PHYSICAL REVIEW A **69**, 032106 (2004)

Informational distance on quantum-state space

Shunlong Luo[1,*] and Qiang Zhang[2,†]
[1]*Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, 100080 Beijing, People's Republic of China*
[2]*Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong*
(Received 3 December 2003; published 16 March 2004)

By formally generalizing the classical Hellinger distance and affinity on the space of probability densities, we introduce their quantum analog on the quantum state space consisting of all density operators. We show that the infinitesimal form of the quantum Hellinger distance is the skew information introduced by Wigner and Yanase in 1963. We compare the Hellinger distance and affinity with the Bures distance and fidelity, and establish a variety of their fundamental properties. We further apply them to characterizing entanglement and to establishing some unusual uncertainty relations relating nonsimultaneous measurements of a single observable in two different quantum states.

DOI: 10.1103/PhysRevA.69.032106 PACS number(s): 03.65.Ta, 03.67.−a

I. INTRODUCTION

In classical statistical theory, a state is defined as a probability distribution on some sample space. One of the most important issues in statistical inference theory is to distinguish two probability distributions based on some observations. In this respect, various distance measures (metrics) play a crucial role. Apart from the relative entropy (Kullback-Leibler entropy) [1], the most fundamental distance is the so-called Hellinger distance [2]

$$H(f,g):=\int[\sqrt{f(x)}-\sqrt{g(x)}]^2dx$$

between two probability densities f and g. In case that f and g are discrete probability distributions, their Hellinger distance is

$$H(f,g):=\sum_x[\sqrt{f(x)}-\sqrt{g(x)}]^2.$$

A closely related concept, affinity, is defined as

$$A(f,g):=\int\sqrt{f(x)}\sqrt{g(x)}dx$$

when f and g are probability densities, and

$$A(f,g):=\sum_x\sqrt{f(x)}\sqrt{g(x)}$$

when f and g are discrete probability distributions. Affinity quantifies how close f and g are, and its application dates back to Bhattacharyya [3], Kakutani [4], and Matusita [5]. It is also widely used in the mathematical literature. One particular virtue of affinity is that it is multiplicative under direct product [4].

In quantum mechanics, a state is described by a density operator, and a key issue in quantum detection and estimation theory is to distinguish two alternative density operators by quantum measurements. A convenient approach to this problem is to introduce various distance measures (metrics) on the space of quantum states, and consequently to geometrize the statistical inference problem. Modern applications of quantum mechanics in quantum information theory have renewed the interests in quantum-state space, whose geometry depends crucially on the metric involved [6]. In this context, Wootters introduced a statistical distance which is essentially the same as the affinity in order to quantifying the discrimination of quantum states [7]. Braunstein and Caves further studied this problem from a geometrical point of view [8].

Inspired by the original Hellinger distance, we introduce a distance between two density operators ρ and σ as

$$D(\rho,\sigma):=\mathrm{tr}(\sqrt{\rho}-\sqrt{\sigma})^2$$

with tr denoting operator trace. We shall call this the (quantum) Hellinger distance, in accordance to the classical case. Here we also define a related notion, the (quantum) affinity, as

$$A(\rho,\sigma):=\mathrm{tr}\sqrt{\rho}\sqrt{\sigma}.$$

From the cyclic property of trace, we easily see that the affinity is non-negative because

$$A(\rho,\sigma)=\mathrm{tr}[(\rho^{1/4}\sigma^{1/4})(\rho^{1/4}\sigma^{1/4})^\dagger].$$

Furthermore, we also have

$$D(\rho,\sigma)=2-2A(\rho,\sigma).$$

While many distance measures, such as the relative entropy [9–11], the trace distance [12], the Hilbert-Schmidt distance [13], and the Bures distance [14–20], have been extensively studied and applied to quantum measurement theory, it seems that the (quantum) Hellinger distance, which is the simplest and most straightforward generalization of the widely used classical Hellinger distance, has received almost no attention. An exception is Ref. [21] in which Albrecht actually used the affinity in discussing issues related with the black-hole information problem. However, he did not study

*Email address: luosl@mail.amt.ac.cn
†Email address: mazq@math.cityu.hk

the informational properties of the affinity in detail. We will uncover the statistical and physical contents of the Hellinger distance and affinity, and in particular, demonstrate that the infinitesimal form of the Hellinger distance is exactly the skew information introduced by Wigner and Yanase in 1963 [22]. Moreover, we will apply this distance measure to characterizing entanglement and to establishing some unusual uncertainty relations.

Before we proceed, let us briefly recall some of the commonly used distance measures for the convenience of comparisons.

(1) Relative entropy:

$$S(\rho,\sigma):=\mathrm{tr}[\rho(\ln\,\rho-\ln\,\sigma)].$$

This nonsymmetric distance is widely used in quantum statistical mechanics [9,10] and recently in quantifying entanglement [11]. Its classical analog is the Kullback-Leibler entropy, which is widely used in mathematical statistics [1].

(2) Trace distance induced by the trace norm:

$$D_{\mathrm{tr}}(\rho,\sigma):=\mathrm{tr}\sqrt{(\rho-\sigma)^2}.$$

(3) Hilbert-Schmidt distance induced by the Hilbert-Schmidt norm:

$$D_{\mathrm{hs}}(\rho,\sigma):=\sqrt{\mathrm{tr}(\rho-\sigma)^2}.$$

These two distances, though formally simple, have some drawbacks from an informational perspective. For instance, $D_{\mathrm{tr}}(\rho,\sigma)$ has the desirable property of being monotone [12], that is, contracts under quantum operations (also called stochastic maps, or completely positive, trace preserving maps between the spaces of density operators), but is not Riemannian, while $D_{\mathrm{hs}}(\rho,\sigma)$ is Riemannian, but not monotone [13].

Comparing the definitions of the Hellinger distance, the trace distance, and the Hilbert distance, we see that they are strikingly similar, except that the position of the square root is different. This difference entails quite different properties for the three distance measures. If we recall that the probability amplitude, that is, the square root of the quantum state, is of fundamental importance in quantum mechanics as a primitive object, we would intuitively expect that the Hellinger distance has some advantages over the other two. This is indeed the case as we shall demonstrate in later sections.

(4) Bures distance:

$$D_{\mathrm{b}}(\rho,\sigma):=2-2\mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}.$$

In quantum information theory, this is a far more interesting and useful metric than those induced by the trace norm and the Hilbert-Schmidt norm. It distinguishes itself by its remarkable properties of being both Riemannian and monotone. The closely related notion, fidelity,

$$F(\rho,\sigma):=\mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}$$

is widely used in quantum information theory [17–19]. Clearly,

$$D_{\mathrm{b}}(\rho,\sigma)=2-2F(\rho,\sigma).$$

The fidelity has the advantage of a direct physical interpretation in terms of the concept of purification of mixed states and of being a generalization of the usual transition probability for pure states.

The formal similarity between the Hellinger distance and the widely studied Bures distance is striking. The former is even formally simpler than the latter, and in the meantime enjoys almost all remarkable informational properties the latter has.

Since the Bures distance is so closely related with the Hellinger distance, and in order to put the Hellinger distance into an appropriate perspective, we will review some fundamental properties of the Bures distance and fidelity in the following section.

II. BURES DISTANCE AND FIDELITY

The Bures distance and fidelity have a variety of nice properties. Some of their fundamental properties are as follows [14–20].

(1) $0\leqslant F(\rho,\sigma)\leqslant 1$, and $F(\rho,\sigma)=1$ if and only if $\rho=\sigma$. Moreover, $F(\rho,\sigma)=F(\sigma,\rho)$, and

$$F(U\rho U^\dagger,U\sigma U^\dagger)=F(\rho,\sigma)$$

for any unitary operator U.

(2) If $\rho=|\psi\rangle\langle\psi|$ and $\sigma=|\phi\rangle\langle\phi|$ are pure states, then $F(\rho,\sigma)=|\langle\psi|\phi\rangle|$ is the magnitude of transition amplitude.

(3) The fidelity $F(\cdot,\cdot)$ is multiplicative under tensor product:

$$F(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2)=F(\rho_1,\sigma_1)F(\rho_2,\sigma_2).$$

(4) The fidelity $F(\cdot,\cdot)$ is jointly concave with respect to its two arguments. That is,

$$F\left(\sum_j\lambda_j\rho_j,\sum_j\lambda_j\sigma_j\right)\geqslant\sum_j\lambda_jF(\rho_j,\sigma_j).$$

Here $\sum_j\lambda_j=1$, $\lambda_j\geqslant 0$.

(5) Let $E=\{E_j\}$ be any measurement, i.e., E_j is positive, and $\sum_j E_j=I$. Let $E_\rho=\{\mathrm{tr}(\rho E_j)\}$ and $E_\sigma=\{\mathrm{tr}(\sigma E_j)\}$ be the corresponding classical probability distributions of the measurement outcomes, then

$$F(\rho,\sigma)\leqslant A(E_\rho,E_\sigma).$$

Therefore, the fidelity is bounded above by the (classical) affinity.

In addition, the fidelity has the following variational characterizations [15,17,18,20].

(1) $F(\rho,\sigma)=\min_E A(E_\rho,E_\sigma)$, that is, the fidelity is the minimum of the (classical) affinity of the measurement outcome probability distributions, the minimum being taken over all quantum measurement.

(2) $F(\rho,\sigma)=\max_{\Psi,\Phi}|\langle\Psi|\Phi\rangle|$, where the maximum is taken over all purifications Ψ and Φ of ρ and σ, respectively.

(3) $F(\rho,\sigma)=\max|\langle\Psi|\Phi\rangle||$, where the maximum is taken over all quantum pure states Ψ and Φ such that there exists a quantum operation mapping $|\Psi\rangle$ and $|\Phi\rangle$ to ρ and σ, respectively. This implies in particular that no quantum opera-

tion can increase the distinguishability of two quantum states [20].

(4) $F(\rho,\sigma)=\max_U \mathrm{tr}(U\sqrt{\rho}\sqrt{\sigma})$, where the maximum is taken over all unitary operators U.

(5) $F(\rho,\sigma)=\frac{1}{2}\min_X \mathrm{tr}(X\rho+X^{-1}\sigma)$, where the minimum is taken over all invertible positive observables X.

The infinitesimal form of the Bures distance is the quantum Fisher information defined via the symmetric logarithmic derivative [16]. We will see that the infinitesimal form of the Hellinger distance is the Wigner-Yanase skew information, a version of quantum Fisher information defined via commutator derivative of the square root of the quantum density operator [23]. Further, both the Bures distance and the (quantum) Hellinger distance reduce to the classical Hellinger distance when ρ and σ are simultaneously diagonal.

III. HELLINGER DISTANCE AND AFFINITY

We now list the fundamental properties of the Hellinger distance and the affinity.

(1) $0 \leqslant A(\rho,\sigma) \leqslant 1$ and $A(\rho,\sigma)=1$ if and only if $\rho=\sigma$. Moreover, $A(\rho,\sigma)=A(\sigma,\rho)$, and

$$A(U\rho U^\dagger, U\sigma U^\dagger) = A(\rho,\sigma)$$

for any unitary operator U.

(2) If $\rho=|\psi\rangle\langle\psi|$ and $\sigma=|\phi\rangle\langle\phi|$ are pure states, then $A(\rho,\sigma)=|\langle\psi|\phi\rangle|$ is the magnitude of transition amplitude.

(3) The affinity $A(\cdot,\cdot)$ is multiplicative under tensor product:

$$A(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2) = A(\rho_1,\sigma_1)A(\rho_2,\sigma_2).$$

(4) The affinity $A(\cdot,\cdot)$ is jointly concave with respect to its two arguments:

$$A\left(\sum_j \lambda_j\rho_j, \sum_j \lambda_j\sigma_j\right) \geqslant \sum_j \lambda_j A(\rho_j,\sigma_j).$$

(5) Let $E=\{E_j\}$ be any measurement, i.e., E_j is positive, and $\sum_j E_j = I$. Let $E_\rho=\{\mathrm{tr}(\rho E_j)\}$ and $E_\sigma=\{\mathrm{tr}(\sigma E_j)\}$ be the corresponding probability distributions of the measurement outcomes, then

$$A(\rho,\sigma) \leqslant A(E_\rho,E_\sigma).$$

That is, the (quantum) affinity is bounded above by the (classical) affinity.

Items (1)–(3) can be easily verified, and (4) follows from Corollary 1.1 of Lieb [24]. Item (5) follows from Theorem 1 and property (4) of the fidelity.

We have seen that the affinity and the Hellinger distance possess almost all desirable information theoretic properties the fidelity and the Bures distance have, and are apparently easier to compute.

Theorem 1. The affinity $A(\cdot,\cdot)$ is always dominated by the fidelity $F(\cdot,\cdot)$, and accordingly, the Hellinger distance $D(\cdot,\cdot)$ always dominates the Bures distance $D_{\mathrm{b}}(\cdot,\cdot)$:

$$A(\rho,\sigma) \leqslant F(\rho,\sigma), \quad D(\rho,\sigma) \geqslant D_{\mathrm{b}}(\rho,\sigma).$$

The conclusions follow from the following expressions for the affinity $A(\cdot,\cdot)$ and the fidelity $F(\cdot,\cdot)$:

$$A(\rho,\sigma) = \mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}),$$

$$F(\rho,\sigma) = \max_U \mathrm{tr}(U\sqrt{\rho}\sqrt{\sigma}).$$

Let us see an example. Consider a qubit system in which a general density operator is of the form

$$\rho_{\mathbf{a}} = \frac{1}{2}(I+\mathbf{a}\vec{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+a_3 & a_1-ia_2 \\ a_1+ia_2 & 1-a_3 \end{pmatrix},$$

with

$$\mathbf{a}=(a_1,a_2,a_3)\in\mathbb{R}^3, \quad |\mathbf{a}|:=\sqrt{a_1^2+a_2^2+a_3^2}\leqslant 1$$

and

$$\vec{\sigma}=(\sigma_x,\sigma_y,\sigma_z)$$

being the Pauli spin matrices. It is easy to compute $\sqrt{\rho_a}$ as

$$\sqrt{\rho_a} = \frac{1}{\sqrt{2}(\sqrt{1+|\mathbf{a}|}+\sqrt{1-|\mathbf{a}|})} \times \begin{pmatrix} 1+\sqrt{1-|\mathbf{a}|^2}+a_3 & a_1-ia_2 \\ a_1+ia_2 & 1+\sqrt{1-|\mathbf{a}|^2}-a_3 \end{pmatrix}.$$

Now let

$$\rho_{\mathbf{b}} = \frac{1}{2}(I+\mathbf{b}\vec{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+b_3 & b_1-ib_2 \\ b_1+ib_2 & 1-b_3 \end{pmatrix}$$

be another density operator, then direct evaluation yields

$$A(\rho_{\mathbf{a}},\rho_{\mathbf{b}}) = \frac{(1+\sqrt{1-|\mathbf{a}|^2})(1+\sqrt{1-|\mathbf{b}|^2})+\mathbf{a}\cdot\mathbf{b}}{(\sqrt{1+|\mathbf{a}|}+\sqrt{1-|\mathbf{a}|})(\sqrt{1+|\mathbf{b}|}+\sqrt{1-|\mathbf{b}|})}.$$

Here $\mathbf{a}\cdot\mathbf{b}=a_1b_1+a_2b_a+a_3b_3$. In contrast, the fidelity is given by

$$F(\rho_{\mathbf{a}},\rho_{\mathbf{b}}) = \sqrt{\tfrac{1}{2}[1+\mathbf{a}\cdot\mathbf{b}+\sqrt{(1-|\mathbf{a}|^2)(1-|\mathbf{b}|^2)}]}.$$

The following result is a kind of quantum data processing inequality.

Theorem 2. Let ρ and σ be quantum states on a composite system $\mathcal{H}_1\otimes\mathcal{H}_2$. Let $\rho_1=\mathrm{tr}_2\rho$ and $\rho_2=\mathrm{tr}_1\rho$ be the partial trace of ρ over $\mathcal{H}_2$ and $\mathcal{H}_1$, respectively, and σ_1,σ_2 are defined similarly. Then

$$A(\rho_j,\sigma_j) \geqslant A(\rho,\sigma), \quad j=1,2.$$

For an elementary proof, see the Appendix.

IV. RELATIONS WITH FISHER INFORMATION

Both the Bures distance and the Hellinger distance have a natural statistical interpretation. To motivate this, let us first consider a natural problem in the theory of classical statistical estimation: Suppose that $\{p_\theta : \theta\in\mathbb{R}\}$ is a family of probability densities on $\mathbb{R}$ parametrized by θ, and we have observed samples $x_1, \ldots, x_n$, where each is a random variable independently distributed according to p_θ for some fixed unknown θ. Our objective is to estimate this θ as faithfully as possible by virtue of the available samples.

In this context, the Fisher information plays a fundamental role [25,26]. Recall that the Fisher information of a parametrized family of probability densities $\{p_\theta: \theta \in \mathbb{R}\}$ on $\mathbb{R}$ is defined as

$$I(p_\theta) = \int \left(\frac{\partial \ln p_\theta(x)}{\partial \theta}\right)^2 p_\theta(x)dx \tag{1}$$

$$=4 \int \left(\frac{\partial \sqrt{p_\theta(x)}}{\partial \theta}\right)^2 dx, \tag{2}$$

provided that the integrals exist. In particular, when $p_\theta(x) = p(x-\theta)$ (hence θ is a location parameter), by the translation invariance of the Lebesgue integral, we have

$$I(p_\theta) = \int \left(\frac{\partial \ln p(x)}{\partial x}\right)^2 p(x)dx \tag{3}$$

$$=4 \int \left(\frac{\partial \sqrt{p(x)}}{\partial x}\right)^2 dx. \tag{4}$$

Thus in this circumstance, $I(p_\theta)$ is independent of the parameter θ, and we will denote $I(p_\theta)$ simply by $I(p)$. This is the Fisher information of the probability p with respect to the location. When p_θ is distributed on a discrete set, all the integrals above should be replaced by summations.

The notion of Fisher information plays a basic role in the theory of maximum likelihood estimation (in fact, this is the origin of Fisher information). The asymptotic normality of a maximum likelihood estimator is characterized by the Fisher information as the inverse variance. The celebrated Cramér-Rao inequality characterizes the intrinsic uncertainty of parameter estimation by virtue of the Fisher information [26].

Now, we see how we can intuitively generalize the notion of Fisher information to the quantum setting. The quantum analog of probability densities are density operators, and that of integral is trace. Usually the natural extensions of a classical quantity to the quantum case are not unique due to the noncommutativity of operators, but depend on the ordering of the operators. In the circumstance here, we have two simple and heuristic routes. The first is to define a quantum analog of the statistical score function (logarithmic derivative) $l_\theta := \partial \ln p_\theta / \partial\theta$ and formally generalizes the expression

$$I(p_\theta) = \int l_\theta^2(x) p_\theta(x) dx$$

defined by Eq. (1). Replacing the parametrized probability densities p_θ by a parametrized density operator ρ_θ, the integration by trace, and the logarithmic derivative l_θ by the symmetric logarithmic derivative L_θ determined by

$$\frac{\partial \rho_\theta}{\partial \theta} = \frac{1}{2}(L_\theta \rho_\theta + \rho_\theta L_\theta), \quad \theta \in \mathbb{R},$$

which is formally motivated by the identity

$$\frac{\partial p_\theta}{\partial \theta} = \frac{1}{2}(l_\theta p_\theta + p_\theta l_\theta), \quad \theta \in \mathbb{R},$$

we come to the quantum Fisher information defined via the symmetric logarithmic derivative

$$I_{\mathrm{b}}(\rho_\theta) := \mathrm{tr}(L_\theta^2 \rho_\theta). \tag{5}$$

The subscript b anticipates that this is related to the infinitesimal form of the Bures distance D_{b}. (Of course, one can also write

$$\frac{\partial p_\theta}{\partial \theta} = p_\theta l_\theta = l_\theta p_\theta,$$

and consider a version of quantum Fisher information defined via the right/left logarithmic derivative, as studied in Refs. [27–29]). This seemingly sophisticated scheme was actually first done in a quantum estimation setting and plays an important role in quantum estimation and detection theory [27–29]. Note that this is different from the simpler and more naive generalization

$$I_{\mathrm{s}}(\rho_\theta) := \mathrm{tr}\left[\left(\frac{\partial \ln \rho_\theta}{\partial \theta}\right)^2 \rho_\theta\right],$$

which seems to never attract anyone's attention. We remark that this definition is not without informational or physical contents, but we will not pursue them here.

In particular, one may also generalize Eq. (2) in a parameter-free manner. Let ρ be any quantum state and X any observable, then we define

$$I_{\mathbf{b}}(\rho, X) := \mathrm{tr}(L^2 \rho), \tag{6}$$

where L is the logarithmic derivative determined by

$$d_X \rho = \frac{1}{2}(L\rho + \rho L)$$

and d_X is the quantum commutator derivative defined by the commutator

$$d_X \rho := i[\rho, X] = i(\rho X - X\rho).$$

The above two generalizations, Eqs. (5) and (6), are consistent in the following sense: If ρ_θ satisfies the Landau–von Neumann equation

$$i\frac{\partial \rho_\theta}{\partial \theta} = [X, \rho_\theta], \quad \rho_0 = \rho,$$

then

$$I_{\mathrm{b}}(\rho_\theta) = I_{\mathrm{b}}(\rho, X).$$

The verification is straightforward by noting

$$\rho_\theta = e^{-i\theta X} \rho e^{i\theta X}.$$

The second quantum extension of the classical Fisher information arises when one formally generalizes the expression

$$I(p_\theta) = 4\int\left(\frac{\partial\sqrt{p_\theta(x)}}{\partial\theta}\right)^2 dx$$

defined by Eq. (2). Replacing the integration by trace, the parametrized probability densities p_θ by a parametrized density operator ρ_θ on some Hilbert space, we can heuristically define

$$I(\rho_\theta) := 4\mathrm{tr}\left(\frac{\partial\sqrt{\rho_\theta}}{\partial\theta}\right)^2. \tag{7}$$

On the other hand, along the line of Eq. (4) in a parameter-free case, we may define

$$I(\rho,X) := 4\mathrm{tr}(d_X\sqrt{\rho})^2. \tag{8}$$

$I(\rho,X)$ is, up to an inessential constant factor, the Wigner-Yanase skew information [22].

The definitions of $I(\rho_\theta)$ by Eq. (7) and $I(\rho,X)$ by Eq. (8) are also consistent in the following sense.

If ρ_θ satisfies the Landau–von Neumann equation

$$i\frac{\partial\rho_\theta}{\partial\theta} = [X,\rho_\theta], \quad \rho_0 = \rho,$$

then [23]

$$I(\rho_\theta) = I(\rho,X).$$

Remark. In the study of information contents of distributions (quantum-mechanical density operators), Wigner and Yanase [22] introduced the quantity $I(\rho,X)$ which they called skew information, to measure the information content of the density operator ρ with respect to the self-adjoint operator X, which serves as a conserved quantity such as a Hamiltonian or a momentum. Alternatively, $I(\rho,X)$ may also be interpreted as a measure of the noncommutativity between ρ and X. Wigner and Yanase argued and proved that this quantity satisfies a variety of the desirable intuitive requirements of an information measure. Furthermore, for any $\alpha\in[0,1]$, they also introduced the quantity

$$I_\alpha(\rho,X) := \frac{1}{\alpha(1-\alpha)}\mathrm{tr}[(d_X\rho^\alpha)(d_X\rho^{1-\alpha})],$$

which is now called the Wigner-Yanase-Dyson information, as a general quantity to measure information content of ρ with respect to X. Lieb proved the convexity of the Wigner-Yanase-Dyson information [24].

Apart from the formal similarity between the Bures distance and the Hellinger distance, they also have the intrinsic relations manifested in the forms of quantum Fisher information: The infinitesimal change of the Bures distance is the quantum Fisher information determined by the symmetric logarithmic derivative, while the infinitesimal change of the Hellinger distance is the quantum Fisher information determined by the commutator derivative, which happens to be the Wigner-Yanase skew information. These are more precisely stated as follows.

Theorem 3. If ρ_θ satisfies the Landau–von Neumann equation

$$i\frac{\partial\rho_\theta}{\partial\theta} = [X,\rho_\theta], \quad \rho_0 = \rho.$$

Then

(i) $\partial^2 D(\rho_\theta,\rho_\zeta)/\partial\theta\partial\zeta|_{\theta=\zeta=0} = 1/2\, I(\rho,X)$.
(ii) $\partial^2 D(\rho_\theta,\rho)/\partial\theta^2|_{\theta=0} = 1/2\, I(\rho,X)$.
(iii) $\partial^2 D_{\mathrm{b}}(\rho_\theta,\rho_\zeta)/\partial\theta\partial\zeta|_{\theta=\zeta=0} = 1/2\, I_{\mathrm{b}}(\rho,X)$.
(iv) $\partial^2 D_{\mathrm{b}}(\rho_\theta,\rho)/\partial\theta^2|_{\theta=0} = 1/2\, I_{\mathrm{b}}(\rho,X)$.

Proof. (1) Note

$$\rho_\theta = e^{-i\theta X}\rho e^{i\theta X}, \quad \rho_\zeta = e^{-i\zeta X}\rho e^{i\zeta X},$$

we have

$$\frac{\partial\sqrt{\rho_\theta}}{\partial\theta} = ie^{-i\theta X}[\sqrt{\rho},X]e^{i\theta X},$$

$$\frac{\partial\sqrt{\rho_\zeta}}{\partial\zeta} = ie^{-i\zeta X}[\sqrt{\rho},X]e^{i\zeta X}.$$

Therefore

$$\frac{\partial^2 D(\rho_\theta,\rho_\zeta)}{\partial\theta\,\partial\zeta} = -2\frac{\partial^2\mathrm{tr}(\sqrt{\rho_\theta}\sqrt{\rho_\zeta})}{\partial\theta\,\partial\zeta} = -2\mathrm{tr}\left(\frac{\partial\sqrt{\rho_\theta}}{\partial\theta}\frac{\partial\sqrt{\rho_\zeta}}{\partial\zeta}\right).$$

Now the conclusion follows from direct evaluation. (2) follows from similar calculations. (3) and (4) are implicit in Ref. [16].

V. QUANTIFYING ENTANGLEMENT

According to the general entanglement quantification procedure proposed by Vedral *et al.* [30], we may use the Hellinger distance as a measure to quantify entanglement. For a quantum state ρ on a composite quantum system $\mathcal{H}_1\otimes\mathcal{H}_2$, let us define

$$\mathcal{E}(\rho) := \inf_\sigma D(\rho,\sigma),$$

where the infimum is taken over all separable states (a state σ is separable if it can be expressed as $\sigma=\sum_j\lambda_j\rho_j^{(1)}\otimes\rho_j^{(2)}$ for some density operators $\rho_j^{(1)}$ and $\rho_j^{(2)}$ on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, $\sum_j\lambda_j=1,\lambda_j\geq 0$. Otherwise it is entangled or nonseparable). This entanglement measure possesses the following basic properties.

(1) $\mathcal{E}(\rho)=0$ if and only if ρ is separable.
(2) $\mathcal{E}(\rho)$ is invariant under local unitary operations, i.e.,

$$\mathcal{E}(\rho) = \mathcal{E}(U_1 \otimes U_2 \rho U_1^\dagger \otimes U_2^\dagger)$$

for any unitary operators U_1 and U_2 on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively.

(3) $\mathcal{E}(\rho)$ is convex with respect to ρ, i.e.,

$$\mathcal{E}(\sum_j \lambda_j \rho_j) \leqslant \sum_j \lambda_j \mathcal{E}(\rho_j).$$

(4) $\mathcal{E}(\rho)$ decreases under quantum operations, i.e., for

$$\sum_j V_j V_j^\dagger = I, \quad \lambda_j = \mathrm{tr}(V_j \rho V_j^\dagger), \quad \rho_j = \frac{V_j \rho V_j^\dagger}{\lambda_j},$$

we have

$$\mathcal{E}(\rho) \geqslant \sum_j \lambda_j \mathcal{E}(\rho_j).$$

VI. UNCERTAINTY RELATIONS

The conventional Heisenberg uncertainty relation relates the uncertainties of two observables in a single quantum state, and its standard form is

$$V(\rho,X)V(\rho,Y) \geqslant \tfrac{1}{4}|\langle [X,Y]\rangle_\rho|^2.$$

Here $V(\rho,X) := \mathrm{tr}(\rho X^2) - [\mathrm{tr}(\rho X)]^2$ is the variance, and $\langle [X,Y]\rangle_\rho := \mathrm{tr}(\rho[X,Y])$ is the average of the commutator. The following uncertainty relations concerning nonsimultaneous measurements of a single observable are of different nature. They relate the quantum uncertainties of a single observable in two different quantum states. To appreciate them, one should recall that $I(\rho,X)$ is the information content of observables not commuting with (skew to) X in the state ρ, and therefore by the Bohr complementary principle, $I(\rho,X)$ may also be interpreted as certain uncertainty of X in ρ. Indeed, when ρ is pure, $I(\rho,X) = 8V(\rho,X)$ is essentially the usual variance. For general mixed state, $I(\rho,X) \leqslant 8V(\rho,X)$.

Theorem 4. Let ρ and σ be two quantum states (generally mixed), and let X be an observable. Then

$$\sqrt{I(\rho,X)} \geqslant \frac{2|\mathrm{tr}([\sqrt{\rho},\sqrt{\sigma}]X)|}{\sqrt{1-A^2(\rho,\sigma)}},$$

$$\sqrt{I(\sigma,X)} \geqslant \frac{2|\mathrm{tr}([\sqrt{\rho},\sqrt{\sigma}]X)|}{\sqrt{1-A^2(\rho,\sigma)}}.$$

In particular,

$$\sqrt{I(\rho,X)}\sqrt{I(\sigma,X)} \geqslant \frac{4|\mathrm{tr}([\sqrt{\rho},\sqrt{\sigma}]X)|^2}{1-A^2(\rho,\sigma)}.$$

Proof. Define

$$\sqrt{\rho_X} := \frac{2d_X\sqrt{\rho}}{\sqrt{I(\rho,X)}}.$$

Then $\sqrt{\rho_X^\dagger} = \sqrt{\rho_X}$, and

$$\mathrm{tr}(\sqrt{\rho_X^2}) = 1, \quad \mathrm{tr}(\sqrt{\rho}\sqrt{\rho_X}) = 0.$$

From the definition of $\sqrt{\rho_X}$, we have

$$X\sqrt{\rho} - \sqrt{\rho}X = \frac{i}{2}\sqrt{I(\rho,X)}\sqrt{\rho_X},$$

which implies

$$\sqrt{\sigma}(X\sqrt{\rho} - \sqrt{\rho}X) = \frac{i}{2}\sqrt{I(\rho,X)}\sqrt{\sigma}\sqrt{\rho_X}.$$

Taking trace, we have

$$\mathrm{tr}([\sqrt{\rho},\sqrt{\sigma}]X) = \frac{i}{2}\sqrt{I(\rho,X)}\mathrm{tr}(\sqrt{\sigma}\sqrt{\rho_X}).$$

But by the abstract Parseval inequality we have

$$1 = \mathrm{tr}(\sqrt{\sigma^2}) \geqslant |\mathrm{tr}(\sqrt{\sigma}\sqrt{\rho})|^2 + |\mathrm{tr}(\sqrt{\sigma}\sqrt{\rho_X})|^2.$$

Therefore

$$|\mathrm{tr}(\sqrt{\sigma}\sqrt{\rho_X})| \leqslant \sqrt{1-A^2(\rho,\sigma)}.$$

The first conclusion now follows. The second result follows similarly. The third conclusion follows from the combination of the first two.

There are some interesting consequences for the above uncertainty relations.

(1) First, there is a trivial consequence. If $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ are two eigenvectors of X with different eigenvalues, we see readily from the first uncertainty relation in Theorem 4 that $|\psi\rangle$ and $|\phi\rangle$ will be orthogonal.

(2) Second, let Y be another observable, and let ρ_θ satisfy the Landau–von Neumann equation

$$i\frac{\partial \rho_\theta}{\partial \theta} = [Y,\rho_\theta], \quad \rho_0 = \rho,$$

then for θ near zero,

$$\mathrm{tr}([\sqrt{\rho_\theta},\sqrt{\rho}]X) = -2i\theta\,\mathrm{Re}\,\mathcal{C}_\rho(X,Y) + o(\theta).$$

Here Re denotes the real part, and

$$\mathcal{C}_\rho(X,Y) := \mathrm{tr}(\rho XY) - \mathrm{tr}(\sqrt{\rho}X\sqrt{\rho}Y)$$

is a correlation measure between observables X and Y in ρ. Note in particular, $\mathcal{C}_\rho(X,X) = \frac{1}{8}I(\rho,X)$ is the Wigner-Yanase skew information.

On the other hand,

$$A(\rho_\theta,\rho) = 1 - \frac{I(\rho,Y)}{8}\theta^2 + o(\theta^2),$$

and thus

$$\sqrt{1-A^2(\rho_\theta,\rho)} = \frac{\sqrt{I(\rho,Y)}}{2}\theta + o(\theta).$$

Substituting the above identities into the first uncertainty relation of Theorem 4, and letting $\theta \to 0$, we obtain

$$I(\rho,X)I(\rho,Y) \geq 8^2|\mathrm{Re}\mathcal{C}_\rho(X,Y)|^2,$$

and in particular,

$$V(\rho,X)V(\rho,Y) \geq |\mathrm{Re}\mathcal{C}_\rho(X,Y)|^2$$

because $I(\rho,X)\leq 8V(\rho,X), I(\rho,Y)\leq 8V(\rho,Y)$. This last uncertainty relation, which employs a kind of correlation of two observables to set a lower bound of the product of variances, complements the original Heisenberg uncertainty relation, which employs the commutator to set a lower bound of the product of variances. It is of particular relevance when the observables X and Y are belonging to two different quantum subsystems, and thus they commute.

VII. CONCLUSIONS

In summary, by analogy with the classical notions, we have introduced the Hellinger distance and affinity on the quantum-state space consisting of all density operators. We have investigated their fundamental properties, and in particular, have made extensive comparisons with the Bures distance and fidelity. The Hellinger distance and the associated affinity are easier to compute than the Bures distance and the fidelity, while enjoy almost all nice informational properties of the latter. Actually, we have put the Hellinger distance and the Bures distance on equal footing by showing that they are intrinsically related to two different but natural quantum extensions of the classical Fisher information.

We have also employed the Hellinger distance to quantifying entanglement. We have established some uncertainty relations in purely quantum terms such as affinity and the skew information. This shed some light on the informational aspect of quantum measurement theory.

ACKNOWLEDGMENTS

S. Luo was supported by NSF of China, Grant No. 10131040. Q. Zhang was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, Project CityU 1221/02P.

APPENDIX

In this appendix, we present a proof of Theorem 2, and in the meantime, we also present an elementary constructive procedure of representing the partial trace as a quantum operation, which may be of independent interest.

We only need to prove the first inequality $(j=1)$. Suppose the dimensions of $\mathcal{H}_1$ and $\mathcal{H}_2$ are m and n, respectively. Let U_2 be the cyclic unitary matrix (permutation matrix) on $\mathcal{H}_2$ defined as

$$U_2 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then for any *diagonal* matrix

$$X = \mathrm{diag}(x_1, x_2, \ldots, x_n)$$

on $\mathcal{H}_2$, we have

$$\mathrm{tr}X = \sum_{j=0}^{n-1} U_2^j X U_2^{\dagger j}.$$

Now for any $mn\times mn$ matrix $\mathbf{X}$ on $\mathcal{H}_1\otimes\mathcal{H}_2$ with the partition

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mm} \end{pmatrix},$$

where each X_{ij} is an $n\times n$ *diagonal* matrix on $\mathcal{H}_2$, let us define an operation $\mathcal{S}$ as

$$\mathcal{S}(\mathbf{X}) = (\mathrm{tr}_2\mathbf{X}) \otimes I_2.$$

Here

$$\mathrm{tr}_2\mathbf{X} = \begin{pmatrix} \mathrm{tr}X_{11} & \mathrm{tr}X_{12} & \cdots & \mathrm{tr}X_{1m} \\ \mathrm{tr}X_{21} & \mathrm{tr}X_{22} & \cdots & \mathrm{tr}X_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{tr}X_{m1} & \mathrm{tr}X_{m2} & \cdots & \mathrm{tr}X_{mm} \end{pmatrix}$$

is the partial trace of $\mathbf{X}$ over $\mathcal{H}_2$.

Let $\mathbf{U}=I_1\otimes U_2$. Then $\mathbf{U}$ is unitary, and $\mathbf{U}^\dagger=I_1\otimes U_2^\dagger$. For the above $\mathbf{X}$ with *diagonal* X_{ij}, we have

$$\mathcal{S}(\mathbf{X}) = \frac{1}{n}\sum_{j=0}^{n-1} \mathbf{U}^j\mathbf{X}\mathbf{U}^{\dagger j}.$$

Let $\omega=e^{2\pi i/n}$ be a primitive root of 1. We form the diagonal unitary matrix

$$V_2 = \mathrm{diag}\{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$$

on $\mathcal{H}_2$.

For any matrix $Y=(y_{ij})$ on $\mathcal{H}_2$, let

$$\mathcal{D}(Y) = \mathrm{diag}(y_{11}, y_{22}, \ldots, y_{nn})$$

be the diagonal matrix obtained from Y by keeping only the diagonal elements and setting all off-diagonal elements to zero. Then we have

$$\mathcal{D}(Y) = \frac{1}{n}\sum_{j=0}^{n-1} V_2^j Y V_2^{\dagger j}.$$

Now for any $mn\times mn$ matrix $\mathbf{Y}$ on $\mathcal{H}_1\otimes\mathcal{H}_2$ partitioned as

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mm} \end{pmatrix},$$

with each Y_{ij} being an $n\times n$ matrix on $\mathcal{H}_2$, let us define an operation $\mathcal{T}$ as

$$\mathcal{T}(\mathbf{Y}) = \begin{pmatrix} \mathcal{D}(Y_{11}) & \mathcal{D}(Y_{12}) & \cdots & \mathcal{D}(Y_{1m}) \\ \mathcal{D}(Y_{21}) & \mathcal{D}(Y_{22}) & \cdots & \mathcal{D}(Y_{2m}) \\ \vdots & \vdots & \cdots & \vdots \\ \mathcal{D}(Y_{m1}) & \mathcal{D}(Y_{m2}) & \cdots & \mathcal{D}(Y_{mm}) \end{pmatrix}.$$

This operation maps an $mn \times mn$ matrix into a matrix with the same order.

Put $\mathbf{V}=I_1 \otimes V_2$, then $\mathbf{V}$ is unitary, and $\mathbf{V}^\dagger = I_1 \otimes V_2^\dagger$. We have

$$\mathcal{T}(\mathbf{Y}) = \frac{1}{n}\sum_{j=0}^{n-1} \mathbf{V}^j \mathbf{Y} \mathbf{V}^{\dagger j}.$$

Now we proceed to prove the desired result. By the concavity of the affinity, we have

$$\begin{aligned} A(\mathcal{S}(\mathcal{T}(\rho)),\mathcal{S}(\mathcal{T}(\sigma))) &= A\left(\frac{1}{n}\sum_{j=0}^{n-1} \mathbf{U}^j \mathcal{T}(\rho)\mathbf{U}^{\dagger j}, \frac{1}{n}\sum_{j=0}^{n-1} \mathbf{U}^j \mathcal{T}(\sigma)\mathbf{U}^{\dagger j}\right) \\ &\geq \frac{1}{n}\sum_{j=0}^{n-1} A(\mathbf{U}^j \mathcal{T}(\rho)\mathbf{U}^{\dagger j}, \mathbf{U}^j \mathcal{T}(\sigma)\mathbf{U}^{\dagger j}) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} A(\mathcal{T}(\rho),\mathcal{T}(\sigma)) = A(\mathcal{T}(\rho),\mathcal{T}(\sigma)). \end{aligned}$$

Furthermore,

$$\begin{aligned} A(\mathcal{T}(\rho),\mathcal{T}(\sigma)) &= A\left(\frac{1}{n}\sum_{j=0}^{n-1} \mathbf{V}^j \rho \mathbf{V}^{\dagger j}, \frac{1}{n}\sum_{j=0}^{n-1} \mathbf{V}^j \sigma \mathbf{V}^{\dagger j}\right) \\ &\geq \frac{1}{n}\sum_{j=0}^{n-1} A(\mathbf{V}^j \rho \mathbf{V}^{\dagger j}, \mathbf{V}^j \sigma \mathbf{V}^{\dagger j}) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} A(\rho,\sigma) \\ &= A(\rho,\sigma). \end{aligned}$$

Now the desired inequality follows from

$$\mathcal{S}(\mathcal{T}(\rho)) = \mathrm{tr}_2 \mathcal{T}(\rho) \otimes \frac{I_2}{n} = \rho_1 \otimes \frac{I_2}{n},$$

$$\mathcal{S}(\mathcal{T}(\sigma)) = \mathrm{tr}_2 \mathcal{T}(\sigma) \otimes \frac{I_2}{n} = \sigma_1 \otimes \frac{I_2}{n},$$

and

$$A(\rho_1 \otimes I_2/n, \sigma_1 \otimes I_2/n) = A(\rho_1,\sigma_1).$$

[1] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).
[2] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory* (Springer-Verlag, New York, 1986).
[3] A. Bhattacharyya, Bull. Calcutta Math. Soc. **35**, 99 (1943).
[4] S. Kakutani, Ann. Math. **49**, 214 (1948).
[5] K. Matusita, Ann. Math. Stat. **26**, 631 (1955); Ann. Inst. Stat. Math. **23**, 137 (1971).
[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[7] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).
[8] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[9] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).
[10] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, Berlin, 1993).
[11] V. Vedral, Rev. Mod. Phys. **74**, 197 (2002).
[12] M. B. Ruskai, Rev. Math. Phys. **6**, 1147 (1994).
[13] M. Ozawa, Phys. Lett. A **268**, 158 (2000).
[14] D. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).
[15] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976); **24**, 229 (1986); **33**, 253 (1993); Phys. Rev. A **62**, 032307 (2000).
[16] M. Hubner, Phys. Lett. A **163**, 239 (1992); **179**, 226 (1993).
[17] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[18] C. A. Fuchs and C. M. Caves, Open Syst. Inf. Dyn. **3**, 1 (1995).
[19] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996); H. Barnum, E. Knill, and M. A. Nielsen, IEEE Trans. Inf. Theory **46**, 1317 (2000); J Dittmann, J. Phys. A **32**, 2663 (1999).
[20] J. L. Dodd and M. A. Nielsen, e-print, quant-ph/0111053.
[21] A. Albrecht, Phys. Rev. D **50**, 2744 (1994).
[22] E. P. Wigner and M. M. Yanase, Proc. Natl. Acad. Sci. U.S.A. **49**, 910 (1963).
[23] S. Luo, Phys. Rev. Lett. **91**, 180403 (2003).
[24] E. H. Lieb, Adv. Math. **11**, 267 (1973).
[25] R. A. Fisher, Proc. Cambridge Philos. Soc. **22**, 700 (1925).
[26] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1974).
[27] H. P. Yuen and M. Lax, IEEE Trans. Inf. Theory **19**, 740 (1973); H. P. Yuen, R. S. Kennedy, and M. Lax, *ibid.* **21**, 125 (1975).
[28] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[29] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[30] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997); V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).