



Geometry of quantum inference

Samuel L. Braunstein

Universität Ulm, Abteilung Quantenphysik, 89069 Ulm, Germany

Received 20 April 1996; accepted for publication 2 May 1996

Communicated by V.M. Agranovich

Abstract

The principle of maximum entropy has been a central tool in thermodynamics, data reduction and inverse problems with noisy data. Unfortunately, it has a natural bias towards the uniform distribution, which it picks out in the absence of any other information. For classical systems the generalization to the principle of minimum Kullback information allows for arbitrary prior information. Even though the quantum mechanical version of the Kullback information has no known closed form it is still a useful tool: we use it to define an information theoretic measure of the quantum “entanglement” of a pair of systems in a pure state; further, using the geometry of statistical correlations we derive trajectories which closely approximate the optimal quantum inference from prior to posterior; finally, we use this geometry to obtain a near optimal detection scheme in binary communication.

PACS: 89.70.+c; 03.65.-w; 89.80.+h

1. Introduction

Shannon’s celebrated definition of information is

$$H(p) = - \sum_j p^j \ln p^j, \quad (1)$$

where $p = \{p^j\}$ is a discrete probability distribution with outcome j having probability p^j . It has found wide application in many fields, most notably: computing, physics and telecommunications. The principle of maximum entropy, first introduced by Jaynes [1] (for some subtleties, see Ref. [2]), was posed as a general method of inference. This principle gives the most likely distribution, subject to some constraints, when we have no specific bias towards any particular distribution (more precisely when we would choose the uniform distribution in the absence of any other information). Originally proposed for thermodynamics, today it is widely used in a variety of

fields principally for data reduction and solving inverse problems with noisy data [3,4].

In 1959 Kullback [5] proposed a somewhat different measure, the Kullback information,

$$K(p|p_0) \equiv \sum_j p^j \ln(p^j/p_0^j), \quad (2)$$

this is a measure of negative relative entropy of one distribution p relative to another p_0 . If $p_0 = \{1/n, \dots, 1/n\}$ is the uniform distribution then this information reduces to the negative Shannon information $H(p)$ up to an additive constant $K(p|\{1/n, \dots, 1/n\}) = \ln n - H(p)$. This measure leads to a generalized method of inference via the principle of minimum Kullback information [5], yielding the most likely distribution p (the posterior) when we have a *prior* bias towards the distribution p_0 (the prior). The Kullback information is the mean

information required to turn the prior p_0 into the posterior p [6].

The Shannon information for a quantum system having density operator $\hat{\rho}$ is simply $H(p)$, with $p^j \equiv \text{tr}(\hat{P}_j \hat{\rho})$, when observed in basis $\{\hat{P}_j\}$. Because a bad choice of basis may throw away information the optimal measure of Shannon information for a quantum system is

$$\min_{\{\hat{P}_j\}} H(p) = -\text{tr}(\hat{\rho} \ln \hat{\rho}) \equiv S(\hat{\rho}), \quad (3)$$

which is just the system's von Neumann entropy. (It may be thought of as the least missing information about the quantum system.)

One of the reasons why the Kullback information has not been more widely used in physics is because it has not been successfully generalized to quantum mechanics. In a fixed basis the Kullback information for state $\hat{\rho}$ relative to $\hat{\rho}_0$ is

$$K_{\{\hat{P}_j\}}(\hat{\rho} | \hat{\rho}_0) = \sum_j \text{tr}(\hat{P}_j \hat{\rho}) \ln \left[\frac{\text{tr}(\hat{P}_j \hat{\rho})}{\text{tr}(\hat{P}_j \hat{\rho}_0)} \right]. \quad (4)$$

Since a bad choice of basis might make $\hat{\rho}$ and $\hat{\rho}_0$ less distinguishable we need to *maximize* Eq. (4) over all bases to obtain the optimal Kullback information, i.e., the quantum Kullback is $K_Q(\hat{\rho} | \hat{\rho}_0) \equiv \max_{\{\hat{P}_j\}} K_{\{\hat{P}_j\}}(\hat{\rho} | \hat{\rho}_0)$. This approach is sufficient for obtaining numerical results, but somewhat detracts from the general utility one might have hoped for by forcing us to explicitly introduce a basis. We might have expected the quantum generalization of Eq. (2) to take the form $K_Q(\hat{\rho} | \hat{\rho}_0) = \text{tr}[\mathcal{O}(\hat{\rho} \ln \hat{\rho} / \hat{\rho}_0)]$ where \mathcal{O} is some ordering operator. Unfortunately, so far the correct choice for this ordering is unknown [7]. (It has even been suggested that no closed form expression exists [8].)

We start by using the quantum Kullback information K_Q to derive a quantitative measure of quantum entanglement between pairs of systems (despite the lack of a closed form). Next, we rewrite classical inference as a trajectory from prior to posterior and show that a simple geometric structure underlies this formulation. We show that this classical geometric structure may be generalized to quantum mechanics and this allows us to perform quantum inference directly. We describe the cases where this procedure leads to optimal inference by minimizing K_Q . Fi-

nally, this geometric structure suggests an observable for “optimally” distinguishing states. We show that this observable may be used to design a near-optimal detection scheme in communication employing pairs of non-orthogonal states.

2. Measure of quantum entanglement

One of the most intriguing features of quantum mechanics is that widely separated systems can exhibit “non-local” influences, suggesting that the measurement of one system affects the other. This effect is due to the “entanglement” of the two separated systems, which means, in the formalism of quantum mechanics, that the two systems are actually one. The entire system is described by a single wave-function.

In quantum mechanics an arbitrary two-particle pure state may be written as $|\Psi\rangle \equiv \sum_j \alpha_j |\psi_j\rangle_1 \otimes |\phi_j\rangle_2$ decomposed here in the Schmidt basis. Measurements on either particle separately will yield reduced density operators for particle 1 and 2 as $\hat{\rho}_1 = \sum_j |\alpha_j|^2 |\psi_j\rangle_1 \langle \psi_j|$ and $\hat{\rho}_2 = \sum_j |\alpha_j|^2 |\phi_j\rangle_2 \langle \phi_j|$ respectively.

Thus, to characterize the difference between the uncorrelated systems $\hat{\rho}_1 \otimes \hat{\rho}_2$ and the entangled systems $|\Psi\rangle$, we would like to quantify the mean information required to infer the complete entangled state from the uncorrelated states. Following an interpretation of the classical version [6] that amount of information will be characterized by the quantum Kullback information $K_Q(|\Psi\rangle \langle \Psi| | \hat{\rho}_1 \otimes \hat{\rho}_2)$. The maximization is easily carried out in this case with the Schmidt basis itself being optimal. The result is

$$K_Q(|\Psi\rangle \langle \Psi| | \hat{\rho}_1 \otimes \hat{\rho}_2) = - \sum_j |\alpha_j|^2 \ln |\alpha_j|^2. \quad (5)$$

This then is an information theoretic measure of quantum entanglement and, as intuitively chosen by others [9]¹, is just the entropy of either subsystem taken separately.

¹ Ref. [9] demonstrates that in more than two dimensions many “measures” of entanglement fail to be even monotonic functions of each other.

3. Classical inference

Consider the classical problem of finding the minimum Kullback information under the constraint $\langle A \rangle = \sum_j p^j A_j = c$. The variational problem can be solved quite easily, yielding $p^j(t) \propto e^{-A_j t} p_0^j$ where t is the Lagrange multiplier. Nonetheless, it is convenient to write the solution in differential form

$$dp^j(t)/dt = -p^j(t) \Delta A_j, \quad (6)$$

where $\Delta A = A - \langle A \rangle$ (with the expectation calculated for $p(t)$). Integrating this trajectory from an initial distribution $p(0) = p_0$ at $t = 0$, we reach the minimum Kullback information when the trajectory passes through the “surface” of distributions which satisfies our constraint $\langle A \rangle = c$.

4. Geometry of classical inference

The above description is simple enough, however, in order to generalize it to quantum systems we find it convenient to rewrite it in terms of geometric concepts. We may achieve this by introducing the metric tensor $g_{jk} \equiv \delta_{jk}/p^j$ and by treating the infinitesimal increments in a distribution dp^j as contravariant vectors (having raised indices), and the deviations of random variables from the mean ΔA_j as covariant vectors (having lowered indices). In this way, Eq. (6) becomes [10]

$$\frac{dp^j(t)}{dt} = - \sum_k g^{jk} \Delta A_k, \quad (7)$$

where $(g^{jk}) = (g_{jk})^{-1}$ is the so-called raising operator that converts covariant vectors to contravariant ones (lowered indices to raised ones). The trajectory described by this equation is determined by the constraint itself. This recasting of an already simple problem yields a surprisingly powerful approach that may be used for more complicated constraints. In particular, this type of approach yields the fastest numerical method for minimizing Kullback information in real-world problems such as image reconstruction or generic inverse problems in the presence of statistical noise [4].

To complete our discussion of the classical geometry we note that the form of the inner product

between a pair of contravariant vectors is $dp \cdot dq = \sum_{jk} g_{jk} dp^j dq^k = \sum_{jk} dp^j dq^k / p^j$ which directly yields the Fisher information matrix [11], an important tool in estimation theory. Similarly, for covariant vectors

$$\Delta A \cdot \Delta B = \sum_{jk} g^{jk} \Delta A_j \Delta B_k = \langle \Delta A \Delta B \rangle, \quad (8)$$

yielding the matrix of statistical correlations between random variables.

We have shown that the classical Kullback information may be constructed in terms of much more elementary geometric objects. This gives us hope that the same approach may lead to useful results in the quantum mechanical case.

5. Quantum geometry

In the classical statistical geometry presented above, each distribution $p = \{p^j\}$ corresponds to a point while the infinitesimal increments $dp = \{dp^j\}$, correspond to contravariant vectors and may be conveniently described with the traditional index notation. By contrast, in the quantum geometry, where density operators $\hat{\rho} = \sum_{jk} \rho^{jk} |j\rangle\langle k|$ are points, the contravariant vectors $d\hat{\rho}$ are no longer conveniently written with a single index. We choose to abandon the index notation as much as possible. Instead of viewing two dual spaces of vectors (the contravariant and covariant vectors) we now view these same objects in a more pictorial manner.

The points remain as they were and we modify our language for the infinitesimal increments $d\hat{\rho}$ by calling them tangent vectors. The biggest change, however, is in our treatment of the dual space: Each observable \hat{A} is viewed as a family of surfaces (called a 1-form); each surface in its family comprises all the points $\hat{\rho}$ that yield a given expectation value, say, $\langle \hat{A} \rangle = c_2$. This is illustrated in Fig. 1. Strictly, we should only describe the zero-mean observables $\Delta \hat{A}$ as 1-forms, however, it is convenient to affix $\mathbb{1}$ to the space of 1-forms, thus incorporating all observables as we have done. (To be consistent, we must then affix the density operators $\hat{\rho}$ to the space of tangent vectors; this is equivalent to including unnormalized density operators as points.)

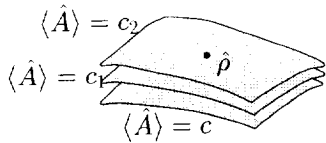


Fig. 1. A family of surfaces of constant expectation value correspond to a 1-form. Here the observable \hat{A} is represented by a family of surfaces. Each surface in this family contains all states $\hat{\rho}$ for which $\text{tr}(\hat{\rho}\hat{A})$ is some constant. The state $\hat{\rho}$ here is a point in the surface $\langle \hat{A} \rangle = c_2$.

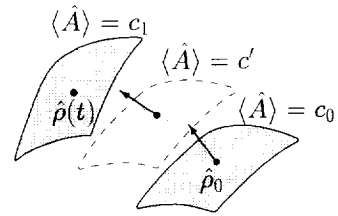


Fig. 3. By moving orthogonally to the family of surfaces \hat{A} from the point $\hat{\rho}_0$, we construct a unique trajectory $\hat{\rho}(t)$. We use this trajectory for quantum inference in the presence of prior information.

We pick our metric so as to describe statistical correlations by the “inner product” between observables (in analogy with Eq. (8)). Thus, $A \cdot B \equiv (\frac{1}{2}(\hat{A}\hat{B} + \hat{B}\hat{A})) = \text{tr}[\hat{A}\frac{1}{2}(\hat{\rho}\hat{B} + \hat{B}\hat{\rho})] = \text{tr}[\hat{A}\mathcal{R}_\rho(\hat{B})]$ which yields the raising operator (generalizing g^{jk})

$$\mathcal{R}_\rho(\hat{B}) = \frac{1}{2}(\hat{\rho}\hat{B} + \hat{B}\hat{\rho}). \tag{9}$$

The lowering operator \mathcal{L}_ρ is just the inverse of \mathcal{R}_ρ [12]. These operators are important, because they tell us how to construct tangent vectors that are orthogonal to surfaces and conversely, as illustrated in Fig. 2.

Finally, $ds^2 \equiv d\hat{\rho} \cdot d\hat{\rho} = \text{tr}[d\hat{\rho}\mathcal{L}_\rho(d\hat{\rho})]$ is the line element. This is precisely the distinguishability metric for density operators, obtained by Braunstein and Caves [13] by optimizing over all generalized quan-

tum measurements for distinguishing neighboring quantum states $\hat{\rho}$ and $\hat{\rho} + d\hat{\rho}$. It is also the infinitesimal form of a distance function introduced by Bures [14], which Uhlmann [15] interpreted as a generalization of transition probabilities to mixed states.

6. Quantum inference

Having specified the geometry, we now derive the generalization to the trajectories which described classical inference. Following the lead given in Eq. (7), the quantum trajectories are given by

$$\frac{d\hat{\rho}}{dt} = -\mathcal{R}_\rho(\Delta\hat{A}), \tag{10}$$

with $\hat{\rho}(0) = \hat{\rho}_0$ at $t = 0$. These trajectories describe paths that pass orthogonally through the family of surfaces, corresponding to some observable \hat{A} , as illustrated in Fig. 3. Using the trivial identity $\mathcal{R}_\rho(\hat{A}) = \mathcal{R}_{\hat{A}}(\hat{\rho})$ we derive the form of the trajectories as

$$\hat{\rho}(t) \propto \exp(-\hat{A}t/2)\hat{\rho}_0 \exp(-\hat{A}t/2). \tag{11}$$

This result may be thought of as a sort of symmetrized Boltzmann distribution at inverse temperature $\beta = t$ in the presence of prior information (i.e., with a bias towards $\hat{\rho}_0$). We know that these trajectories yield the optimal state with respect to $K_Q(\hat{\rho} | \hat{\rho}_0)$ for three cases: (i) in the classical case where $[\hat{A}, \hat{\rho}_0] = 0$; (ii) when the prior distribution is uniform $\hat{\rho}_0 \propto \hat{1}$; and (iii) when the posterior $\hat{\rho}$ is “near” the prior $\hat{\rho}_0$ [13].

Having found the quantum trajectories, we now consider the inverse problem of finding the family of surfaces \hat{A} whose trajectories connect the pair of

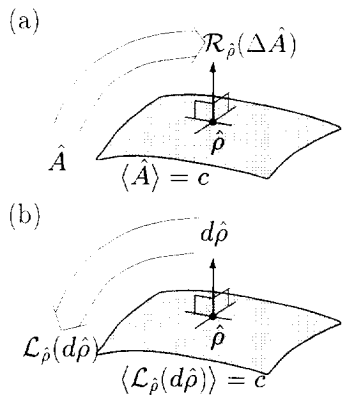


Fig. 2. Illustration of the geometric action of the raising and lowering operators. (a) The raising operator \mathcal{R}_ρ converts a family of surfaces of constant expectation value for an observable \hat{A} into the (tangent) vector $d\hat{\rho} = \mathcal{R}_\rho(\Delta\hat{A})$, orthogonal to this family at the point $\hat{\rho}$. (b) The lowering operator \mathcal{L}_ρ converts a vector $d\hat{\rho}$ at the point $\hat{\rho}$ into the family of surfaces $\hat{A} = \mathcal{L}_\rho(d\hat{\rho})$, orthogonal to this vector. (Note that for the surface shown here, $c = 0$.)

Table 1
Comparison of $\langle I_{\text{traj}} \rangle$ with $\langle I_{\text{acc}} \rangle$ (in nats) for various signal probabilities q . Pairs of 2×2 reduced density operators were calculated from pairs of randomly selected pure states in a D -dimensional Hilbert spaces

q	$D = 4$		$D = 6$		$D = 8$	
	$\langle I_{\text{traj}} \rangle$	$\langle I_{\text{acc}} \rangle$	$\langle I_{\text{traj}} \rangle$	$\langle I_{\text{acc}} \rangle$	$\langle I_{\text{traj}} \rangle$	$\langle I_{\text{acc}} \rangle$
0.01	0.00830	0.00868	0.00523	0.00531	0.00389	0.00392
0.1	0.0676	0.0691	0.0465	0.0469	0.0343	0.0344
0.2	0.1164	0.1178	0.0804	0.0807	0.0593	0.0594
0.5	0.1735	0.1742	0.1228	0.1229	0.0916	0.0916

states $\hat{\rho}_0$ and $\hat{\rho}_1$. This may be done by inspection, giving the observable

$$\hat{A} \equiv 2 \ln \left(\hat{\rho}_1^{-1/2} \sqrt{\hat{\rho}_1^{1/2} \hat{\rho}_0 \hat{\rho}_1^{1/2}} \hat{\rho}_1^{-1/2} \right). \quad (12)$$

This expression has been derived independently by Fuchs and Caves [16] in a different context. As an indirect test of our approach we will now use this observable as the basis for a detection scheme in communication employing non-orthogonal states.

7. Binary communication

A binary communication channel consists of a pair of signal states $\hat{\rho}_0$ and $\hat{\rho}_1$ which are sent with probabilities $1 - q$ and q respectively. Shannon's mutual information describes the mean information that can be transmitted down the channel per symbol; expressed in terms of the Kullback information it takes the form [17]

$$I = (1 - q)K(p_0 | p) + qK(p_1 | p), \quad (13)$$

where $p_0^j = \text{tr}(\hat{P}_j \hat{\rho}_0)$, $p_1^j = \text{tr}(\hat{P}_j \hat{\rho}_1)$ and $p^j = (1 - q)p_0^j + qp_1^j$ for basis $\{\hat{P}_j\}$. The eigenvalues of \hat{A} in Eq. (12) define a basis with which a detection scheme could in principle be constructed. We call the mutual information associated with this trajectory observable I_{traj} . The maximum mutual information (over choice of basis) is the maximum *accessible information* I_{acc} available and has been notoriously difficult to calculate [18], though it has been explicitly solved for in the case of 2×2 density operators [17]; we shall use this result as a bench mark for the performance of I_{traj} .

The numerical experiments we performed consisted of randomly selecting a pair of 2×2 density operators and calculating both I_{traj} and I_{acc} . The

subtle part of the calculations consisted of choosing a suitable measure for randomly sampling the density operators in order to obtain a meaningful average. Building on the work of uniform measures for pure states, a random state was chosen as the reduced density operator in a 2-dimensional subspace of a randomly selected pure state in a $2n$ -dimensional Hilbert space [19]. Table 1 shows the results for samplings made from a $D = 4, 6$ and 8 dimensional Hilbert space for various values of the signal probability q . We see that the mean $\langle I_{\text{traj}} \rangle$ is very close to its upper bound $\langle I_{\text{acc}} \rangle$ and indeed the relative discrepancy decreases as the dimensionality of the Hilbert space is increased.

The work reported here was based on techniques that are simple to manipulate. Their remarkable performance is encouraging for the future of quantum inference and suggests that they may be expanded to characterizing entanglement for mixed states and improving inference from non-trivial prior information. These methods can be applied to improve detection design in communication.

References

- [1] E.T. Jaynes, Phys. Rev. 106 (1957) 620; 108 (1957) 171; A. Shimony, Synthese 63 (1985) 35.
- [2] K. Friedman and A. Shimony, J. Stat. Phys. 3 (1973) 381; A. Shimony, in: Search for a naturalistic world view, Vol. 1 (Cambridge Univ. Press, Cambridge, 1993) pp. 162–179; T. Seidenfeld, Theory of Decision 11 (1979) 413.
- [3] J.E. Shore and R.W. Johnson, IEEE Trans. Inf. Theory IT-26 (1980) 26.
- [4] J. Skilling and S.F. Gull, in: Maximum-entropy and Bayesian methods in inverse problems, eds. C.R. Smith and W.T. Grandy Jr. (Reidel, Dordrecht, 1985), pp. 83–132.
- [5] S. Kullback, Information theory and statistics (Wiley, New York, 1959).

- [6] A. Hobson, *J. Stat. Phys.* 1 (1969) 383.
- [7] G. Lindblad, *Commun. Math. Phys.* 39 (1974) 111;
H. Araki, *Publ. Res. Inst. Math. Sci. (Kyoto)* 11 (1976) 809;
13 (1977) 173;
M.J. Donald, *Commun. Math. Phys.* 105 (1986) 13;
M.J. Donald, *Math. Proc. Cambridge Philos. Soc.* 101 (1987) 363; F. Hiai and D. Petz, *Commun. Math. Phys.* 143 (1991) 99;
M.J. Donald, *Found. Phys.* 22 (1992) 1111.
- [8] C.A. Fuchs, Ph.D. thesis, University of New Mexico, Albuquerque (1995).
- [9] A. Shimony, in: *Sixty years of EPR*, eds. A. Mann and M. Revzen (IOP, Bristol, 1996), to appear.
- [10] L.L. Campbell, *Inform. Sciences* 35 (1985) 199.
- [11] R.A. Fisher, *Proc. Cambridge Philos. Soc.* 22 (1925) 700.
- [12] S.L. Braunstein and C.M. Caves, in: *Fundamental problems in quantum theory*, ed. D. Greenberger (*Ann. NY Acad. Sci.*), to be published.
- [13] S.L. Braunstein and C.M. Caves, *Phys. Rev. Lett.* 72 (1994) 3439.
- [14] D.J.C. Bures, *Trans. Am. Math. Soc.* 135 (1969) 199.
- [15] A. Uhlmann, *Rep. Math. Phys.* 9 (1976) 273;
M. Hübner, *Phys. Lett. A* 163 (1992) 239; 179 (1993) 226;
R. Jozsa, *J. Mod. Opt.* 41 (1994) 2315.
- [16] C.A. Fuchs and C.M. Caves, *J. Open Syst. Inform. Dyn.*, to be published.
- [17] C.A. Fuchs and C.M. Caves, *Phys. Rev. Lett.* 73 (1994) 3047.
- [18] A.S. Holevo, *Probl. Inform. Transmi.* 9 (1973) 177;
E.B. Davies, *IEEE Trans. Inf. Theory* IT-24 (1978) 596.
- [19] S. Lloyd and H. Pagels, *Ann. Phys.* 188 (1988) 186–213;
K.R.W. Jones, *J. Phys. A* 23 (1990) L1247.