to characterize quantum phase transitions [32].

The first step toward this geometric approach to quantum states is to define the line element ds or (infinitesimal) distance between two neighboring "points" ρ and $\rho-d\rho$. All local properties follow from this definition. More precisely, they follow from the metric, i.e., from the set of coefficients of ds^2 when written as a quadratic form in the differentials of the coordinates (parameters) that specify the quantum states. There is, however, no unique choice of ds unless some monotonicity conditions are invoked.

For classical probability distributions $\{p(b)\}$, a line element is singularized (up to a proportionality factor) by imposing that it be nonincreasing under stochastic maps. It is the well known Fisher metric (in what follows the terms metric and line element will be used interchangeably):

$$ds_F^2 = \frac{1}{4}\sum_b \frac{[dp(b)]^2}{p(b)}. \quad (36)$$

In contrast to the classical case, the monotonicity condition under completely positive (quantum stochastic) maps does not define a metric uniquely, which explains why a substantial body of research on quantum metrics has emerged over the last years. Among the main developments, Petz [33] has characterized the family of quantum contractive metrics by establishing a correspondence with operator-monotone functions (from which they also take the name of monotone metrics).

[4]For the sake of clarity, in this section we assume a finite sample space, but the results also hold for general probability measures over continuous spaces.

An alternative, more physical approach is to define a line element from a suitable distinguishability measure between infinitesimally close states. A remarkable example is given in Ref. [34]. In this seminal paper Braunstein and Caves consider a one-parameter family of states $\rho(\theta)$ and map the problem of distinguishability to that of estimating the parameter θ optimally. They define a line element ds_{BC}^2 as $d\theta^2$ expressed in the appropriate units of statistical deviation (roughly speaking, $d\theta^2$ divided by the minimal error in the estimation of θ). By making use of classical statistical methods (Cramér-Rao bound) they find

$$ds_{\mathrm{BC}}^2 = 4\max_{\{E_b\}} ds_F^2 = \max_{\{E_b\}} I_F \; d\theta^2, \quad (37)$$

where $I_F=\Sigma_b[dp(b)/d\theta]^2/p(b)$ (it is the so called Fisher information), with $p(b)=\mathrm{tr}[E_b\rho(\theta)]$, and the maximization is over all possible POVM measurements $\{E_b\}$ on a single copy of $\rho(\theta)$. They also succeed in giving a closed expression for ds_{BC}^2 and show that their metric coincides up to a factor with that induced by the Bures-Uhlmann distance [35,36]

$$d_{\mathrm{BU}}(\rho_0,\rho_1) = \sqrt{2}[1-\sqrt{F(\rho_0,\rho_1)}]^{1/2}. \quad (38)$$

More precisely, they show that $ds_{\mathrm{BC}}^2=4ds_{\mathrm{BU}}^2$, where

$$ds_{\mathrm{BU}}^2 \equiv [d_{\mathrm{BU}}(\rho,\rho-d\rho)]^2 \quad (39)$$

[see also Eq. (69) below] and a series expansion to $O(d\rho^2)$ is understood in the right-hand side of this equation. We note in passing that for commuting states, i.e., classical probability distributions, the Bures-Uhlmann line element ds_{BU}^2 coincides with the Fisher metric (36). A quantum metric with such normalization is said to be Fisher adjusted.

Although one can obtain a finite distance $d_{\mathrm{BC}}(\rho_0,\rho_1)$ for arbitrary states ρ_0 and ρ_1 by integrating ds_{BC} along geodesics, it is important to notice that the operational meaning of the Braunstein and Caves metric is lost in the process. In the spirit of the Braunstein and Caves physical approach to metrics, we next consider the distinguishability measures D_{QC} and D_{CC}, discussed in Sec. IV, for infinitesimally close states and derive line elements with the same operational meaning, which we call ds_{QC} and ds_{CC}, respectively. For ds_{QC} we also give the volume element and the prior probability distribution, whereas those corresponding to the metric ds_{CC} can be easily found in the literature since, as will be shown, ds_{CC}^2 is proportional to the widely studied Bures metric ds_{BU}^2.

Before we start we would like to point out that one could also consider line elements induced by other quantities, such as the quantum relative entropy, which, as we saw above, also has a clear operational interpretation. The quantum relative entropy induces the so-called Kubo-Mori metric [37], which has the drawback of being singular for pure states.

A. Quantum Chernoff metric

For neighboring density matrices ρ and $\rho-d\rho$ (e.g., those for which their independent matrix elements differ by an infinitesimal amount) the distinguishability measure $D(\rho,\rho-d\rho)$ defines a metric, as in Eq. (39). For the quantum Chernoff measure D_{QC}, this metric can be computed from Eq. (27) [38]:

$$ds^2_{\rm QC} = 1 - \min_{s\in[0,1]} {\rm tr}[\rho^s(\rho - d\rho)^{1-s}] + \cdots, \tag{40}$$

where the dots stand for higher order terms in $d\rho$ that will not contribute to ds^2 and we have also used that $\log y = y - 1 + \cdots$. We now recall the integral representation

$$a^t = \frac{\sin(t\pi)}{\pi}\int_0^\infty dx \frac{ax^{t-1}}{a+x}; \quad 0 < t < 1 \tag{41}$$

and its derivative

$$ta^{t-1} = \frac{\sin(t\pi)}{\pi}\int_0^\infty dx \frac{x^t}{(a+x)^2}; \quad -1 < t < 1. \tag{42}$$

These representations hold for $a>0$ and can be straightforwardly extended to positive matrices. In particular, using Eq. (41) and the convergent sequence

$$\frac{1}{a-b} = a^{-1} + a^{-1}ba^{-1} + a^{-1}ba^{-1}ba^{-1} + \cdots, \tag{43}$$

which also holds for matrices provided $a>b$, one can write, up to second order in $d\rho$,

$$\begin{aligned}(\rho - d\rho)^{1-s} &= c_s\int_0^\infty dx(\rho - d\rho)\frac{x^{-s}}{\rho - d\rho + x} \\ &\approx c_s\int_0^\infty dx\, x^{-s}(\rho - d\rho)\left(\frac{1}{\rho+x} + \frac{1}{\rho+x}d\rho\frac{1}{\rho+x}\right. \\ &\quad \left. + \frac{1}{\rho+x}d\rho\frac{1}{\rho+x}d\rho\frac{1}{\rho+x}\right),\end{aligned} \tag{44}$$

where $c_s = \pi^{-1}\sin(s\pi)$. Inserting this expansion in Eq. (40) one finds

$$\begin{aligned}ds^2_{\rm QC} = \max_{s\in(0,1)} c_s\int_0^\infty dx\, {\rm tr}\left(\frac{x^{1-s}}{(\rho+x)^2}\rho^s d\rho\right. \\ \left. + \frac{x^{1-s}}{(\rho+x)^2}\rho^s d\rho\frac{1}{\rho+x}d\rho\right).\end{aligned} \tag{45}$$

The first term in the integrand vanishes, as can be seen by using Eq. (42) and ${\rm tr}\, d\rho = 0$, while the second term can be computed in the eigenbasis $\{|i\rangle\}$ of ρ; $\rho = \Sigma_i \lambda_i |i\rangle\langle i|$:

$$\begin{aligned}ds^2_{\rm QC} &= \max_{s\in(0,1)} \sum_{ij} c_s\int_0^\infty dx\, x^{1-s}\frac{\lambda_i^s |\langle i|d\rho|j\rangle|^2}{(\lambda_i + x)^2(\lambda_j + x)} \\ &= \frac{1}{2}\max_{s\in(0,1)} \sum_{ij}\frac{|\langle i|d\rho|j\rangle|^2}{(\lambda_i - \lambda_j)^2}(\lambda_i + \lambda_j - \lambda_i^s\lambda_j^{1-s} - \lambda_j^s\lambda_i^{1-s}) \\ &= \frac{1}{2}\sum_{ij}\frac{|\langle i|d\rho|j\rangle|^2}{(\lambda_i - \lambda_j)^2}(\lambda_i + \lambda_j - 2\sqrt{\lambda_i\lambda_j}),\end{aligned} \tag{46}$$

where in the second equality we have taken into account that $d\rho = d\rho^\dagger$, which enabled us to symmetrize the expression in parenthesis that multiplies $|\langle i|d\rho|j\rangle|^2$ in the sum (this symmetrization gives the factor 1/2). The quantum Chernoff metric can be finally written as

$$ds^2_{\rm QC} = \frac{1}{2}\sum_{ij}\frac{|\langle i|d\rho|j\rangle|^2}{(\sqrt{\lambda_i} + \sqrt{\lambda_j})^2}. \tag{47}$$

The quantum Chernoff metric belongs to the family of contractive quantum metrics, as it should, since by construction the probability of error cannot be improved by a preprocessing of the states. In fact the quantum Chernoff metric coincides with a member of this family that has been explicitly written by Petz in Ref. [39] and with the so-called Wigner-Yanase metric, which has been recently studied in depth by the authors of Ref. [40]. In particular, the geodesic distance, the geodesic path, and the scalar curvature of the quantum Chernoff metric can be read off from their Eqs. (5.1)–(5.3). In Ref. [41] its relation to the uncertainty principle is discussed.

By separating diagonal from off-diagonal terms, the metric in Eq. (47) can also be written as

$$ds^2_{\rm QC} = \sum_i \frac{(d\lambda_i)^2}{8\lambda_i} + \sum_{i<j}\frac{|\langle i|d\rho|j\rangle|^2}{(\sqrt{\lambda_i} + \sqrt{\lambda_j})^2}. \tag{48}$$

Next, we wish to identify the degrees of freedom in the off-diagonal terms. We will see that they correspond to infinitesimal unitary transformations acting on ρ (which leave its eigenvalues unchanged). This is most conveniently done by parametrizing ρ by its eigenvalues and eigenvectors, namely, by λ_i and the components of $|i\rangle$ onto a given canonical basis $\{|\alpha_k\rangle\}$:

$$U_{ki} \equiv \langle\alpha_k|i\rangle = \langle\alpha_k|U|\alpha_i\rangle \tag{49}$$

(naturally, it also holds that $U_{ki} = \langle k|U|i\rangle$). A neighboring density matrix $\rho' = \Sigma_i \lambda_i' |i'\rangle\langle i'|$ is thus parametrized by $\lambda_i' = \lambda_i + d\lambda_i$ and $U_{ki}' = U_{ki} + dU_{ki} = \langle\alpha_k|i'\rangle$. We further note that $|i'\rangle = (\mathbb{1} + \delta T)|i\rangle$, where δT is anti-Hermitian, $\delta T^\dagger = -\delta T$. It is actually the infinitesimal generator along the direction in parameter space that takes $\{|i\rangle\}$ into $\{|i'\rangle\}$. It follows that $dU_{ki} = \langle\alpha_k|\delta T|i\rangle$. The matrix elements of $d\rho$ can be expressed as

$$\begin{aligned}\langle i|d\rho|j\rangle &= \langle i|(\rho' - \rho)|j\rangle = \sum_k \langle i|k'\rangle\langle k'|j\rangle\lambda_k' - \lambda_i\delta_{ij} \\ &= d\lambda_i\delta_{ij} + (\lambda_j - \lambda_i)\langle i|\delta T|j\rangle + O(\delta T^2)\end{aligned} \tag{50}$$

and those of δT as

$$\begin{aligned}\langle i|\delta T|j\rangle &= \sum_k \langle i|\alpha_k\rangle\langle\alpha_k|\delta T|j\rangle = \sum_k U_{ki}^* dU_{kj} \\ &= \sum_k \langle\alpha_i|U^\dagger|\alpha_k\rangle\langle\alpha_k|dU|\alpha_j\rangle = \langle\alpha_i|U^\dagger dU|\alpha_j\rangle \\ &\equiv (U^\dagger dU)_{ij},\end{aligned} \tag{51}$$

where we have used Eq. (49) in going from the first to the second line [the very same matrix elements of δT can also be written as $(dUU^\dagger)_{ij}$ in the eigenbasis of ρ]. Substituting these relations back into Eq. (48) we obtain

$$ds^2_{\rm QC} = \sum_i \frac{(d\lambda_i)^2}{8\lambda_i} + \sum_{i<j}(\sqrt{\lambda_i} - \sqrt{\lambda_j})^2|(U^\dagger dU)_{ij}|^2. \tag{52}$$

The same expression can also be derived by differentiating

$$\rho = U^\dagger \rho^{(0)} U, \tag{53}$$

where $\rho^{(0)} \equiv \Sigma_i \lambda_i |\alpha_i\rangle\langle\alpha_i|$ is diagonal in the canonical basis and has the spectrum of ρ.

Equation (52) displays the metric ds^2_{QC} in a very suggestive form. Any density matrix can be parametrized by its eigenvalues $\{\lambda_i\}$ and the unitary matrix U that diagonalizes it. Equation (52) expresses the infinitesimal distance between two such matrices in terms of these very parameters. The first term is immediately recognized as the (Fisher) metric on the $(d-1)$-dimensional simplex of eigenvalues of ρ, which is assumed to be $d \times d$ throughout the rest of this section (note that $\Sigma_i \lambda_i = 1$, which implies $\Sigma_i d\lambda_i = 0$). Thus, *stricto senso*, it should be expressed in terms of a set of $d-1$ independent eigenvalues. If we choose this set to be $\{\lambda_i\}_{i=1}^{d-1}$ the first term in Eq. (52) becomes

$$\frac{1}{8} \sum_{i,j}^{d-1} g_F^{ij} d\lambda_i d\lambda_j, \tag{54}$$

where the subscript F stands for Fisher and

$$g_F^{ij} = \frac{\delta^{ij}}{\lambda_i} + \frac{\Phi^{ij}}{1 - \sum_i^{d-1} \lambda_i}; \quad \Phi^{ij} = 1 \quad \text{for} \quad 1 \le i,j \le d-1. \tag{55}$$

It follows that the determinant of g_F, which we will need below, is

$$\det g_F = (\lambda_1 \cdots \lambda_{d-1} \lambda_d)^{-1}. \tag{56}$$

The second term in Eq. (52) contains the factors $|(U^\dagger dU)_{ij}|^2$, which are invariant under left-multiplication [since the left-hand side of Eq. (51) is independent of the choice of basis $\{|\alpha\rangle\}$]. Hence, the normalized volume element induced by these terms will coincide with the (unique) Haar measure dV_H of $U(d)/[U(1)]^d$, known as the flag manifold $\mathrm{Fl}_{\mathbb{C}}^{(d)}$ (see, e.g., Ref. [42], and references therein). Using the wedge product of differential forms, this Haar measure can be written as

$$dV_H = \frac{1}{C_H} \Big| \bigwedge_{i<j} \mathrm{Re}(U^\dagger dU)_{ij} \wedge \mathrm{Im}(U^\dagger dU)_{ij} \Big|, \tag{57}$$

where C_H is a normalization constant so that $\int dV_H = 1$. Note that the one-form basis in Eq. (57) contains $2 \times [d(d-1)/2]$ (real and independent) elements, which indeed coincides with the (d^2-d)-independent parameters of $U(d)/[U(1)]^d$.

Volume elements (derived from metrics) are of great interest because they give a canonical way of defining prior probability distributions on continuous sets. According to this approach, Eqs. (52)–(57) provide a means to define such probability distribution for general density matrices: if $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ is a set of independent real parameters that specifies the density matrices as $\rho(\boldsymbol{\theta})$ and the metric is written as $ds^2 = d\boldsymbol{\theta} \boldsymbol{g}\, d\boldsymbol{\theta}^t$ (i.e., $\boldsymbol{g}$ is the metric tensor), then we can define the prior $\mathcal{P}[\rho(\boldsymbol{\theta})]$ through the relation $\mathcal{P}[\rho(\boldsymbol{\theta})] \Pi_\alpha d\theta_\alpha = dV / \int dV$, where $dV = \sqrt{\det \boldsymbol{g}} \Pi_\alpha d\theta_\alpha$. It follows from Eq. (52) that $\mathcal{P}[\rho(\boldsymbol{\theta})]$ is the product of two independent probability distributions: one that depends exclusively on the parameters encoded in the unitary matrix U and expresses the fact that they are simply distributed according to the Haar measure dV_H; and one, denoted as $\mathcal{P}(\{\lambda_i\})$, that gives the probability distribution of eigenvalues. The latter can be written as

$$\mathcal{P}(\{\lambda_i\}) = \frac{1}{C_d} \prod_i^d \frac{1}{\sqrt{\lambda_i}} \delta\Big(1 - \sum\nolimits_j \lambda_j\Big) \prod_{i<j} (\sqrt{\lambda_i} - \sqrt{\lambda_j})^2, \tag{58}$$

where for a given dimension d the constant C_d is chosen to ensure that probability adds up to 1.

The prior distribution on the simplex of eigenvalues of ρ for the Bures metric (see below), analogous to $\mathcal{P}(\{\lambda_i\})$ in Eq. (58), was proposed in Ref. [43], but it took considerable efforts to compute the right normalization constant. Slater [44] gave values for dimensions $d = 3, 4, 5$ and finally Sommers and Życzkowski [45] managed to give a general expression for arbitrary finite dimensions. Here we will compute C_d following similar techniques. Let us also point out that the volumes for various monotone metrics have been computed for the specific case of qubits in Ref. [46].

The coefficient C_d is defined by the normalization condition $\int \mathcal{P}(\{\lambda_i\}) \Pi_i^d d\lambda_i = 1$. Thus, $C_d = I(1)$, where

$$I(r) = \int_0^\infty \prod_i^d \frac{d\lambda_i}{\sqrt{\lambda_i}} \delta\Big(r^2 - \sum\nolimits_j \lambda_j\Big) \prod_{i<j} (\sqrt{\lambda_i} - \sqrt{\lambda_j})^2. \tag{59}$$

Although we only need this integral for $r = 1$, the introduction of this radial parameter r enables us to compute the normalization $I(1)$ more easily. We first note that by rescaling $\lambda_i \to r^2 \lambda_i$ one gets

$$I(r) = r^{d^2-2} I(1) \tag{60}$$

[i.e., $I(r)$ is a homogeneous function of r of degree d^2-2], and thus

$$\int_0^\infty dr\, r e^{-r^2} I(r) = I(1) \int_0^\infty dr\, r^{d^2-1} e^{-r^2}. \tag{61}$$

It follows from this equation that

$$C_d = I(1) = \frac{2^d}{\Gamma(d^2/2)} \int_0^\infty \prod_i^d \frac{d\lambda_i}{2\sqrt{\lambda_i}} e^{-\Sigma_i \lambda_i} \times \prod_{i<j} (\sqrt{\lambda_i} - \sqrt{\lambda_j})^2. \tag{62}$$

This expression can be further simplified by the change of variables $\lambda_i \to t_i = \sqrt{\lambda_i}$, which leads to

$$C_d = \frac{2^d}{\Gamma(d^2/2)} \int_0^\infty \prod_i^d dt_i e^{-t_i^2} \prod_{i<j} (t_i - t_j)^2. \tag{63}$$

By expanding the square of the Vandermonde determinant $\Pi_{i<j}(t_i - t_j)$, one could, in principle, compute C_d in terms of Euler gamma functions. However, this is very impractical since the number of terms in such an expansion grows exponentially with d. A much more efficient way to proceed is as follows. Let $\{P_k(t) = a_k t^k + a_{k-1} t^{k-1} + \cdots + a_1 t + a_0\}$, $a_k \neq 0$, be a family or orthonormal polynomials in the set $[0, \infty)$ with a weight function of Hermite type, so that

$$\int_0^\infty dt\, e^{-t^2} P_k(t) P_l(t) = \delta_{kl}. \tag{64}$$

Note that $\{P_k(t)\}$ are not Hermite polynomials, since the integration range is $[0,\infty)$ instead of $(-\infty,\infty)$. Now, if we define the renormalized polynomials $Q_k(t) \equiv P_k(t)/a_k$ it is not hard to show that

$$\prod_{i<j}(t_i - t_j) = \begin{vmatrix} Q_{d-1}(t_1) & Q_{d-2}(t_1) & \cdots & Q_0(t_1) \\ Q_{d-1}(t_2) & Q_{d-2}(t_2) & \cdots & Q_0(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ Q_{d-1}(t_d) & Q_{d-2}(t_d) & \cdots & Q_0(t_d) \end{vmatrix}. \tag{65}$$

Substituting in to Eq. (63) and using the orthonormality of P_k, one has

$$C_d = \frac{2^d d!}{\Gamma(d^2/2)} \prod_{k=0}^{d-1} a_k^{-2}. \tag{66}$$

In contrast to the examples considered in Ref. [45], and as far as we are aware, there is no known closed expression for the leading coefficients a_k for the case at hand. However, Eq. (66) provides an efficient way of computing the quantum Chernoff normalization constant C_d; e.g., by applying the Gram-Schmidt orthogonalization algorithm [with the internal product defined in Eq. (64)] one easily obtains the coefficients a_k, and thereby C_d. We give the value of this constant for $d \leq 6$:

$$C_2 = \pi - 2,$$

$$C_3 = \frac{8}{35}(\pi - 3),$$

$$C_4 = \frac{6\pi^2 - 29\pi + 32}{6720},$$

$$C_5 = \frac{128(72\pi^2 - 435\pi + 656)}{21082276215},$$

$$C_6 = \frac{9(480\pi^3 - 3747\pi^2 + 9352\pi) - 65536}{2023466257612800}. \tag{67}$$

B. Classical Chernoff and Bures metric

From the measure $D_{\mathrm{CC}}(\rho_0,\rho_1)$ in Eq. (32), based on local measurements, one can readily obtain the corresponding local metric. If $0 < p(b) = \mathrm{tr}(\rho E_b) < 1$ for every measurement outcome b, direct differentiation of $D_{\mathrm{CC}}(\rho,\rho - d\rho)$ leads to

$$ds^2_{\mathrm{CC}} = \frac{1}{2} \max_{\{E_b\}} ds^2_F, \tag{68}$$

where ds^2_F is the Fisher metric (36), with $p(b) = \mathrm{tr}(\rho E_b)$, $dp(b) = \mathrm{tr}(d\rho E_b)$, and $s^* = 1/2$ being the value of s that achieves this minimum in Eq. (32). The maximization of Eq. (36) over the local measurements $\{E_b\}_{b=1}^M$, which commutes with the minimization over s as long as $p(b) \neq 0,1$, results in [34]

$$ds^2_{\mathrm{BU}} = \frac{1}{2} \sum_{ij} \frac{|\langle i|d\rho|j\rangle|^2}{\lambda_i + \lambda_j} \tag{69}$$

Or, equivalently,

$$ds^2_{\mathrm{BU}} = \sum_i \frac{(d\lambda_i)^2}{4\lambda_i} + \sum_{i<j} \frac{(\lambda_i - \lambda_j)^2}{\lambda_i + \lambda_j} |(U^\dagger dU)_{ij}|^2, \tag{70}$$

where we use the same notation as in Eqs. (47) and (52), respectively. This is the Bures-Uhlmann metric, which, as mentioned above, can be also obtained from the Bures distance (38) [47]. From Eq. (68) we then have that the local metric is proportional to the Bures metric for strictly mixed states and

$$ds^2_{\mathrm{CC}} = \frac{1}{2} ds^2_{\mathrm{BU}} = \frac{1}{2}[1 - F(\rho,\rho - d\rho)], \tag{71}$$

where the last equality holds to order $d\rho^2$. The corresponding prior probability distribution (quantum Jeffreys prior) was derived and calculated in Refs. [43–45].

If one of the states is pure (say ρ_0, as in previous sections) then the classical distribution $p(b)$ becomes degenerate $[p(0)=1]$ for the optimal choice $E_0 = \rho_0$ (recall the last comments in Sec. IV C), and the previous derivation does not hold. In this case, the optimal choice of s in Eq. (1) is obtained by taking the limit $s \to 0$, as we already discussed in Sec. IV C. Recalling the first equality in Eq. (35), we obtain $D_{CC}(\rho,\rho - d\rho) = -\log[p(0) - dp(0)] = dp(0)$ [note that $dp(0) \geq 0$ since $1 \geq p(0) - dp(0) = 1 - dp(0)$], which is linear in $dp(b)$ and therefore does not define a proper metric in probability space. From the results of Sec. IV C we also know that if one of the states is pure then $D_{\mathrm{CC}}(\rho_0,\rho_1) = -\log F(\rho_0,\rho_1)$ and therefore

$$ds^2_{\mathrm{CC}} = 1 - F(\rho,\rho - d\rho) = ds^2_{\mathrm{BU}} \tag{72}$$

for pure states. This agrees with the previous discussion since $dp(0) = 1 - F(\rho,\rho - d\rho)$ if ρ is a pure state. Equation (72) has to be taken with special care. It gives a valid metric for the set of pure states (which only includes variations in the unitary parameters), i.e., when $\rho - d\rho$ is also a pure state $(\rho - d\rho = U\rho U^\dagger)$. Moreover, for pure states ds^2_{CC} coincides with the Fubini-Study metric [recall that the Bures-Uhlmann metric is Fubini-Study-adjusted [45], hence this statement follows from Eq. (72)].

By combining Eqs. (71) and (72), we see that local metric ds^2_{CC} shows a discontinuity when the mixed state ρ approaches the set of pure states. In contrast, the metric based on collective measurements, that is, the quantum Chernoff metric (47), does not have this pathology. This can be seen by comparing the $i<j$ $(d\lambda_i = 0)$ terms in Eq. (52) with those in Eq. (70) (the diagonal terms $i=j$ coincide). As $\lambda_j \to \delta_{1j}$ (ρ approaches a pure state), we readily see that $ds^2_{\mathrm{QC}} \to ds^2_{\mathrm{BU}}$. In the opposite situation, when ρ approaches the completely mixed state $\mathbb{1}/d$, we can write $\lambda_i = 1/d + \epsilon_j$, where ϵ_j approaches zero. Expanding the $i<j$ terms in both Eqs. (52) and (70) we can check that $ds^2_{\mathrm{QC}} = \frac{1}{2} ds^2_{\mathrm{BU}}$ up to terms of order

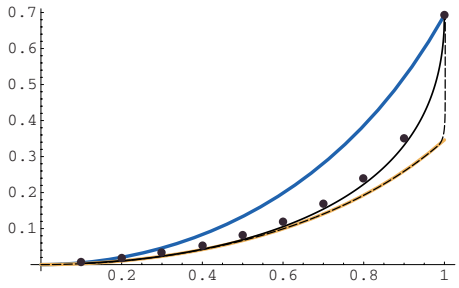

FIG. 2. (Color online) Measures of distinguishability between two-qubit states with relative angle $\theta=\pi/2$ for different values of $r=r_0=r_1$: Values extrapolated from exact evaluation of the probability of error for $30\leq N\leq 35$ (dots); bounds provided by the fidelity (28) (shaded); measure based on identical local measurements, i.e., $D_{\mathrm{CC}}(\rho_0,\rho_1)$ (dashed); measure based on collective measurements, i.e., $D_{\mathrm{QC}}(\rho_0,\rho_1)$ (solid line). In this plot the logarithm is taken to the natural base.

ϵ^3. We conclude that the quantum Chernoff metric smoothly interpolates between the two components (that on strictly mixed states and that on pure states) of the local metric ds^2_{CC}. We will come back to this point in the next section, where qubit states are discussed as an example to illustrate the results in this and in previous sections.

VI. QUBIT STATES

In this section we apply our results to qubit mixed states, that is, general two-dimensional states. We will first study the distinguishability measures D_{QC} and D_{CC} and then move on to the corresponding metrics and priors.

For qubits one has $\rho_i=(\mathbb{1}+\vec{r}_i\cdot\vec{\sigma})/2$, $i=0,1$, where $\vec{r}_i$ is the Bloch vector of ρ_i, $0\leq|\vec{r}_i|\equiv r_i\leq 1$. The eigenvalues of ρ_i are $\wp_i=(1+r_i)/2$ and $\bar{\wp}_i\equiv 1-\wp_i$. It is straightforward to obtain

$$Q_s\equiv \mathrm{tr}\,\rho_0^s\rho_1^{1-s}=(\wp_0^s\wp_1^{1-s}+\bar{\wp}_0^s\bar{\wp}_1^{1-s})\cos^2\frac{\theta}{2}+(\wp_0^s\bar{\wp}_1^{1-s}+\bar{\wp}_0^s\wp_1^{1-s})\sin^2\frac{\theta}{2}, \tag{73}$$

where θ is the angle between $\vec{r}_0$ and $\vec{r}_1$. The value of s that minimizes Q_s and hence gives Eqs. (14) and (27) is in general a function of r_i and θ. However, one can check that in the particular case $r_0=r=r_1$ the minimum is at $s^*=1/2$.[5]

In Fig. 2 we plot the quantum Chernoff distinguishability measure $D_{\mathrm{QC}}(\rho_0,\rho_1)$ and the measure based on local measurements $D_{\mathrm{CC}}(\rho_0,\rho_1)$ together with the bounds (28) provided by the fidelity, for states of equal purity $r_0=r_1=r$ and for $\theta=\pi/2$. Notice that in general local measurements perform much worse than the collective ones and $D_{\mathrm{CC}}(\rho_0,\rho_1)$ runs remarkably close to (actually, coincides with) the fidelity lowerbound (28) for most values of r. However, as it approaches the pure-state regime ($r\to 1$) it rapidly increases toward its upper bound. The reason for this rapid change can be understood by recalling the unanimity vote protocol discussed in Sec. IV C. For two pure states, $\rho_i=|\psi_i\rangle\langle\psi_i|$ (as corresponds to $r=1$), it boils down to Ref. [30] projecting along one of the states, say $|\psi_0\rangle$, and its orthogonal $|\psi_0^\perp\rangle$. After performing this measurement on each of the N copies, if all of them project on $|\psi_0\rangle$, one claims that the unknown state is $|\psi_0\rangle$ (hypothesis H_0). However, if at least one of them projects on $|\psi_0^\perp\rangle$ the guess is $|\psi_1\rangle$ (one accepts H_1). This corresponds to $\xi=1$ in Eq. (7). For pure states it reaches the joint-measurement Chernoff bound by making use of a much less demanding local-measurement protocol (see also Refs. [30,48] for the optimal local strategy for finite N).

In contrast, near the completely mixed state $\mathbb{1}/2$, for low r, the optimal local strategy consists in choosing the measurement $\{E_0,E_1\}$ such that $p=p_0(0)=\mathrm{tr}(\rho_0E_0)=\mathrm{tr}(\rho_1E_1)=p_1(1)=\bar{q}$, with $p>1/2$. In this case, the acceptance of either H_0 or H_1 is done on the basis of a majority vote protocol: H_0 is accepted if the outcome 0 occurs more times than the outcome 1 does, i.e., $N_0=N/2$ [see also Eq. (7)]. It follows from Eq. (4) that $s^*=1/2$. Therefore, the lower-bound provided by the fidelity (28) is saturated [$s=s^*=1/2$ saturates the second inequality in Eq. (33) and thus it also saturates Eq. (34)]. This protocol is optimal up to a given value of the purity, i.e., for $r\leq r^*(\theta)$. For larger values of r the "voting rule" (given by ξ) starts changing and so does s^*. Accordingly, $D_{\mathrm{CC}}(\rho_0,\rho_1)$ moves away from its lower bound to end up saturating its upper bound at $r=1$.

We next consider the metrics induced by local and by joint measures. The former, in particular, requires special attention because of the abrupt behavior of $D_{\mathrm{CC}}(\rho_0,\rho_1)$ near the set of pure states. Indeed the critical value $r^*(\theta)$, beyond which majority vote is no longer optimal, goes to one as the relative angle θ between the Bloch vectors of the states becomes smaller; $r^*(\theta)\to 1$ as $\theta\to 0$. As a result, the sudden increase of $D_{\mathrm{CC}}(\rho_1,\rho_2)$ develops into a jump discontinuity at $r=1$ [from $-(1/2)\log F(\rho_0,\rho_1)$ if $r<1$ to $-\log F(\rho_0,\rho_1)$ if $r=1$]. For this reason, when defining the corresponding metric we have to distinguish these two regions: the set of strictly mixed states ($r<1$) and the set of pure states ($r=1$).

In the region $r<1$ the outcome probabilities will never be degenerate and the metric reduces to the Fisher metric, which upon optimization over local measurements coincides with one-half the Bures metric:

$$ds^2_{\mathrm{CC}}=\frac{1}{2}ds^2_{\mathrm{BU}}=\frac{1}{8}\left(\frac{dr^2}{1-r^2}+r^2d\Omega^2\right), \tag{74}$$

where $d\Omega^2=d\theta^2+\sin^2\theta d\phi^2$ is the usual metric on the two-sphere. In the region $r=1$ (pure states), the abovementioned unanimity vote protocol is optimal and the resulting metric is

$$ds^2_{\mathrm{CC}}=\frac{1}{4}d\Omega^2=ds^2_{\mathrm{FS}}, \tag{75}$$

where ds^2_{FS} is the well known Fubini-Study metric, which, as mentioned above, also coincides with the Bures metric ds^2_{BU} in the limiting case $r\to 1$. We notice again that ds^2_{CC} in Eq. (75) is a factor 2 larger than $\lim_{r\to 1} ds^2_{\mathrm{CC}}$ in Eq. (74), where

[5]Qubit states are an example for which the doubly stochastic matrix $D_{ij}=|\langle i|U|j\rangle|^2$ is symmetric ($D_{ij}=D_{ji}$). Therefore, for isospectral states, $Q_s(\rho,U\rho U^\dagger)=\Sigma_{ij}\lambda_i^s\lambda_j^{1-s}D_{i,j}=\Sigma_{ij}(\lambda_i^s\lambda_j^{1-s}+\lambda_j^s\lambda_i^{1-s})D_{ij}$, which has its minimum at $s^*=1/2$.

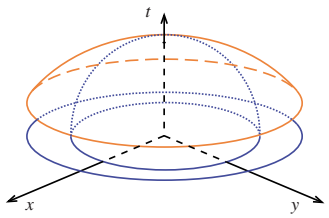

FIG. 3. (Color online) Uhlmann representation of the set of single qubit states according to metric ds^2_{CC}, based on P_e^{loc} for local repeated measurements (A and B), and according to the quantum Chernoff metric ds^2_{QC}, based on P_e for general joint measurements (C).

the limit is taken along the lines $dr=0$. The local distinguishability measure thus induces a discontinuous metric or, phrased in a different way, two different metrics for pure states or for strictly mixed states.

This can be visualized using the Uhlmann representation, that is, by embedding the Bloch sphere $r\leq 1$ in $\mathbb{R}^4$. To this end, one simply needs to define the new coordinate as $t=\cos\tau$, where $\sin\tau\equiv r$. In spherical coordinates one has

$$ds^2_{\mathrm{CC}}=\begin{cases}\frac{1}{8}(d\tau^2+\sin^2\tau d\Omega^2), & 0\leq\tau<\pi/2,\\ \frac{1}{4}d\Omega^2, & \tau=\pi/2,\end{cases} \tag{76}$$

where the first line correspond to strictly mixed states and the second to pure states. We note that in the second (first) line ds^2_{CC} is nothing but the standard metric on a two-sphere (the top half of a three-sphere) of radius 2^{-1} ($2^{-3/2}$).

In Fig. 3, A and B represent (the slice $z=0$ of) these two manifolds. One readily sees that the radius of B (pure states) is a factor $\sqrt{2}$ larger than that of the limiting circle of A (for $r\to 1\Leftrightarrow t\to 0$, i.e., $\tau\to\pi/2$).

The quantum Chernoff (collective-measurement based) metric can be readily obtained from Eq. (27) [or Eq. (52) particularized to qubit mixed states]:

$$ds^2_{\mathrm{QC}}=\frac{1}{8}\left[\frac{dr^2}{1-r^2}+2(1-\sqrt{1-r^2})d\Omega^2\right]. \tag{77}$$

This metric quantifies distinguishability of qubit states in a precise and operational way, and encapsulates the full power of quantum mechanics. It approaches the Fubini-Study metric ds^2_{FS} for pure states and also ds^2_{CC} for very mixed states, i.e., for small r. The metric smoothly interpolates between the two regimes. By defining $r\equiv\sin 2\tau$ with $0\leq\tau\leq\pi/4$ we obtain again the standard metric on a three-sphere but this time of radius $1/\sqrt{2}$:

$$ds^2_{\mathrm{QC}}=\frac{1}{2}(d\tau^2+\sin^2\tau d\Omega^2). \tag{78}$$

The corresponding manifold is denoted by C in Fig. 3. Geometrically the space of states endowed with the quantum Chernoff metric ds^2_{QC} is a spherical cap defined by $0\leq\tau\leq\pi/4$ whose radius is twice that of the Bures-like hemisphere A. In order to emphasize that the two metrics, are equal up to order r^3 at $\tau\approx 0$, i.e., $r\approx 0$ (near $1/2$), in the figure we have shifted the center of the larger sphere so as to make the two manifolds tangent at $\tau=0$. The fact that $ds^2_{\mathrm{CC}}=\frac{1}{2}ds^2_{\mathrm{BU}}=ds^2_{\mathrm{QC}}+O(r^4)$ is a particular example of a general relation that we discussed at the end of Sec. V B.

From the quantum Chernoff metric one can obtain a proper finite distance (satisfying the triangle inequality) by, for example, computing the geodesic distance

$$d_{\mathrm{QC}}(\rho_0,\rho_1)=\frac{\arccos(\cos\tau_0\cos\tau_1\cos\theta+\sin\tau_0\sin\tau_1)}{\sqrt{2}}, \tag{79}$$

where $r_i\equiv\sin 2\tau_i$ and θ is the relative angle between the respective Bloch vectors. The volume element and the prior distribution of density matrices for qubit mixed states, which we here denote as $\mathcal{P}[\rho(\vec{r})]$, can be easily obtained from the above metrics. According to the local and quantum Chernoff metrics we have, respectively,

$$\mathcal{P}_{\mathrm{CC}}[\rho(\vec{r})]=\frac{\sin\theta}{\pi^2}\frac{r^2}{\sqrt{1-r^2}}, \tag{80}$$

$$\mathcal{P}_{\mathrm{QC}}[\rho(\vec{r})]=\frac{\sin\theta}{2\pi(\pi-2)}\frac{1-\sqrt{1-r^2}}{\sqrt{1-r^2}}, \tag{81}$$

where it is understood that r and θ are the length and the azimuthal angle of the Bloch vector of ρ. Since the Haar volume density on the two-sphere is $\sin\theta/(4\pi)$, we see that the eigenvalues of ρ, $\lambda_\pm=(1\pm r)/2$ are distributed according to

$$\mathcal{P}_{\mathrm{CC}}(\lambda_\pm)=\frac{4}{\pi}\frac{r^2}{\sqrt{1-r^2}}, \tag{82}$$

$$\mathcal{P}_{\mathrm{QC}}(\lambda_\pm)=\frac{2}{\pi-2}\frac{1-\sqrt{1-r^2}}{\sqrt{1-r^2}}. \tag{83}$$

(One can check that the latter agrees with our results in Sec. V.) This have been recently used in Ref. [49] to assess the accuracy of different quantum tomographic measurements.

VII. GAUSSIAN STATES

We now illustrate our results with infinite-dimensional systems. In particular we will focus on the family of single-mode Gaussian states. This is a very significant class of quantum states mainly for two reasons. First, it has a very simple mathematical characterization that allows for the derivation of otherwise highly nontrivial results and, second, it describes accurately states of light that are realized with current technology. In the following we show that the quantum Chernoff information, in addition to being the natural distinguishability measure, has the advantage of being relatively easy to compute. The calculation of the fidelity, for instance, is much more involved, as is apparent from Refs. [50–54], where one can find such calculations for different classes of Gaussian states.

Gaussian states are by definition those that have a Gaussian characteristic function. The (symmetrically ordered) characteristic function of one such state ρ is

$$\chi(u) \equiv \mathrm{tr}[\mathcal{D}(u)\rho] = \exp\left(-iu^t\sigma\xi - \frac{1}{4}u^t\sigma^t\Gamma\sigma u\right), \quad (84)$$

where t denotes transposition and σ is the symplectic matrix

$$\sigma = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (85)$$

and $\mathcal{D}(u)=\exp[i(u_2\hat{q}-u_1\hat{p})]$ is the displacement operator, with $u=(u_1,u_2)^t$ and with position and momentum operators satisfying $[\hat{q},\hat{p}]=i$. The annihilation and creation operators, defined as $a=(\hat{q}+i\hat{p})/\sqrt{2}$ and $a^\dagger=(\hat{q}-i\hat{p})/\sqrt{2}$, fulfill the canonical commutation relations. The positivity of ρ implies that the 2×2 covariance matrix Γ is real symmetric and satisfies $\Gamma+i\sigma\geq 0$. A symplectic transformation is a linear transformation $S^t(\hat{q},\hat{p})$ that preserves the commutation relations or, more succinctly, $S\sigma S^t=\sigma$. Under such a transformation the displacement vector $\xi=(q,p)^t$ and the covariance matrix transform as $\tilde{\xi}=S\xi$ as $\tilde{\Gamma}=S\Gamma S^t$, respectively.

An equivalent, more physical, definition can be given by the action of the squeezing operator $\mathcal{S}(r,\phi)=\exp\{\frac{r}{2}[e^{-i2\phi}a^2-e^{i2\phi}(a^\dagger)^2]\}$ and the displacement operator $\mathcal{D}(u)$ defined above, on a thermal state $\rho_\beta=(1-e^{-\beta})\Sigma_n e^{-\beta n}|n\rangle\langle n|$, where the Fock states $|n\rangle$ satisfy $a^\dagger a|n\rangle=n|n\rangle$:

$$\rho(\beta,\xi,r,\phi) = \mathcal{D}(\xi)^\dagger\mathcal{S}(r,\phi)^\dagger\rho_\beta\mathcal{S}(r,\phi)\mathcal{D}(\xi). \quad (86)$$

The covariance matrix of a thermal state is simply $\Gamma_\beta=\gamma_\beta\mathbb{I}$, with $\gamma_\beta^{-1}=\tanh(\beta/2)$. The squeezing operator $\mathcal{S}(r,\phi)$ induces the symplectic transformation $S_{r,\phi}=O_\phi D_r O_\phi^t$, where

$$D_r = \begin{pmatrix} e^r & 0 \\ 0 & e^{-r} \end{pmatrix}, \quad O_\phi = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}, \quad (87)$$

and the latter corresponds to a rotation in phase space, i.e., to the unitary operation $\mathcal{O}(\phi)=\exp[i\phi a^\dagger a]$. One thus finds that the covariance matrix can be written as $\Gamma=\gamma_\beta S_{r,\phi}S_{r,\phi}^t$.

In order to calculate the Chernoff bound it is sufficient to realize that any power ρ^s of any Gaussian state ρ is also a Gaussian (unnormalized) state with a rescaled temperature

$$\begin{aligned}\rho(\beta,\xi,r,\phi)^s &= \mathcal{D}(\xi)^\dagger\mathcal{S}(r,\phi)^\dagger\rho_\beta^s\mathcal{S}(r,\phi)\mathcal{D}(\xi) \\ &= N_{\beta,s}\mathcal{D}(\xi)^\dagger\mathcal{S}(r,\phi)^\dagger\rho_{s\beta}\mathcal{S}(r,\phi)\mathcal{D}(\xi) \\ &= N_{\beta,s}\rho(s\beta,\xi,r,\phi), \end{aligned} \quad (88)$$

where we have used the relation

$$\rho_\beta^s = (1-e^{-\beta})^s\sum_n e^{-s\beta n}|n\rangle\langle n| = N_{\beta,s}\rho_{s\beta}, \quad (89)$$

with $N_{\beta,s}=(1-e^{-\beta})^s/(1-e^{-\beta s})$. Recall now that given any two Gaussian states ρ_A and ρ_B, one can write the inner product $\mathrm{tr}\,\rho_A\rho_B$ in terms of their displacement vectors and covariance matrices as

$$\mathrm{tr}(\rho_A\rho_B) = 2[\det(\Gamma_A+\Gamma_B)]^{-1/2}e^{-\delta^t(\Gamma_A+\Gamma_B)^{-1}\delta}, \quad (90)$$

where $\delta=\xi_A-\xi_B$. Using this equation we find that the quantum Chernoff bound (14) is $Q=\min_s Q_s$ with

$$Q_s = \mathrm{tr}(\rho_0^s\rho_1^{1-s}) = 2N_{\beta_0,s}N_{\beta_1,1-s}[\det(\tilde{\Gamma}_0+\tilde{\Gamma}_1)]^{-1/2}e^{\delta^t(\tilde{\Gamma}_0+\tilde{\Gamma}_1)^{-1}\delta}, \quad (91)$$

where $\tilde{\Gamma}_i=\gamma_{s\beta_i}S_{r_i,\phi_i}S_{r_i,\phi_i}^t$, $i=0,1$, and $\delta=\xi_0-\xi_1$. To simplify the notation we will denote the covariance matrix of the Gaussian state with $\beta=0$ as $A=S_{r,\phi}S_{r,\phi}^t$.

A. States with equal covariance matrices

If two general Gaussian states ρ_0 and ρ_1 are identical modulo a relative displacement δ, i.e., $\rho_1=\mathcal{D}(\delta)\rho_0\mathcal{D}(\delta)^\dagger$ we find that

$$Q_s = e^{-\delta^t(\tilde{\Gamma}_1+\tilde{\Gamma}_2)^{-1}\delta} = e^{-(\gamma_{s\beta}+\gamma_{(1-s)\beta})^{-1}\delta^tA^{-1}\delta}, \quad (92)$$

where in the first equality we used the fact that the factor multiplying the exponential in Eq. (92) must be equal to 1, since it is independent of δ and for $\delta=0$ one must have $\rho_0=\rho_1$, which implies that $Q_s=1$. That is,

$$2N_{\beta,s}N_{\beta,1-s} = [\det(\gamma_{s\beta}A+\gamma_{(1-s)\beta}A)]^{1/2} = \gamma_{s\beta}+\gamma_{(1-s)\beta}, \quad (93)$$

where we have used that symplectic transformations have unit determinant, i.e., $\det A=\det(SS^t)=1$. One readily sees that Q_s, Eq. (92), attains its minimum at $s^*=1/2$, hence we find that in this case the Chernoff measure is

$$\begin{aligned} Q = \min_s Q_s &= \exp\left(-\frac{1}{2\gamma_{\beta/2}}\delta^tA^{-1}\delta\right) \\ &= \exp\left(-\frac{1}{2}\delta^tO_\phi D_{2r}^{-1}O_\phi^t\delta\tanh\frac{\beta}{4}\right) \\ &= \exp\left[-\frac{|\delta|^2}{2}(e^{-2r}\cos^2\theta+e^{2r}\sin^2\theta)\tanh\frac{\beta}{4}\right], \end{aligned} \quad (94)$$

where θ is the relative angle between the squeezing axis and the displacement vector, i.e., if $\delta=O_\varphi(|\delta|,0)^t$ then $\theta=\varphi-\phi$.

B. States with the same temperature

We can generalize the previous result to states that have the same spectra, i.e., the same temperature ($\beta_0=\beta_1=\beta$). In this case we can use Eq. (93) to find

$$\begin{aligned} Q_s = (\gamma_{s\beta}+\gamma_{(1-s)\beta})&\det[\gamma_{s\beta}A_0+\gamma_{(1-s)\beta}A_1]^{-1/2} \\ &\times\exp[\delta^t(\gamma_{s\beta}A_0+\gamma_{(1-s)\beta}A_1)^{-1}\delta]. \end{aligned} \quad (95)$$

The determinant can be explicitly written in a compact form as

$$\det[\gamma_{s\beta}\mathbb{I}+\gamma_{(1-s)\beta}A] = \gamma_{s\beta}^2+\gamma_{(1-s)\beta}^2+2\gamma_{s\beta}\gamma_{(1-s)\beta}\cosh(2R), \quad (96)$$

where we have defined

$$\mathcal{A} = S^{-1}_{r_0,\phi_0} S_{r_1,\phi_1} (S^{-1}_{r_0,\phi_0} S_{r_1,\phi_1})^t \equiv S_{R,\Phi} S^t_{R,\Phi}, \tag{97}$$

with

$$\cosh 2R = \cos^2(\phi_0 - \phi_1)\cosh[2(r_0 - r_1)] + \sin^2(\phi_0 - \phi_1)\cosh[2(r_0 + r_1)]. \tag{98}$$

With this generality s^*, the optimal value of s, is a complicated function of the states' parameters.[6] In the case of $\delta = 0$, i.e., states with no relative displacement and the same temperature, the minimization over s can be done analytically, and one finds $s^* = 1/2$. The quantum Chernoff measure becomes

$$Q = \frac{1}{\cosh R} = [\cosh^2(r_0 - r_1) + \sin^2(\phi_0 - \phi_1)\sinh 2r_0 \sinh 2r_1]^{-1/2}. \tag{99}$$

Notice that this expression is independent of the temperature (or purity) of the states. That is, the distinguishability of two arbitrary Gaussian states with no relative displacement and equal temperature is independent of the degree of mixedness of the states.

C. Chernoff metric for Gaussian states

Following the definition (40) and using the previous results we find that Chernoff metric is

$$ds^2_{\mathrm{QC}} = \frac{d\beta^2}{32 \sinh^2 \frac{\beta}{2}} + \frac{dr^2 + d\phi^2 \sinh^2 2r}{2} + \frac{e^{-2r} dq_\phi^2 + e^{2r} dp_\phi^2}{2} \tanh \frac{\beta}{4}, \tag{100}$$

where we have defined the rotated displacement variables $(q_\phi, p_\phi) = (q,p) O_\phi$ and we have used that for infinitesimal changes $s^* = 1/2$. We find again that the metric is independent of the temperature under variations of the squeezing parameters r and ϕ.

The (unnormalized) quantum Jeffreys prior can be obtained from the metric tensor

$$\mathcal{P}_{\mathrm{QC}}(\rho) \propto \sqrt{|\det \boldsymbol{g}|} = \frac{1}{16\sqrt{2}} \frac{\tanh \beta/4}{\sinh \beta/2} \sinh 2r. \tag{101}$$

The metric induced by the local measure on the set of mixed states is given by one-half the Bures metric [7]

$$ds^2_{\mathrm{CC}} = \frac{d\beta^2}{32 \sinh^2 \frac{\beta}{2}} + \frac{e^{-2r} dq_\phi^2 + e^{2r} dp_\phi^2}{4} \tanh \frac{\beta}{2} + \frac{dr^2 + d\phi^2 \sinh^2 2r}{4} (1 + \operatorname{sech} \beta). \tag{102}$$

We note that $ds^2_{\mathrm{CC}} \to \frac{1}{2} ds^2_{\mathrm{QC}}$ as ρ approaches the set of pure states ($\beta \to \infty$) along the lines $d\beta = 0$, in agreement with the general statement at the end of Sec. V B. In the limit of very mixed states ($\beta \approx 0$) the quantum Chernoff and local metric coincide up to first order in β. In this limit of high temperatures ($\beta \approx 0$, highly mixed states) the quantum Chernoff metric and Jeffreys prior agree with those derived from Bures distance (modulo the omnipresent factor 1/2). In particular this implies that the analysis in Ref. [56] of the Bures volume element in this high-temperature regime also applies here.

VIII. SUMMARY AND CONCLUSIONS

We have analyzed quantum state discrimination (symmetric hypothesis testing) and the classical and quantum Chernoff bound focusing on the link between them and the concept of measures (distances) and metrics on the space of quantum states. More precisely, we have been concerned with defining measures and metrics that have a clear operational meaning, so that they can as a matter of principle be obtained from experiments. The error probability in state discrimination, or rather its asymptotic rate exponent (error exponent), has been shown to provide the natural link. Thus, the concept of distinguishability measure has emerged and has been analyzed in depth throughout the central part of this work. Before doing so, we have reviewed the methods and the main results of classical and quantum hypothesis testing in the first three sections of the paper. Qubit and Gaussian states have provided two excellent, very relevant examples to illustrate our results in the last sections.

Our main points and results are summarized as follows: The quantum Chernoff bound gives an upper bound to the error probability in state discrimination. When the unknown state (which we are asked to identify as either one or the other of two known states) is a tensor product, corresponding to many identical copies, the quantum Chernoff information (which is essentially the log of the quantum Chernoff bound) gives the error exponent of the optimal discrimination protocol. We propose this quantity as a distinguishability measure for general mixed states. We show that the quantum Chernoff measure is not attainable by protocols that use local fixed measurements (those for which the same measurement is performed on each of the individual copies). Given the practical relevance of these types of protocols (they can be realized with current technology), we define a local distinguishability measure as the error exponent of the best such protocol and present its main features. We derive the metrics induced by these measures and their corresponding volume elements. The latter provide a means to define operational prior probability distributions of density matrices. We derive them for general matrices of arbitrary dimension.

Examples of all the above are given in the last part of the paper. For qubit and Gaussian states, we give explicit formulas for the distinguishability measures and their corresponding metrics and volume elements. We give a geometrical picture of the space of qubit states based on those metrics. This space can be viewed as a spherical cap, similar to Uhlmann hemisphere, with the pure states sitting on the rim. These examples also illustrate the fact that the quan-

[6] In contrast to the claims in Exercise 3.9, p. 77 of Ref. [17], it is not generally the case that for states with equal spectra the minimum of Q_s is reached for $s^* = 1/2$.

[7] There seems to be a typo in Ref. [55] in the contribution of small displacements of Eq. (13).

tum Chernoff measure, in addition to being the most natural distance between general states, is conveniently easy to compute relative to other distances, such as the widely used fidelity.

ACKNOWLEDGMENTS

We are grateful to Montserrat Casas, Juli Céspedes, Alex Monràs, Sandu Popescu, and Andreas Winter for discussions. We are specially grateful to Koenraad Audenaert and Frank Verstraete for their collaboration at the early stages of this work. We acknowledge financial support from the Spanish MEC, through the Ramón y Cajal program (J.C.), the travel Grant No. PR2007-0204 (E.B.), Contracts No. FIS2005-01369, No. FIS2004-05639 (A.C.), and project QOIT (Consolider-Ingenio 2010), from the Generalitat de Catalunya, Contract No. CIRIT SGR-00185, and from the EU QAP project (A.C.).

[1] H. Chernoff, Ann. Math. Stat. **23**, 493 (1952).
[2] T. Ogawa and M. Hayashi, IEEE Trans. Inf. Theory **50**, 1368 (2004).
[3] K. M. R. Audenaert, J. Calsamiglia, R. Munoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).
[4] M. Nussbaum and A. Szkola, e-print arXiv:quant-ph/0607216v1.
[5] J. A. Bergou, U. Herzog, and M. Hillery, Lect. Notes Phys. **649**, 417 (2004).
[6] A. Chefles, Contemp. Phys. **41**, 401 (2000).
[7] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[8] M. Hayashi and H. Nagaoka, IEEE Trans. Inf. Theory **49**, 1753 (2003).
[9] D. Bacon, A. Childs, and W. V. Dam (unpublished).
[10] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).
[11] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).
[12] H. Nagaoka, e-print arXiv:quant-ph/0611289v1.
[13] M. Hayashi, e-print arXiv:quant-ph/0611013v2.
[14] K. M. R. Audenaert, M. Nussbaum, A. Szkola, and F. Verstraete, e-print arXiv:0708.4282v1[quant-ph].
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications (Wiley, New York, 1991).
[16] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976), Vol. 123.
[17] M. Hayashi, *Quantum Information: An Introduction* (Springer-Verlag, Berlin, 2006).
[18] C. A. Fuchs and J. V. de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).
[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[20] V. Kargin, Ann. Stat. **33**, 959 (2005).
[21] D. Bacon, I. L. Chuang, and A. W. Harrow, Phys. Rev. Lett. **97**, 170502 (2006).
[22] A. S. Kholevo, Theory Probab. Appl. **23**, 411 (1979).
[23] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[24] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).
[25] A. W. Harrow and A. Winter, e-print arXiv:quant-ph/0606131v1.
[26] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[27] C. Fuchs, e-print arXiv:quant-ph/9601020.
[28] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).
[29] J. Lee, M. S. Kim, and C. Brukner, Phys. Rev. Lett. **91**, 087902 (2003).
[30] A. Acin, E. Bagan, M. Baig, L. Masanes, and R. Munoz-Tapia, Phys. Rev. A **71**, 032338 (2005).
[31] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).
[32] P. Zanardi, L. C. Venuti, and P. Giorda, e-print arXiv:quant-ph/0707.2772v2.
[33] D. Petz, Numer. Linear Algebra Appl. **244**, 81 (1996).
[34] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[35] D. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).
[36] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[37] D. Petz, J. Phys. A **35**, 929 (2002).
[38] K. M. R. Audenaert, J. Calsamiglia, L. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).
[39] D. Petz and C. Sudar, in *Geometry in Present Days Science*, edited by O. E. Barndorff-Nielsen and E. Jensen (World Scientific, Singapore, 1999), pp. 21–34.
[40] P. Gibilisco and T. Isola, J. Math. Phys. **44**, 3752 (2003).
[41] P. Gibilisco, D. Imparato, and T. Isola, J. Math. Phys. **48**, 072109 (2007).
[42] K. Zyczkowski and H. J. Sommers, J. Phys. A **34**, 7111 (2001).
[43] M. J. Hall, Phys. Lett. A **242**, 123 (1998).
[44] P. B. Slater, J. Phys. A **32**, 8231 (1999).
[45] H.-J. Sommers and K. Zyczkowski, J. Phys. A **36**, 10083 (2003).
[46] A. Andai, J. Phys. A **39**, 13641 (2006).
[47] M. Hubner, Phys. Lett. A **163**, 239 (1992).
[48] D. Brody and B. Meister, Phys. Rev. Lett. **76**, 1 (1996).
[49] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, e-print arXiv:0706.3756v1.
[50] S. Abe, Phys. Lett. A **254**, 149 (1999).
[51] J. Twamley, J. Phys. A **29**, 3723 (1996).
[52] G.-S. Paraoanu and H. Scutaru, Phys. Rev. A **58**, 869 (1998).
[53] X. B. Wang, C. H. Oh, and L. C. Kwek, Phys. Rev. A **58**, 4186 (1998).
[54] P. B. Slater, J. Phys. A **29**, L601 (1996).
[55] L. C. Kwek, C. H. Oh, and X. B. Wang, J. Phys. A **32**, 6613 (1999).
[56] P. B. Slater, Phys. Rev. E **61**, 6087 (2000).