Towards a geometrical interpretation of quantum-information compression

Graeme Mitchison[1] and Richard Jozsa[2]
[1]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom*
[2]*Department of Computer Science, University of Bristol, Merchant Venturers Building, Bristol BS8 1UB, United Kingdom*
(Received 30 October 2003; published 9 March 2004)

Let S be the von Neumann entropy of a finite ensemble $\mathcal{E}$ of pure quantum states. We show that S may be naturally viewed as a function of a set of geometrical volumes in Hilbert space defined by the states and that S is monotonically increasing in each of these variables. Since S is the Schumacher compression limit of $\mathcal{E}$, this monotonicity property suggests a geometrical interpretation of the quantum redundancy involved in the compression process. It provides clarification of previous work in which it was shown that S may be increased while increasing the overlap of each pair of states in the ensemble. As a by-product, our mathematical techniques also provide an interpretation of the subentropy of $\mathcal{E}$.

DOI: 10.1103/PhysRevA.69.032304 PACS number(s): 03.67.−a

I. INTRODUCTION

One of the most satisfying results of quantum-information theory is Schumacher's source coding theorem [1–5], which says that a message of length K from a source of pure quantum states with density matrix ρ can be compressed to $KS(\rho)$ qubits (asymptotically for large K), where $S(\rho)$ is the von Neumann entropy

$$S(\rho) = -\operatorname{Tr}(\rho \ln \rho).$$

Conceptually we can associate the possibility of compression with the presence of a degree of redundancy in the source. Suppose ρ_1,ρ_2 are the density matrices of two sources and

$$S(\rho_1) \leqslant S(\rho_2). \tag{1}$$

Then the first source can be compressed further than the second, so it has greater redundancy. If the two sources both have k states $|\psi_i\rangle$ and $|\phi_i\rangle$, say, with the same probabilities p_i, then this increased redundancy does not lie in the classical probabilities p_i of emitting the states but in the properties of the states themselves: it is true "quantum redundancy." Intuitively, one expects that, if the states $|\psi_i\rangle$ are more similar to each other than the states $|\phi_i\rangle$, then the quantum redundancy of the $|\psi_i\rangle$ will be greater. So if the pairwise overlaps of the $|\psi_i\rangle$ are larger than those of the $|\phi_i\rangle$, i.e.,

$$|\langle\psi_i|\psi_j\rangle| \geqslant |\langle\phi_i|\phi_j\rangle| \quad \text{for all } i,j, \tag{2}$$

then ρ_1 should have more quantum redundancy than ρ_2, and consequently Eq. (1) should hold. Jozsa and Schlienz [6] showed this is indeed true for a source with two states. However, they produced a counterexample consisting of a set of three states $|\phi_i\rangle$ in three dimensions and a slight perturbation of them $|\psi_i\rangle$ that has greater overlaps but *larger* entropy. This phenomenon raises the question of whether compression and quantum redundancy can be understood in *geometrical* terms, i.e., in terms of the geometry of the source's states in Hilbert space. This question has also been recently raised in Ref. [7]. In this paper we will establish a connection between quantum redundancy and the volumes in Hilbert space defined by the source states.

One reason the above example appears paradoxical is that, in real three-dimensional geometry, the pairwise inner products of three unit vectors determine the figure the vectors make, up to an orthogonal transformation, and they also determine the entropy in this real-valued setting. An analogous result holds in complex geometry if the complex inner product is specified, but not if only its absolute value, the overlap, is given. In fact, taking into account the freedom in specifying phase for a quantum mechanical state, it turns out that there are *four* real-valued degrees of freedom in specifying three states in three complex dimensions up to a unitary transformation (as will be shown shortly). Specifying the three pairwise overlaps $|\langle\psi_1|\psi_2\rangle|$, $|\langle\psi_1|\psi_3\rangle|$, and $|\langle\psi_2|\psi_3\rangle|$ therefore leaves a further degree of freedom that can be used to adjust the entropy.

What is this extra degree of freedom for three states in three (complex) dimensions? Here it is useful to introduce the Gram matrix G, with entries $G_{ij}=\sqrt{p_ip_j}\langle\psi_i|\psi_j\rangle$ and the matrix A with entries $a_{ij}=\langle\psi_i|\psi_j\rangle$. G has the same eigenvalues as the density matrix $\rho=\Sigma p_i|\psi_i\rangle\langle\psi_i|$ (see Ref. [6]) so the von Neumann entropy

$$S(\rho) = -\sum x_i \ln x_i \tag{3}$$

can be computed from the eigenvalues x_i of the Gram matrix

$$G = \begin{pmatrix} p_1a_{11} & \sqrt{p_1p_2}a_{12} & \sqrt{p_1p_3}a_{13} \\ \sqrt{p_2p_1}a_{21} & p_2a_{22} & \sqrt{p_2p_3}a_{23} \\ \sqrt{p_3p_1}a_{31} & \sqrt{p_3p_2}a_{32} & p_3a_{33} \end{pmatrix}.$$

G has the characteristic equation

$$x^3 - s_1x^2 + s_2x - s_3 = 0, \tag{4}$$

where the s_i are the symmetric polynomial functions of the eigenvalues, with $s_1=\Sigma x_i=1$, and $s_2=\Sigma_{i<j}x_ix_j$, and $s_3=x_1x_2x_3$. Expanding $\det(G-Ix)$ one finds that

$$s_2 = \sum_{i<j} p_ip_j(1-|a_{ij}|^2) \quad \text{and} \quad s_3 = p_1p_2p_3 \det A. \tag{5}$$

Since A can be written as $A=BB^\dagger$, where B is the matrix

$$\begin{pmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{pmatrix}$$

of coordinates ψ_{ij} of the states $|\psi_i\rangle$ in any orthonormal basis, it follows that $\det A = |\det B|^2$, and we can regard $\det A$ as the squared modulus of the *complex volume* spanned by the $|\psi_i\rangle$. Note also that $(1-|a_{ij}|^2)=\det A_{i,j}$, where $A_{i,j}$ is the submatrix of A obtained by striking out rows and columns having labels not in the set $\{i,j\}$. Thus we can likewise regard $1-|a_{ij}|^2$ as the squared modulus of the complex volume spanned by $|\psi_i\rangle$ and $|\psi_j\rangle$. We can write the terms of Eq. (5) that depend on the states as

$$\alpha_{12} = \det A_{1,2}, \quad \alpha_{13} = \det A_{1,3}, \qquad \alpha_{23} = \det A_{2,3}, \quad \alpha_{123} = \det A, \tag{6}$$

and s_2 and s_3 then appear as positive linear combinations of these volume variables. Thus $S(\rho)$ is determined by the probabilities p_i and the four squared volumes in Eq. (6). If the states are perturbed in such a way that three of these four parameters are fixed but one of them varies, then the entropy changes as intuition would dictate: increasing A_{12} (i.e., decreasing the overlap between states 1 and 2) increases the entropy (decreases the redundancy), and increasing the volume (i.e., spreading the states apart) also increases the entropy. In this sense we can regard the four parameters $\alpha_{12},\alpha_{13},\alpha_{23},\alpha_{123}$ as measures of quantum redundancy for the set of states. This behavior of the entropy follows from a theorem proved later, which tells us that $\partial S/\partial s_i > 0$ for $i \geqslant 2$. Consequently when none of the p_i is zero (which we shall assume throughout), $\partial S/\partial\alpha_{ij} = p_i p_j \partial S/\partial s_2 > 0$ and $\partial S/\partial\alpha_{123} = p_1 p_2 p_3 \partial S/\partial s_3 > 0$.

II. *k* STATES IN *n* DIMENSIONS

We now investigate the situation for any number k of states which span a space of n dimensions (so $k \geqslant n$). We can try to imitate in n dimensions the procedure that gave the variables in Eq. (6). Let the states be $|\psi_i\rangle$ with probabilities p_i for $i=1,\ldots,k$. As above we introduce the Gram matrix with entries $G_{ij}=\sqrt{p_i p_j}\langle\psi_i|\psi_j\rangle$ and the A matrix with entries $a_{ij} = \langle\psi_i|\psi_j\rangle$, both being $k\times k$ matrices. The eigenvalues of the Gram matrix are those of the $n\times n$ density matrix ρ padded out with $k-n$ zeros [6]. Thus the characteristic equation $\det(G-xI)=0$ for the Gram matrix for k states in n dimensions has the form

$$(-1)^{k-n}x^{k-n}\sum_{i=0}^{n}(-1)^i s_{n-i}x^i = 0, \tag{7}$$

where s_i is the ith elementary symmetric polynomial of the eigenvalues $x_1,\ldots,x_n$ of the density matrix, defined by

$$s_0 = 1, \quad s_i = \sum_{u_1<\cdots<u_i} x_{u_1}\cdots x_{u_i} \quad \text{for } i=1,\ldots,n. \tag{8}$$

The ith symmetric polynomial may be expressed as

$$s_i = \sum_{u_1<\cdots<u_i} p_{u_1}\cdots p_{u_i}\alpha_{u_1,\ldots,u_i}, \tag{9}$$

where $\alpha_{u_1,\ldots,u_i}=\det A_{u_1,\ldots,u_i}$ and $A_{u_1,\ldots,u_i}$ is the submatrix of the matrix A obtained by striking out all rows and columns with labels not in the set $\{u_1,\ldots,u_i\}$; i.e., $A_{u_1,\ldots,u_i}$ is the $i\times i$ A matrix constructed from the subset $\{|\psi_{u_1}\rangle,\ldots,|\psi_{u_i}\rangle\}$ of the states. This follows because s_i is the coefficient of $(-x)^{k-i}$ in $\det(G-xI)$, which is obtained by picking $k-i$ elements on the diagonal, corresponding to rows $v_1,\ldots,v_{k-i}$ say, and, for each such choice of v's, constructing the determinant of the submatrix $G_{u_1,\ldots,u_i}$ of G, where $u_1,\ldots,u_i$ is the complementary set to the v's. Since the Gram matrix G is related to the corresponding A matrix by $G=QAQ$, where $Q=\mathrm{diag}(\sqrt{p_1},\ldots,\sqrt{p_k})$, we get $\det G_{u_1,\ldots,u_i}=p_{u_1}\cdots p_{u_i}\det A_{u_1,\ldots,u_i}$. Equation (9) is then obtained by summing over all sets of u's.

Note that, though the individual entries in $A_{u_1,\ldots,u_i}$ are dependent on a choice of phase for the states $|\psi_i\rangle$, $\alpha_{u_1,\ldots,u_i}$ is invariant under phase choices. Furthermore, $\alpha_{u_1,\ldots,u_i}$ is real, since $A_{u_1,\ldots,u_i}$ is Hermitian. Thus $\alpha_{u_1,\ldots,u_i}$ is a real-valued unitary invariant, and the complete set of all $\alpha_{u_1,\ldots,u_i}$ for all sets $u_1,\ldots,u_i$ can be regarded as the analogues of the invariants (6) in the case $k=n=3$. Also, as in the case of $k=n=3$, $A_{u_1,\ldots,u_i}=B_{u_1,\ldots,u_i}B^\dagger_{u_1,\ldots,u_i}$, where $B_{u_1,\ldots,u_i}$ is the $i\times i$ matrix whose rows are the components of the i states $|\psi_{u_1}\rangle,\ldots,|\psi_{u_i}\rangle$ (expanded in any choice of orthonormal basis in the span of these i states). Thus $\alpha_{u_1,\ldots,u_i}$ may be identified as the squared modulus of the complex volume determined by $|\psi_{u_1}\rangle,\ldots,|\psi_{u_i}\rangle$.

Let $x_1,\ldots,x_n$ be any probability distribution and let $s_i = s_i(x_1,\ldots,x_n)$ for $i=1,\ldots,n$ be the corresponding symmetric polynomials. For any symmetric function $f(x_1,\ldots,x_n)$ (e.g., the entropy S) we consider a change of variables from the x_i's to the s_j's. Note that the probability condition $\Sigma x_i = 1$ corresponds to $s_1=1$, and lifting this condition we get n variables in each case. Then the Jacobian is readily seen to be $\Pi_{i<j}(x_i-x_j)$, so the change of variables is valid if the x_i's are all different. For simplicity we will work within this restriction but expect that our results will have suitable (finite) limiting behavior for coincident values $x_i \to x_j$. Furthermore we will be interested primarily in partial derivatives $\partial f/\partial s_i$ for $i\geqslant 2$ (which have s_1 held constant), so our results will also remain valid if we impose the probability constraint $s_1 = 1 = $constant at the start. We have the following fundamental property of the entropy.

Theorem 1. If $S=-\Sigma x_i \ln x_i$ is viewed as a function of the symmetric polynomials $s_1,\ldots,s_n$ then $\partial S/\partial s_q > 0$ for $q = 2,\ldots,n$.

Two proofs of this theorem are given in Sec. V below after a discussion of some of its fundamental consequences.

III. GEOMETRICAL INTERPRETATION OF QUANTUM REDUNDANCY

Equation (9) gives an expression for the symmetric polynomials s_i that is canonically determined by the state set $\{|\psi_1\rangle,\ldots,|\psi_k\rangle\}$ and probabilities $p_1,\ldots,p_k$. Thus from

$S=S(s_1,\ldots,s_n)$ (with $s_1=1$) we can view the von Neumann entropy of the source in a natural way as being a function of the probabilities and all the squared volumes $\alpha_{i_1 i_2},\ldots,\alpha_{i_1 i_2\ldots i_k}$. By theorem 1, $\partial S/\partial s_q>0$ for $q\geqslant 2$ and by Eq. (9), each s_q is a positive linear combination of the α variables, so we conclude that $\partial S/\partial\alpha>0$ for each squared volume variable α.

This suggests a geometrical interpretation of the quantum redundancy in the ensemble of quantum states. If the probabilities are held fixed and the states are deformed then the change in entropy can be seen as an accumulation of monotonic effects arising from the changes induced in each of the squared volumes α. Since $\partial S/\partial\alpha>0$ the set of these squared volumes can be regarded as a geometric measure of the quantum redundancy associated to a set of states alone.

Note, however, that, for $k>3$, there are more α's than degrees of freedom needed to fix k states up to overall unitary equivalence. Indeed, let $v(k,n)$ denote the number of degrees of freedom in specifying k states in n dimensions up to unitary transformation. To specify k states requires $k(2n-2)$ real parameters (as each state is defined only up to overall phase). The unitary group $U(n)$ has n^2 parameters, but because of the overall phase freedom in each state, U and $e^{ix}U$ have the same action for any x. Thus, unitary action on the states eliminates n^2-1 parameters from the $k(2n-2)$, giving $v(k,n)=k(2n-2)-(n^2-1)$.

The following table shows $v(k,n)$ for small values of k and n and $k\geqslant n$. The bracketed numbers are the total number $\tau(k,n)$ of terms $\alpha_{u_1,\ldots,u_i}$, ignoring those with $i>n$, which are zero since more than n states must be linearly dependent and therefore have zero determinant. The numbers in brackets are therefore $\tau(k,n)=\sum_{i=2}^{n}\binom{k}{i}$.

k	$n=2$	$n=3$	$n=4$	$n=5$
2	1 (1)			
3	3 (3)	4 (4)		
4	5 (6)	8 (10)	9 (11)	
5	7 (10)	12 (20)	15 (25)	16 (26)

For $k\leqslant 3$ the two sets of numbers agree, so the α's can be used to parametrize the sets of states. For $k>3$ there are always too many α's. Thus viewing the entropy S as a function of the $\tau(k,n)$ α's (for fixed probabilities) amounts to a nontrivial extension of S to a larger space of variables: not every $\tau(k,n)$-tuple of α values is geometrically realizable by an ensemble of k states in n dimensions, and when an actual ensemble is deformed, the α variables are constrained to lie on a surface of dimension $v(k,n)$ in the ambient space of dimension $\tau(k,n)$. But the virtue of this nonphysical extension of the number of parameters is that we are able to attribute compressibility of the source to geometrical constructs, viz. the α's. In any deformation of actual states, each α varies positively or negatively and the compressibility varies by a corresponding accumulation of monotonic positive and negative effects.

IV. MINIMAL SETS OF MONOTONIC PARAMETERS?

Since $\tau(k,n)>v(k,n)$ for $k>3$, it is interesting to ask whether we can find some alternative set of parameters $\beta_1,\ldots,\beta_{v(k,n)}$ which is in 1-1 correspondence with the set of k states and has some of the desirable properties possessed by the α's in the case $k\leqslant 3$. In particular, to make a connection with the phenomenon of compression, we would like the monotonicity property $\partial S/\partial\beta_i>0$ to hold for any choice of probabilities p_i. Intuitively, this means that we can regard the β's as measures of quantum entropy.

Consider the case $k=n=4$. Here the set of states is nine dimensional, whereas there are 11 α's. Can we perhaps keep some of the α's, taking, say, the overlap-related terms α_{ij} as $\beta_1,\ldots,\beta_6$, and adding a further 3 β's? [For instance, one might consider adding the three distances between subspaces generated by disjoint pairs of the four states, i.e., $d(12,34)$, $d(13,24)$, and $d(14,23)$, where $d(ij,kl)$ is some measure of the distance between the subspace spanned by $|\psi_i\rangle,|\psi_j\rangle$ and that spanned by $|\psi_k\rangle,|\psi_l\rangle$.] However, it turns out that *any* set of parameters that includes the six terms α_{ij} cannot have the desired monotonicity property. To see this, write $r_{ij}=|a_{ij}|$ and define $u=\arg(a_{12}a_{23}a_{31})$, $v=\arg(a_{14}a_{21}a_{42})$, and $w=\arg(a_{13}a_{34}a_{41})$. Then we can write all the α's in terms of the six r_{ij}'s and u, v, and w, giving nine parameters in all. For instance, $\alpha_{123}=1-r_{12}^2-r_{23}^2-r_{31}^2+2r_{12}r_{23}r_{31}\cos u$.

Now pick one of the β_i for $i>6$, and call it x. Taking the partial derivative with respect to x, the fact that $\beta_1,\ldots,\beta_6$ are constant implies $\partial s_2/\partial x=0$ [since by Eq. (9) s_2 is a function only of $\beta_1,\ldots,\beta_6$]. Furthermore, if we choose a set of states with $r_{ij}>0$ for all i,j and $u=v=w=\pi/2$ we find

$$\begin{aligned}\partial s_3/\partial x = &-2(p_1p_2p_3)(r_{12}r_{23}r_{31})u_x-2(p_1p_2p_4)(r_{14}r_{21}r_{42})v_x\\ &-2(p_1p_3p_4)(r_{13}r_{34}r_{41})w_x+2(p_2p_3p_4)(r_{23}r_{34}r_{42})\\ &\times(u_x+v_x+w_x).\end{aligned}$$

Suppose the first three terms in the above expression for $\partial s_3/\partial x$ are positive. Since $\beta_1,\ldots,\beta_6$ are real and positive, this means $u_x<0, v_x<0, w_x<0$. So the fourth term is negative. By taking p_1 small enough, we can ensure that $\partial s_3/\partial x<0$ and also that $\partial s_4/\partial x$ is sufficiently small to ensure that the term with $q=3$ dominates the sum $\partial S/\partial x=\Sigma(\partial S/\partial s_q)(\partial s_q/\partial x)$. So $\partial S/\partial x<0$ and monotonicity fails. Suppose on the other hand that at least one of the first three terms is negative, the term with u_x say, so $u_x>0$. Then taking p_4 small enough leads to the same conclusion.

This result suggests that it may be difficult to construct a minimal set of $v(k,n)$ monotonic parameters which also has a simple geometrical interpretation, thus further underlining the benefits of considering the nonminimal parameter set discussed in Sec. III.

V. TWO PROOFS OF THEOREM 1

We give here two proofs of the theorem that $\partial S/\partial s_q>0$ for $2\leqslant q\leqslant n$.

First proof. We will prove a slightly stronger result, giving a positive lower bound for $\partial S/\partial s_q$ [see Eq. (19)]. Note first that the x_i can be implicitly defined as functions

$x_i(s_1, \ldots, s_n)$ of the symmetric polynomials s_i by

$$\sum_{i=0}^{n} (-1)^{n-i} s_{n-i} x^i = 0 \tag{10}$$

(with $s_0=1$). Differentiating this equation with respect to s_q, we get

$$\frac{\partial x_k}{\partial s_q}\left[\sum_{i=0}^{n} (-1)^{n-i} s_{n-i} i x_k^{i-1}\right] + (-1)^q x_k^{n-q} = 0. \tag{11}$$

Since the expression in square brackets is the derivative of $\Pi(x-x_i)$ with x set to x_k, we have

$$\frac{\partial x_k}{\partial s_q} = \frac{(-1)^{q+1} x_k^{n-q}}{\prod_{i\neq k}(x_k - x_i)}. \tag{12}$$

From the chain rule, for any function f of the s_q,

$$\frac{\partial f}{\partial s_q} = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial s_q}. \tag{13}$$

Taking $f=s_1$ we get

$$\sum_{k=1}^{n} \frac{x_k^{n-q}}{\prod_{i\neq k}(x_k - x_i)} = 0 \quad \text{if } 2 \leqslant q \leqslant n. \tag{14}$$

Then, taking $f=S=-\Sigma x_k \ln x_k$ (and assuming that all $x_k>0$) gives

$$\frac{\partial S}{\partial s_q} = (-1)^q \sum_{k=1}^{n} \frac{x_k^{n-q}(1+\ln x_k)}{\prod_{i\neq k}(x_k - x_i)}, \tag{15}$$

which in view of Eq. (14) implies

$$\frac{\partial S}{\partial s_q} = (-1)^q \sum_{k} \frac{x_k^{n-q}\ln x_k}{\prod_{i\neq k}(x_k - x_i)} \quad \text{for } 2 \leqslant q \leqslant n. \tag{16}$$

Define

$$W_q(a) = (-1)^q \sum_{k} \frac{(x_k+a)^{n-q}\ln(x_k+a)}{\prod_{i\neq k}(x_k - x_i)},$$

so $W_q(0)=\partial S/\partial s_q$. Then

$$W_q(a) = (-1)^q a^{n-q} \sum_{k} \frac{(1+x_k/a)^{n-q}\ln(1+x_k/a)}{\prod_{i\neq k}(x_k - x_i)},$$

where we have omitted a term with factor $\ln(a)$ since Eq. (14) implies that this term is zero. Using Eq. (14) again, we can approximate $W_q(a)$ for large a by the term in x_k^{n-1} in the expansion of $(1+x_k/a)^{n-q}\ln(1+x_k/a)$, giving

$$\begin{aligned} W_q(a) &\simeq (-1)^q a^{1-q} \\ &\quad \times [\text{coefficient of } x^{n-1} \text{ in } (1+x)^{n-q}\ln(1+x)] \\ &= a^{1-q}\int_0^1 y^{q-2}(1-y)^{n-q}dy \\ &\to 0 \text{ as } a\to\infty \text{ for } 2 \leqslant q \leqslant n. \end{aligned}$$

Using Eq. (14) again we find

$$\frac{\partial W_q(a)}{\partial a} = -(n-q)W_{q+1}(a), \tag{17}$$

for $2\leqslant q<n$, and applying Eq. (14) to the set $a, x_1, \ldots, x_n$

$$\frac{\partial W_n(a)}{\partial a} = -1\Big/\prod (a+x_k). \tag{18}$$

Using $\Pi_{k=1}^{n}(a+x_k)\leqslant(a+1/n)^n$ in the preceding equation,

$$W_n(x) \geqslant \int_x^\infty \frac{da}{(a+1/n)^n} = \frac{1}{(x+1/n)^{n-1}(n-1)},$$

so $\partial S/\partial s_n = W_n(0) \geqslant n^{n-1}/(n-1)$.

Equation (17) for $q=n-1$ then implies

$$\begin{aligned} W_{n-1}(x) &= \int_x^\infty W_n(a)da \geqslant \int_x^\infty \frac{da}{(n-1)(a+1/n)^{n-1}} \\ &= \frac{1}{(x+1/n)^{n-2}(n-1)(n-2)}, \end{aligned}$$

so $\partial S/\partial s_{n-1} = W_n(0) \geqslant n^{n-2}/(n-1)(n-2)$. And continuing this way we find

$$\partial S/\partial s_{n-q+1} \geqslant \frac{n^{n-q}}{q\binom{n-1}{q}}. \tag{19}$$

So all the partial derivatives are bounded away from zero, which proves the theorem.

Second proof. A second proof involves using a theorem from numerical analysis—the so-called Hermite-Gennochi theorem [10]—to replace the explicit derivation above from Eq. (16) onwards. (This theorem was also used in Ref. [8], end of Appendix A.) Thus we begin as above, deriving the expression in Eq. (16) for $\partial S/\partial s_q$.

Now if $f(x)$ is any function whose values are known only at n points $x_1, \ldots, x_n$ then there is a unique polynomial of degree $n-1$, the Lagrange interpolating polynomial, that agrees with the function at these points. The coefficient of x^{n-1} is called the Newton divided difference of f and has standard explicit formula $\Sigma_i f(x_i)/\Pi_{k\neq i}(x_k-x_i)$. Thus Eq. (16) states that $\partial S/\partial s_q$ the Newton divided difference for the function $f(x)=(-1)^q x^{n-q}\ln x$. Now the Hermite-Gennochi theorem asserts that the Newton divided difference is also given by the integral over the probability simplex $\{(p_1, \ldots, p_n): p_i \geqslant 0, \Sigma_i p_i = 1\}$ of $f^{(n-1)}(p_1x_1+\cdots p_nx_n)$ where $f^{(n-1)}$ is the $(n-1)$th derivative of f. Taking f to be $(-1)^q x^{n-q}\ln x$ it is straightforward to check that $f^{(n-1)}(x)>0$ for all $0<x<1$. Hence the integral over the probability simplex is positive and we get $\partial S/\partial s_q>0$.

VI. A REMARK ON SUBENTROPY

It is curious that our proof of the theorem that the entropy is an increasing function of the symmetric functions s_q (Sec. V) depends upon an algebraic expression closely related to a quantity called the subentropy. Given any density matrix ρ, the subentropy $Q(\rho)$ [8] is the greatest lower bound on the accessible information of any ensemble of pure states $|\phi_1\rangle,\ldots,|\phi_m\rangle$ with probabilities $p_1,\ldots,p_m$, for which $\rho=\Sigma p_i|\phi_i\rangle\langle\phi_i|$. In terms of the eigenvalues $x_1,\ldots,x_n$ of ρ (or indeed for any classical probability distribution) the subentropy can be written

$$Q(\rho)=-\sum_k \frac{x_k^n \ln x_k}{\prod_{i\neq k}(x_k-x_i)},$$

which would correspond to our Eq. (16) if we could put $q=0$ in that formula. For this to make sense, we need to introduce a variable which plays the role of s_0, i.e., a coefficient of the term x^n in Eq. (10). We therefore define

$$t_1=\frac{1}{s_1},\quad t_2=\frac{s_2}{s_1},\ldots,\quad t_n=\frac{s_n}{s_1},$$

and divide Eq. (10) through by s_1 to obtain

$$t_1x^n-x^{n-1}+\cdots+(-1)^q t_q x^{n-q}+\cdots+(-1)^n t_n=0. \tag{20}$$

This equation defines the roots x_i in terms of $t_1,\ldots,t_n$. Following the same method used to derive Eq. (16), we can carry out an implicit differentiation with respect to t_1, noting that the equivalent of the expression in square brackets in Eq. (11) is now the derivative by x of $(1/s_1)\Pi(x-x_i)$ with x set to x_k. This gives

$$\frac{\partial x_k}{\partial t_1}=-\frac{x_k^n s_1}{\prod_{k\neq i}(x_k-x_i)}$$

and

$$\frac{\partial S}{\partial t_1}=s_1\sum_k \frac{(1+\ln x_k)x_k^n}{\prod_{k\neq i}(x_k-x_i)}. \tag{21}$$

Finally, applying the chain rule $\partial f/\partial t_1=\Sigma_k(\partial f/\partial x_k)(\partial x_k/\partial t_1)$ to $f=s_1$, and noting that $\partial s_1/\partial t_1=-s_1^2$, we obtain the identity

$$\sum_{k=1}^n \frac{x_k^n}{\prod_{k\neq i}(x_k-x_i)}=s_1,$$

so Eq. (21) can be written

$$\frac{\partial S}{\partial t_1}=s_1(s_1-Q).$$

If $x_1,\ldots,x_n$ is a probability distribution, so $s_1=1$, then we get $\partial S/\partial t_1=1-Q$. Thus we have proved the following.

Theorem 2. Let $S=-\Sigma x_i \ln x_i$ be the Shannon entropy function defined on $\{(x_i,\ldots,x_n):x_i>0 \text{ all } i\}$ (i.e., we lift the probability condition $\Sigma x_i=1$). If S is viewed as a function of $t_1=1/s_1, t_2=s_2/s_1,\ldots,t_n=s_n/s_1$ then at points with $s_1=\Sigma x_i=1$ the subentropy is given by $Q(x_1,\ldots,x_n)=1-\partial S/\partial t_1$.

Note that the above mathematical characterization of subentropy applies equally well within *classical* information theory (as it is a derivative property of the Shannon entropy function), in contrast to all previous work on subentropy [8,9] where it relates only to *quantum* mechanical considerations (especially the theory of information gain from quantum measurements).

Finally we also note that there are other possible ways of getting a nontrivial coefficient of x^n in Eq. (10). For example, instead of dividing through by s_1 we could divide through by s_n and introduce the variables $r_q=(s_n-q)/s_n=q$th symmetric polynomial of $1/x_1,\ldots,1/x_n$. We then get the equation

$$\frac{1}{s_n}p(x)=r_nx^n+\cdots+(-1)^q r_{n-q}x^{n-q}+\cdots+(-1)^n$$
$$=(x/x_1-1)\cdots(x/x_n-1)=0$$

leading to an alternative characterization of subentropy Q as

$$\frac{\partial S}{\partial r_n}=s_n(s_1-Q)$$

and the condition for $x_1,\ldots,x_n$ to be a probability distribution is now $s_1=r_{n-1}/r_n=1$, i.e., $r_n=r_{n-1}$.

VII. DISCUSSION

The problem we have addressed in this paper is whether there are real-valued functions α_q of k states in n dimensions that together characterize those states up to a unitary transformation and are also “measures of quantum redundancy” in the sense that $\partial S/\partial\alpha_q>0$ for each α_q. In other words, increasing one α while holding the others fixed increases the entropy S and hence reduces the redundancy of the set of states. We would also like the α_q to have an interpretation in terms of the Hilbert space geometry of the states.

We use the term “quantum redundancy” here because we require that $\partial S/\partial\alpha_q>0$ holds for any choice of probabilities p_i of the states $|\psi_i\rangle$, and the p_i can be thought of as embodying the classical aspect of redundancy. Of course, one might ask whether there are joint functions of the states and their probabilities that characterize the entropy, and one example of this is the “perimeter” considered recently by Hartley and Vedral [11]. The question then is why one such function should be preferred to another; after all, the symmetric functions s_q trivially determine the entropy via the characteristic equation (7). The functions in Ref. [11] are motivated by the possibility of experimental measurement whereas our considerations are motivated by a desire to geometrically characterize a notion of quantum redundancy in quantum information compression.

Our main conclusion is that there is a natural set of measures (in our sense) for sets of two or three states, but for four or more states the corresponding parameters—the determinants of square submatrices of the matrix $(\langle\psi_i|\psi_j\rangle)$—outnumber the degrees of freedom in the sets of states, and the more obvious ways of carrying over the results from two or three states fail. Nevertheless the simple geometrical interpretation of these parameters, in terms of Hilbert space volumes defined by the states, makes it appealing to consider an extension of the entropy function to the full space of these variables, and the entropy of any physical ensemble of states then appears as a special case satisfying some extra algebraic constraint equations.

ACKNOWLEDGMENT

R.J. was supported by the U.K. Engineering and Physical Sciences Research Council.

[1] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).
[2] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).
[3] M. Horodecki, Phys. Rev. A **57**, 3364 (1998).
[4] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).
[5] A. Winter, Ph.D. thesis, Fakultät für Mathematik, University of Bielefeld, e-print quant-ph/9907077.
[6] R. Jozsa and J. Schlienz, Phys. Rev. A **62**, 012301 (2000).
[7] R. Jozsa, IBM J. Res. Dev. (to be published), e-print quant-ph/0305114.
[8] R. Jozsa, D. Robb, and W. K. Wootters, Phys. Rev. A **49**, 668 (1994).
[9] S. Nichols and W. K. Wootters, Quantum Inf. Comput. **3**, 14 (2003).
[10] K. E. Atkinson, *An Introduction to Numerical Analysis* (Wiley, New York, 1978), pp. 107–123.
[11] J. Hartley and V. Vedral, e-print quant-ph/0309088.