Convex probability domain of generalized quantum measurements

1998 J. Phys. A: Math. Gen. 31 L671

(http://iopscience.iop.org/0305-4470/31/38/003)

J. Phys. A: Math. Gen. **31** (1998) L671–L675. Printed in the UK PII: S0305-4470(98)94982-4

LETTER TO THE EDITOR

Convex probability domain of generalized quantum measurements

Asher Peres† and Daniel R Terno‡

Department of Physics, Technion—Israel Institute of Technology, 32000 Haifa, Israel

Received 9 June 1998, in final form 10 July 1998

Abstract. Generalized quantum measurements with N distinct outcomes are used for determining the density matrix, of order d, of an ensemble of quantum systems. The resulting probabilities are represented by a point in an N-dimensional space. It is shown that this point lies in a convex domain having at most $d^2 - 1$ dimensions.

In elementary quantum measurement theory, a test performed on a quantum system is represented by a complete set of orthogonal projection operators $\mathbf{P}_m$, where the label m takes at most d different values (d is the dimensionality of the Hilbert space, assumed finite). The probability of obtaining outcome m of that test, following the preparation of a quantum ensemble in a state ρ, is $p_m = \mathrm{tr}(\rho \mathbf{P}_m)$. If ρ is arbitrary, the only constraint on these probabilities is $\sum_m p_m = 1$.

It is well known that this type of test is not optimal if only a finite number of quantum systems can be observed. (As a concrete example, we receive five photons from a distant source, and we want a good estimate of their polarization. What is the best strategy?) In such a case, more information may be derived from a *positive operator valued measure* (POVM) [1, 2] with $N > d$ different outcomes. Such a POVM is a set of N positive matrices $\mathbf{A}_\mu$, which in general do not commute, but still satisfy $\sum_\mu \mathbf{A}_\mu = \mathbb{1}$, where $\mathbb{1}$ is the unit matrix in d dimensions, and μ is an arbitrary label running from 1 to N. If the quantum system is prepared in state ρ, the probability to get outcome μ is

$$p_\mu(\rho) \equiv p(\mathbf{A}_\mu|\rho) = \mathrm{tr}(\rho \mathbf{A}_\mu). \tag{1}$$

To each preparation ρ of the system, we thus associate N probabilities, $p_1, p_2, \ldots, p_N$. We refer to this set of positive numbers as a point $P(\rho)$ in probability space. The set of all output points is labelled by P. Density matrices form a convex set whose extreme points are pure states [3]. The linear relation (1) between input states and output probabilities implies that the set of points P is also convex:

$$P(\rho) \equiv P(x\rho_1 + (1-x)\rho_2) = x\,P(\rho_1) + (1-x)\,P(\rho_2). \tag{2}$$

Thus the shape of the hypersurface that bounds the domain of the points $P(\rho)$, for all possible preparations of the system, is determined by the outputs for the pure states only.

Obviously $\sum_\mu p_\mu = 1$, so that the points $P(\rho)$ lie on a hyperplane of dimension $(N-1)$. However, the results of generalized measurements are subject to stronger constraints (which

† E-mail address: peres@photon.technion.ac.il
‡ E-mail address: terno@physics.technion.ac.il

0305-4470/98/380671+05$19.50

may be important for the statistical analysis of experimental results). Let D be the number of linearly independent parameters in ρ (for a generic density matrix in a d-dimensional complex Hilbert space, $D = d^2 - 1$). The following proposition will now be proved: *If $N > (D + 1)$, the output of any POVM is confined to a D-dimensional subspace.*

Indeed, let us write the elements of a generic density matrix in terms of real (symmetric) and imaginary (antisymmetric) parts,

$$\rho_{mn} = \xi_{mn} + \mathrm{i}\eta_{mn}. \tag{3}$$

There are $d(d-1)/2$ independent elements η_{mn} and $(d+2)(d-1)/2$ independent ξ_{mn}, because of the condition $\mathrm{tr}\,\rho = 1$ which can be written

$$\xi_{dd} = 1 - \sum_{n=1}^{d-1} \xi_{nn}. \tag{4}$$

Likewise, the elements of each POVM matrix $\mathbf{A}_\mu$, of order d, can be written as $x^\mu_{mn} + \mathrm{i}y^\mu_{mn}$ in terms of d^2 real parameters. We thus obtain from equation (1),

$$p_\mu(\rho) = \sum_{m=1}^{d-1}(x^\mu_{mm} - x^\mu_{dd})\xi_{mm} + 2\sum_{m=1}^{d}\sum_{n>m}(x^\mu_{mn}\xi_{mn} + y^\mu_{mn}\eta_{mn}) + x^\mu_{dd}. \tag{5}$$

Thus $P(\rho)$ is obtained from ρ by an affine transformation [4]

$$\boldsymbol{p} = \mathbf{M}\boldsymbol{r} + \boldsymbol{c} \tag{6}$$

where $\boldsymbol{p}$ is a 'vector' consisting of any $N-1$ components p_μ (the remaining component is obtained from $\sum_\mu p_\mu = 1$). Likewise $\boldsymbol{r}$ is a vector of D linearly independent parameters of ρ. The matrix $\mathbf{M}$, with $N-1$ rows and D columns, depends only on the POVM used for the test; and $\boldsymbol{c}$ is a vector whose $N-1$ components are x^μ_{dd}, which also are parameters of the POVM. Explicitly, the μth row of $\mathbf{M}$, which is

$$(x^\mu_{11} - x^\mu_{dd}) \ldots (x^\mu_{d-1,d-1} - x^\mu_{dd}) \quad 2x^\mu_{12} \ldots 2x^\mu_{d-1,d} \quad 2y^\mu_{12} \ldots 2y^\mu_{d-1,d} \tag{7}$$

and

$$\boldsymbol{r}^T = (\xi_{11} \ldots \xi_{d-1,d-1} \quad \xi_{12} \ldots \xi_{d-1,d} \quad \eta_{12} \ldots \eta_{d-1,d}) \tag{8}$$

have $D = d^2 - 1$ real components.

If $N - 1 > D$, the rank of $\mathbf{M}$ is at most D, and any $D+1$ vectors $\tilde{\boldsymbol{p}} = \mathbf{M}\boldsymbol{r}$ are linearly dependent. A translation by the constant vector $\boldsymbol{c}$ in the $(N-1)$-dimensional vector space transforms a D-dimensional subspace into another D-dimensional subspace. Thus the output of any POVM on a system whose density matrix has D linearly independent parameters is confined to a D-dimensional subspace of the probability space.

Next, let us examine the shape of the surface that encloses the domain of $P(\rho)$. The set of density operators, and therefore the set of probabilities, are convex. The extreme points of these sets are the pure states, which are defined by $2(d-1)$ real parameters, and the probabilities corresponding to these pure states, respectively. Thus any interior point of the D-dimensional set P is a convex combination of the extreme points of that set, which lie on a $2(d-1)$-dimensional hypercurve.

Note that any density matrix ρ of rank d can be written as a convex combination of no more than d pure density matrices, corresponding to the eigenvectors of ρ. As a result, any interior point of P can be obtained from at most d extreme points. This result ought to be compared with Caratheodory's theorem [4], which states that any interior point of an arbitrary convex set of dimension D can be obtained as a convex combination of $D+1$ (or fewer) extreme points of that set. Here, $D + 1 = d^2$. The smaller number of extreme

points needed in the present case is due to the fact that density matrices are not an *arbitrary* convex set (they are positive and have unit trace).

As a simple example, consider the case of spin-$\frac{1}{2}$ systems. Their states can be described by means of a Bloch sphere. The pure states correspond to points on the surface of the sphere, and mixed states lie in its interior. With our notations, we have

$$rho = \begin{pmatrix} x_{11} & x_{12} + \mathrm{i}y_{12} \\ x_{12} - \mathrm{i}y_{12} & 1 - x_{11} \end{pmatrix} \tag{9}$$

where the three parameters are subject to the positivity condition

$$x_{11}(1 - x_{11}) - x_{12}^2 - y_{12}^2 \geqslant 0. \tag{10}$$

The transformation (6) is linear. Therefore the Bloch sphere is transformed into another quadratic surface, usually an ellipsoid. Exceptionally, if a POVM element has unit norm (so that the corresponding p_μ can be equal to 1, and then all the other p_μ vanish), we have a straight segment.

In particular, consider a POVM with four elements, $\mathbf{A}_\mu = (\mathbb{1} + \boldsymbol{a}_\mu \cdot \boldsymbol{\sigma})/4$, where the four unit vectors $\boldsymbol{a}_\mu$ form a regular tetrahedron in a real three-dimensional Euclidean space, and $\boldsymbol{\sigma}$ denotes the three Pauli matrices. Likewise, any state ρ can be written as $\rho = (\mathbb{1} + \boldsymbol{n} \cdot \boldsymbol{\sigma})/2$. We thus have

$$p_\mu = \mathrm{tr}(\rho \mathbf{A}_\mu) = (\mathbb{1} + \boldsymbol{a}_\mu \cdot \boldsymbol{n})/4 \tag{11}$$

whence

$$\sum_{\mu=1}^{4} (p_\mu - \tfrac{1}{4})^2 = \boldsymbol{n}^2/12. \tag{12}$$

The Bloch sphere is thus mapped into a three-dimensional sphere of radius $1/\sqrt{12}$, centred at $p_\mu = \frac{1}{4}$, and lying in the hyperplane $\sum_\mu p_\mu = 1$. If we want to parametrize that hyperplane with three of the p_μ, we substitute in the above equation $p_4 = 1 - p_1 - p_2 - p_3$. We then obtain an ellipsoid in a three-dimensional space, as shown in figure 1. It is also possible to use as coordinates suitable linear combinations of the p_μ, such as

$$\begin{aligned} x &= p_1 + p_2 - p_3 - p_4 \\ y &= p_1 - p_2 + p_3 - p_4 \\ z &= p_1 - p_2 - p_3 + p_4. \end{aligned} \tag{13}$$

The Bloch sphere is then mapped into a sphere $x^2 + y^2 + z^2 \leqslant \frac{1}{3}$.

The case of spin-1 systems is more complicated. A generic density matrix can be written in terms of its eigenstates as

$$\rho = \sum_{j=1}^{3} \lambda_j |v_j\rangle\langle v_j|. \tag{14}$$

This is as a convex combination of three extreme points. Any pure state, such as the above eigenstates, can be parametrized, with a suitable choice of its phase, as

$$|v\rangle = (\sin\theta \cos\phi \mathrm{e}^{\mathrm{i}\alpha}, \sin\theta \sin\phi \mathrm{e}^{\mathrm{i}\beta}, \cos\theta) \tag{15}$$

where

$$0 \leqslant \theta,\ \phi \leqslant \pi/2 \qquad \text{and} \qquad 0 \leqslant \alpha,\ \beta < 2\pi. \tag{16}$$

All the components of the corresponding pure ρ, which is a matrix of rank 1, are functions of the four parameters θ, ϕ, α, and β. Thus all the probabilities $p_\mu = \mathrm{tr}(\rho \mathbf{A}_\mu)$ are also

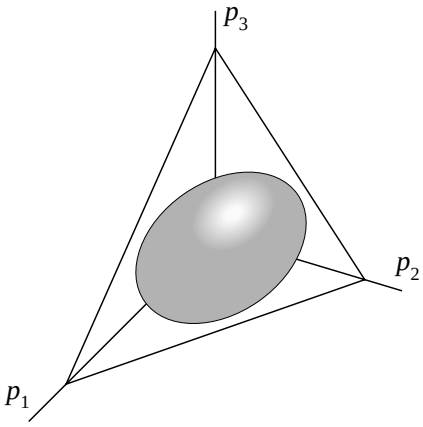

Figure 1. The probability ellipsoid that corresponds to the Bloch sphere is tangent to the plane $p_1 + p_2 + p_3 = 1$ (that is $p_4 = 0$) at the point $p_1 = p_2 = p_3 = \frac{1}{3}$ and likewise it is tangent to each coordinate plane at the point where the two coordinates in that plane are $\frac{1}{3}$.

functions of these four angles. This gives the extreme points of the set P: they form a four-parameter hypersurface in an eight-dimensional space S (which is itself embedded in the N-dimensional space of the p_μ). The rest of the boundary of S, corresponding to density matrices of rank two, lies on the segments between any pair of extreme points. The interior points of S can be obtained by a convex combination of three suitably chosen extreme points, as in equation (14). All these considerations are readily extended to quantum systems whose Hilbert spaces have more than three dimensions: there are $(d-1)$ polar angles like θ and ϕ, and $(d-1)$ phases like α and β.

Finally, let us consider potential applications of the above results to the analysis of experimental data. The probabilities p_μ cannot be measured exactly, as this would require testing an infinite number of quantum sytems. If only n systems are available, and the μth outcome is found to occur n_μ times (so that the experimenter records a set of N integers or zeros), then the N ratios $q_\mu = n_\mu/n$ are the only data available for evaluating the true p_μ. Obviously, $\sum_\mu q_\mu = 1$, just like $\sum_\mu p_\mu$, but the other constraints on p_μ may not be satisfied. In particular, if $N > D+1$, the point $Q = \{q_\mu\}$ will not in general lie in the hyperplane of dimension D to which the point P is restricted.

How far can Q be from the true P? Each one of the experimental data n_μ has an expected binomial distribution with dispersion

$$\Delta n_\mu = [n\, p_\mu(1-p_\mu)]^{1/2} \simeq [n_\mu(n-n_\mu)/n]^{1/2} \tag{17}$$

where the last expression is valid if $n_\mu \gg 1$. We can imagine an error box with sides equal to Δq_μ, centred at the point Q, and we then have to examine where that error box overlaps with the hyperplane to which P is constrained.

Obviously, it is best to design the experiment so as to have $N = D+1 = d^2$, and not more than that. A larger value of N leads to a less efficient use of the experimental data. This result is reminiscent of Davies's theorem [5] which deals with a related question, namely how to maximize the mutual information obtainable from a set of non-orthogonal signals. The theorem asserts that no more than d^2 outputs are needed.

If $N = D+1$, the only question is whether Q lies in the convex domain of P, namely whether the resulting ρ, obtained by solving equation (1), is a positive matrix. If it is, then Q is the best estimate for the true P. In the opposite case, the experimenter should either test a larger number of samples, or use maximum likelihood or similar nonlinear methods [6] to analyse the existing data.

DRT was supported by a grant from the Technion Graduate School. Work by AP was supported by the Gerard Swope Fund, and the Fund for Encouragement of Research.

References

[1] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[2] Peres A 1993 *Quantum Theory: Concepts and Methods* (Dordrecht: Kluwer)
[3] Davies E B 1976 *Quantum Theory of Open Systems* (New York: Academic)
[4] Kelly P J and Weiss M L 1979 *Geometry and Convexity* (New York: Wiley)
[5] Davies E B 1978 *IEEE Trans. Inf. Theory* **IT-24** 239
[6] Hradil Z 1997 *Phys. Rev.* A **55** 1561