Information and Quantum Measurement

E. B. DAVIES

Abstract—**Given a finite number of quantum states with *a priori* probabilities, the positive operator-valued measure that maximizes the Shannon mutual information is investigated. The group covariant case is examined in detail.**

I. Introduction

In THE QUANTUM detection problem, the input is defined to be a set $\rho_1,\cdots,\rho_m$ of states (positive trace class operators of trace one) on a Hilbert space $\mathcal{H}$, together with *a priori* probabilities $\lambda_1,\cdots,\lambda_m$ such that $\lambda_i \geqslant 0$ and $\Sigma\lambda_i = 1$. The output is defined to be a positive operator-valued measure $A_1,\cdots,A_n$ that is a set of nonzero positive self-adjoint operators A_i on $\mathcal{H}$ such that

$$A_1+\cdots+A_n=1. \tag{1.1}$$

The input and output together determine an $m\times n$ matrix

$$P_{ij}=\lambda_i \operatorname{tr}\left[\rho_i A_j\right] \tag{1.2}$$

called a probability matrix since $P_{ij}\geqslant 0$ and $\Sigma P_{ij}=1$. This matrix gives the joint distribution of input and output and is used to calculate their mutual information [2, p. 50], [12, p. 26], [13, p. 38] according to the formula

$$I(P)=\sum_{ij} H(P_{ij})-\sum_i H\Big(\sum_j P_{ij}\Big)-\sum_j H\Big(\sum_i P_{ij}\Big) \tag{1.3}$$

where

$$H(\alpha)=-\alpha\log\alpha. \tag{1.4}$$

The information-theoretic optimum detection problem consists of determining the output that maximizes the mutual information for given input.

Because of the nonlinear nature of H, the above problem is harder than one solved in [6], [7], [9], [10], [15] where one minimizes

$$C(P)=\sum_{ij} C_{ij}P_{ij}$$

for a specified Bayes cost matrix C_{ij}. The two problems generally have different solutions and are relevant in different contexts [11], [15]. The Bayes cost problem applies when one has to make a decision after a single measurement, while the information-theoretic problem is more relevant when one has a sequence of measurements with coding and decoding, as in communication systems [2], [13]. In the information-theoretic problem, the parameter n is free to vary and must be optimized along with the operators $A_1,\cdots,A_n$, while in the Bayes cost problem, n is often fixed.

In this paper we report some progress towards the solution of the information-theoretic optimum detection problem that reduces the problem to computable proportions, particularly in the group covariant case. As an illustration of the methods, we compute the maximum mutual information for four equiprobable input states with tetrahedral symmetry on a two-dimensional Hilbert space. While many of the results and methods apply to infinite-dimensional spaces with infinite, even continuous, input and output values, we restrict attention to the finite case where sharper results are available.

II. Constraints on the Optimal Measurement

We shall need some elementary lemmas, whose proofs we omit.

Lemma 1: Let X be the convex set of $m\times n$ probability matrices P with fixed row sums

$$\lambda_i=\sum_j P_{ij}.$$

Then $I(P)$ is a continuous convex function on X.

Lemma 2: Let P be a $m\times n$ probability matrix and P' be the $m\times(n-1)$ probability matrix obtained by replacing two columns by their column sum. Then

$$I(P')\leqslant I(P)$$

with equality in case the chosen columns are proportional.

Theorem 3: Let $\mathcal{H}$ have finite dimension d, and let an input be given. Then the mutual information is maximized for an output $A_1,\cdots,A_n$ such that each operator A_i has rank one, that is

$$A_i=|\psi_i\rangle\langle\psi_i| \tag{2.1}$$

for some vector ψ_i with $\|\psi_i\|\leqslant 1$, where $|v\rangle$ and $\langle v|$ denote the vector v in Dirac's bra-ket notation. Moreover the constant n can be made to satisfy

$$d\leqslant n\leqslant d^2. \tag{2.2}$$

Proof: For any output $A_1,\cdots,A_n$, we may write

$$A_i=\sum_{r=1}^{d}\lambda_{ir}|\psi_{ir}\rangle\langle\psi_{ir}|$$

where $\|\psi_{ir}\|=1$ and $\lambda_{i1},\cdots,\lambda_{id}$ are the eigenvalues of A.

Manuscript received October 1, 1976; revised February 16, 1978.

The author is with the Mathematical Institute, Oxford University, Oxford, England.

0018-9448/78/0900-0596$00.75 ©1978 IEEE

The new output

$$B_{ir}=\lambda_{ir}|\psi_{ir}\rangle\langle\psi_{ir}|$$

has mutual information no less than the old output by Lemma 2. Also by Lemma 2, we may suppose no two of the ψ_{ir} are proportional. Changing notation we need therefore only consider outputs of the form

$$B_i=\mu_i d|\psi_i\rangle\langle\psi_i|$$

where ψ_i are all different, $\|\psi_i\|=1$, and $\mu_i\geqslant 0$. By taking the trace of the equation $\Sigma\, B_i=1$, we see that μ_i must also satisfy

$$\Sigma\,\mu_i=1.$$

If X is the compact convex set of operators

$$X=\{A\geqslant 0:\mathrm{tr}\,[A]=1\},$$

then the extreme points of X are the operators of the form $|\psi\rangle\langle\psi|$ with $\|\psi\|=1$. Therefore the outputs we are considering are in one–one correspondence with those probability measures (of finite support) on the set ∂X of extreme points whose barycenters are the point $d^{-1}1\in X$ in the sense of [1]. Such probability measures themselves form a compact convex set Y. By Lemma 1 the mutual information is a convex function Y and, therefore, takes its maximum value at an extreme point of Y. By a slight modification of the proof of Caratheodory's theorem in [1], [5] every extreme point of Y is a probability measure whose support has $\leqslant(1+\dim X)$ points. Since

$$\dim X=d^2-1,$$

it follows that we may take $n\leqslant d^2$.

If $n<d$, then there exists a nonzero vector ψ such that $\langle\psi,\psi_i\rangle=0$ for all i. Then

$$\begin{aligned}\langle\psi,\psi\rangle&=\sum_i\langle B_i\psi,\psi\rangle\\&=\sum_i\mu_i d|\langle\psi,\psi_i\rangle|^2\\&=0.\end{aligned}$$

The contradiction implies that $n\geqslant d$.

III. The Group Covariant Case

It is known [4, p. 25], [14, p. 173] that every affine automorphism α of the convex set S of all states on $\mathcal{H}$ is representable in the form

$$\alpha(\rho)=U\rho U^*$$

where U is a unitary or anti-unitary operator on $\mathcal{H}$ and U^* is its adjoint. We shall not, however, use this fact, preferring to regard a representation of a group G as a homomorphism π from G to the affine automorphisms of S. (At the Hilbert space level, one would have to consider projective unitary–antiunitary representations.) A representation of G is called irreducible if the only G-invariant point of S is $d^{-1}1$ where d is the dimension of $\mathcal{H}$.

We say that the input is group covariant if there is a finite group G and a representation π of G on S such that the states ρ_g are parametrized by $g\in G$ and obey

$$\pi_g\rho_h=\rho_{gh}$$

for all $g,h\in G$. We also demand that the *a priori* probabilities all equal $|G|^{-1}$ where $|G|$ is the number of elements of G.

A covariant output is defined as a collection of operators $A_g\geqslant 0$ on $\mathcal{H}$ parametrized by $g\in G$ and satisfying

$$\pi_g^*A_h=A_{gh}$$

for all g, $h\in G$, where π^* is the representation of G on $\mathcal{L}(\mathcal{H})$ dual to π. We also require

$$\sum_{g\in G}A_g=1.$$

See [3], [4], [9] for investigations of this notion.

Theorem 4: If the input is covariant with respect to a group G that has an irreducible representation π on S, then there exists a unit vector $\psi\in\mathcal{H}$ such that the mutual information is maximized by the covariant output

$$A_g=|G|^{-1}d\pi_g^*\{|\psi\rangle\langle\psi|\}. \tag{3.1}$$

Comment: If $I(P)$ had been concave instead of convex as shown in Lemma 1, we could have deduced this result from [9]. We use instead a special lemma.

Lemma 5: Let $P_1,\cdots,P_r$ be $m\times n$ probability matrices whose row sums are all equal to m^{-1}, and suppose that each P_i is obtained from P_1 by some permutation of the rows and columns. If Q is the $m\times rn$ probability matrix

$$Q=r^{-1}[P_1,P_2,\cdots,P_r], \tag{3.2}$$

then

$$I(Q)=I(P_1)=\cdots=I(P_r). \tag{3.3}$$

Proof: This is a straightforward computation.

Proof of Theorem: For any output $\{A_1,\cdots,A_n\}$, we write $I(A)$ for the mutual information of the probability matrix P associated with A according to (1.2). If

$$B_{ig}=|G|^{-1}\pi_g^*(A_i),$$

then B is an output whose probability matrix is obtained from that of A according to Lemma 5. Therefore

$$I(B)=I(A).$$

Since we always suppose that the operators A_i are positive and nonzero, tr $[A_i]\neq 0$, and we can define

$$C_g^i=|G|^{-1}d\pi_g^*(A_i)/\mathrm{tr}\,[A_i].$$

Now $\Sigma_g C_g^i$ has trace d and is group invariant, so by irreducibility

$$\sum_g C_g^i=1.$$

That is C^i is a covariant output for each i. Since B is obtained from $C^1,\cdots,C^n$ by averaging and the information is a convex function, it follows that

$$I(B) \leq \max_i I(C^i).$$

If the maximum value β of the mutual information is achieved for the output A, then

$$\beta = I(A) \leq \max_i I(C^i) \leq \beta$$

so the maximum is also achieved for one of the covariant outputs.

For computational purposes Theorem 5 needs to be supplemented by the following lemma.

Lemma 6: For a covariant input and the covariant output of the form of (3.1), the mutual information is

$$I = \log d + |G|^{-1} d \sum_g \langle \rho_g \psi, \psi \rangle \log \langle \rho_g \psi, \psi \rangle. \tag{3.4}$$

Proof: The probability matrix is

$$P(g,h) = |G|^{-2} d \operatorname{tr} \left[\rho_g \pi_h^* \{ |\psi\rangle\langle\psi| \} \right] = |G|^{-2} d \operatorname{tr} \left[\rho_{h^{-1}g} \psi, \psi \rangle \right]$$

from which it follows that

$$\sum_g P(g,h) = \sum_h P(g,h) = |G|^{-1}.$$

The result follows by substitution into (1.3).

Since one may have $|G| > d^2$, the outputs determined by Theorem 3 and 4 need not coincide, although both maximize the mutual information. In a situation where several different outputs yield the same mutual information, one might attempt to choose between them according to the complexity of the coding–decoding procedures required to approach the theoretically attainable channel capacity [2], [13]. We content ourselves here, however, with investigating to what extent the maximum is unique in the simplest possible case where $d = 2$.

If $d = 2$ or $\mathcal{H} = \mathbb{C}^2$, then by use of the Pauli spin matrices it may be seen that S is isomorphic to the unit ball in Euclidean three-space $\mathbb{R}^3$. The full symmetry group $O(3)$ of S contains the reflection

$$R(x) = -x$$

with determinant -1, which is induced by the anti-unitary map

$$R_0(\alpha, \beta) = (\bar{\beta}, -\bar{\alpha})$$

on $\mathbb{C}^2$.

Theorem 7: If the input is covariant with respect to a group G with $R \in \pi(G)$, then there exist orthogonal unit vectors ψ_1, $\psi_2 \in \mathbb{C}^2$ such that the mutual information is maximized by the projection-valued measure

$$A_1 = |\psi_1\rangle\langle\psi_1|, \qquad A_2 = |\psi_2\rangle\langle\psi_2|.$$

Proof: Let $\psi \in \mathcal{H}$ be a unit vector of the type specified in Theorem 4. Since $R \in \pi G$, the operators A_g may be broken up into pairs at opposite ends of a diameter. Each such pair determines two orthogonal unit vectors and hence a projection-valued measure. The output A is therefore an average of projection-valued measure outputs P^i. By convexity

$$I(A) \leq \max_i I(P_i)$$

so one of the $I(P_i)$ also equals the maximum mutual information.

We note that if the states ρ_i lie on the vertices of a regular cube, octahedron, dodecahedron, or icosahedron in $\mathbb{R}^3$ with center at the origin, then the corresponding symmetry group G does have $R \in \pi(G)$. This is not the case for the tetrahedral group, however.

Proposition 8: If the input consists of four pure states with tetrahedral symmetry on $\mathcal{H} = \mathbb{C}^2$, then the mutual information I is not maximized by any projection-valued measure. It is maximized by taking a covariant output of the form of Lemma 5 where the vector ψ must be orthogonal to one of the input states, and the maximum value is

$$I = \log(4/3).$$

Proof: We put

$$\psi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \psi_2 = \begin{pmatrix} (1/3)^{1/2} \\ (2/3)^{1/2} \end{pmatrix}$$

$$\psi_3 = \begin{pmatrix} (1/3)^{1/2} \\ (2/3)^{1/2} e^{2\pi i/3} \end{pmatrix} \quad \psi_4 = \begin{pmatrix} (1/3)^{1/2} \\ (2/3)^{1/2} e^{4\pi i/3} \end{pmatrix}$$

so that

$$|\langle \psi_i, \psi_j \rangle|^2 = \begin{cases} 1 & \text{if } i = j \\ 1/3 & \text{if } i \neq j. \end{cases}$$

The input states are taken to be

$$\rho_i = |\psi_i\rangle\langle\psi_i|$$

with *a priori* probabilities 1/4, and they form the vertices of a regular tetrahedron in S. We index by vertices of the tetrahedron rather than by elements of the tetrahedral group since this is easier and gives the same answers.

The projection-valued measures in $\mathcal{H}$ form a two-parameter family

$$P_1 = |\xi_1\rangle\langle\xi_1| \qquad P_2 = |\xi_2\rangle\langle\xi_2|$$

where

$$\xi_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \, e^{i\varphi} \end{pmatrix} \qquad \xi_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \, e^{i\varphi} \end{pmatrix}.$$

The mutual information can be represented as a two-parameter function

$$I_1(\theta, \varphi) = I(P)$$

where

$$P_{ij} = \frac{1}{4} |\langle \psi_i, \xi_j \rangle|^2.$$

The function $I_1(\theta, \varphi)$ may be compared with the function $I_2(\theta, \varphi)$ for covariant outputs calculated from Lemma 6 by putting

$$\psi=\begin{pmatrix}\cos\theta\\ \sin\theta\, e^{i\varphi}\end{pmatrix}.$$

Computer calculations show that the maximum of I_1 is strictly less than the maximum value of I_2, which is attained for the stated values of ψ, one of these being

$$\psi=\begin{pmatrix}0\\ 1\end{pmatrix}.$$

The quantities $\langle\rho_i\psi,\psi\rangle$ in (3.4) are

$$\langle\rho_i\psi,\psi\rangle=|\langle\psi_i,\psi\rangle|^2=0,\frac{2}{3},\frac{2}{3},\frac{2}{3}$$

so the maximum mutual information I is

$$I=\log 2+\frac{1}{4}\cdot 2\cdot 3\cdot\frac{2}{3}\log\frac{2}{3}=\log\frac{4}{3}.$$

IV. Conclusion

We have shown how to compute the maximum mutual information for given input at least in the group covariant case. If the input may also be varied to maximize the mutual information, then one has quite a different problem. If $\psi_1,\cdots,\psi_d$ is an orthonormal basis of $\mathcal{H}$ and we put

$$\rho_i=|\psi_i\rangle\langle\psi_i|$$

with equal *a priori* probabilities d^{-1} and if we take the output to be the projection-valued measure

$$A_j=|\psi_j\rangle\langle\psi_j|,$$

then a simple calculation shows that the mutual information is

$$I=\log d.$$

The fact that this is the maximum possible value of the mutual information may be deduced from [8, (1)].

Acknowledgment

The author would like to thank M. Gerzon for many stimulating conversations, P. Palmer for carrying out the computations mentioned above, and A. S. Holevo for drawing his attention to his paper [8].

References

[1] E. M. Alfsen, *Compact Convex Sets and Boundary Integrals*. New York: Springer-Verlag, 1971.

[2] R. Ash, *Information Theory*. New York: Interscience, 1965.

[3] E. B. Davies, "On the repeated measurement of continuous observables in quantum mechanics," *J. Funct. Anal.*, vol. 6, 318–346, 1970.

[4] ——, *Quantum Theory of Open Systems*. New York: Academic, 1976.

[5] B. Grunbaum, *Convex Polytopes*. New York: Interscience, 1967.

[6] C. W. Helstrom and R. S. Kennedy, "Noncommuting observables in quantum detection and estimation theory," *IEEE Trans. Inform. Theory*, vol. 20, 16–24, 1974.

[7] C. W. Helstrom, J. W. S. Liu, and J. P. Gordon, "Quantum mechanical communication theory," *Proc. IEEE*, vol. 58, 1578–1598, 1970.

[8] A. S. Holevo, "Some estimates of the quantity of information broadcast in a quantum communication channel" (Russian), *Prob. Transmission of Inform.*, vol. 9, 3–11, 1973.

[9] ——, "Statistical decision theory for quantum systems," *J. Multivariate Anal.*, vol. 3, 337–394, 1973.

[10] J. W. S. Liu, "Reliability of quantum mechanical communication systems," *IEEE Trans. Inform. Theory*, vol. 16, 319–329, 1970.

[11] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960.

[12] D. B. Osteyee and I. J. Good, *Information, Weight of Evidence, the Singularity Between Probability Measures, and Signal Detection*. Lecture Notes in Math. 376, New York: Springer-Verlag, 1974.

[13] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. of Illinois, 1949.

[14] V. S. Varadarajan, *Geometry of Quantum Theory*, vol. 1. Princeton, NJ: Van Nostrand, 1968.

[15] P. M. Woodward, *Probability and Information Theory, with Applications to Radar*, 2nd ed. New York: Pergamon, 1964.

[16] H. P. Yuen, R. S. Kennedy, and M. Lax, "Optimal testing of multiple hypotheses in quantum detection theory," *IEEE Trans. Inform. Theory*, vol. 21, 125–134, 1975.