View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository of the Academy's Library

arXiv:1304.5723v2 [cs.IT] 25 Sep 2014

CLASSICAL INFORMATION STORAGE IN AN n-LEVEL QUANTUM SYSTEM

PÉTER E. FRENKEL AND MIHÁLY WEINER

Dedicated to Professor Andor Frenkel on the occasion of his 80^{th} birthday

Abstract. A game is played by a team of two — say Alice and Bob — in which the value of a random variable x is revealed to Alice only, who cannot freely communicate with Bob. Instead, she is given a quantum n-level system, respectively a classical n-state system, which she can put in possession of Bob in any state she wishes. We evaluate how successfully they managed to store and recover the value of x by requiring Bob to specify a value z and giving a reward of value $f(x, z)$ to the team.

We show that whatever the probability distribution of x and the reward function f are, when using a quantum n-level system, the maximum expected reward obtainable with the best possible team strategy is equal to that obtainable with the use of a classical n-state system. The proof relies on mixed discriminants of positive matrices and — perhaps surprisingly — an application of the Supply–Demand Theorem for bipartite graphs.

As a corollary, we get an infinite set of new, dimension dependent inequalities regarding positive operator valued measures and density operators on complex n-space.

As a further corollary, we see that the greatest value, with respect to a given distribution of x, of the mutual information $I(x; z)$ that is obtainable using an n-level quantum system equals the analogous maximum for a classical n-state system. We propose a natural conjecture that would imply both this result and Holevo's inequality.

1. Introduction

In contrast to a classical bit which has only 2 pure states, a qubit has infinitely many. However, this does not necessarily mean that we can store more (classical) information in a qubit than in a classical bit. The point is that although the qubit has infinitely many different pure states, it is impossible to distinguish these states with certainty. This is a fundamental fact, and cannot be circumvented by some better measuring device.

The first author's research is partially supported by MTA Rényi "Lendület" Groups and Graphs Research Group. The second author's research is supported in part by the ERC Advanced Grant 227458 "Operator Algebras and Conformal Field Theory", OTKA grant no. 104206 and by the "Bolyai János" Research Scholarship of the Hungarian Academy of Sciences.

In the case of a qubit, one can distinguish with certainty between at most 2 states. In general, in the case of an n-level system (whose state space is modelled by the set of density operators of an n-dimensional complex Hilbert space), one can distinguish with certainty between at most n states. So in this respect a quantum n-level system performs like a classical n-state system. However, this does not make them necessarily equivalent. Perhaps it is possible to distinguish between $k > n$ states of a quantum n-level system not with certainty, but in a way that is — in some sense — "closer" to certainty than what we can achieve with a classical n-state system.

How to define "closer to certainty"? In general, we may view our qubit as a memory — or as is often done in the literature: as a channel[1] — in which there is an ingoing and an outgoing information: Alice chooses a certain state from a previously fixed set of states and puts the system into the selected state, then passes it to Bob, who will have to try to figure out the chosen state. So one may investigate the issue from the point of view of some kind of (classical) *channel capacity*. One idea is of course to use Shannon-type informational capacity which is well-investigated in the quantum setting; see e.g. [13] on quantum information theory and [11] on quantum entropy; see also Section 4 of the present paper.

However, here we argue that this in itself cannot fully settle the problem. As an example, suppose that the following game is played.

A \$1 bill is put randomly and with equal probability into one of 3 boxes. Bob will pick one of the boxes and he will get what is inside that box. Alice knows where the \$1 bill has been put, and she wishes to help Bob. However, Alice is not allowed to directly tell Bob where the money is (in which case Bob could always get the \$1 bill with certainty). Instead, she is only allowed to send to Bob a classical bit, respectively a qubit (not previously entangled to anything else) whose state she can manipulate as she wishes. That is, she is allowed to send a classical or a quantum bit of information.

They may agree on some scheme beforehand. For example, played with a classical bit, Alice and Bob can agree that the bit-value 0 will mean that the money is in box nr. 1 (in which case Bob will pick box nr. 1) and the bit-value 1 will mean that the money is either in box nr. 2 or in box nr. 3 (in which case Bob will toss a coin and accordingly pick box nr. 2 or 3). The question then becomes: after the game is played once, what is the expected value of the money won?

[1] Usually by specifying a *channel* one means to fix a collection of states corresponding to encoding, while the measurement on the decoding side remains unspecified. For us neither encoding nor decoding is fixed as both are to be optimalized; only the level n of the system (i.e. the dimension of the Hilbert space) is fixed. For this reason we prefer to talk about a *memory unit* rather than a channel.

Every team-strategy leads to a specific *channel matrix*, i.e. a collection of conditional probabilities of the type a_{ij} where

$$a_{ij} = P(\text{Bob chooses the } i^{\text{th}} \text{ box given that the money is in the } j^{\text{th}} \text{ box}),$$

so that for example the previously described simple strategy using a classical bit gives the channel matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

In general, the channel matrix A will be a *stochastic matrix*: its entries are all nonnegative, and each column sums to 1. The above channel matrix will allow Bob to have an expected win of 2/3 dollars, so we may say that its "money capacity" is $c_\$ = 2/3$. It is an elementary exercise to show that this is the best we can get using a classical bit.

Note that in general $c_\$(A) = \frac{1}{3}\mathrm{tr}(A)$ whereas the "usual" (informational) capacity $c(A)$ would be the maximum *mutual information* between Bob's choice and the actual place of the money; for the above channel matrix c is precisely 1 bit. Now, by Holevo's celebrated theorem [7], even if Alice and Bob used a quantum bit instead of a classical one, the informational capacity c could not get above 1 bit. This is in accordance with the (common) belief that a single qubit (on its own) is worth no more than a classical one.

The fact that in *superdense coding* [3] Alice manages to transmit 2 bits of information to Bob by physically sending only 1 qubit is no contradiction to what was said. Indeed, in superdense coding 2 qubits are used: for the *decoding* part both of them are necessary. However, for the *encoding* part only one of them is needed; this is achieved by previously entangling the 2 qubits. So the 2 classical bits are actually stored in 2 quantum bits; the surprising feature is rather that it is a kind of 2-bit memory made out of 2 qubits where for the "read out" we need both, but for the "write in" we only need to get in touch with one of them.

However, here we are interested by how much (if any) better is a qubit *on its own* (not entangled to other qubits) than a classical one. So can we apply Holevo's theorem to conclude that in the described game, even by using a qubit, Alice and Bob cannot win more than 2/3 of a dollar (i.e. the maximum amount possible when a classical bit is used)? The answer is negative. In fact, consider the stochastic matrix

$$A = \begin{pmatrix} 3/4 & 1/8 & 1/8 \\ 1/8 & 3/4 & 1/8 \\ 1/8 & 1/8 & 3/4 \end{pmatrix}.$$

Its "money capacity" $c_\$(A) = \frac{1}{3}\operatorname{tr} A = \frac{3}{4}$ is larger than what can be achieved by using a classical bit. However, elementary arguments together with a straightforward computation show that $c(A) = \frac{7}{4}\log(3) -$

$\frac{9}{4}\log(2)$, which is *smaller* than $\log(2)$, i.e., smaller than one bit. This should not be considered a surprise: in Shannon's *Noisy Channel Coding Theorem*, the channel capacity as reliable transmission rate is only achieved in the "long run"; with a single use of the channel, things can go different.

Thus, as shown by the previous example, Holevo's theorem cannot rule out the existence of a strategy by which using a qubit Alice and Bob can win more money in this game (in terms of expected values) than what is possible using a classical bit.

Nevertheless, for this actual game it is not difficult to come up with an argument [9] to show that $c_\$ \leq 2/3$ holds even if Alice and Bob use a qubit. However, why sticking exactly to this game, and why investigating only the case of 2-level systems (i.e. single bits)?

In general we might consider the following scheme. The value of a random variable x is revealed to Alice but not to Bob. Though previous to the game they can meet and agree to follow any kind of strategy they like, during the game Alice cannot freely communicate with Bob. Instead, she is given a quantum n-level system (or alternatively: a classical n-state system) which she can put in possession of Bob in any state she wishes. Bob is then required to specify a value z. We evaluate how successfully they managed to store and recover the value of x by giving a reward of value $f(x, z)$ to the team, where f is some previously fixed "reward function" (e.g. in our original game the reward was \$1 if $x = z$ and zero otherwise).

Consider the sets $\mathcal{C}_n(k, l)$ and $\mathcal{Q}_n(k, l)$ defined as convex hulls of all possible $k \times l$ channel matrices that can be obtained by Alice passing to Bob a classical n-state system or a quantum n-level system, respectively. We postpone a more detailed description of these sets to the next section, but note that obviously $\mathcal{C}_n(k, l) \subseteq \mathcal{Q}_n(k, l)$.

Since we think of the distribution of x as fixed with the rules of the game, the expected reward $\mathbb{E}(f(x, z))$ is just an arbitrary affine linear functional of the channel matrix of Alice and Bob (e.g. in our original game it was $1/3$ times the trace). Thus, there would exist a game of the specified type in which it is more efficient to use a quantum n-level system than a classical one if and only if we had $\mathcal{C}_n(k, l) \neq \mathcal{Q}_n(k, l)$. However, our main result is that $\mathcal{C}_n(k, l) = \mathcal{Q}_n(k, l)$ for all values of n, k, and l.

Note that this result is trivial for $n \geq \min(k, l)$, since then both sets consist of all stochastic $k \times l$ matrices. However, in general the equality is nontrivial in the sense that it results in some new, dimension dependent inequalities regarding positive operator valued measures and density matrices.

It is worthwile to make a remark on the specific type of game we are dealing with. Suppose for a change that x and y are drawn uniformly and independently from the set $\{A, B, C\}$ and x is revealed to Alice

(but not to Bob) whereas the value of y is revealed to Bob (but not to Alice). This time they get rewarded if Bob correctly guesses whether $x = y$ or not, but again, they can only communicate in a certain restricted way: Alice is allowed to pass to Bob a single quantum bit (or alternatively: a single classical bit). Now it is easy to show that in this game, using a quantum bit rather than a classical one is indeed more advantageous. This does not contradict our result — this game is not of the discussed type: here Bob recieves information from *two* separate sources (apart from what he gets from Alice, he also receives the value of y). So again we stress that our result concerns the capacity of a quantum n-level system when it is considered *on its own*, without further supporting classical or quantum information (like the value of y in the new game). The new game is in fact a *quantum fingerprinting* problem. For such problems, a quantum system is known [4] to perform exponentially better than the corresponding classical system.

Notations and terminology. The set $\{1, \ldots, k\}$ is denoted by $[k]$. We write e_{ij} for the matrix that has an entry 1 at position (i, j) and all other entries zero. The identity matrix is $\mathbf{1}$. A matrix is *stochastic* if all entries are nonnegative reals and each column sums to 1. A complex matrix A is *psdh* if it is positive semidefinite Hermitian, written $A \geq 0$. A *positive operator valued measure (POVM),* also called a *partition of unity,* is a sequence $E_1, \ldots, E_k$ of psdh matrices summing to $\mathbf{1}$. A *density matrix* is a psdh matrix with trace 1.

2. Sets of channel matrices

Throughout this section, let n, k and l be fixed positive integers.

Suppose that a letter of an alphabet containing l letters is to be encoded by Alice in a classical n-state system, whereas on the decoding side Bob uses an alphabet containing k letters. Which channel matrices can Alice and Bob realize by a suitable strategy?

Every strategy is a convex combination of *pure* strategies; strategies — we are in the classical setting — in which no randomness appears. Thus in the case of a pure strategy, the channel matrix is a $k \times l$ matrix such that

(1) each entry is either 1 or 0,
(2) in each column there is exactly one 1,
(3) the number of nonzero rows is at most n.

This last property is due to the fact that Alice and Bob use a classical n-state system: if no randomness is used, then at decoding at most n different things can happen, regardless of the number k of letters that Bob has in his alphabet. Thus we make the following

Definition 1. Let $\mathcal{C} = \mathcal{C}_n(k, l)$ be the convex hull of $k \times l$ matrices satisfying properties (1), (2) and (3) above.

Then $\mathcal{C}$ is a convex polytope whose vertices are the $k \times l$ matrices satisfying (1), (2) and (3). Note that $\mathcal{C}$ is also the convex hull of $k \times l$ stochastic matrices with at most n nonzero rows.

Now suppose that instead of a classical n-state system, Alice can use an n-level quantum system. Its state space can be identified with the set of complex $n \times n$ *density matrices*: the set of matrices $\rho \in M_n(\mathbb{C})$ such that

$$\rho \geq 0,\ \mathrm{tr}(\rho) = 1.$$

A specific measurement scheme with k possible outcomes gives rise to an affine map from this state space to the set of classical probability distributions on the set $[k] = \{1, 2, \ldots, k\}$. Such an affine map is always given in the following way: we have a *positive operator valued measure* (POVM) $E_1, E_2, \ldots, E_k \in M_n(\mathbb{C})$, i.e. $E_i \geq 0$ for each $i = 1, 2, \ldots, k$ and $E_1 + E_2 + \ldots + E_k = \mathbf{1}$ (identity matrix), and the map in question is

$$\rho \mapsto (\mathrm{tr}(E_1\rho), \mathrm{tr}(E_2\rho), \ldots \mathrm{tr}(E_k\rho)),$$

see details in [8, Section 1.6]. Thus the most general measurement with k outcomes is a POVM E_1, E_2, $\ldots$, E_k in the sense that if the state of the system was described by the density operator ρ then the measurement will result in the i^{th} outcome with probability $\mathrm{tr}(E_i\rho)$.

Now let us return to the set of possible channel matrices. On the encoding side, Alice needs to choose a state for each letter to be encoded. On the decoding side, Bob needs to choose a measurement whose result will be interpreted as "read out". Thus a specific strategy is given by the choice of l density matrices $\rho_1, \rho_2, \ldots, \rho_l \in M_n(\mathbb{C})$ and a POVM $E_1, E_2, \ldots, E_k \in M_n(\mathbb{C})$ resulting in the channel matrix A with entries

$$a_{ij} = \mathrm{tr}(E_i\rho_j).$$

The set of matrices A we can obtain depends on n, k and l. We make the following

Definition 2. Let $\mathcal{Q} = \mathcal{Q}_n(k, l)$ be the convex hull of $k \times l$ matrices of the form $(\mathrm{tr}(E_i\rho_j))$, where $E_1, \ldots, E_k \in M_n(\mathbb{C})$ is a POVM and $\rho_1, \ldots, \rho_l \in M_n(\mathbb{C})$ are density matrices.

It may not be obvious that $\mathcal{Q}$ is a polytope, but in fact, our main result is

Theorem 3. $\mathcal{C} = \mathcal{Q}$.

We start by proving the trivial inclusion.

Proof of $\mathcal{C} \subseteq \mathcal{Q}$. Assuming $A \in \mathcal{C}$, we show that $A \in \mathcal{Q}$. We know that A is a stochastic matrix, and we may assume that only the first $\min(n, k)$ rows of A can be nonzero.

Put

$$\rho_j = \sum_{i=1}^{\min(n,k)} a_{ij} e_{ii}.$$

These are density matrices.

When $n \leq k$, put $E_i = e_{ii}$ for $i \leq n$ and $E_i = 0$ otherwise. When $n \geq k$, put $E_i = e_{ii}$ for $i \leq k-1$ and $E_k = \sum_{i=k}^{n} e_{ii}$. In either case, this is a POVM. We have $\mathrm{tr}(E_i \rho_j) = a_{ij}$, whence $A \in \mathcal{Q}$. □

For the proof of the reverse inclusion, we recall the definition and the positivity property of mixed discriminants.

The determinant is a homogeneous polynomial of degree n on $M_n(\mathbb{C})$. Therefore, there exists a unique symmetric n-linear function D such that

$$D(X, \ldots, X) = \det X$$

for all $X \in M_n(\mathbb{C})$. This function D is the *mixed discriminant*. By [1, Lemma 2(vi)], if $E_1, \ldots, E_n$ are all positive semidefinite Hermitian matrices, then

$$D(E_1, \ldots, E_n) \geq 0.$$

Proof of $\mathcal{Q} \subseteq \mathcal{C}$. Assume that $A \in \mathcal{Q}$. We prove that $A \in \mathcal{C}$. We may assume that $a_{ij} = \mathrm{tr}(E_i \rho_j)$, where $E_1, \ldots, E_k \in M_n(\mathbb{C})$ is a POVM and $\rho_1, \ldots, \rho_l \in M_n(\mathbb{C})$ are density matrices.

For $I = (i_1, \ldots, i_n) \in [k]^n$, put

$$p_I = D(E_{i_1}, \ldots, E_{i_n}),$$

where D is the mixed discriminant. We have $p_I \geq 0$ for all I. Thus, we get a measure P on $[k]^n$ defined by $P(S) = \sum_{I \in S} p_I$. Using the multilinearity of D and the assumption that $E_1, \ldots, E_k$ is a partition of unity, we see that

$$P([k]^n) = D(\mathbf{1}, \ldots, \mathbf{1}) = \det \mathbf{1} = 1,$$

so P is a probability measure. In fact, for any $R \subseteq [k]$, we may put $E_R = \sum_{i \in R} E_i$, and then we have

$$P(R^n) = \det E_R.$$

Since $0 \leq E_R \leq \mathbf{1}$, all eigenvalues of E_R are in $[0, 1]$. Thus, $\det E_R$, the product of eigenvalues, does not exceed the smallest eigenvalue. Hence, $(\det E_R)\mathbf{1} \leq E_R$, so, for all j,

$$\mathrm{tr}(E_R \rho_j) \geq \mathrm{tr}((\det E_R)\mathbf{1}\rho_j) = \det E_R.$$

The left hand side here is $A_j(R)$, where A_j is the probability measure on $[k]$ given by the j-th column of A. So we have

$$A_j(R) \geq P(R^n) \qquad \text{for all } R \subseteq [k].$$

Let us connect $I \in [k]^n$ to $i \in [k]$ by an edge if i occurs in I. This gives us a bipartite graph. The neighborhood of any set $S \subseteq [k]^n$ is

the set $R \subseteq [k]$ of indices occurring in some element of S. We always have $S \subseteq R^n$, whence $A_j(R) \geq P(R^n) \geq P(S)$. Thus, by the Supply–Demand Theorem, there exists a probability measure $\tilde{P}_j$ on $[k]^n \times [k]$ which is supported on the edges of the graph and has marginals P and A_j. Whenever $p_I \neq 0$, let $B(I)$ be the $k \times l$ stochastic matrix whose j-th column is given by the conditional distribution $\tilde{P}_j|I$ on $[k]$. Then $B(I)$ has at most n nonzero rows, and $A = \int B dP \in \mathcal{C}$. □

Remark 4. (Infinite channel matrices.) We may replace $[l]$ by any subset J of an affine space. In this setting, the polyhedron of $k \times l$ stochastic matrices is replaced by the set of those nonnegative functions on $[k] \times J$ that sum to 1 on $[k]$ for any fixed $j \in J$ and are piecewise linear on J for any fixed $i \in [k]$. Let $\mathcal{C}_n(k, J)$ be the convex hull of elements that are supported on $N \times J$ for some $N \subseteq [k]$ of cardinality at most n. Let $\mathcal{Q}_n(k, J)$ be the convex hull of all elements of the form $(i, j) \mapsto \operatorname{tr}(E_i \rho(j))$, where $E_1, \ldots, E_k$ is a POVM and ρ is a piecewise linear map from J to the set of $n \times n$ density matrices. Then $\mathcal{C}_n(k, J) = \mathcal{Q}_n(k, J)$. This follows easily from the proof of Theorem 3, so this is left to the reader.

Going one step further, we may also replace $[k]$ by a separable metric space X. Let $\mathcal{M}(X)$ be the set of probability measures on the Borel sets of X, endowed with the Prokhorov metric. In this setting, the polyhedron of $k \times l$ stochastic matrices is replaced by the set of piecewise linear functions $J \to \mathcal{M}(X)$, endowed with the supremum metric. Let $\mathcal{C}_n(X, J)$ be the closure of the convex hull of elements whose range is contained in $\mathcal{M}(N)$ for some $N \subseteq X$ of cardinality at most n. Let $\mathcal{Q}_n(X, J)$ be the closure of the convex hull of all elements of the form $j \mapsto \operatorname{tr}(E\rho(j))$, where $E : \mathcal{B}(X) \to M_n(\mathbb{C})$ is a positive operator valued probability measure on the Borel sets of X and ρ is a piecewise linear map from J to the set of $n \times n$ density matrices. Then $\mathcal{C}_n(X, J) = \mathcal{Q}_n(X, J)$. This follows easily from the previous paragraph and the fact that finitely supported POVMs are dense (due to the separability of X).

Everything said above remains true if J is any set, resp. a topological space, and the words 'piecewise linear' are erased, resp. replaced by 'continuous' on all occurrences.

3. Inequalities for POVM's and density matrices

As before, let n, k and l be positive integers. If a linear inequality is satisfied by the entries of any $k \times l$ stochastic 0-1 matrix with at most n nonzero rows, then we can use Theorem 3 to deduce that it is also satisfied by the entries of any $A \in \mathcal{Q}_n(k, l)$. This is a way to get new inequalities for POVM's and density matrices. Therefore, we want to find inequalities satisfied by all $A \in \mathcal{C}_n(k, l)$.

When $n \geq \min(k,l)$, the polytope $\mathcal{C}_n(k,l)$ is obviously the set of all stochastic $k \times l$ matrices and we do not get anything interesting.

In general, we are not able to describe the faces of the polytope $\mathcal{C}_n(k,l)$. However, it is clear that a $k \times l$ real matrix A belongs to the polytope if and only if it satisfies all linear inequalities

$$\operatorname{tr}(CA) \geq c \qquad (C \in \mathbb{R}^{l\times k}, c \in \mathbb{R}) \tag{3.1}$$

that the vertices satisfy. The vertices are the stochastic 0-1 matrices A with at most n nonzero rows, and all of them satisfy (3.1) if and only if for all $N \subseteq [k]$, $|N| = n$, we have

$$\sum_r \min_{i\in N} c_{ri} \geq c \tag{3.2}$$

for the entries of the matrix $C = (c_{ri})$. For example, if C is a 0-1 matrix such that any n columns have at least one 1 at the same position, and $c = 1$, then the inequalities (3.2) hold, and therefore $\operatorname{tr}(CA) \geq 1$ for all A in the polytope. E.g., let $n = 2$ and

$$C = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 1/6 & 0 & 1/2 & 1/2 \\ 1/6 & 1/2 & 0 & 1/2 \\ 1/6 & 1/2 & 1/2 & 0 \end{pmatrix}, \tag{3.3}$$

then $\operatorname{tr}(CA) = 1/2$, so A is not in the polytope $\mathcal{C}_2(4,4) = \mathcal{Q}_2(4,4)$.

Now observe that if C is a matrix of size $m \times k$, with arbitrary m, such that the inequalities (3.2) hold, then any vertex, and therefore any point A of the polytope $\mathcal{C}$ satisfies

$$\sum_r \min_{j\in[l]} (CA)_{rj} \geq c, \tag{3.4}$$

which is stronger than (3.1). Note that (3.4) can be rewritten as a system of l^r linear inequalities holding simultaneously. Namely, we choose a j_r for each r and write

$$\sum_r (CA)_{rj_r} \geq c. \tag{3.5}$$

As an example, choose $n \leq r \leq k$, let $m = \binom{k}{r}$, and let the rows of C be indexed by the r-element subsets S of $[k]$. Let $c_{S,i} = 1$ if $i \in S$ and zero otherwise. Then the inequalities (3.2) hold with

$$c = \binom{k-n}{r-n} = \binom{k-n}{k-r}$$

and therefore (3.4) also holds, i.e.,

$$\sum_{|S|=r} \min_{j\in[l]} \sum_{i\in S} a_{ij} \geq \binom{k-n}{k-r}$$

for all A in the polytope. Replacing r by $k-r$ and S by $[k]-S$, and using that A is stochastic, we get

$$\sum_{|S|=r} \max_{j\in[l]} \sum_{i\in S} a_{ij} \leq \binom{k}{r} - \binom{k-n}{r} \tag{3.6}$$

for all A in the polytope, whenever $0 \leq r \leq k-n$. When $r=1$, this is immediate for $A \in \mathcal{Q}_n(k,l)$ from the fact that for any density matrix ρ, we have $\rho \leq \mathbf{1}$ and so $\operatorname{tr}(E\rho) \leq \operatorname{tr} E$ for any psdh matrix E. However, when $r \geq 2$, the inequalities (3.6) seem nontrivial for $A \in \mathcal{Q}_n(k,l)$, and they only follow from Theorem 3. Note that the inequalities (3.6) are not sufficient to describe the polytope. Indeed, for $k=l=4$ and $n=2$, the matrix A in (3.3) satisfies (3.6) for all r, but is not in the polytope.

These examples already show that we have a huge freedom in choosing C, and the combinatorics of families of subsets of $[k]$ enters into the subject of finding dimension dependent linear inequalities for values $\operatorname{tr}(E_i\rho_j)$. This could turn out to be interesting. For example, think of the famous question of *mutually unbiased bases*. A complete set of mutually unbiased bases can also be described as a set of density operators $\rho_1, \ldots, \rho_{n(n+1)} \in M_n(\mathbb{C})$ with $(\operatorname{tr}(\rho_i\rho_j))_{i,j=1}^{n(n+1)}$ being certain prescribed values; see e.g. [2] for a good review or [12] for some recent development. If n is a power of a prime, then such systems exist, while for other dimensions such systems are believed to not to exist, although this has not yet been proved. Thus, to prove nonexistence, one will have to use inequalities (or other tools) that show nontrivial dependence on the dimension n.

To close this section, we mention the slightly related open question of describing the cone of *completely positive semidefinite* matrices. These are $k \times k$ matrices of the form $X = (\operatorname{tr}(A_iA_j))$, where k is fixed, n is arbitrary, and $A_1, \ldots, A_k$ are positive semidefinite $n \times n$ matrices. (Note that requiring the A_i to be *real* matrices rather than *complex* ones leads to the same notion of complete positive semidefiniteness.) Clearly, X is positive semidefinite, with nonnegative real entries. However, not all positive semidefinite and entrywise nonnegative matrices are completely positive semidefinite, even though n was arbitrary in the definition; see [5, 6, 10].

4. Mutual information

We now wish to place Theorem 3 in context with respect to Holevo's inequality. First, we recall some basic information theory. The *Shannon entropy* of a sequence $p = (p_1, \ldots, p_n)$ of nonnegative reals summing to 1 (i.e., a probability distribution) is defined to be

$$H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

We have

$$0 \le H(p) \le \log n. \tag{4.1}$$

The *von Neumann entropy* $S(\rho)$ of a density matrix ρ is the Shannon entropy of its eigenvalues. Von Neumann entropy is a concave function on density matrices. For density matrices $\rho_1, \ldots, \rho_l$ of size $n \times n$ and a probability distribution $q = (q_1, \ldots, q_l)$, define

$$\chi = S\left(\sum q_j \rho_j\right) - \sum q_j S(\rho_j).$$

We have

$$0 \le \chi \le \min(H(q), \log n). \tag{4.2}$$

When x is a discrete random variable with distribution p, we write $H(x) = H(p)$. Given a matrix M whose elements are nonnegative and sum to 1, we may view M as the distribution of a pair (z, x). The mutual information between z and x is defined to be

$$I = H(z) + H(x) - H(z, x),$$

which is symmetric with respect to z and x. We have

$$0 \le I \le \min(H(z), H(x)). \tag{4.3}$$

In quantum information theory, we are interested in the case

$$M = (q_j \operatorname{tr}(E_i \rho_j)),$$

where $E_1, \ldots, E_k \in M_n(\mathbb{C})$ is a POVM, $\rho_1, \ldots, \rho_l \in M_n(\mathbb{C})$ are density matrices, and $q = (q_1, \ldots, q_l)$ is a probability distribution. (Cf. the game discussed in the Introduction. The distribution q is the distribution of the random variable x whose value is revealed to Alice.) We then have

Holevo's inequality [7]. $I \le \chi$.

From Theorem 3, we can deduce a different upper bound for I. Let $m = \min(n, k)$ and

$$\psi = \max H(Q_1, \ldots Q_m), \tag{4.4}$$

where the maximum is taken over all partitions of $[l]$ into m pairwise disjoint subsets $L_1, \ldots, L_m$ (empty set allowed), and

$$Q_r = \sum_{j \in L_r} q_j \qquad (r = 1, \ldots, m).$$

We have

$$0 \le \psi \le \min(H(q), \log m). \tag{4.5}$$

Theorem 5. $I \le \psi$.

Proof. It is well known that the mutual information for the matrix $M = (q_j a_{ij})$ is a convex function of the stochastic matrix $A = (a_{ij})$ if the probability distribution $q = (q_1, \dots, q_l)$ is fixed. Thus, for $a_{ij} = \mathrm{tr}(E_i \rho_j)$, in which case we have $A \in \mathcal{Q}_n(k,l) = \mathcal{C}_n(k,l)$, the maximum of I is attained when A is a vertex of this polytope, i.e., a stochastic 0-1 matrix with at most m nonzero rows. For such A, we have

$$I = H(z) + H(x) - H(z,x) = H(z) \le \psi,$$

since x and (z,x) both have distribution q and z has a distribution of the form $Q_1, \dots, Q_m$. □

Remark 6. In view of (4.2) and (4.5), either one of Holevo's inequality and Theorem 5 implies $I \le \log n$.

It is obvious that Theorem 5 does not imply Holevo's inequality, i.e., it is possible to have $\chi < \psi$.

We now show that Holevo's inequality does not imply Theorem 5, even if we assume that $n \le k$, so that $m = n$. For example, let $n = 2$, $l = 3$, and let ρ_j $(j = 1, 2, 3)$ be projections onto three lines in the plane $\mathbb{R}^2$ that pass through the origin and mutually form angles of $\pi/3$. Let $q_j = 1/3$ $(j = 1, 2, 3)$. Then $S(\rho_j) = 0$, so

$$\chi = S\left(\sum_{j=1}^{3} q_j \rho_j\right) = S(\mathbf{1}/2) = \log 2,$$

while

$$\psi = H\left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} = \log\frac{3}{\sqrt[3]{4}},$$

whence $\psi < \chi$ for this example.

We now have two upper bounds for the mutual information I, of very different nature: Holevo's inequality is analytic, Theorem 5 is combinatorial. A common lower bound for χ and ψ is

$$\omega = \max\left(S\left(\sum_{r=1}^{m} Q_r \bar{\rho}_r\right) - \sum_{r=1}^{m} Q_r S(\bar{\rho}_r)\right),$$

where the maximum and the Q_r are to be understood as in the definition (4.4) of ψ, and

$$\bar{\rho}_r = \frac{1}{Q_r}\sum_{j \in L_r} q_j \rho_j \qquad (1 \le r \le m, \quad Q_r \ne 0).$$

Indeed, for all partitions $L_1, \dots, L_m$, we have

$$S\left(\sum_{r=1}^{m} Q_r \bar{\rho}_r\right) - \sum_{r=1}^{m} Q_r S(\bar{\rho}_r) \le H(Q_1, \dots, Q_m)$$

by (4.2), whence $\omega \le \psi$. Also, $\omega \le \chi$ by concavity of S. Thus, both of Holevo's inequality and Theorem 5 would be implied by

Conjecture 7. $I \leq \omega$.

Acknowledgements. The authors wish to thank Aidan J. Klobuchar (a student of the second author) for writing up the game with the $1 bill [9], and Márton Naszódi and Gábor Tardos for useful comments and conversations.

References

[1] R. B. Bapat: Mixed discriminants of positive semidefinite matrices. *Linear Algebra Appl.* **126** (1989), 107–124.

[2] I. Bengtsson and Å. Ericsson: Mutually unbiased bases and the complementarity polytope. *Open Syst. Inf. Dyn.* **12** (2005), pg. 107–120.

[3] C. H. Bennett and S. J. Wiesner: Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.* **69** (1992), pg. 2881–2884.

[4] H. Buhrman, R. Cleve, J. Watrous and R. de Wolf: Quantum Fingerprinting. *Phys. Rev. Lett.* **87** (2001).

[5] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, R. R. Thomas: Positive semidefinite rank, `arXiv:1407.4095`

[6] P. E. Frenkel and M. Weiner: On vector configurations that can be realized in the cone of positive matrices. *Linear Alg. Appl.* **459** (2014), 465–474.

[7] A. S. Holevo: Bounds for the quantity of information transmitted by a quantum communication channel. *Problems Inform. Transmission* **9** (1973), pg. 177–183.

[8] A. S. Holevo: Probabilistic and statistical aspects of quantum theory. Springer, 2011.

[9] A. J. Klobuchar: Classical capacities of a qubit. Notes, BSM fall semester 2010. `www.renyi.hu/~mweiner/qubit.pdf`

[10] M. Laurent and T. Piovesan, Conic approach to quantum graph parameters using linear optimization over the completely positive semidefinite cone, `arXiv:1312.6643v4`

[11] M. Ohya and D. Petz: Quantum Entropy and Its Use. Springer-Verlag, Heidelberg, 1993. Second edition 2004.

[12] M. Weiner: A gap for the maximum number of mutually unbiased bases. *Proc. Amer. Math. Soc.* **141** (2013), 1963-1969.

[13] M. M. Wilde: Quantum Information Theory. Cambridge University Press, 2013. Available online with title "From classical to quantum Shannon theory" as a preprint, `arXiv:1106.1445`.

Eötvös Loránd University, Department of Algebra and Number Theory, H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary

E-mail address: `frenkelp@cs.elte.hu`

Budapest University of Technology and Economics, Department of Analysis, H-1111 Budapest, Műegyetem rkp. 3–9, Hungary

E-mail address: `mweiner@renyi.hu`