FOURIER ANALYSIS FOR PROBABILISTIC COMMUNICATION COMPLEXITY

Ran Raz

Abstract. We use Fourier analysis to get general lower bounds for the probabilistic communication complexity of large classes of functions. We give some examples showing how to use our method in some known cases, and for some new functions.

Our main tool is an inequality by Kahn, Kalai and Linial, derived from two lemmas of Beckner.

Key words. Lower-bound, Complexity, Fourier-Analysis.

Subject classifications. 68Q99.

1. Introduction and Notations

1.1. Communication Complexity. The model of **communication complexity** was first introduced by Yao [9]. In this model, we consider two finite sets X and Y, and a function $f : X \times Y \to \{0,1\}$. Given such a function, consider the following game between players I and II: for $(x,y) \in X \times Y$, give x to player I, and y to player II. Their goal is to compute the value of $f(x,y)$. Initially, each player has no information about the input of the other. In each step, one of the players sends one bit of information (about his input) to the other player. In the end, they both have to know the value of $f(x,y)$.

A **strategy** for the two players describes (in each step): 1) A way for the two players to agree on a player that will speak in this step (this decision is based only on the messages already transmitted - so it is a function from the set of all possible messages to the set {I,II}). 2) A way for this player to send one bit of information (based on his input, and on all the messages already transmitted).

A **protocol** for the game is a strategy for the two players, such that after the last step they both know the value of $f(x, y)$. The maximum number of steps in the protocol is called the **communication complexity of the protocol** (where the maximum is taken over all the possible inputs). This is because this number is the maximal number of bits exchanged between the two players, by the protocol.

The **deterministic communication complexity** of f is the minimum number of bits the players must exchange between them (for the worst case input), in any deterministic protocol (i.e. the communication complexity of the best protocol). The probabilistic case differs from the deterministic case in allowing the protocol to depend on flips of a coin (so the functions of the strategy may be probabilistic), and in allowing the protocol to make errors (in computing the value of $f(x, y)$). The probability of error must be smaller than a constant ϵ, where $0 \leq \epsilon < \frac{1}{2}$, **for every input pair** (x, y). As long as ϵ is a constant, and $0 < \epsilon$, its exact value is of less importance, and may affect the communication complexity only by a constant.

The model of communication complexity was studied in many papers, and seems to have close relationships with certain other complexity issues (see [7] for a survey). A lot of effort went into proving lower bounds in the different cases.

In the theory of deterministic communication complexity there are many general methods for proving lower bounds. When faced with a new problem, one can usually just check if one of these methods is applicable. The situation in the theory of probabilistic communication complexity is much worse: there are some very impressive lower bounds (like [4], [6], [8]), but most of the techniques work only for very specific functions. There are no powerful general methods, and almost no results about large classes of natural functions. Moreover, all the known lower bounds for functions are proved only in cases in which there are no big monochromatic sub-matrices of zeros, or no big monochromatic sub-matrices of ones (in the input matrix of the function f). Almost all the known methods strongly depend on the property that for some measure on the inputs for f, the function f cannot be approximated by any deterministic (or non deterministic) communication protocol (the "meta-method" below). These methods cannot prove lower bounds for functions that do not have this property.

The sequence of bits, transmitted by a deterministic communication protocol, for the input pair (x, y), is called the **history** of (x, y). The set of all input pairs that have the same history is a sub-matrix of $X \times Y$. A deterministic communication protocol naturally partitions the set $X \times Y$ into sub-matrices, corresponding to the different possible histories of the protocol. In the same way, a probabilistic protocol creates a **cover** of $X \times Y$ by sub-matrices, corresponding to the different possible histories (here we consider the random string of the protocol as a part of the history). Now each sub-matrix has a weight (its probability to appear in the protocol, i.e. the probability of the corresponding random string), and the sum of the weights of all the sub-matrices, covering one point, is always one. One can easily show that the communication complexity of a deterministic protocol is at least $\log m$, where m is the number of sub-matrices in the cover. For probabilistic protocols the equivalent bound is $\log \sum w_i$, where w_i is the weight of the i'th sub-matrix. So $\sum w_i$ is the probabilistic equivalent to the number of sub-matrices. Dealing with the probabilistic case, we will sometimes talk about $\sum w_i$ as the number of sub-matrices in the cover.

Denote by r_i the characteristic function of the i'th sub-matrix. Denote by R the set of indices of all sub-matrices corresponding to histories on which the protocol answers by 1. Define

$$h = \sum_{i \in R} w_i r_i.$$

Then, by the definition of probabilistic communication complexity, it is easy to prove that $\| f - h \|_\infty$ is the maximal error of the protocol, and therefore

$$\| f - h \|_\infty \leq \epsilon. \tag{1.1}$$

The main "meta-method" for proving lower bounds for probabilistic communication complexity was introduced in [10],[2]. Almost all previous lower bounds for probabilistic communication complexity use this "meta-method". According to this "meta-method", the lower bound is based on the following main steps:

(1) Fix some probability measure μ on the inputs for f, and assume (for a contradiction) that we have a probabilistic communication protocol for f, with communication complexity K, and error ϵ. By (1.1) we have

$$\| f - h \|_1 \leq \| f - h \|_\infty \leq \epsilon, \tag{1.2}$$

where the 1-norm is taken according to the probability measure μ.

(2) From (1.2), it is easy to conclude the existence of (at least one) history sub-matrix $A \times B$, such that:

$$\mu(A \times B) \geq 2^{-(K+1)}$$

(i.e. $A \times B$ is large), and there exists $\delta \in \{-1, 1\}$ s.t.

$$Prob_{\mu}[(2f-1)r_{A\times B} = \delta] \geq (1-2\epsilon)\mu(A \times B),$$

where $r_{A\times B}$ is the characteristic function of $A \times B$. (i.e. the f-discrepancy of $A \times B$ is large, comparing to its size (according to the measure μ)).

(3) A contradiction is then derived by proving the non-existence of a large sub-matrix with large f-discrepancy, compared to its size (according to the measure μ) - usually, this is the main part of the proof. In almost all previous lower bounds, this is actually proved by proving the non-existence of sub-matrices with large f-discrepancy.

The previous lower-bound-technique leaves two main gaps: The first gap is that the existence of a large sub-matrix with large f-discrepancy is a much weaker assumption than the existence of a probabilistic communication protocol. First, (1.2) is weaker than (1.1). Also, the existence of a large sub-matrix with large f-discrepancy is weaker than (1.2). What would be most desirable is to find a way to use (1.1) directly. However, this looks harder, because sub-matrices are much simpler combinatoric objects than protocols. Therefore, it is not surprising that most lower bounds were achieved by analyzing sub-matrices. The reason for using (1.2) instead of (1.1) is that in the 1-norm, the contribution of the discrepancy of the different sub-matrices is additive, and therefore is easier to analyze.

The second gap in the previous lower-bound-technique is in the third step. Most methods derived so far actually prove the non-existence of sub-matrices with large f-discrepancy, but cannot prove lower bounds for cases in which there are sub-matrices with large f-discrepancy, but with small f-discrepancy compared to their size.

In this paper we use Fourier analysis to get general lower bounds for the probabilistic communication complexity of functions. We introduce two methods. Both methods address the second gap in the previous technique. The first method (the main one) address also the first gap. The main method we introduce, described in section 2, gives explicit lower bounds as a function of the sum of some Fourier coefficients. These lower bounds do not necessarily depend on the non-existence of large sub-matrices with large f-discrepancy (the "meta-method" above), but on the weaker property that some subset of the Fourier coefficients of f cannot be approximated. The method we use is based on the following main steps:

(1) Assume (for a contradiction) that we have a probabilistic communication protocol for f, with complexity K, and error ϵ. By (1.1) we have

$$\| f - h \|_2 \leq \| f - h \|_\infty \leq \epsilon. \tag{1.3}$$

(2) Denote by f_S the Fourier coefficient (corresponding to S) of the function f, and assume that for some set E : $\sum_{S \in E} |f_S|$ is large. Then it is proved (using the identity of Parseval) that $\sum_{S \in E} |h_S|$ is also large. More specifically, it is proved that $\sum_{S \in E} |h_S|$ is bounded from below by $2C$, where C depends mainly on $\sum_{S \in E} |f_S|$.

(3) Now we concentrate on some specific sets E (we consider sets that are subsets of $\{(S, S') : S = S'\}$, however, as explained in the next section, the use of other sets is also possible), and assume (a theorem assumption) that $\sum_{S \in E} |f_S|$ is large (say for one such E). Therefore by the previous step, we know that $\sum_{S \in E} |h_S|$ is also large. To derive a contradiction, we prove that the contribution of every sub-matrix to $\sum_{S \in E} |h_S|$ is small compared to its size, and therefore $\sum_{S \in E} |h_S|$ cannot be large.

The "meta-method" we use can be described as a generalization of the [10],[2] "meta-method" described above: instead of proving that the contribution of a large sub-matrix to the f-discrepancy is small (compared to its size), we prove that the contribution of a large sub-matrix to some set of Fourier coefficients is small (compared to its size). (The use of (1.3) instead of (1.2) is of less importance. We use (1.3), because this way we are able to use the identity of Parseval, and to analyze the contribution of the different Fourier coefficients). Another important difference is that (as described before) in most of the previous bounds, what people really proved is that the **absolute** discrepancy of every sub-matrix (large or small) is small (see [4]). In our case, by using the non-trivial inequality of Kahn Kalai, and Linial [5], we are able to deal also with large sub-matrices that have relatively large **absolute** contribution (but relatively small contribution compared to their size).

For instance, we prove in section 2 that: $C(f) = \Omega(\lambda / \log n)$, where $C(f)$ denotes the probabilistic communication complexity of f, and

$$\lambda = \sum_i \left| \mathrm{Prob}[f(x,y) = 0 \mid x_i = y_i] - \mathrm{Prob}[f(x,y) = 0 \mid x_i \neq y_i] \right|.$$

Moreover, if $\lambda = \Omega(\sqrt{n})$ then $C(f) = \Omega(\lambda)$.

We describe some applications in section 3. In section 4 we apply the Kahn-Kalai-Linial inequality in a different manner. The results (of section 4) will be usually weaker, and more difficult to use, although we obtain better bounds in some special cases.

1.2. Fourier Analysis. In this paper we consider functions $f : \{0,1\}^n \to R$. We think of the elements in $\{0,1\}^n$ as n-dimensional vectors or as sets. We identify each set with its characteristic vector, and use them interchangeably. On $\{0,1\}^n$ we take the uniform distribution, and define inner products according to this distribution:

$$< f, g >= \frac{1}{2^n} \sum_T f(T)g(T).$$

We identify the cube $\{0,1\}^n$ with the group Z_2^n, and work with the Fourier transform according to this group. The characters for Z_2^n are the functions $\{U_S\}_{S\subset[n]}$, where $U_S : \{0,1\}^n \to R$ is defined by

$$U_S(T) = (-1)^{|S\cap T|}$$

(here and throughout $[n]$ denotes $\{1,\ldots,n\}$). The Fourier transform of f : $\{0,1\}^n \to R$ is defined by

$$f = \sum_S f_S U_S,$$

where f_S is the Fourier coefficient

$$f_S =< f, U_S > .$$

We also consider functions $f : \{0,1\}^n \times \{0,1\}^n \to R$. The elements of $\{0,1\}^{n+n}$ are the pairs (T, T'), where $T, T' \subset [n]$. Again we identify each set with its characteristic vector, and use them interchangeably. On $\{0,1\}^{n+n}$ we take again the uniform distribution, and define inner products according to it. The character $U_{S,S'}$ is defined by

$$U_{S,S'}(T, T') = (-1)^{|S\cap T|+|S'\cap T'|}.$$

The Fourier transform is now

$$f = \sum_{S,S'} f_{S,S'} U_{S,S'},$$

where

$$f_{S,S'} =< f, U_{S,S'} > .$$

To every possible history in a communication protocol there corresponds a sub-matrix of input pairs that have this history. Given $A \times B \subset \{0,1\}^n \times \{0,1\}^n$, we denote by I_A and I_B the characteristic functions of A and B, and by $\{\alpha_S\}, \{\beta_S\}$ the Fourier coefficients of these functions. By α and β we denote the probability of the sets A and B, i.e. $\alpha = \alpha_0 = |A|/2^n$ and $\beta = \beta_0 = |B|/2^n$.

One of the basic tools in our analysis is the following lemma:

LEMMA 1.1. *[5]*
Let f be a function $f : \{0,1\}^n \to \{-1,0,1\}$ (for example the characteristic function of a set). Let t be the probability that $f \neq 0$. Then for every $0 \leq \delta \leq 1$

$$\sum_S \delta^{|S|} f_S^2 \leq t^{\frac{2}{1+\delta}}.$$

2. Our general Method

Special subsets of the set of all the pairs (S, S') are the sets $V = \{(S, S') : S = S'\}$, and $V_k = \{(S, S) : |S| = k\} \subset V$. We think of V, V_k as sets of indices of characters. The method we introduce in this section considers subsets $E \subset V_k$ (in subsections 2.1,2.2), or $E \subset V$ (in subsection 2.3). Basically, we prove that it is hard to approximate the Fourier coefficients, corresponding to these sets E, by communication protocols. We prove that if f has large Fourier coefficients in the set E, then the probabilistic communication complexity of f is large. The method can be generalized to different kinds of sets E (and not only $E \subset V_k$ or $E \subset V$). As an example, one can apply any linear permutation on the set of all possible inputs of one of the players. This doesn't change the nature (and the communication complexity) of the function f. However, it changes the possible sets E that we are able to consider. In fact, one can also apply non-linear permutations, and obtain different Fourier coefficients.

2.1. The Main Theorem. Define a function $L_k : R_+ \to R_+$ in the following way: for $x \geq e^k$ $\quad L_k(x) = kx^{1/k}/e$, and for $x \leq e^k$ $\quad L_k(x) = ln(x)$.

THEOREM 2.1.
Let f be a function $f : \{0,1\}^{n+n} \to \{0,1\}$, and let $E \subset V_k$. Define

$$\sigma_0 = |E| \quad , \quad \sigma_1 = \sum_{(S,S') \in E} |f_{S,S'}|.$$

Then if $\sigma_1 \geq \sqrt{\sigma_0}$

$$C(f) \geq \Omega(L_k(\sigma_1)),$$

and if $\sigma_1 < \sqrt{\sigma_0}$

$$C(f) \geq \Omega(L_k(\sigma_1)/(\log \sqrt{\sigma_0} - \log \sigma_1 + 1)),$$

where $C(f)$ *is the probabilistic communication complexity of* f*.*

PROOF. Let P be a protocol for f that makes an error with probability at most 2^{-d}, for every input. Let m be the complexity of P. Denote by $h : \{0,1\}^{2n} \to R$ the function $h(x,y) = \mathrm{Prob}[P(x,y) = 1]$. As described in the introduction, the function h is created by a weighted sum of the characteristic functions of all the history sub-matrices of the protocol, that give an answer 1 (the sum is weighted because the protocol is probabilistic). The protocol P makes an error with probability at most 2^{-d}, and hence

$$\| f - h \| \leq \mathrm{MAX}|f(x,y) - h(x,y)| \leq 2^{-d},$$

where the norm is the 2-norm. By the identity of Parseval

$$\sum_{S,S'} (f_{S,S'} - h_{S,S'})^2 = \| f - h \|^2 \leq 2^{-2d}.$$

First, we would like to derive a lower bound for $\sum_E |h_{S,S'}|$. We are interested in the smallest $\sum_E |h_{S,S'}|$ for which the inequality can still hold. For fixed $\sum_E |h_{S,S'}|$, the minimum of $\sum_E (f_{S,S'} - h_{S,S'})^2$ is achieved when all the terms $|f_{S,S'} - h_{S,S'}|$ are equal (denote their value by λ). Then $\sigma_0 \lambda^2 \leq 2^{-2d}$, and we have $\lambda \leq 2^{-d}/\sqrt{\sigma_0}$. Hence it is clear that always

$$\sum_E |h_{S,S'}| \geq \sum_E (|f_{S,S'}| - \lambda) = \sum_E |f_{S,S'}| - \sigma_0 \lambda \geq \sigma_1 - 2^{-d}\sqrt{\sigma_0}.$$

Define $C = (\sigma_1 - 2^{-d}\sqrt{\sigma_0})/2$ to get

$$\sum_E |h_{S,S'}| \geq 2C.$$

The term $\sum_E |h_{S,S'}|$ is created by contributions from all the sub-matrices (corresponding to all the different possible histories). Let $A \times B \subset \{0,1\}^n \times \{0,1\}^n$ be such a sub-matrix. Its contribution to $\sum_E |h_{S,S'}|$ is $\sum_{(S,S) \in E} |\alpha_S \beta_S|$ (weighted by the sub-matrix weight). We call $A \times B$ regular if

$$\sum_{(S,S) \in E} |\alpha_S \beta_S| \geq C \alpha \beta,$$

i.e. if its contribution is not small compared to its size. This definition ensures that at least half of $\sum_E |h_{S,S'}|$ is created by the contribution of regular sub-matrices.

We examine two cases:

Case a: $1 < C \leq e^{2k}$. This is the simpler case. By the identity of Parseval, we have $\sum \alpha_S^2 \leq 1$, $\sum \beta_S^2 \leq 1$, and hence by the Cauchy-Schwartz inequality $\sum_E |\alpha_S \beta_S| \leq 1$. This is true for every sub-matrix (corresponding to every possible history of the protocol), i.e. there must exist many sub-matrices. We know that $\sum_E |h_{S,S'}| \geq 2C$, and therefore the lower bound we get (by a standard argument) is $m \geq \Omega(\log C)$.

Case b: $C \geq e^{2k}$. Let $A \times B \subset \{0,1\}^n \times \{0,1\}^n$ be a regular sub-matrix. Define c by the equation $\sum_{(S,S) \in E} |\alpha_S \beta_S| = c\alpha\beta$. So $c \geq C$. By the Cauchy-Schwartz inequality we get

$$\sqrt{\sum_{(S,S)\in E} \alpha_S^2}\sqrt{\sum_{(S,S)\in E} \beta_S^2} \geq \sum_E |\alpha_S \beta_S| = c\alpha\beta.$$

So $\sum_E \alpha_S^2 \geq c\alpha^2$, or $\sum_E \beta_S^2 \geq c\beta^2$. W.l.o.g assume $\sum_E \alpha_S^2 \geq c\alpha^2$. By Lemma 1.1, for every $0 \leq \delta \leq 1$

$$\alpha^{\frac{2}{1+\delta}} \geq \delta^k \sum_E \alpha_S^2 \geq c\delta^k \alpha^2.$$

Hence, $\alpha \leq (c\delta^k)^{-\frac{1}{2\delta}}$. By substituting $\delta = e/c^{1/k}$, we get $\alpha \leq e^{-kc^{1/k}/2e}$. So the sub-matrix is small ! and its contribution $\sum_E |\alpha_S \beta_S|$ is at most $ce^{-kc^{1/k}/2e}$. In this case the fact that the function $g(c) = ce^{-kc^{1/k}/2e}$ decreases in the range $c \geq e^{2k}$ implies that this contribution is at most $Ce^{-kC^{1/k}/2e}$ (weighted by the sub-matrix weight). But, at least half of $\sum_E |h_{S,S'}|$ (which is at least C) is created by the contribution of regular sub-matrices (we defined them to satisfy this property), so their number must be large and we get by a standard argument a lower bound of $\Omega(\log(e^{-kC^{1/k}/2e})) = \Omega(kC^{1/k})$.

If $C \geq e^{2k}$ (*case (b)*) then $L_k(C) = kC^{1/k}/e$. If $1 < C \leq e^{2k}$ (*case (a)*) then $\log C = \Omega(L_k(C))$. So, in any case we have

$$m \geq \Omega(L_k(C)).$$

Suppose now that P is a protocol that makes an error with probability at most $1/4$, for every input. Let m be the complexity of P. Apply the protocol P $O(d)$ times to get a protocol that makes an error with probability at most 2^{-d}. Denote $C_d = (\sigma_1 - 2^{-d}\sqrt{\sigma_0})/2$ to get

$$\forall C_d > 1 \qquad O(d)m \geq \Omega(L_k(C_d)).$$

If $\sigma_1 \geq \sqrt{\sigma_0}$ we choose $d = 1$ and get

$$m \geq \Omega(L_k(\sigma_1)).$$

If $\sigma_1 < \sqrt{\sigma_0}$ we choose $d \approx \log \sqrt{\sigma_0} - \log \sigma_1 + 1$ to get

$$m \geq \Omega(L_k(\sigma_1)/(\log \sqrt{\sigma_0} - \log \sigma_1 + 1)).$$

□

2.2. The Special Case $k = 1$. Let f be a function $f : \{0,1\}^{2n} \rightarrow \{0,1\}$. Define

$$P_i = |\mathrm{Prob}[f(x,y) = 0, x_i \oplus y_i = 0] - \mathrm{Prob}[f(x,y) = 0, x_i \oplus y_i = 1]|,$$

where the probability space is the set of possible inputs. P_i measures the influence of $x_i \oplus y_i$ on the probability that $f = 0$. Define $\lambda = \sum_i |P_i|$, then by taking in Theorem 2.1 $E = V_1$, we have $\sigma_0 = n$, $\sigma_1 = \lambda$, and we get the lower bound

$$C(f) = \Omega(\lambda / \log n)$$

always, and the lower bound

$$C(f) = \Omega(\lambda)$$

if $\lambda = \Omega(\sqrt{n})$.

In the special case where $\forall i P_i = P$, the lower bound is

$$C(f) = \Omega(nP/\log n)$$

always, and

$$C(f) = \Omega(nP)$$

if $nP = \Omega(\sqrt{n})$. An example for this special case is when $f = \tilde{f}(x \oplus y)$, or $f = \tilde{f}(x \wedge y)$, or $f = \tilde{f}(x \vee y)$, where $\tilde{f}$ is any function with a transitive automorphism group (i.e. for any two coordinates $1 \leq i, j \leq n$, there exists a permutation $\pi \in S_n$ such that $\pi(i) = j, \pi(j) = i$, and such that permuting the coordinates of z, according to π, doesn't change $f(z)$). This can give lower bounds for functions f that are properties of the XOR, or the $Union$, or the $Intersection$ of two graphs (by definition, because the associated function $\tilde{f}$ is a property of a graph it doesn't change by permuting the vertices of the graph).

In the special case where $f = \tilde{f}(x \oplus y)$ and $\tilde{f}$ is monotone, P_i is the influence on $\tilde{f}$ in the sense of [3] and [5].

2.3. The Case $E \subset V$. In this subsection, we consider subsets $E \subset V$. For this case we prove a version of Theorem 2.1, that gives better lower bounds in some cases. Actually, the proof of this theorem is contained in the proof of Theorem 2.1, and is much simpler. It uses only *case (a)* of the proof of Theorem 2.1.

THEOREM 2.2.
Let f be a function $f : \{0,1\}^{n+n} \to \{0,1\}$, and let $E \subset V$. Define

$$\sigma_0 = |E| \ , \ \ \sigma_1 = \sum_{(S,S')\in E} |f_{S,S'}|.$$

Then if $\sigma_1 \geq \sqrt{\sigma_0}$

$$C(f) \geq \Omega(\log(\sigma_1)),$$

and if $\sigma_1 < \sqrt{\sigma_0}$

$$C(f) \geq \Omega(\log(\sigma_1)/(\log\sqrt{\sigma_0} - \log\sigma_1 + 1)).$$

PROOF. Again, let P be a protocol for f that makes an error with probability at most 2^{-d}, for every input. Let m be the complexity of P. Denote by $h : \{0,1\}^{2n} \to R$ the function $h(x,y) = \text{Prob}[P(x,y) = 1]$. Define $C = (\sigma_1 - 2^{-d}\sqrt{\sigma_0})/2$ to get (in the same way as in Theorem 2.1)

$$\sum_E |h_{S,S'}| \geq 2C.$$

By the identity of Parseval, we have $\sum \alpha_S^2 \leq 1$, $\sum \beta_S^2 \leq 1$, and hence $\sum_E |\alpha_S \beta_S| \leq 1$. This is true for every sub-matrix (corresponding to every possible history of the protocol), i.e. there must exist many sub-matrices. We know that $\sum_E |h_{S,S'}| \geq 2C$, and therefore the lower bound we get (by a standard argument) is $m \geq \Omega(\log C)$.

Suppose now that P is a protocol that makes an error with probability at most $1/4$, for every input. Let m be the complexity of P. Apply the protocol P $O(d)$ times to get a protocol that makes an error with probability at most 2^{-d}. Denote $C_d = (\sigma_1 - 2^{-d}\sqrt{\sigma_0})/2$, to get

$$\forall C_d > 1 \qquad O(d)m \geq \Omega(\log(C_d)).$$

If $\sigma_1 \geq \sqrt{\sigma_0}$ we choose $d = 1$, and get

$$m \geq \Omega(\log(\sigma_1)).$$

If $\sigma_1 < \sqrt{\sigma_0}$ we choose $d \approx \log\sqrt{\sigma_0} - \log\sigma_1 + 1$ to get

$$m \geq \Omega(\log(\sigma_1)/(\log\sqrt{\sigma_0} - \log\sigma_1 + 1)).$$

□

3. Some Examples

The theorems we just proved can give lower bounds for many functions. Sometimes these lower bounds are tight, and sometimes they are not. Here we describe three examples in detail, showing the generality of our method, and demonstrating the way to use it.

3.1. The size of set intersection. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the following:

$$f(T, T') = 1 \quad \text{iff} \quad |T \cap T'| > n/4.$$

The communication complexity of this function is higher than the communication complexity of the *Disjointness* function (by a simple reduction). It was proved that the probabilistic communication complexity of *Disjointness* is $\Omega(\sqrt{n})$ by [2], and $\Omega(n)$ by [6] (see also [8]). Here we prove that the probabilistic communication complexity $C(f)$ is $\Omega(\sqrt{n})$ (a weaker bound), as an application of Theorem 2.1:
It is an easy calculation to check that if $|S| = 1$ then $f_{S,S} = \Omega(1/\sqrt{n})$. Hence, taking in Theorem 2.1 $E = V_1$, we have $\sigma_0 = n$, $\sigma_1 = \Omega(\sqrt{n})$, and the lower bound we get is $C(f) \geq \Omega(\sqrt{n})$. As we said, this lower bound is not tight.

3.2. The Hamming weight of the XOR of two vectors. Assume that n is divisible by four. Denote $m = n/2$.
Let $\tilde{f} : \{0,1\}^n \to \{0,1\}$ be defined by:

$$\tilde{f}(T) = 1 \quad \text{iff} \quad |T| = m.$$

Let $f : \{0,1\}^{n+n} \to \{0,1\}$ be defined by

$$f(T, T') = \tilde{f}(T \oplus T').$$

A tight lower bound of $\Omega(n)$ can be proved for the probabilistic communication complexity $C(f)$, by a reduction from the *Disjointness* function (tight lower bounds for the *Disjointness* function where proved in [6],[8]). Here we show how to derive a weaker $\Omega(n/\log n)$ bound, as an application of Theorem 2.1.
For every k, $\tilde{f}_S$ is the same for all S with $|S| = k$. $2^n \tilde{f}_S$ is simply the number of subsets of size m that intersect a fixed subset of size k in an even number of points, minus the number of subsets of size m that intersect this fixed subset of size k at odd number of points. Summing over all S, with $|S| = k$, we have the number of pairs of subsets (size m,size k) that intersect in an even number

of points, minus the number of pairs of subsets (size m,size k) that intersect at odd number of points. Therefore a straight forward calculation gives:

$$2^n \sum_{S:|S|=k} \tilde{f}_S = \binom{n}{m} \sum_{i=0}^{k} \binom{m}{i} \binom{m}{k-i} (-1)^i.$$

$\sum_{i=0}^{k} \binom{m}{i} \binom{m}{k-i} (-1)^i$ is the coefficient of x^k in the polynomial $(x+1)^m(1-x)^m = (1-x^2)^m$. This coefficient is 0 if k is odd, and its absolute value is $\binom{m}{k/2}$ if k is even. By the definition of f: $f_{S,S'} = \tilde{f}_S \delta_{S,S'}$. Take in Theorem 2.1 $k = m$, and $E = V_k$, to get

$$\sigma_0 = \binom{n}{m} \quad , \quad \sigma_1 = \binom{n}{m} \binom{m}{k/2} / 2^n.$$

Then $\log(\sigma_1) = \Omega(n)$, and $\log(\sigma_1) - \log(\sqrt{\sigma_0}) = O(\log n)$, and the lower bound we get is $C(f) = \Omega(n/\log n)$. As we said, this is not tight.

3.3. The scalar product of the XOR of two pairs of vectors. Let $\tilde{f} : \{0,1\}^{2n} \to \{-1,1\}$ be defined by

$$\tilde{f}(T,Q) = (-1)^{|T \cap Q|}$$

(up to constants, this is just the scalar product of T and Q) .
Let $f : \{0,1\}^{2n+2n} \to \{-1,1\}$ be defined by

$$f(T,Q,T',Q') = \tilde{f}(T \oplus T', Q \oplus Q').$$

We will prove here a tight lower bound of $\Omega(n)$ for the probabilistic communication complexity $C(f)$, as an application of Theorem 2.2 (the matching upper bound is achieved by a trivial protocol). By definition $\tilde{f}(T,Q) = \prod_{i=1}^{n} g(T_i, Q_i)$, where $g : \{0,1\}^2 \to \{-1,1\}$ is defined by $g(0,0) = g(0,1) = g(1,0) = 1$; $g(1,1) = -1$. The Fourier coefficients of g are $g_{00} = g_{01} = g_{10} = \frac{1}{2}$; $g_{11} = -\frac{1}{2}$. Hence,

$$\tilde{f}_{S,R} = 2^{-n}(-1)^{|S \cap R|},$$

i.e. $|\tilde{f}_{S,R}| = 2^{-n}$. By the definition of f: $f_{S,R,S',R'} = \tilde{f}_{S,R}\delta_{S,S'}\delta_{R,R'}$, and therefore

$$\forall S,R \quad |f_{S,R,S,R}| = 2^{-n}.$$

Take in Theorem 2.2 $E = V$, to get $\sigma_0 = 2^{2n}$, $\sigma_1 = 2^n$, and the lower bound we get is $C(f) \geq \Omega(n)$.
(Notice that f gets the values $\{-1,1\}$, instead of $\{0,1\}$. So actually we use Theorem 2.2 for the function $h = (f+1)/2$. Each Fourier coefficient of h is half of the corresponding coefficient of f, except for the first coefficient. So the calculation is the same.)

4. A Slightly Different Method

4.1. One More Theorem. For a function $f : \{0,1\}^{n+n} \to R$, define

$$f_{S,*} = \sum_{S'} |f_{S,S'}| \quad ; \quad f_{*,S'} = \sum_{S} |f_{S,S'}|.$$

In the special case where $f(T,T') = \tilde{f}(T \oplus T')$ (for some $\tilde{f} : \{0,1\}^n \to R$) $f_{S,S'} = \tilde{f}_S \delta_{S,S'}$. Therefore $f_{S,*} = |\tilde{f}_S|$ and $f_{*,S'} = |\tilde{f}_{S'}|$. The method we introduce in this section considers the coefficients $f_{S,*}$, $f_{*,S'}$, and is based on Theorem 4.1. Roughly speaking the method implies that if for small S, S' we have $f_{S,*} < m^{-|S|}$, $f_{*,S'} < m^{-|S'|}$, and for large S, S' we have $f_{S,*} < m^{|S|-n}$, $f_{*,S'} < m^{|S'|-n}$ then usually the probabilistic communication complexity of f is $\Omega(m)$. This method works basically for functions of the XOR of the two input vectors, but can also work for other functions.

THEOREM 4.1.
Let f be a function $f : \{0,1\}^{n+n} \to R$, and let $0 < d_1, d_2, d_3, c_1, c_2, c_3, \delta \le 1$, such that

$$\begin{aligned} f_{S,*}\ ,\ f_{*,S} &\le c_1\delta^{|S|} + c_2 + c_3\delta^{n-|S|}, \\ f_{\bar{0},*}\ ,\ f_{*,\bar{0}} &\le c_1\delta^0 - d_1 + c_2 - d_2 + c_3\delta^n = c_1 - d_1 + c_2 - d_2 + c_3\delta^n, \\ f_{\bar{1},*}\ ,\ f_{*,\bar{1}} &\le c_1\delta^n + c_2 + c_3\delta^0 - d_3 = c_1\delta^n + c_2 + c_3 - d_3. \end{aligned}$$

Let $A, B \subset \{0,1\}^n$ such that $\alpha_{\bar{0}}^2 = \alpha_{\bar{1}}^2$ and $\beta_{\bar{0}}^2 = \beta_{\bar{1}}^2$, and let $\epsilon > 0$ such that

$$\left| \sum_{A \times B} f \right| \ge \epsilon \alpha_0 \beta_0 = \epsilon \alpha \beta$$

then

$$\alpha\beta = \alpha_0\beta_0 \le MAX \left[\left(\frac{c_1}{d_1 + \frac{1}{3}\epsilon} \right)^{\frac{1}{2\delta}} \ , \ \frac{c_2}{d_2 + \frac{1}{3}\epsilon} \ , \ \left(\frac{c_3}{d_3 + \frac{1}{3}\epsilon} \right)^{\frac{1}{2\delta}} \right].$$

PROOF. The Fourier coefficient of $I_{A\times B}$ corresponding to (S, S') is $\alpha_S \beta_{S'}$. Therefore, by the identity of Parseval

$$\epsilon\alpha_0\beta_0 \le |\sum_{A\times B} f| = | < f, I_{A\times B} > | = |\sum_{S,S'} f_{S,S'}\alpha_S\beta_{S'}|,$$

By the Cauchy-Schwartz inequality

$$\left(\sum_{S,S'} f_{S,S'}\alpha_S\beta_{S'}\right)^2 \le$$

$$\left(\sum_{S,S'} |f_{S,S'}|\alpha_S^2\right)\left(\sum_{S,S'} |f_{S,S'}|\beta_{S'}^2\right) = \left(\sum_S f_{S,*}\alpha_S^2\right)\left(\sum_{S'} f_{*,S'}\beta_{S'}^2\right) \le$$

$$\left(c_1\sum_S \delta^{|S|}\alpha_S^2 - d_1\alpha_0^2 + c_2\sum_S \alpha_S^2 - d_2\alpha_0^2 + c_3\sum_S \delta^{n-|S|}\alpha_S^2 - d_3\alpha_{\bar{1}}^2\right)$$

$$\left(c_1\sum_{S'} \delta^{|S'|}\beta_{S'}^2 - d_1\beta_0^2 + c_2\sum_{S'} \beta_{S'}^2 - d_2\beta_0^2 + c_3\sum_{S'} \delta^{n-|S'|}\beta_{S'}^2 - d_3\beta_{\bar{1}}^2\right).$$

By Lemma 1.1 we have

$$\sum_S \delta^{|S|}\alpha_S^2 \le \alpha_0^{\frac{2}{1+\delta}}, \quad \text{and} \quad \sum_{S'} \delta^{|S'|}\beta_{S'}^2 \le \beta_0^{\frac{2}{1+\delta}}.$$

And clearly, we also have

$$\sum_S \delta^{n-|S|}\alpha_S^2 \le \alpha_0^{\frac{2}{1+\delta}}, \quad \text{and} \quad \sum_{S'} \delta^{n-|S'|}\beta_{S'}^2 \le \beta_0^{\frac{2}{1+\delta}}.$$

By the identity of Parseval, $\sum_S \alpha_S^2 = \alpha_0$, and $\sum_{S'} \beta_{S'}^2 = \beta_0$. Recall also that $\alpha_{\bar{1}}^2 = \alpha_0^2$, and $\beta_{\bar{1}}^2 = \beta_0^2$. We get

$$\begin{aligned}\epsilon^2\alpha_0^2\beta_0^2 \le\ & (c_1\alpha_0^{\frac{2}{1+\delta}} - d_1\alpha_0^2 + c_2\alpha_0 - d_2\alpha_0^2 + c_3\alpha_0^{\frac{2}{1+\delta}} - d_3\alpha_0^2)\\ & (c_1\beta_0^{\frac{2}{1+\delta}} - d_1\beta_0^2 + c_2\beta_0 - d_2\beta_0^2 + c_3\beta_0^{\frac{2}{1+\delta}} - d_3\beta_0^2).\end{aligned}$$

Therefore at least one of the following six inequalities is true:

$$\begin{aligned}
\frac{1}{3}\epsilon\alpha_0^2 &\le c_1\alpha_0^{\frac{2}{1+\delta}} - d_1\alpha_0^2,\\
\frac{1}{3}\epsilon\alpha_0^2 &\le c_2\alpha_0 - d_2\alpha_0^2,\\
\frac{1}{3}\epsilon\alpha_0^2 &\le c_3\alpha_0^{\frac{2}{1+\delta}} - d_3\alpha_0^2,\\
\frac{1}{3}\epsilon\beta_0^2 &\le c_1\beta_0^{\frac{2}{1+\delta}} - d_1\beta_0^2,\\
\frac{1}{3}\epsilon\beta_0^2 &\le c_2\beta_0 - d_2\beta_0^2,\\
\frac{1}{3}\epsilon\beta_0^2 &\le c_3\beta_0^{\frac{2}{1+\delta}} - d_3\beta_0^2.
\end{aligned}$$

The first inequality gives

$$\alpha_0^{2\delta} \le \alpha_0^{\frac{2\delta}{1+\delta}} = \alpha_0^{2-\frac{2}{1+\delta}} \le \frac{c_1}{d_1 + \frac{1}{3}\epsilon}.$$

The third inequality gives in the same way $\alpha_0^{2\delta} \le \frac{c_3}{d_3+\frac{1}{3}\epsilon}$.
The fourth $\beta_0^{2\delta} \le \frac{c_1}{d_1+\frac{1}{3}\epsilon}$.
The sixth $\beta_0^{2\delta} \le \frac{c_3}{d_3+\frac{1}{3}\epsilon}$.
The second $\alpha_0 \le \frac{c_2}{d_2+\frac{1}{3}\epsilon}$.
And the fifth $\beta_0 \le \frac{c_2}{d_2+\frac{1}{3}\epsilon}$.

Always $\alpha_0, \beta_0 \le 1$ holds, and therefore

$$\alpha\beta \le \mathrm{MAX}\left[\left(\frac{c_1}{d_1+\frac{1}{3}\epsilon}\right)^{\frac{1}{2\delta}}, \frac{c_2}{d_2+\frac{1}{3}\epsilon}, \left(\frac{c_3}{d_3+\frac{1}{3}\epsilon}\right)^{\frac{1}{2\delta}}\right].$$

□

4.2. Some Examples. Theorem 4.1 implies that under some conditions a protocol can approximate the function f only by using small sub-matrices, therefore we get a lower bound for the probabilistic communication complexity $C(f)$. We give here two examples, describing applications of Theorem 4.1. We remark that one can get a lower bound for a given function g, also by using Theorem 4.1 for $f = g\mu$, where μ is any probability measure on the inputs.

4.3. The *Majority* of *XOR* of two vectors. Let $\tilde{f} : \{0,1\}^n \to \{-1,1\}$ be the *Majority* function. Let $f : \{0,1\}^{n+n} \to \{-1,1\}$ be defined by $f(T,T') = \tilde{f}(T \oplus T')$. By definition: $\| \tilde{f} \|_2 = 1$, and therefore $\sum_S \tilde{f}_S^2 = 1$. by the symmetry of $\tilde{f}$, we have $\tilde{f}_S = \tilde{f}_{S'}$ if $|S| = |S'|$. Therefore

$$|\tilde{f}_S| \le \binom{n}{|S|}^{-1/2}.$$

It is easy to check that $\tilde{f}_{\bar{1}} = o(1)$. By the definition of f: $f_{S,S'} = \tilde{f}_S \delta_{S,S'}$, and therefore

$$f_{\bar{0},*} = f_{*,\bar{0}} = f_{\bar{0},\bar{0}} = \tilde{f}_{\bar{0}} = 0.$$
$$f_{\bar{1},*} = f_{*,\bar{1}} = |f_{\bar{1},\bar{1}}| = |\tilde{f}_{\bar{1}}| = o(1).$$
$$f_{S,*} = f_{*,S} = |\tilde{f}_S| \le \binom{n}{|S|}^{-1/2}.$$

For $|S| \le \sqrt{n}$ $\qquad \binom{n}{|S|}^{-\frac{1}{2}} \le (n^{-\frac{1}{4}})^{|S|}$.

For $\sqrt{n} \le |S| \le n - \sqrt{n}$ $\qquad \binom{n}{|S|}^{-\frac{1}{2}} \le (n^{-\frac{1}{4}})^{\sqrt{n}}$.

For $n - \sqrt{n} \le |S|$ $\qquad \binom{n}{|S|}^{-\frac{1}{2}} \le (n^{-\frac{1}{4}})^{n-|S|}$.

Therefore, we can take in Theorem 4.1: $\delta = n^{-\frac{1}{4}}$, $c_1 = d_1 = 1$, $c_2 = d_2 = n^{-\frac{1}{4}\sqrt{n}}$, $c_3 = 1$, $d_3 = 1 - o(1)$, and get that if for A, B : $\alpha_1^2 = \alpha_0^2$, $\beta_1^2 = \beta_0^2$, and if $|\sum_{A\times B} f| \ge \frac{1}{4}\alpha_0\beta_0$ then $\alpha_0\beta_0 \le 2^{-\Omega(n^{\frac{1}{4}})}$. If we restrict ourselves to vectors of even cardinality then always $\alpha_1^2 = \alpha_0^2$, and $\beta_1^2 = \beta_0^2$, and the conclusion is that in this region only very small sub-matrices can approximate f. Therefore, there are many of them, and a standard argument gives a lower bound of $\Omega(-\log 2^{-\Omega(n^{\frac{1}{4}})}) = \Omega(n^{1/4})$ for $C(f)$. A better lower bound of $\Omega(n^{1/2})$ for this function can be proved by using the first method of this paper with $k = 1$. A lower bound of $\Omega(n/\log n)$ can be proved taking $k = n/2$. A tight lower bound of $\Omega(n)$ can be proved by a reduction from the *Disjointness* function (tight lower bounds for the *Disjointness* function where proved in [6],[8]).

4.4. The scalar product of the XOR of two pairs of vectors. Let $\tilde{f} : \{0,1\}^{2n} \to \{-1,1\}$ be defined by

$$\tilde{f}(T,Q) = (-1)^{|T\cap Q|}.$$

Let $f : \{0,1\}^{2n+2n} \to \{-1,1\}$ be defined by

$$f(T,Q,T',Q') = \tilde{f}(T \oplus T', Q \oplus Q').$$

By definition $\tilde{f}(T,Q) = \prod_{i=1}^{n} g(T_i,Q_i)$, where $g : \{0,1\}^2 \to \{-1,1\}$ is defined by $g(0,0) = g(0,1) = g(1,0) = 1$; $g(1,1) = -1$. The Fourier coefficients of g are $g_{00} = g_{01} = g_{10} = \frac{1}{2}$; $g_{11} = -\frac{1}{2}$. Hence,

$$\tilde{f}_{S,R} = 2^{-n}(-1)^{|S\cap R|},$$

i.e. $|\tilde{f}_{S,R}| = 2^{-n}$. This shows that $f_{S,R,*,*} = f_{*,*,S',R'} = 2^{-n}$ for all S, R, S', R'. We can now take in Theorem 4.1: $c_2 = 2^{-n}$; $d_2 = 0$ and $c_1 = d_1$, $c_3 = d_3$, δ very close to 0, to get that if $|\sum_{A\times B} f| \ge \frac{1}{4}\alpha_0\beta_0$ and $\alpha_0^2 = \alpha_1^2$; $\beta_0^2 = \beta_1^2$ then $\alpha_0\beta_0 \le 12 \cdot 2^{-n}$. By a standard argument, the probabilistic communication complexity of f is $C(f) = \Omega(-\log(12 \cdot 2^{-n})) = \Omega(n)$. This lower bound is tight.

Acknowledgements

I would like to thank Laszlo Lovász and Avi Wigderson for helpful discussions, and the two anonymous referees for many important comments that significantly improved the presentation of the paper.

References

[1] W. Beckner, Inequalities in Fourier Analysis. *Ann. Math.* **102** (1975), 159–182.

[2] L. Babai, P. Frankl, J. Simon, Complexity classes in communication complexity. In *Proc. 27th Ann. IEEE Symp. Found. Comput. Sci.*, 1986, 337–347.

[3] M. Ben-Or, N. Linial, *Collective coin flipping in randomness and computation.* Tech. Rep. Leibnitz Center Report TR 87-2, Computer Science Department, Hebrew University, 1987.

[4] L. Babai, N. Nisan, M. Szegedy, Multiparty protocols and logspace-hard pseudorandom sequences. In *Proc. Twenty-first Ann. ACM Symp. Theor. Comput.*, 1989, 1–11.

[5] J. Kahn, G. Kalai, N. Linial, The influence of variables on boolean functions. In *Proc. 29th Ann. IEEE Symp. Found. Comput. Sci.*, 1988, 68–80.

[6] B. Kalyanasundaram, G. Schnitger, The probabilistic communication complexity of set intersection. In *Proc. 2nd Ann. Symp. Structure in Complexity Theory*, 1987, 41–49.

[7] L. Lovász, *Communication complexity: A survey.* Tech. Rep. Princeton University Report CS-TR-204-89, Princeton, 1989.

[8] A. Razborov, On the distributional complexity of disjointness. *Theoret. Comput. Sci.* **106** (1992), 385–390.

[9] A. Yao, Some complexity questions related to distributive computing. In *Proc. Eleventh Ann. ACM Symp. Theor. Comput.*, 1979, 209–213.

[10] A. Yao, Lower bounds by probabilistic arguments. In *Proc. Fifteenth Ann. ACM Symp. Theor. Comput.*, 1983, 420–428.

Manuscript received 8 September 1994

Ran Raz
Department of Computer Science
Princeton University

Current address :
Department of Applied Math.
Weizmann Institute
Rehovot 76100, Israel
`ranraz@wisdom.weizmann.ac.il`