Public vs. Private coin flips in one round communication games (Extended Abstract)

Ilan Newman * *and* Mario Szegedy †

Abstract

We study 1-round two parties communication complexity games, where private random bits are used. We observe that the existence of good protocols for such games is related to a notion of approximating the matrix that represents the function by a certain low rank matrix. This gives rise to a new notion of rank, analogous of positive rank for 1-round communication complexity. We prove that the identity matrix is non approximable by low rank matrices in this sense. As a corollary we prove that any randomized protocol for the equality function requires $\Omega(\sqrt{n})$ complexity in the 1-round, private bits model, answering an open question raised by Yao [8].

A corollary is an answer to the following graph theoretic question: Assume a graph on n vertices has a set of $N = N(n)$ cliques and so that for each two different cliques of this set, the number of edges between them is at most 0.1 of the product of their sizes. How large can N be ? We show that $N = n^{\theta(\log n)}$ is the right order of magnitude.

1 Introduction

In this work we investigate randomized 1-round communication complexity, in particular we are interested in the way the parties access the random bits.

*Dept. of Mathematic and Comp. Sci, Haifa U. Israel, email: ilan@mathcs.haifa.ac.il

†AT&T Bell Labs, email: ms@research.att.com

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. In communication complexity games there are typically two players that get partial information on the input. Player A gets $x \in \{0,1\}^n$ and Player B gets $y \in \{0,1\}^n$. Their goal is to compute $f(x,y)$. A protocol for the game specifies for each player what message it should send at any stage of the computation. The players are assumed to be computationally unbounded. The complexity of a protocol is the total number of bits that is transmitted for the worst case input.

A randomized protocol is a protocol in which each player may decide what is the next message to transmit in a random way. In this case, the output of the computation becomes a random variable. An ϵ-error protocol for f is said to have complexity c if the error for every input is bounded by ϵ and the (maximum) communication is bounded by c. For further details as well as a comprehensive survey on communication complexity the reader is referred to [4, 3].

In the communication complexity model, as well as in any multi party computational scenario, two ways of access to random bits are considered - the model where the random bits are common, and the more restrictive model in which each party has its own source of randomness. The first, is less natural, but is more suitable for lower bounds. The second, in which each party has its own independent random source is more natural from the algorithmic point of view.

In [5] it was proved that for any function $f : X \times Y \longrightarrow \{0,1\}$, any protocol with common random bits can be simulated by a protocol with private random bits with extra cost of $O(\log n)$ bits of communication (while maintaining the same error bound). However this requires some extra interaction. Focusing on 1-rounds games, this is not applicable anymore. Recently, Kremer, Nisan and Ron, [2], explicitly restated an open problem of Yao [8] concerning common versus private random bits for 1-round, simultaneous protocols. In this scenario there is only one round of communication in which each player sends a message to a referee that initially has no knowledge of the input (x, y). The referee decides (possibly by a randomized procedure) the value of $f(x, y)$ based on the information he receives from the players. The equality function $eq : \{0,1\}^n \times \{0,1\}^n \longrightarrow \{0,1\}$ is defined by $eq(x, y) = 1$ iff $x = y$. In [2] it was observed that the following protocol for the equality function has constant error, and $O(1)$ bits of communication using common bits: The players share n random bits, each player sends the inner product (mod 2) of its input vector with the n-bits random vector. The referee decides equality if it gets the same result from both players. What is an obvious upper bound for equality with private bits ? Several people observed that there is a constant error protocol of complexity $O(\sqrt{n})$. On the other hand, the question of lower bound (better than $\Omega(\log n)$) was already left open by Yao in [8].

We observe some general properties of 1-round randomized protocols with private random bits. We show that the referee can be restricted to be deterministic and that the players use the same strategy. Our main theorem is that any randomized protocol for equality requires $\Omega(\sqrt{n})$ communication complexity, which settles the question of [8, 2]. In particular it reveals a difference between 1-round and multi round protocols; unlike in multi rounds protocols, there is a large gap between common and private random bits for 1-round protocols.

In [2] it is shown that for common bits, there is a close relation between the complexity of simultaneous one round protocols to the complexity of the stronger one round model in which one player sends the information to the other. Our lower bound also shows that for private bits, there is no such relation.

Our main theorem is presented in an algebraic formulation; it states that a certain approximation of the identity matrix must have a relatively large rank. Assume that for $N \times m$ stochastic matrices A, B and $m \times m$ matrix M whose entries are in the $[0,1]$ interval, AMB^T has entries of at least 0.99 on the diagonal and at most 0.01 everywhere else. We prove in such case that $m = 2^{\Omega(\sqrt{\log N})}$. This 'rank' like measure is similar to the positive rank, we discuss this issue in section 5.

A key argument in the proof, which may be of independent interest relates various measures of divergence between probability distributions. Let $P = (p_1, ..., p_r)$ and $Q = (q_1, ..., q_r)$ be two probability distributions on the set $\{1, ..., r\}$. We can measure the difference between P and Q by the 'information divergence', $\delta(P, Q)$, which is defined by $\delta(P, Q) = \sum_{i=1}^{r} p_i \log(p_i/q_i)$. Another way to exhibit the difference between P and Q is to find a vector $X = (x_1, ..., x_r)$ with entries in the interval $[0,1]$ for which $\sum p_i x_i \geq 1 - \epsilon$ and $\sum q_i x_i \leq \epsilon$. These two measures are related to statistical tests that measure the error made when an element is drawn by either P or Q on the hypothesis that it was actually P. We relate in Lemmata 4 and 6 the two measures.

A corollary of the lower bound gives an answer to the following question: Let G_n be an n-vertices graph with the following property: G_n a set of N cliques so that for each two different cliques in this set, the number of edges between them is at most a fraction (say 0.01) of the product of their sizes. How large can N be ? It could be verified that for the random graph $N = n^{\theta(\log n)}$. Our theorem states that this is best possible (even for the case that each 'clique' misses a small fraction of edges).

We note here that recently, Bourgain and Wigderson gave a much simpler proof of our main theorem [1]. We hope however, that the machinery developed here will be of independent interest.

2 Definitions and notation

For a natural number k, we denote the set $\{1, ..., k\}$ by $[k]$. All vectors in this paper are row vectors. For two vectors of the same dimension $v = (v_1, ..., v_r)$, $u = (u_1, ..., u_r)$ we denote their standard inner product by $< v, u >= \sum_{i=1}^{r} v_i u_i$. For a vector v we denote its L_2 norm $|v|_2 =< v, v >^{1/2}$. For a vector v or a matrix A we say that v, (A) is 0-1 vector, (matrix) if every entry of v, (A) belongs to $\{0, 1\}$.

Let D be a probability distribution on a ground set $X = \{x_1, ..., x_r\}$. We identify D with the row vector of dimension r, $D = (p_D(x_1), ..., p_D(x_r)) = (p_1, ..., p_r)$. The entropy function $H(D)$, is defined by $H(D) = -\sum_i p_i \log p_i$. A *stochastic* matrix is a matrix in which each row is a probability distribution. For a stochastic matrix A we will denote by A_i its ith row.

Let $f : X \times Y \rightarrow \{0, 1\}$. In the sequel we always take $X, Y \subseteq \{0, 1\}^n$ and assume that $|X| = |Y| = N$. We associate with f a $N \times N$ matrix F that is defined by $F_{x,y} = f(x, y)$.

A *1-round, two parties randomized protocol for* f (abv. protocol) is the following: There are two players A and B and a referee. An input is a pair $(x, y) \in X \times Y$ so that A gets x and B gets y. Their goal is to send a minimal amount of information on input (x, y) just enough for the referee to compute $f(x, y)$. Player A sends a message according to a probability distribution that depends on x. Player B sends a message according to a probability distribution that depends on y. The referee decides the output according to the messages it received. We consider here the case where the referee can be randomized too.

Formally, a protocol is specified by three matrices A, B, M: A, B are two stochastic matrices of dimensions $2^n \times m = N \times m$. We assume that the rows of A and B are indexed by the elements of $\{0, 1\}^n$. For each $x \in \{0, 1\}^n$ the x row in A, A_x, and the x row in B, B_x, are probability distributions on $[m]$ (the set of messages). The matrices A and B are called the strategy matrices of the first and the second player respectively. M is a $m \times m$ matrix whose entries are in the interval $[0, 1]$ and represent the probability that the referee decides '1' on the corresponding messages. The matrix M is called the referee matrix.

Let $f : X \times Y \longrightarrow \{0, 1\}$ be a Boolean function and F its associated matrix. We say that the triplet $P = (A, B, M)$ is an ϵ-error, 1-round protocol for a function f using private random bits if $AMB^T = F^*$ where F^*_{xy} is at most ϵ for every entry (x, y) for which $F_{xy} = 0$ and at least $1 - \epsilon$ for every entry for which $F_{xy} = 1$. The complexity of the protocol is $\log m$. We note that this correspond to the maximum complexity rather then the average however, it is folklore that for constant error, these are of the same order (see also [3]).

This definition should be interpreted as follows: On input $(x, y) \in \{0, 1\}^n$, the first player chooses a message $i \in \{1, ..., m\}$ according to the distribution A_x (the xth row of A). Similarly, the second player chooses a message $j \in [m]$ according to the distribution B_y. Since the number of distinct messages is m, the communication complexity is $\log m$. The referee decides that the input is '1' with probability M_{ij}. If M is a 0-1 matrix then the referee is deterministic. The probability that the referee decides '1' on input (x, y) is given by $\sum_{i,j} A_{xi} B_{yj} M_{ij} = A_x M B_y^T$. This probability is also referred as the acceptance probability for (x, y). The acceptance probability could be written for every (x, y) in a matrix form by $AMB^T = F^*$ which differs from F (in every entry) by at most ϵ.

Let $P = (A, B, M)$ be a 1-round protocol for f, we denote by $\epsilon_P(1), \epsilon_P(0)$ the maximal error probability of P for 1 inputs and 0 inputs respectively. If $\epsilon_P(1) = 0$ the protocol is said to have *1-sided error* (abv. 1-sided).

In this paper we consider only constant error protocols. As amplification is trivial in this model (at the cost of multiplying the complexity), we may assume that if there is a protocol with error bounded by a constant ϵ, then ϵ is an arbitrarily small constant.

3 Some general facts about 1-round protocols

In this section we show that we can restrict our attention to protocols for which the two players use the same strategy matrix and the matrix M is $0-1$ matrix (namely, the referee is deterministic). Although we will not use this in our lower bound for equality, this inspires part of the intuition behind the proof of Lemma 6, and is an interesting fact by its own.

Proposition 1 *Let P be a protocol for f then there is a protocol P' for f with the same complexity, whose referee is deterministic and for which $\epsilon_{P'}(0) \leq 2\epsilon_P(0)$ and $\epsilon_{P'}(1) \leq 2\epsilon_P(1)$.*

Proof: Let $P = (A, B, M)$ be a protocol for f. Let M' be the matrix defined by $M'_{ij} = 0$ if $M_{ij} < 0.5$ and $M'_{ij} = 1$ otherwise. It can be verified that M' defines a protocol P' for a deterministic referee with $\epsilon_{P'}(0) \leq 2\epsilon_P(0)$ and $\epsilon_{P'}(1) \leq 2\epsilon_P(1)$. We omit further details here ■.

We also note that any protocol with strategy matrices A and B, can be transformed to a protocol where the two players use the same strategy by letting each player send two messages simulating itself and the second player. This doubles the communication while keeping the same error probability. In addition f can be 'symmetrized' by by taking disjoint copies of X and Y and defining f' on $(X \cup Y) \times (X \cup Y)$ so that its projection on $X \times Y$ is f. For symmetric functions we may assume that the referee matrix is symmetric. This is by taking $M' = (M + (M)^T)/2$, where M is the original referee matrix.

Finally we note that if one starts with 1-sided error protocol, the transformations above as well as the transformation defined by Proposition 1, produce a 1-sided error protocol.

4 Lower bound for equality

4.1 Main Theorem and Intuition

The equality function on $2n$ variables is represented by the 2^n identity matrix I_{2^n}. Let $N = 2^n$ and let P be a 2-sided error protocol for equality, with $\epsilon(1), \epsilon(0) \leq 0.01$. Let A, B the $N \times m$ strategy matrices and M the $m \times m$ referee matrix. As we explained AMB^T has diagonal entries at least 0.99, and non-diagonal entries at most 0.01. Our goal is to prove that $m \geq 2^{\Omega(\sqrt{\log N})}$. In turn, we fix m and prove that $N = m^{O(\log m)}$.

Theorem 2 *Let A, B be $N \times m$ stochastic matrices, M a $m \times m$ matrix with entries in $[0, 1]$ and so that AMB^T has diagonal entries at least 0.99, and non-diagonal entries at most 0.01, then $N \leq m^{O(\log m)}$.*

The proof uses a variant of the method, that we may call "subsequent conditioning." The intuition behind the proof is the following: We impose a probability distribution D on $[N] = \{1, ..., N\}$ that represents the set of rows of A (the inputs). We start with the uniform distribution, namely of L_2-norm of $1/\sqrt{N}$. Then by a conditioning process explained below, we transform the distribution, until we get a constant L_2-norm distribution. By analyzing the amount of increase in norm the upper bound on N follows.

Let $[m]$ denote the set $\{1, ..., m\}$. A probability distribution D on $[N]$ together with the stochastic matrix A defines a probability distribution P on $[N] \times [m]$ in a natural way; $P(i, j) = D(i)A_{ij}$. This corresponds to the following 'trial'; first pick $i \in [N]$ according to D, then pick $j \in [m]$ according to A_i (the ith row of A). The marginal of P on $[m]$ is given by the vector $DA = (q_1, ..., q_m)$ (D is taken here as a row vector). One may define a collection of m conditional probabilities $D_j(i) = P(i, j)/q_j, \ j = 1, ..., m$. Then, $D(i) = \sum_{j=1}^{j=m} P(i, j) = \sum_{j=1}^{j=m} q_j D_j(i)$. Namely the vector D is a convex combination of the vectors D_j given by $D = \sum_{j=1}^{j=m} q_j D_j$.

We start with D being the uniform distribution on $[N]$. We compare the entropy of the marginal distribution DA to these of the conditionals $D_j A$. We will show that there is a drop in the entropy by a constant for at least one of the js. On the other hand, we show that the L_2-norm of D_j increases by at most $\sim \sqrt{m}$. Replacing D by D_j, we repeat the process. As $H(DA) = O(\log m)$, this process terminates after at most $O(\log m)$ iterations, after which we show that we end up with a distribution D' of constant L_2-norm. Thus we conclude the upper bound on N.

Technically, we break the proof into three lemmata stated below, of which only the first uses the premises asserted by the theorem. The second lemma, as well as Lemma 6 might have independent interest outside the context of 1-round communication complexity.

For two distributions Δ, Δ_1 we measure their "divergence" in two different ways. We can measure the 'information divergence' $\delta(\Delta_1, \Delta)$. Another way to detect that Δ_1 diverges from Δ is if we find a vector X with entries in the $[0, 1]$-interval such that $\Delta X^T \leq \epsilon$ but $\Delta_1 X^T \geq 1 - \epsilon$. In Lemma 3 we make use of an appropriate conditioning in order to decompose the distribution DA into a collection of distributions that diverge from it in the second sense. In Lemma 4 we relate the two different notions of divergence. Lemma 5 then draws the conclusion of the theorem.

Let $\mathcal{S}_1$ be the set of distributions on $[N]$, and $\mathcal{S}_2$ be the set of distributions on $[m]$.

Lemma 3 *Let A, B be $N \times m$ stochastic matrices, M a $m \times m$ matrix with entries in $[0, 1]$ and so that AMB^T has diagonal entries at least 0.99, and non-diagonal entries at most 0.01. Let $D \in \mathcal{S}_1$ with $|D|_2 \leq 0.01$. Then for $i = 1, \ldots, m$ there are non-negative numbers q_i, probability distributions $D_i \in \mathcal{S}_1$, and vectors $X_i \in [0, 1]^m$ with the following properties:*

1. $\sum_{i=1}^{m} q_i = 1$,
2. $\sum_{i=1}^{m} q_i D_i = D$,
3. $\sum_{i=1}^{m} q_i (DA) X_i^T \leq 0.02$,
4. $\sum_{i=1}^{m} q_i (D_i A) X_i^T \geq 0.99$.

Lemma 4 *Let $\Delta, \Delta_j \in \mathcal{S}_2$, $0 < q_j \ j = 1, ..., \ell$ be such that $\Delta = \sum_{j=1}^{\ell} q_j \Delta_j$. If there are vectors $X_j, \ j = 1, ..., \ell$ of length m with coordinates in $[0, 1]$ such that*

1. $\sum_{j=1}^{\ell} q_j \Delta X_j^T \leq \epsilon$
2. $\sum_{j=1}^{\ell} q_j \Delta_j X_j^T \geq 1 - \epsilon$.

Then $H(\Delta) \geq \sum_{j=1}^{\ell} q_j H(\Delta_j) + C(\epsilon)$ Where $C(\epsilon)$ is a constant for which $C(\epsilon) \geq \frac{1-16\epsilon}{2} \log \frac{1}{8\epsilon} - \frac{1-8\epsilon}{2} \log \frac{1}{1-8\epsilon}$

We note that for ϵ small enough $C(\epsilon) > 0$, e.g for $\epsilon \leq 1/32$, $C(\epsilon) > \frac{\log 3}{4}$.

Lemma 5 *Let A be a $N \times m$ stochastic matrix with the following property: Every probability distribution $D \in \mathcal{S}_1$ with $|D|_2 < 0.01$ can be expressed as a convex combination of r probability distributions, $D = \sum_{i=1}^{r} q_i D_i$, such that there is a constant C so that, $\sum_{i=1}^{r} q_i H(D_i A) \leq H(DA) - C$. Then $N \leq (r \log m)^{O(\log m)}$.*

4.2 Proofs

Proof of Lemma 3:

Let $D \in \mathcal{S}_1$ with $|D|_2 \leq 0.01$. We define a probability distribution P on $[N] \times [m]$ (as explained in the previous subsection) by, $P(i,j) = D(i)B_{ij}$. P is indeed a probability distribution since B is stochastic. Let $\overline{q} = (q_1, ..., q_m)$ be the marginal distribution of P on $[m]$, namely $\overline{q} = DA$ (D is taken here as a row vector). Let D_j, $j = 1, ..., m$ be the conditional probability distributions, namely $D_j(i) = P(i,j)/q_j$, $j = 1, ..., m$. Then, (1) and (2) follows by the definition.

Let X_i^T, $i = 1, ..., m$ be the ith column of M. Condition (4) corresponds to the expected acceptance probability over the diagonal, with respect to D. To see that recall that for every $x \in [N]$ $A_x M B_x^T \geq 0.99$, where A_x, B_x are the x rows of A and B respectively. By the definition of X_i this is equal to $A_x \sum_{i=1}^{m} X_i^T B_{xi}$. Now summing this for every x with weights from D we get $\sum_{x \in [N]} D(x) A_x \sum_{i=1}^{m} X_i^T B_{xi} \geq 0.99$. Rearranging and plugging in the definition of $P(x,i)$ we get

$$0.99 \leq \sum_{x \in [N]} D(x) A_x \sum_{i=1}^{m} X_i^T B_{xi} =$$

$$\sum_{x \in [N]} \sum_{i=1}^{m} P(x,i) A_x X_i^T = \sum_{x \in [N]} \sum_{i=1}^{m} q_i D_i(x) A_x X_i^T =$$

$$\sum_{i=1}^{m} q_i D_i A X_i^T$$

For (3) recall that $|D|_2 \leq 0.01$ (this is the only place where we use this assumption). Now write $M_A P M_B$ in the form $I_N + E$, where I_N is the identity matrix and for all $1 \leq i,j \leq N$ $|E(ij)| \leq 0.01$. Then $DAMB^T D^T = D(I + E)D^T = DD^T + DED^T \leq |D|_2^2 + 0.01 \leq 0.02$. However, by plugging in the definition of q_i and rearranging, the left hand side of the above becomes $DAMB^T D^T = DAM\overline{q}^T = \sum_{i=1}^{m} q_i DAX_i^T$.

Proof of Lemma 4:

Definition 4.1 *Let $P = (p_1, \ldots, p_m)$, and $Q = (q_0, \ldots, q_m)$ be two probability distributions. The* information divergence *of P with respect to Q is: $\delta(P,Q) = \sum_{i=1}^{m} p_i \log p_i/q_i$, where $0 \log 0/0$ is defined as 0.*

The information divergence is a measure of distance between distributions. The most important property of the divergence is its positivity. Indeed, from the convexity of the $-x \log x$ function $-\sum_{i=1}^{m} p_i \log p_i/q_i =$
$-\sum_{i=1}^{m} q_i (p_i/q_i) \log p_i/q_i \leq -1 \log 1 = 0$.

Definition 4.2 *Two probability distributions P, Q are $1-\epsilon$ apart if there is a vector X with entries in $[0,1]$ such that $PX^T \geq 1 - \epsilon/2$ and $QX^T \leq \epsilon/2$. In this case we also say that such a vector X separates P from Q by $1-\epsilon$.*

Next we relate the two measures.

Lemma 6 *If P, Q are $1-\epsilon$ apart, then*

$$\delta(P,Q) \geq (1-2\epsilon) \log \frac{1}{\epsilon} - (1-\epsilon) \log \frac{1}{1-\epsilon}$$

Proof:

Claim 7 *If a vector X separates P from Q by $1-\epsilon$ then there is a 0-1 vector X' that separates P from Q by $1-2\epsilon$.*

Proof: Let $P = (p_1, ..., p_m)$ and $Q = (q_1, ..., q_m)$, and assume that X separates P from Q by $1-\epsilon$. Define a vector X' by $X'_i = 1$ if $p_i \geq q_i$ and $X'_i = 0$ otherwise. Then $1-\epsilon \leq (P-Q)X^T \leq \sum_{i:p_i \geq q_i}(p_i - q_i) = (P-Q)(X')^T$. In particular $P(X')^T \geq 1-\epsilon$ and $Q(X')^T \leq \epsilon$, hence the 0-1 vector X' separates P from Q by at least $1-2\epsilon$.

Claim 8 *Assume that a 0-1 vector X separates P from Q by $1-2\epsilon$ then*

$$\delta(P,Q) \geq (1-2\epsilon)\log\frac{1}{\epsilon} - (1-\epsilon)\log\frac{1}{1-\epsilon}$$

Proof: Let $P = (p_1, ..., p_m)$ and $Q = (q_1, ..., q_m)$. X defines a set of indices $S = \{i|\ X_i = 1\}$. Let $p = \sum_{i \in S} p_i$, $\overline{p} = 1-p$ and $q = \sum_{i\in S} q_i$, $\overline{q} = 1-q$. We show that unifying the points in S to one point and the points in $\overline{S}$ to another point does not increase the divergence:

$$\delta(P,Q) = \sum_{i\in S} p_i \log\frac{p_i}{q_i} + \sum_{i \notin S} p_i \log\frac{p_i}{q_i} =$$

$$p\sum_{i\in S}\frac{p_i}{p}\log\frac{p_i q}{p q_i} + p\log\frac{p}{q} + \overline{p}\sum_{i\notin S}\frac{p_i}{\overline{p}}\log\frac{p_i\overline{q}}{\overline{p}q_i} + \overline{p}\log\frac{\overline{p}}{\overline{q}} \geq$$

$$p\log\frac{p}{q} + \overline{p}\log\frac{\overline{p}}{\overline{q}}$$

The last inequality follows by the positivity of the divergence.

By the fact that X separates P from Q by $1-2\epsilon$ we have $p \geq 1-\epsilon$ and $q \leq \epsilon$. Thus,

$$\delta(P,Q) \geq p\log\frac{p}{q} + \overline{p}\log\frac{\overline{p}}{\overline{q}} \geq p\log p + \overline{p}\log\overline{p} +$$

$$p\log\frac{1}{q} \geq p\log\frac{1}{q} - \epsilon\log\frac{1}{\epsilon} - (1-\epsilon)\log\frac{1}{1-\epsilon} \geq$$

$$(1-\epsilon)\log\frac{1}{\epsilon} - \epsilon\log\frac{1}{\epsilon} - (1-\epsilon)\log\frac{1}{1-\epsilon} =$$

$$(1-2\epsilon)\log\frac{1}{\epsilon} - (1-\epsilon)\log\frac{1}{1-\epsilon}$$

The second inequality follows from the monotonicity of the binary entropy function $-p\log p - (1-p)\log(1-p)$ in $[\frac{1}{2}, 1]$. Putting claims 7 and 8 together ends the proof of Lemma 6 ■.

Proposition 9 *Let $P = (p_1, ...p_m)$ be a probability distribution and $Q_j = (q_{j1}, ..., q_{jm})$, $j = 1, ..., \ell$ be probability distributions such that P is the convex combination of Q_j's: $P = \sum \alpha_j Q_j$, with $\sum \alpha_j = 1$. Then*

$$H(P) = \sum_{j=1}^{\ell} \alpha_j H(Q_j) + \sum_{j=1}^{\ell}\alpha_j \delta(Q_j, P). \quad (1)$$

Proof: We have:

$$H(P) = -\sum_{i=1}^{m} p_i \log p_i = -\sum_{i=1}^{m}\sum_{j=1}^{\ell}\alpha_j q_{ji}\log p_i =$$

$$-\sum_{i=1}^{m}\sum_{j=1}^{\ell}\alpha_j q_{ji}\log\frac{p_i}{q_{ji}}q_{ji} = -\sum_{i=1}^{m}\sum_{j=1}^{\ell}\alpha_j q_{ji}\log q_{ji} +$$

$$\sum_{i=1}^{m}\sum_{j=1}^{\ell}\alpha_j q_{ji}\log\frac{q_{ji}}{p_i} = \sum_{j=1}^{\ell}\alpha_j H(Q_j) + \sum_{j=1}^{\ell}\alpha_j\delta(Q_j, P).$$

We now return to the proof of Lemma 4. We have that $\Delta = \sum_{j=1}^{\ell} q_j \Delta_j$, thus by Proposition 9, $H(\Delta) = \sum_{j=1}^{\ell} q_j H(\Delta_j) + \sum_{j=1}^{\ell} q_j \delta(\Delta_j, \Delta)$, and we need to bound $\sum_{j=1}^{\ell} q_j\delta(\Delta_j, \Delta)$. By the premises of Lemma 4, $\sum_{j=1}^{\ell} q_j \Delta X_j^T \leq \epsilon$ and $\sum_{j=1}^{\ell} q_j \Delta_j X_j^T \geq 1-\epsilon$, where X_j, $j = 1, ..., \ell$ are vectors with entries in $[0,1]$. Let $\beta(j) = (\Delta_j - \Delta)X_j^T$. By the above

$$\sum_j q_j \beta(j) \geq 1 - 2\epsilon \quad (2)$$

Let $U = \{j|\ \beta(j) \geq 1-4\epsilon\}$. As $|\beta(j)| \leq 1$ for every j, (2) implies that $\sum_{j\in U} q_j \geq 1/2$. Moreover, Δ, Δ_j are $1-8\epsilon$ apart for each $j \in U$. Thus by Lemma 6 for each $j \in U$, $\delta(\Delta, \Delta_j) \geq (1-16\epsilon)\log\frac{1}{8\epsilon} - (1-8\epsilon)\log\frac{1}{1-8\epsilon}$, thus $\sum q_j \delta(\Delta_j, \Delta) \geq \frac{1-16\epsilon}{2}\log\frac{1}{8\epsilon} - \frac{1-8\epsilon}{2}\log\frac{1}{1-8\epsilon}$ ■.

Proof of Lemma 5: Let C be the constant in the premise of the Lemma. We start with $D^{(0)} = (1/N, ..., 1/N)$. We will inductively apply the premise of the Lemma to construct a sequence of probability distributions $D^{(0)}, D^{(1)}, \ldots, D^{(k)} \in \mathcal{S}_1$; with the properties,

$$|D^{(0)}|_2 = 1/\sqrt{N} \tag{3}$$

$$100 \log m\sqrt{r}|D^{(i)}|_2 \geq |D^{(i+1)}|_2 \tag{4}$$

$$H(D^{(i+1)}A) \leq H(D^{(i)}A) - C/2 \text{ for } \quad i = 0, 1, \ldots \tag{5}$$

Then, as $(D^{(0)}A)$ is a distribution on $[m]$ we have that $H(D^{(0)}A) \leq \log m$. Thus by property (5) this sequence can be constructed for at most $k = O(\log m)$ (after which the entropy would be negative). However, this would mean that the premises of the Lemma does not hold for this k, namely $|D^{(k)}|_2 \geq 0.01$. Hence, by property (4) we get

$$1/\sqrt{N} \geq \left(\frac{1}{100 \log m\sqrt{r}}\right)^{O(\log m)}$$

and the Lemma follows.

The construction is by an inductive argument, let $D^{(0)} = (1/N, ..., 1/N)$ and assume that we have already constructed $D^{(\ell)}$ such that the sequence $D^{(0)}, D^{(1)}, \ldots, D^{(\ell)}$ satisfies properties (4) and (5). Assume also that $|D^{(\ell)}| < 0.01$ (for $D^{(0)}$ this is satisfied or $1/\sqrt{N} \geq 0.01$ and we are already done). We construct $D^{(\ell+1)}$ in the following way:

Let us denote $D = D^{(\ell)}$. By the premise of the Lemma for D we can write D as a convex combination of r probability distributions $D = \sum_{i=1}^{r} q_i D_i$, and so that $\sum_{i=1}^{r} q_i H(D_i A) \leq H(DA) - C$. Let $L = \{i| \ H(D_i A) \leq H(DA) - C/2\}$. Then,

$$H(DA) - C \geq \sum_{i=1}^{r} q_i H(D_i A) \geq$$

$$\sum_{i \notin L} q_i H(D_i A) \geq (H(DA) - C/2) \sum_{i \notin L} q_i$$

It follows that $H(DA) \sum_{i \in L} q_i \geq \frac{C}{2}$.

As (DA) is a distribution on $[m]$ we have that $H(DA) \leq \log m$. Hence $\sum_{i \in L} q_i \geq \frac{C}{2 \log m}$. On the other hand

$$\sum_{i \in L} q_i |D_i|_2 \leq \sum_{i=1}^{r} q_i |D_i|_2 \leq \sqrt{r}(\sum_{i=1}^{r}(|q_i D_i|_2)^2)^{1/2} \leq$$

$$\sqrt{r}(< \sum_{i=1}^{r} q_i D_i, \sum_{i=1}^{r} q_i D_i >)^{1/2} = \sqrt{r}|D|$$

The first inequality is by Cauchy-Schwarz inequality, the last inequality is by the positivity of q_i's and D_i's. Combining the above we get:

$$\frac{\sum_{i \in L} q_i |D_i|_2}{\sum_{i \in L} q_i} \leq \frac{\sqrt{r}|D|_2}{\sum_{i \in L} q_i} \leq 2 \log m\sqrt{r}|D|_2.$$

But, the left hand side of the above is a convex combination of the $|D_i|$'s, hence we conclude that for some $i \in L$, $|D_i| \leq 2 \log n\sqrt{r}|D|_2$. We define $D^{(\ell+1)} := D_i$ for this particular i. Since $i \in L$ we have $H(D^{(\ell+1)}A) \leq H(D^{(\ell)}A) - C/2$. Thus conditions (2) and (3) hold for $D^{(\ell+1)}$. ∎

Proof of Theorem 2:

We turn back to the proof of the theorem. It follows directly from the combination of Lemmata 3, 4 and 5.

Let A, B, M be matrices as in the theorem. We wish to apply Lemma 5 on A. Let $D \in \mathcal{S}_1$ be any distribution with $|D|_2 < 0.01$. By Lemma 3 for D and the matrices A, B and M we are asserted that D can be written as a convex combination $D = \sum_{i=1}^{m} q_i D_i$ of distributions D_i $i = 1, ..., m$ and so that

1. $\sum_{i=1}^{m} q_i (DA) X_i^T \leq 0.02$,
2. $\sum_{i=1}^{m} q_i (D_i A) X_i^T \geq 0.99$.

Thus applying Lemma 4 on $\Delta = DA$, $\Delta_i = D_i A$, $i = 1, ..., \ell$ and the vectors X_i's with $\ell = m$, and $\epsilon = 0.02$, we conclude that

$\sum_{i=1}^{m} q_i H(D_i A) \leq H(DA) - C(\epsilon)$ for $C(\epsilon) \geq 1/4$. Hence we can apply Lemma 5 for D with $r = m$ and the above C and conclude that $N = m^{O(\log m)}$ as required ■.

5 Further Discussion

5.1 Upper bounds

Here we show how to construct a protocol for equality that meets the lower bound. This construction was noticed earlier by several people, we bring it here for sake of completeness.

The idea is based on a theorem in [5]. A randomized algorithm may be viewed as sampling from a collection of deterministic algorithms. The following is (implicitly) proven in [5], let $\mathcal{A}$ be a constant error randomized algorithm. Assume also that there are altogether $2^{O(n)}$ inputs, then it is enough to sample form a set of $O(n)$ deterministic algorithms in order to maintain the same (constant) error bound.

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, and assume that there is an ϵ-error protocol for f that uses **common** random bits, and whose complexity is bounded by $z = O(\sqrt{n})$ for every input and every random string. Then by the above, there are $O(n)$ deterministic algorithms, such that their average error is bounded by ϵ for every input. This suggests the following protocol with **private** random bits, and complexity $O(z\sqrt{n})$: We index the $O(n)$ deterministic algorithms by $[k] \times [k]$ for appropriate $k = O(\sqrt{n})$. Player A chooses a random index $i \in [k]$ and transmits the output of all the protocols indexed by $\{(i,j) | \ j \in [k]\}$. Similarly, player B chooses a random index $r \in [k]$ and transmits the output of all the protocols indexed by $\{(j,r) | \ j \in [k]\}$. The referee decides according to the output of the protocol indexed by (i,r) which is included in both messages. Clearly the error bound is the same as in the 'common' protocol.

For equality, the protocol that was described in the introduction has $z = O(1)$ maximum complexity for arbitrary constant error. Thus by the above we get a protocol for equality that uses private random bits, has arbitrary constant error, and complexity $O(\sqrt{n})$. A different construction of such a protocol is described in the next subsection.

5.2 A graph theoretic problem

For a graph $G = (V,E)$ and two subsets $S, T \subseteq V$ we denote $E(S,T) = \{(s,t) | \ s \in S, \ t \in T$ and either $s = t$ or $(s,t) \in E\}$. Let $e(S,T) = |E(S,T)|$.

Consider the following question: Let G be a graph on m vertices, with a set of $N = N(m)$ cliques, $\mathcal{Q}$, and so that for each two different cliques $S, T \in \mathcal{Q}$ $e(S,T) \leq \epsilon |S||T|$. How large can N be for a given constant ϵ ?

For the random graph on m vertices with edge probability 0.5 it can be verified that $N = m^{\theta(\log m)}$ (ϵ can be chosen to be arbitrary small constant by choosing appropriate edge probability).

A construction of G implies a 1-sided error protocol for equality with complexity $\log m$, for m such that $N(m) = 2^n$. Let $G = (V,E)$ with $|V| = m$, a collection of cliques $\mathcal{Q}$ of size 2^n, and with parameter ϵ. Then, the following is a protocol for equality whose complexity is $\log m$ and has 0 error for 1-inputs. Every input $x \in \{0,1\}^n$ is associated with a clique $Q_x \in \mathcal{Q}$, in an arbitrary way. Each of the players on input x choose a random vertex $v \in Q_x$ and sends it to the referee. The referee gets two vertices u, v and decides '1' if $(u,v) \in E(G)$ or if $u = v$. This protocol has no error on '1' inputs and has error bounded by ϵ for '0' inputs. If we take G to be that with $N(m) = m^{\theta(\log m)}$ (as guaranteed by the random graph) we get $\theta(\sqrt{n})$ protocol for equality.

By Theorem 2, we conclude that the bound

$N = m^{\theta(\log m)}$ is optimal.

5.3 Positive and Convex ranks

We discuss here some 'rank' related measures for bounds on communication complexity games. For background material the reader is referred to [4, 3].

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, and let F be the $2^n \times 2^n$ associated matrix. A lower on the deterministic communication complexity for multi rounds game is $\log(rank(F))$. A (possibly) better bound that was defined by Yannakakis, [6], is the 'positive-rank'. For a positive $n \times n$ matrix F, the positive-rank of F, denoted by $prank(F)$ is the smallest integer r for which F can be expressed as $F = A \times B^T$ where A, B are positive $m \times r$ matrices.

One can relate these measures to the randomized complexity too. Let F be an $n \times n$ non-negative matrix. We say that a non-negative matrix F' is an *ϵ-approximation* of F if for every (i,j), $|F_{ij} - F'_{ij}| \leq \epsilon$. Let P be a (multi rounds) randomized protocol for a function that is represented by a $2^n \times 2^n$ matrix F, with error probability bounded by ϵ and complexity bounded by c. By [5], we may assume that there are $m = O(n)$ deterministic protocols, $P_1, ..P_m$ whose average error for each input is bounded by ϵ. Let the matrix F_i represents the function that is computed by the ith protocol. By the assumptions $F' = \frac{1}{m}\sum F_i$ is an ϵ-approximation of F. As the complexity is bounded by c, it is easy to see that $prank(F') \leq m \cdot 2^c$. Alternatively, if every ϵ approximation F' of F has $prank(F') \geq r$ then it follows that $c \geq \Omega(\log r - \log n)$.

Our main theorem suggest another measure on non-negative matrices. It bounds the positive rank from above and it lower bounds the complexity of 1-rounds protocols with private bits. For a non-negative $n \times n$ matrix F, let $crank(F)$ (c for convex) be the smallest integer r for which F can be expressed as $F = AMB^T$ where A, B are $n \times r$ stochastic matrices, and M is any $r \times r$ matrix whose entries is in $[0,1]$. Clearly $rank(F) \leq prank(F) \leq crank(F)$.

As shown in section 2 the $crank$ replaces the positive rank for 1-round protocols. Namely, if any ϵ approximation F' of F has $crank(F') \geq r$ then any 1-round protocol for f using private bits has complexity $\Omega(\log r)$.

Theorem 2 asserts that any ϵ approximation for the $n \times n$ identity matrix I_n has $crank$ of at least $2^{\Omega(\sqrt{\log n})}$. On the other hand, it is interesting to observe that the existence of $O(1)$ communication, 1-round randomized protocols for the equality (with common random bits) implies that I_n can be ϵ-approximated by a matrix of $prank = O(\log n)$.

References

[1] J. Bourgain and A. Wigderson, Personal communication.

[2] I. Kremer, N Nissan, D. Ron, On Randomized One-Round Communication Complexity, STOC95.

[3] E. Kuchilevitz, N. Nisan, A survey on Communication Complexity, manuscript 1995.

[4] L. Lovász, Communication complexity: a survey, in Path flowers and VLSI, (Ed, B. Korte, L.Lovász, A. Schrijver,), Springer Verlag, 1990.

[5] I. Newman, Private vs. Common Random bits in communication complexity, Information Processing Letters 39(1991) 67-71.

[6] M. Yannakakis, Expressing combinatorial optimization problems by linear programs, STOC, 1988 pp.223-228.

[7] A. C. Yao, *Probabilistic computation, towards a unified measure of complexity,* Proc. 18th IEEE Symp. on Foundation of Computer Science, (1977), pp. 222–227.

[8] A. C. Yao, Some complexity questions related to distributive computing, STOC, 1979, pp. 209-213.