Entropy and Geometry of Quantum States

Kumar Shivam\textsuperscript{1}, Anirudh Reddy\textsuperscript{1}, Joseph Samuel\textsuperscript{1} and Supurna Sinha\textsuperscript{1}

\textsuperscript{1}Raman Research Institute, Bangalore 560080, India.

(Dated: September 26, 2016)

It is widely accepted that the Bures-Helstrom (BH) metric and the corresponding Cramer-Rao bound impose a limit on our ability to discriminate between neighbouring quantum states. We explore an alternative metric based on the quantum relative entropy, which imposes a less stringent bound. Specialising to qubits, we derive the explicit form of the metric, which defines a distance on the qubit state space. We then show how the Bures-Helstrom-Cramer-Rao (BHCR) bound can be beaten using quantum entanglement. We propose an experiment using trapped cold atoms to implement this idea. The less stringent bound imposes the absolute limit on state discrimination.

PACS numbers: 03.67.-a, 02.40.-k, 03.65.Ta

I. INTRODUCTION

Given two quantum states, how easily can we tell them apart? Consider for instance, gravitational wave detection which is of considerable interest in recent times\textsuperscript{1,2}. Typically, we expect a weak signal which produces a small change in the quantum state of the detector. The sensitivity of our instrument is determined by our ability to detect small changes in a quantum state. This leads to the issue of distinguishability measures on the space of quantum states\textsuperscript{3,4}. In general, quantum states are represented by density matrices. In this paper, we address this question and arrive at a natural Riemannian metric on the space of density matrices.

In fact, even in the classical domain, one encounters similar questions while considering drug trials, electoral predictions or when we compare a biased coin to a fair one. As the number of trials (or equivalently, the size of the sample) increases, our ability to distinguish between candidate probability distributions improves. Such considerations give rise in a natural and operational manner, to a metric on the space of probability distributions\textsuperscript{7}. This metric is known as the Fisher-Rao metric and plays an important part in the theory of parameter estimation. This metric leads to the Cramer-Rao bound which limits the variance of any unbiased estimator.

Another example of the use of a Riemannian metric to measure distinguishability occurs in the theory of colours\textsuperscript{8,9}. The space of colours is two dimensional (assuming normal vision) and one can see this on a computer screen in several graphics softwares. The sensation of colour is determined by the relative proportion of the RGB values, which gives us two parameters. The extent to which one can distinguish neighbouring colours is usually represented by MacAdam ellipses\textsuperscript{8,10}, which are contours on the chromaticity diagram which are just barely distinguishable from the centre. These ellipses give us a graphical representation of an operationally defined Riemannian metric on the space of colours. The flat metric on the Euclidean plane would be represented by circles, whose radii are everywhere the same. As it turns out, the metric on the space of colours is not flat and the MacAdam ellipses vary in size, orientation and eccentricity over the space of colours. This analogy is good to bear in mind, for we provide a similar visualisation of the geometry of state space based on entropic considerations.

In this paper we use Umegaki’s quantum relative entropy\textsuperscript{11} as a measure of distinguishability between quantum states. We will show that this entropy defines a metric on state space which is the true quantum counterpart of the Fisher-Rao metric. The paper is organized as follows. We review the connection between the Kullback-Leibler (KL) divergence and the theory of statistical inference\textsuperscript{6,12} in Sec II. In Sec III we introduce Umegaki’s quantum relative entropy, which we use as a starting point for deriving a metric on the space of density matrices. We then outline the derivation of this metric for the special case of a qubit. In Sec IV we relate our metric to fluctuations in the free energy and thus give a thermodynamic interpretation of our metric. In Sec V we represent the geometry of the qubit state space by using MacAdam ellipses, make a comparison between the KL divergence and the quantum relative entropy and demonstrate the advantage of the quantum over the classical relative entropy in the context of measurements. We discuss quantum strategies which enable us to go beyond the naive Cramer-Rao bound in Sec VI and finally end the paper with some concluding remarks in Sec VII.

II. KL DIVERGENCE AS MAXIMUM LIKELIHOOD

Let us consider a biased coin for which the probability of getting a head is \( p_H = \frac{1}{3} \) and that of getting a tail is \( p_T = \frac{2}{3} \). Suppose we incorrectly assume that the coin is fair and assign probabilities \( q_H = \frac{1}{2} \) and \( q_T = \frac{1}{2} \) for getting a head and a tail respectively. The question of interest is the number of trials needed to be able to distinguish (at a given confidence level) between our assumed probability distribution and the measured probability distribution. A popular measure for distinguishing between the expected distribution and the measured distribution is given by the relative entropy or the KL
divergence (KLD) which is widely used in the context of distinguishing classical probability distributions [13]. Let us consider \( n \) independent tosses of a coin leading to a string \( S = \{ HTHHTHTHTHTTTT \ldots \} \). What is the probability that the string is generated by the model distribution \( Q = \{ q, 1 - q \} \)? The observed frequency distribution is \( P = \{ p, 1 - p \} \). If there are \( n_H \) heads and \( n_T \) tails in a string then the probability of getting such a string is \( \frac{n!}{n_H!n_T!} q^{n_H} (1 - q)^{n_T} \) which we call the likelihood function \( L(n|Q) \). If we take the average of the logarithm of this likelihood function and use Stirling’s approximation for large \( n \) we get the following expression:

\[
\frac{1}{n} \log L(n|Q) = -D_{KL}(P||Q) + \frac{1}{n} \log \frac{1}{\sqrt{2\pi np(1-p)}}
\]

where \( p = \frac{n_H}{n} \) and \( D_{KL}(P||Q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \). The second term in (1) is due to the sub-leading term \( \frac{1}{2} \log 2\pi n \) of Stirling’s approximation. If \( D_{KL}(P||Q) \neq 0 \) then the likelihood of the string \( S \) being produced by the \( Q \) distribution decreases exponentially with \( n \).

Thus \( D_{KL}(P||Q) \) gives us the divergence of the measured distribution from the model distribution. The KL divergence is positive and vanishes if and only if the two distributions \( P \) and \( Q \) are equal. In this limit, we find that the exponential divergence gives way to a power law divergence, as shown by the subleading term in (1). The arguments above generalize appropriately to an arbitrary number of outcomes (instead of two) and also to continuous random variables.

### III. RELATIVE ENTROPY AS A METRIC

As we saw above, the relative entropy (or KLD) gives an operational measure of how distinguishable two distributions are, quantified by the number of trials needed to distinguish two distributions at a given confidence level. However, the KLD is not a distance function on the space of probability distributions: it is not symmetric between the distributions \( P \) and \( Q \). One may try to symmetrize this function, but then, the result does not satisfy the triangle inequality. However, when \( P \) and \( Q \) approach one another, the relative entropy can be Taylor expanded to second order about \( P \). The Riemannian metric does define a positive definite quadratic form at \( P \) and thus a Riemannian metric on the space of probability distributions. In classical probability theory, this is known as the Fisher-Rao metric [7,14] and this forms the basis of classical statistical inference and the famous \( \chi^2 \)-squared test. The Riemannian metric then defines a distance function, based on the lengths of the shortest curves connecting any two states \( P \) and \( Q \). Our strategy below is to apply the same idea toUmegaki’s quantum relative entropy defined on the space of quantum states.

Consider density matrix \( \rho \) of an \( N \) state system, satisfying \( \rho^1 = \rho \), \( \text{Tr}(\rho) = 1 \) and \( \rho > 0 \), where we assume \( \rho \) to be strictly positive, so that we are not at the boundary of state space. The quantum relative entropy of any two states parametrized by a set of parameters \( \lambda(\lambda^1, \lambda^2, \ldots, \lambda^N) \) is represented in terms of density matrices \( \rho_1(\lambda_1) \) and \( \rho_2(\lambda) \) and is given by:

\[
S(\rho_1(\lambda_1)||\rho_2(\lambda)) = \text{Tr}[\rho_1 \log \rho_1 - \rho_1 \log \rho_2].
\] (2)

\( S \) is positive and vanishes if and only if \( \rho_2 = \rho_1 \) [15]. Let us consider \( S(\rho_1(\lambda_1)||\rho_2(\lambda)) \) as a function of its second argument. If the states \( \rho_1 \) and \( \rho_2 \) are infinitesimally close to each other, we can Taylor expand the relative entropy function.

\[
S(\rho_1||\rho_2) = S(\rho_1||\rho_1) + \frac{\partial S}{\partial \lambda^i} \Delta \lambda^i + \frac{1}{2} \frac{\partial^2 S}{\partial \lambda^i \partial \lambda^j} \Delta \lambda^i \Delta \lambda^j + ..
\] (3)

Notice that \( S(\rho_1||\rho_1) \) is zero and the second term is zero because we are doing a Taylor expansion about the minimum of the relative entropy function. The third term, which is second order in \( \Delta \lambda \), gives us the metric and is positive definite for \( \Delta \lambda \neq 0 \).

\[
g_{ij} = \frac{\partial^2 S}{\partial \lambda^i \partial \lambda^j}.
\] (4)

As is made clear below, the Hessian actually defines a metric tensor. If we consider the relative entropy \( S(\lambda^i) \) as a function of the parameters of the density matrix then the first derivative in (3) transforms like a tensor of rank-1 under a change of co-ordinates. To see this let us consider \( A_i = \frac{\partial S}{\partial x^i} \). When we make a change of coordinates from \( \lambda^i \to \lambda^i \), \( A_i \) transforms as follows:

\[
\bar{A}_i = \frac{\partial \lambda^j}{\partial \lambda^i} \frac{\partial S}{\partial \lambda^j} = T_i^j A_j.
\] (5)

Consider the second derivative term \( g_{ij} = \frac{\partial^2 S}{\partial x^i \partial x^j} = \frac{\partial}{\partial x^j} A_i \). In order to see its transformation property consider,

\[
\bar{g}_{ij} = \frac{\partial}{\partial \lambda^j} \left( \frac{\partial S}{\partial \lambda^i} \right) = \frac{\partial}{\partial \lambda^j} (\bar{A}_i) = \frac{\partial}{\partial \lambda^j} (T_i^k A_k).
\] (6)

\[
\frac{\partial}{\partial \lambda^j} (T_i^k A_k) = T_j^k \left( \frac{\partial A_k}{\partial \lambda^j} \right) + \left( \frac{\partial T_i^k}{\partial \lambda^j} \right) A_k
\]

\[
= T_j^k T_i^l g_{kl} + \left( \frac{\partial T_i^k}{\partial \lambda^j} \right) \frac{\partial S}{\partial \lambda^l}.
\] (7)

The last term in (7) vanishes identically as we are Taylor expanding the relative entropy function about a stationary point and we obtain \( \bar{g}_{ij} = T_j^k T_i^l g_{kl} \). Positivity of the Hessian is guaranteed as the stationary point is the absolute minimum.

Consider two mixed states \( \rho_1 \) and \( \rho_2 \) of a two level quantum system commonly referred to as a qubit. These
can be written as \( \rho_1 = \frac{1 + X.\sigma}{2} \) and \( \rho_2 = \frac{1 + Y.\sigma}{2} \) where \( |X| \) and \( |Y| < 1 \). \( X \) and \( Y \) are three dimensional vectors with components \( x^i \) and \( y^j \). The relative entropy function can be written as follows:

\[
S(\rho_1\|\rho_2) = \text{Tr} \left[ \left( \frac{1 + X.\sigma}{2} \right) \log \left( \frac{1 + X.\sigma}{2} \right) - \frac{1 + Y.\sigma}{2} \right] \tag{8}
\]

We can use the power series expansion of \( \log(1 + r) \) to evaluate the trace of the above expression.

\[
\log(1 + Y.\sigma) = \left( \sum_{m=0}^{\infty} \frac{|Y|^{2m+1}}{2m+1} \right) \frac{Y.\sigma}{|Y|} + \left( \sum_{n=0}^{\infty} \frac{|Y|^{2n}}{2n} \right) I, \tag{9}
\]

where \( f_o(|Y|) \) and \( f_e(|Y|) \) are respectively the odd and even parts of the function \( f(r) = \log(1 + r) \). Notice that the odd part of the expansion is traceless. Making use of the above expansion we can express \( S(\rho_1\|\rho_2) \) as follows

\[
S(\rho_1\|\rho_2) = S(X\|Y) = f_e(|Y|) - \frac{f_o(|Y|)}{|Y|} (X.Y). \tag{10}
\]

In order to compute the Hessian of \( S(\rho_1\|\rho_2) \) we compute the second derivative \( \frac{\partial^2 S}{\partial y^i \partial y^j} \) with respect to \( y^j \) and then set \( y^j = x^i \) and obtain the following metric [3]:

\[
g_{ij} = \frac{\partial^2 S}{\partial x^i \partial x^j} = C(r) \frac{x^i x^j}{r^2} + D(r) \{ \delta_{ij} - \frac{x^i x^j}{r^2} \} \tag{11}
\]

where \( C(r) = \frac{1}{1-rr}, \ D(r) = \frac{1}{2} \log \left( \frac{1+r}{1-r} \right) \) and \( r = |Y| \).

The corresponding line element is given in polar coordinates by:

\[
ds^2 = \frac{dr^2}{1 - r^2} + \left[ \frac{r}{2} \log \left( \frac{1+r}{1-r} \right) \right] \left( d\theta^2 + \sin^2 \theta d\phi^2 \right). \tag{12}
\]

This metric has been discussed earlier by Bogoliubov, Kubo and Mori (BKM) in the context of statistical mechanical fluctuations [13, 17]. We refer to it as the BKM metric.

The scalar curvature \( R \) of the BKM metric is given by:

\[
R = \frac{4r^2 - 4r(1 + r^2) \log \left( \frac{1+r}{1-r} \right) + (1 + 2r^2 - 3r^4) [\log \left( \frac{1+r}{1-r} \right)]^2}{2r^2 (1 - r^2) [\log \left( \frac{1+r}{1-r} \right)]^2} \tag{13}
\]

As we can see from Fig.1, the metric has negative scalar curvature and therefore the geodesics (Fig.2) cannot cross more than once. It follows therefore that any two states, are connected by an unique geodesic. The length of this geodesic gives us a distance on the space of states. This has all the properties expected of a distance function: it is symmetric, strictly positive between distinct points and satisfies the triangle inequality. The scalar curvature is zero near the origin and diverges logarithmically to minus infinity as \( r \) goes to unity. The geodesics of this metric are easily worked out from classical mechanics. The metric has spherical symmetry, because the quantum state space is invariant under unitary transformations.

Setting \( r = \sin \alpha \), we rewrite the metric as

\[
ds^2 = d\alpha^2 + F(\alpha) \left( d\theta^2 + \sin^2 \theta d\phi^2 \right), \tag{14}
\]

where \( F(\alpha) = \frac{\sin \alpha}{2} \log \left[ \frac{1 + \sin \alpha}{1 - \sin \alpha} \right] \). Because of the spherical symmetry, there is a conserved angular momentum vector \( \hat{J} \) and thus the geodesics lie in the plane perpendicular to \( \hat{J} \). Thus we can confine our calculations to a plane, reducing the form of the metric to

\[
ds^2 = d\alpha^2 + F(\alpha) \left( d\phi^2 \right), \tag{15}
\]

where we have set \( \theta = \frac{\pi}{2} \). The Lagrangian of the classical mechanical system is

\[
L = \frac{1}{2} \left( \dot{\alpha}^2 + F(\alpha) \dot{\phi}^2 \right). \tag{16}
\]

The constants of motion for this problem are the energy and the angular momentum, which are given by

\[
E = \frac{1}{2} \left( \dot{\alpha}^2 + F(\alpha) \dot{\phi}^2 \right), \quad P_\phi = J = \frac{\partial L}{\partial \dot{\phi}} = F(\alpha) \dot{\phi}. \tag{17}
\]

Using the above equations we solve for \( \dot{\alpha} \) and \( \dot{\phi} \). Our numerical solution gives us the geodesics of interest. A typical geodesic is displayed in Fig.2. Given any two points in the state space (for example the red dots of Fig.2), the length of the unique geodesic [18] connecting them gives us a distance function. This is very similar in spirit to a construction of Wootters [19], who introduced a metric based on distinguishability for pure states and used this to define a metric on pure states, which ultimately yielded the Fubini study metric.
free energy \( \mathbf{18} \) can be simplified as:

\[
-TS(\rho_1\|\rho_2) = \langle U_1 - F_1 \rangle - \langle U_2 - F_2 \rangle.
\] (19)

Since \( H_2 = H_1 + \epsilon V \) taking the ensemble average on both sides we get \( U_2 = U_1 + \epsilon \langle V \rangle \). Substituting this into (19) we get:

\[
-TS(\rho_1\|\rho_2) = \langle F_2 - F_1 \rangle - \epsilon \langle V \rangle.
\] (20)

Since \( \rho_1 \) and \( \rho_2 \) are infinitesimally close together we can use thermodynamic perturbation theory \( \mathbf{21} \) to show:

\[
F_2 = F_1 + \epsilon F'_1 + \frac{\epsilon^2}{2} F''_1 + O(\epsilon^3),
\] (21)

where,

\[
F'_1 = \frac{d}{d\epsilon} \left( \frac{1}{\beta} \log Z_2 \right) = \langle V \rangle.
\] (22)

From (13) we find

\[
S(\rho_1\|\rho_2) = \frac{1}{2} \frac{\partial^2 S}{\partial \lambda^i \partial \lambda^j} \Delta \lambda^i \Delta \lambda^j.
\] (23)

Substituting (21) and (22) in (20) we get:

\[
g_{ij} \Delta \lambda^i \Delta \lambda^j = \frac{\partial^2 S}{\partial \lambda^i \partial \lambda^j} \Delta \lambda^i \Delta \lambda^j = -\beta \epsilon^2 F''_1.
\] (24)

From the statistical physics perspective, the metric \( \mathbf{14} \) is related to the susceptibility of the quantum state to perturbations \( \mathbf{16}, \mathbf{17} \). The Gibbs state is the state that maximizes its entropy subject to an energy constraint. However, in statistical physics a system makes spontaneous excursions to neighbouring lower entropy states. The size of these fluctuations is determined by the Hessian of the entropy function and thus related to the susceptibility.

V. MEASUREMENTS ON SINGLE QUBITS

Let us consider the problem of distinguishing between two states \( \rho_1 \) and \( \rho_2 \) of a qubit. Suppose that we are given a string of \( n \) qubits all in the same state, which may be either \( \rho_1 \) or \( \rho_2 \). A possible strategy is to make projective measurements on individual qubits, measuring the spin component in the direction \( \hat{m} \). For each choice of \( \hat{m} \) we find \( p_+ = \frac{1 + \hat{m} \cdot \vec{m}}{2} \) and \( p_- = \frac{1 - \hat{m} \cdot \vec{m}}{2} \) and we can compute the KL-Divergence of the two distributions as:

\[
S(\rho_1\|\rho_2) = p_+ \log \frac{p_+}{q_+} + p_- \log \frac{p_-}{q_-}.
\] (25)

We will now choose \( \hat{m} \) in such a way as to maximize \( S(\rho_1\|\rho_2) \). This gives us,

\[
\delta S_m = \frac{\partial S}{\partial \hat{m}} \delta \hat{m} = \lambda \delta \hat{m},
\] (26)
which can be rewritten as

$$\frac{\partial S_m}{\partial a_1} X + \frac{\partial S_m}{\partial a_2} Y = \lambda \delta \mathbf{m}, \quad (27)$$

where $a_1 = \mathbf{m} \cdot X$ and $a_2 = \mathbf{m} \cdot Y$. Since $\delta S_m$ is a linear combination of $X$ and $Y$ we find that $\mathbf{m}$ must lie in the $x$-$y$ plane, as shown in Fig.3. We can replace $\mathbf{m} = (\cos \beta, \sin \beta)$ by the angle $\beta$, which gives us $p_\pm = \frac{1 \pm r_1 \cos \beta}{2}$ and $q_\pm = \frac{1 \pm r_2 \cos (\theta + \beta)}{2}$. Plotting $S(\beta)$ (Fig.4), we find that the maximum distinguishability is attained at $\beta = \beta^*$. This is clearly the most advantageous choice of $\beta$. The value of $S_m$ at the maximum is denoted by $S^*(r_1, r_2, \theta) = S(r_1, r_2, \theta, \beta^*(r_1, r_2, \theta))$. $S^*(r_1, r_2, \theta)$ gives us the optimal choice for state discrimination when we measure qubits, one at a time. As we can see in Fig.4, $S^*(r_1, r_2, \theta)$ is always less than the quantum relative entropy. Equality between $S^*(\rho_1 \parallel \rho_2)$ and $S(\rho_1 \parallel \rho_2)$ happens if and only if $[\rho_1, \rho_2] = 0$ ($\theta = 0, \pi$) [See Fig.5] i.e when the two density matrices commute with each other.

The difference

$$\Omega(\rho_1, \rho_2) = S(\rho_1, \rho_2) - S^*(\rho_1, \rho_2) \quad (28)$$

is always positive and represents the Quantum Advantage: Our main aim is to use the difference $\Omega$ to improve our ability to experimentally distinguish quantum states.

Optimising $S(\rho_1 \parallel \rho_2)$ with respect to $\beta$ is not tractable analytically for general states. But if $\rho_1$ and $\rho_2$ are close together, an analytical treatment is possible. Let

$$r_1 = r, \quad r_2 = r + dr, \quad \theta = d\theta, \quad (29)$$

where $dp_+$ and $dp_-$ are:

$$dp_+ = q_+ - p_+ = \cos \beta dr - r \sin \beta d\theta, \quad (30)$$

$$dp_- = q_- - p_- = \frac{r \sin \beta d\theta - \cos \beta dr}{2}.$$ 

Considering the relative entropy between two neighbouring states and doing a Taylor expansion, similar to the procedure outlined in Sec III gives us the Fisher-Rao distance [2], which is given by

$$ds^2 = \sum \frac{dp_i^2}{p_i},$$

or more explicitly

$$ds^2 = \frac{dp_+^2}{p_+} + \frac{dp_-^2}{p_-}.$$ 

Substituting $dp_+, dp_- , p_+$ and $p_-$, we finally get

$$ds^2 = \frac{(dr - r \tan \beta d\theta)^2}{1 - r^2 + \tan^2 \beta}, \quad (31)$$

Keeping $r, dr, d\theta$ fixed and optimising with respect to $\beta$ we find

$$\tan \beta^* = -\frac{r(1 - r^2)}{dr/d\theta}, \quad (32)$$

and get an expression for the metric

$$ds^2 = \frac{dr^2}{1 - r^2} + r^2 d\theta^2. \quad (33)$$

FIG. 3: The figure shows the measurement direction $\mathbf{m}$ and directions $X$ and $Y$ corresponding to the density matrices $\rho_1(X)$ and $\rho_2(Y)$ respectively.
This is the BH metric \cite{22,24}. Ref.\cite{25} and references therein. It plays the role of the Fisher-Rao metric in quantum physics, if one restricts oneself to measuring one qubit at a time. As we will see in the next section there are strategies involving quantum entanglement that will go beyond the BHCR bound. Fig. 6 shows a graphical representation of the geometry of state space as given by the BH metric (in blue) and the BKM metric \cite{12} (in red).

VI. QUANTUM ADVANTAGE: BEYOND THE BHCR BOUND

Let us consider \( n \) qubits all of which are in the same state, either \( \rho_1 \) or \( \rho_2 \) and make projective measurements on each qubit one at a time. This gives us classical probabilities and we can use the KL divergence to measure the distinguishability. As we saw in the last section the best strategy is to optimize the projective measurement so as to maximise the relative entropy \( S(\rho_1 || \rho_2) \). This strategy is limited by the BHCR bound set by the BH metric. Now let us consider these qubits two at a time. The quantum state of the combined system is now given by \( \hat{\rho} = \rho \otimes \rho \), where \( \rho \) can refer to either \( \rho_1 \) or \( \rho_2 \). In choosing a measurement basis to distinguish \( \tilde{\rho}_1 \) from \( \tilde{\rho}_2 \), we now have the additional advantage that we can choose bases which are not separable. This extra freedom gives us the quantum advantage which comes from entanglement. Such choices allow us to go beyond the bound set by the BH metric.

For example, let us choose \((r_1,r_2,\theta) = (0.9,0.5,\pi/2)\) so that \( X = \{r_1,0,0\} \), \( Y = \{r_2/\sqrt{2},0,r_2/\sqrt{2}\} \) the direction \( \hat{m} \) in the \( x-y \) plane \( \hat{m} = \{\cos \beta, \sin \beta\} \). Let the corresponding 1-qubit basis which diagonalizes \( \hat{m} \sigma \) be \(|+\rangle, |-\rangle \). We now construct the non separable basis \(|b_1\rangle = |++\rangle \), \(|b_2\rangle = |+-\rangle \), \(|b_3\rangle = |--\rangle \), \(|b_4\rangle = |--\rangle \). Note that two of these basis states are maximally entangled Bell states and two are completely separable. (Curiously, using all basis states as Bell states leads to no improvement over the separable states.) We numerically compute the relative entropy and optimize over \( \beta \). This leads to an improvement over the strategy outlined in the earlier section where we considered measurements on one qubit at a time. The improvement is seen in the value of the relative entropy per qubit, which increases from 0.5839 in the one qubit strategy to 0.5856 in the two qubit strategy.

In fact this number can be further improved. By numerical Monte-Carlo searching we have found bases (which don’t have the clean form above) which yield a relative entropy of 0.5863 per qubit. Our Monte-Carlo search is simplified by the observation that one can by a unitary transformation bring any two states described by \( X \) and \( Y \) to the \( x-z \) plane of the Bloch ball, so that we
are working over the real numbers rather than complex numbers. Over the reals, unitary matrices are orthogonal. We start with an initial basis in the four dimensional real Hilbert space of the composite system and then rotate the basis by a random orthogonal matrix close to the identity. We then compute the relative entropy using the new basis and accept the move if the new basis has larger relative entropy and reject it otherwise. This gives us a monotonic rise in the relative entropy and produces for us the optimal basis in the two qubit Hilbert space. The method extends easily to three qubits and more although the searches are more time consuming.

This improvement stems from an enlargement of the measurement basis set. In the strategy of the last section, we were effectively confined to separable bases in the two-qubit Hilbert space. Our numerical analysis which allows for a greater freedom of choice of basis states for measurement in fact improves over the one using separable basis states.

We have numerically observed that measuring three qubits at a time results in a further improvement over the two qubit measurement strategy. However, this number (0.5880) still falls short of the quantum relative entropy which is 0.6385. As the number of qubits $n \to \infty$ and we measure all of them together we approach the quantum relative entropy. Thus the gap between the classically optimised relative entropy and the quantum relative entropy (Fig.4) progressively reduces as one increases the number of qubits measured at a time. This strategy clearly improves on the BHCR bound.

VII. CONCLUSION

We have used the quantum relative entropy as a starting point and derived a metric on the state space which measures our ability to distinguish two nearby quantum states. This leads to a distance function on the state space. Our use of the quantum relative entropy is motivated by the fact that it has both physical and operational significance. Other measures of entropy do exist, like for example, the $\alpha$-divergence. Hasegawa [26] has derived a Riemannian metric using the $\alpha$-divergence as the starting point. As discussed here, relative entropy has an operational significance in the context of quantum measurement which makes it an attractive candidate for deriving a metric on the space of quantum states. Umegaki’s relative entropy is additive for independent systems, (unlike the the $\alpha$-divergence) and also has a well defined physical interpretation in terms of statistical physics.

Questions addressed here were raised but not fully answered in an early paper of Peres and Wootters [27]. At that time it was not fully clear whether there was a one qubit strategy which could compete with the multiqubit strategy. Subsequent work using the machinery of $C^*$ algebras has made it clear [28] that the best one qubit strategy is inferior to the multiqubit strategy. As $n$ goes to infinity we approach the bound set by the BKM metric. Thus the quantum Cramer-Rao bound set by the BKM metric can be approached but not surpassed. In contrast, the BHCR bound can be surpassed, as we have seen in Sec VI. Thus the limit on distinguishability is set by the BKM metric rather than the BH metric.

We have worked out the geodesics of the BKM metric and plotted them numerically. We have noticed that any two points are connected by a unique geodesic. The BKM metric leads to a distance function on the state space that emerges naturally from entropic and geometric considerations. In working out the geodesics, it is easily seen analytically that the geodesics approach the boundary of the state space at right angles. However, this approach is logarithmically slow and is not apparent in Fig.2. The form of the geodesics on state space is reminiscent of the geodesics of the Poincaré metric which also meet the boundary at right angles. However, there are serious differences. While both metrics have negative curvature, the Poincaré metric has constant negative curvature, unlike our metric that has varying curvature, which diverges logarithmically at the boundary.

The quantum advantage described in this paper is strongly connected to the commutation properties of the density matrices $\rho_1$ and $\rho_2$. For commuting matrices, it is evident that the optimal basis should simultaneously diagonalise both the states and then the quantum advantage disappears. In fact, numerical studies show a strong correlation between the quantum advantage and the quantumness defined in [24, 30] as $Q(\rho_1, \rho_2) = 2\text{Tr}([\rho_1, \rho_2]^2)$ in terms of the commutator of two density matrices $\rho_1$ and $\rho_2$, using the Hilbert-Schmidt norm as a distinguishability measure on the space of density matrices. In particular, in the case analysed here, considering $\rho_2$ in the neighbourhood of $\rho_1$, we find that $Q(\rho_1, \rho_2) \propto r^4 d\theta^2$.

Our work suggests that experimental realisations of the quantum advantage are within reach. There have been studies involving measurements for quantum state discrimination [4, 6], where the upper limit of the state distinguishability is set by the BHCR bound. In order to exploit the quantum advantage discussed here and go beyond the BHCR bound, we need to measure in an entangled basis of the two qubit system. The entangled basis $|b_i\rangle$ mentioned in our paper, is related to the separable basis $|+\rangle, |-\rangle, |+\rangle, |+\rangle$ by a unitary transformation $U$ in the four dimensional Hilbert space. One can equivalently apply $U$ to the separable state $\bar{\rho} = \rho \otimes \rho$. This creates an entangled state $U^\dagger \bar{\rho} U$, which can then be measured in the separable basis using a projective measurement. Consider a pair of qubits subject to the Hamiltonian

$$H = \sigma_1 B_1 + \sigma_2 B_2 + J(t) \sigma_1 \sigma_2,$$  \hspace{1cm} (34)

which is a standard Heisenberg Hamiltonian for spins. This Hamiltonian evolution produces the unitary trans-
formation $U$ for suitable choice of $J(t)$.

This entangling unitary transformation $U$ is the square root of the SWAP operation $U = \sqrt{\text{SWAP}}$. $U$ has already been experimentally realised in [31] by creating a system in the laboratory subject to the Hamiltonian [31]. The method used in [31] is to load $^{87}$Rb atoms in pairs into an array of double well potentials. The experimenters have control over all the parameters in the Hamiltonian. They can generate the transformation $U$ at will by using a $\pi/4$ pulse for $J(t)$ by using radio frequency, site selective pulses to address the qubits in pairs (See Table 1 of [31]), thus effecting the entangling unitary transformation $U$. What remains to be done to implement our proposal is to projectively measure each of the qubits separately and thus achieve a violation of the BHC bound in distinguishing states. This would operationally bring out a subtle aspect of quantum information. We hope to interest experimental colleagues in this endeavour.

VIII. ACKNOWLEDGEMENT

It is a pleasure to thank Nomaan Ahmed, Giuseppe Manno, Saverio Pascacio and Rafael Sorkin for discussions. We also acknowledge Saptarishi Chaudhuri and Sanjukta Roy for discussions on possible experimental realisations.
[29] P. Iyengar, G. N. Chandan, and R. Srikanth, ArXiv e-prints (2013), 1312.1329.
[30] L. Ferro, P. Facchi, R. Fazio, F. Illuminati, G. Marmo, V. Vedral, and S. Pascazio, ArXiv e-prints (2015), 1501.03099.
[31] M. Anderlini, P. J. Lee, B. L. Brown, J. Sebby-Strabley, W. D. Phillips, and J. V. Porto, Nature 448(7152), 452 (2007). ISSN 0028-0836, URL http://dx.doi.org/10.1038/nature06011