Introduction to Communication Complexity

Will Rosenbaum
Updated: April 30, 2014

Department of Mathematics
University of California, Los Angeles

Introduction

In this essay, we give a basic introduction to the exciting field of communication complexity. Communication complexity was first formalized by Yao in 1979 [3]. The context is easy enough to understand: two parties (usually named Alice and Bob) wish to compute the value of some function f of two variables. Alice knows the value of one of the variables, while Bob knows the value of the other. How much must Alice and Bob communicate in order to find the value of f on their combined input?

Since Yao's seminal work on the subject, the field of communication complexity has exploded. This growth is driven by the field's diverse applications to computational complexity and connections to other branches of discrete math and computer science. Our brief exposition here is inspired by Yao's original work, but is strongly influenced by Kushilevitz and Nisan's more modern treatise on communication complexity [1]. We focus on a deterministic theory of communication, although we hint at other models in the final section.

This essay is broken into four modules. In the first, we formalize the notion of a communication protocol and define the communication complexity of a function. The bulk of the theoretical material resides in the second section, where we give techniques for proving communication lower bounds. In the third section we apply the techniques of the previous section to prove lower bounds for particular functions. In the final section, we give examples that suggest other models of communication.

1 Communication Protocols

Here, we describe the basic model of communication we will use. Throughout, we assume that f is a boolean function of two variables, say

$$f : X \times Y \longrightarrow \{0,1\}.$$

Typically, we will make the (harmless) assumption that $X = Y = \{0,1\}^n$. We assume that two players, Alice and Bob, hold x and y respectively, with $x \in X$ and $y \in Y$. For the sake of consistency, we will always use the variable x to refer to Alice's input and y to refer to Bob's input. Together, Alice and Bob wish to compute $f(x,y)$. In order to do so, they must communicate some information about their input. We formalize a conversation between Alice and Bob with the following.

Definition 1. A ***communication protocol*** Π for is a binary tree $T = (V,E)$ directed away from the root with the following properties:

1. Each internal vertex $v \in V$ is labeled with a player p (Alice or Bob) and a boolean function $f_v : Z \to \{0,1\}$ where $Z = X$ if $p =$Alice and $Z = Y$ if $p =$Bob.
2. Each leaf $\ell \in V$ is labeled with a value $f_\ell \in \{0,1\}$.
3. The two out edges from each internal node are labeled 0 and 1.

Alice and Bob execute a protocol Π in the following manner. Let v_0 denote the root of the tree T. The label of v_0 determines which player speaks first—Alice or Bob. Suppose v_0 is labeled "Alice." She then computes $f_{v_0}(x)$, and sends this value to Bob. On the tree T (which is known to both Alice and Bob) Alice and Bob follow the edge out of the root labeled $f_{v_0}(x)$, which brings them to a new node v_1. The label of v_1 then determines which player speaks next, and their message is determined by applying f_{v_1} to their input. Alice and Bob continue their conversation in this manner until they reach a leaf ℓ. At this point, the protocol terminates and the output is given by $\Pi(x, y) = f_\ell$.

The path in T traversed by Π on input (x, y) is uniquely determined by (x, y). We refer to the sequence of labels of edges in the traversal as the ***transcript*** of the protocol on input (x, y). Notice that the transcript uniquely determines the output $\Pi(x, y)$.

Definition 2. The ***communication cost*** of a protocol Π, denoted $\mathrm{CC}(\Pi)$ is the depth of the associated tree T, i.e. the length of the longest path from the root of T to a leaf.

Definition 3. We say that a protocol Π ***computes*** a function f if for all $(x, y) \in X \times Y$, $\Pi(x, y) = f(x, y)$.

Definition 4. The ***(deterministic) communication complexity*** of a function f, denoted $\mathrm{D}(f)$ is given by

$$\mathrm{D}(f) = \min \{\mathrm{CC}(\Pi) \mid \Pi \text{ computes } f\}.$$

That is, $\mathrm{D}(f)$ is the communication cost of the best protocol Π which computes f.

Proposition 5. For every function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, we have

$$\mathrm{D}(f) \leq n + 1.$$

Proof. We give a formal description of the following "intuitively obvious" protocol: Alice sends x to Bob, and Bob outputs $f(x, y)$. Formally, the first n layers of the tree T for Π are labeled "Alice," and the function f_v for v a distance i from the root is simply $f_v(x) = x_{i+1}$. The vertices w in the penultimate layer of T are labeled "Bob." If the unique path leading to w corresponds to the transcript $x = x_1 x_2 \cdots x_n$, we define $f_w(y) = f(x, y)$, and label the leaves below w correspondingly. This tree has depth $n + 1$, which gives the desired result. □

Example 6 (Parity function)**.** Consider the function

$$f(x, y) = \sum_{i=1}^{n} (x_i + y_i)$$

where all arithmetic is performed modulo 2. For this function, Alice and Bob need not divulge their entire inputs to compute f. Indeed, Alice computes $f_0(x) = \sum_{i=1}^{n} x_i$ and sends this value to Bob. Bob then computes $f(x, y) = f_0(x) + f_0(y)$ and sends this value to Alice. Thus, $\mathrm{D}(f) \leq 2$. In the following section, we will prove that in fact $\mathrm{D}(f) = 2$.

Remark 7. We note that in the definition of a communication protocol, we do not make any assumptions about the nature of the functions which label the vertices of T. We may interpret this lack of restriction as allowing Alice and Bob to be computationally unbounded. To bring this feature into sharp relief, consider the function $f(x, y)$ whose value is 1 if and only if the Turing machine with description number y halts on input x. The value of $f(x, y)$ is not computable by any Turing machine, yet as a result of the previous proposition, $\mathrm{D}(f) \leq n + 1$. Thus, in our model, Alice and Bob are individually infinitely powerful. The only quantity that we restrict is the amount of communication between Alice and Bob.

2 Lower Bound Techniques

In this section we describe techniques for proving lower bounds for communication complexity. Our first task is to described the combinatorial structure of inputs corresponding to the same transcript in a fixed protocol.

2.1 Combinatorial rectangles

Definition 8. Given a Cartesian product of sets $X \times Y$, a ***combinatorial rectangle***, or simply a ***rectangle*** is a subset of the form $A \times B$ for $A \subset X$ and $B \subset Y$.

We will find the following alternate characterization of combinatorial rectangles useful in what follows.

Proposition 9. A subset $R \subset X \times Y$ is a combinatorial rectangle if and only if for every pair of elements $(x, y), (x', y') \in R$, we also have $(x, y'), (x', y) \in R$.

Proof. Exercise. □

Definition 10. Let $f : X \times Y \to \{0, 1\}$ be a boolean function and $A \times B$ a combinatorial rectangle in $X \times Y$. We say that $A \times B$ is ***monochromatic*** with respect to f if f is constant on $A \times B$. If $f(x, y) = b$ on $A \times B$, we say that $A \times B$ is b-***monochromatic***.

The following observation will be of paramount importance in proving communication lower bounds.

Theorem 11. Let Π be a communication protocol for $f : X \times Y \to \{0, 1\}$, and let $m = m_1 m_2 \cdots m_\ell$ be a transcript of some execution of Π which outputs $b \in \{0, 1\}$. Then the set of all inputs $(x, y) \in X \times Y$ such that Π has transcript m is a b-monochromatic combinatorial rectangle with respect to f. In particular, Π induces a partition of $X \times Y$ into monochromatic rectangles.

Proof. We first prove by induction on the depth of a vertex v in the tree T for Π that the set of inputs for which Π visits v is a rectangle. The base case where $v = v_0$ is the root is trivial, as all inputs in $X \times Y$ visit v_0. For the inductive step, suppose the statement is true up to depth d, and let v' be a vertex at depth $d + 1$. Suppose v is the parent of v', v is labelled by Alice, and the edge (v, v') is labelled 1. By the inductive hypothesis, the set of all inputs that reach v are of the form $A \times B$. Thus the set of inputs that lead to v' is given by

$$A \times B \cap \{(x, y) \mid f_v(x) = 1\} = (A \cap \{x \mid f_v(x) = 1\}) \cap B.$$

This set clearly has the form $A' \times B$, as desired.

Since the output $\Pi(x, y)$ agrees with $f(x, y)$ for all inputs, it must be the case that for each leaf $\ell \in T$, $f(x, y)$ is equal to the label of ℓ for all inputs leading to ℓ. In particular, the inputs leading to ℓ form a monochromatic rectangle. □

Example 12. We revisit the function f from example 6. We can represent f by the "graph" of its values in the form of a $2^n \times 2^n$ matrix. For $n = 2$ we get

	00	01	10	11
00	0	1	1	0
01	1	0	0	1
10	1	0	0	1
11	0	1	1	0

.

The protocol described in example 6 gives rise to the following rectangles. Let

$$A_0 = B_0 = \{00, 11\}$$

and

$$A_1 = B_1 = \{01, 10\}.$$

Then the protocol partitions the input into the rectangles $A_i \times B_j$ for $i, j \in \{0, 1\}$. Notice that these sets are indeed monochromatic.

Remark 13. We note that although every protocol for f induces a monochromatic partition of the input space, not every monochromatic partition by rectangles is induced by a communication protocol. Try to find an example of such a function and partition.

2.2 **Fooling sets** We are now ready to use proposition 9 and theorem 11 to describe a powerful technique for proving communication lower bounds.

Definition 14. We call a set $S \subset X \times Y$ a ***fooling set*** for f if all distinct $(x, y), (x', y') \in S$, one of the following holds:

1. $f(x, y) \neq f(x', y')$
2. $f(x, y) = f(x', y')$ but $f(x', y) \neq f(x, y)$
3. $f(x, y) = f(x', y')$ but $f(x, y') \neq f(x, y)$.

Theorem 15. Suppose S is a fooling set for f. Then $\mathrm{D}(f) \geq \log |S|$.*

* Since we are studying theoretical *computer science*, all logs are assumed to be base 2.

Proof. We will first show that for any partition of $X \times Y$ into monochromatic rectangles, all elements in S must reside in different rectangles. Suppose to the contrary that $(x, y), (x', y') \in S$ are in the same rectangle R. If R is to be monochromatic, then we must have $f(x, y) = f(x', y')$. By proposition 9, we have $(x', y), (x, y') \in R$. But by properties 2 and 3 of fooling sets, this contradicts the monochromaticity of R. Thus any tiling of $X \times Y$ by monochromatic rectangles must contain at least $|S|$ rectangles.

Since the monochromatic rectangles induced by a protocol Π for f correspond to leaves of a binary tree, the tree must have $|S|$ leaves. Since a binary tree of depth d can have at most 2^d leaves, we must have $\mathrm{CC}(\Pi) \geq \log |S|$, as desired. □

Example 16. The "parity" function f from example 6 has $\mathrm{D}(f) = 2$. We already proved that 2 is an upper bound for $\mathrm{D}(f)$, but now we can prove a matching lower bound. Consider the set

$$S = \{(00, 00), (01, 01), (10, 11), (11, 10)\}.$$

The ambitious reader should verify that this is indeed a fooling set of size 4. Hence by theorem 15, $\mathrm{D}(f) \geq \log |S| = 2$.

2.3 **Rank lower bound** Here, we prove a communication lower bound using algebraic considerations.

Theorem 17. Let f be a boolean function $f : X \times Y \to \{0, 1\}$, and let M_f denote the $|X| \times |Y|$ matrix whose (x, y)-entry is $f(x, y)$. Then

$$\mathrm{D}(f) \geq \log \mathrm{rank}(M_f)$$

where $\mathrm{rank}(M_f)$ is the rank of M_f over $\mathbf{R}$.

Proof. As with the fooling set lower bound, the idea is to show that $\operatorname{rank} M_f$ is a lower bound for the number of leaves of any protocol for f. Suppose Π is a protocol for f. Let $L_1 \subset V$ be the set of leaves in Π labeled by 1. For $\ell \in L_1$ let M_ℓ be the $|X| \times |Y|$ matrix given by

$$M_\ell(x,y) = \begin{cases} 1 & \Pi(x,y) \text{ terminates at } \ell \\ 0 & \text{otherwise.} \end{cases}$$

Notice that

$$M_f = \sum_{\ell \in L_1} M_\ell.$$

Further, for each $\ell \in L_1$, $\operatorname{rank} M_\ell = 1$. Thus

$$\operatorname{rank} M_f \leq \sum_{\ell \in L_1} \operatorname{rank} M_\ell = |L_1|.$$

Since Π has at least L_1 leaves, we conclude that

$$\log \operatorname{rank} M_f \leq \mathrm{CC}(\Pi),$$

whence the theorem follows. □

It is widely conjectured that $\log \operatorname{rank} f$ characterizes the communication complexity of f. Specifically, the ***log-rank conjecture*** asserts that there exist absolute constants $c_1, c_2 > 0$ such that

$$\mathrm{D}(f) \leq c_1 (\log \operatorname{rank} f)^{c_2}.$$

The current best upper bound (due to Lovett) is

$$\mathrm{D}(f) \leq \mathcal{O}(\sqrt{\operatorname{rank} f} \log \operatorname{rank} f).$$

See [2] for a very up-to-date survey on this conjecture.

3 Applications

In this section we apply the techniques of the previous section to obtain lower bounds for some natural functions. We begin by showing that "most" boolean functions have a high communication complexity. Then we show that two particular functions, equality and disjointness, are maximally complex.

3.1 Most functions are "hard"

In this section, we will show that a boolean function f generated uniformly at random is overwhelmingly likely to satisfy $\mathrm{D}(f) \geq n - 2$.

Theorem 18. Suppose $|X| = |Y| = N = 2^n$, and let $f : X \times Y \to \{0,1\}$ be chosen uniformly at random. That is, for each (x,y), the value $f(x,y)$ is chosen uniformly and independently from $\{0,1\}$. Then

$$\mathbf{P}(\mathrm{D}(f) \leq n-2) \leq 2^{\mathcal{O}(-N^2)}.$$

Proof. Let $d(f)$ denote the minimum number of f-monochromatic rectangles which partition $X \times Y$. By the discussion above, we have $\mathrm{D}(f) \geq \log d(f)$. Notice that there are at most 2^{2N} rectangles on $X \times Y$. Thus there are at most

$$\binom{2^{2N}}{k} 2^k \leq 2^{2kN+k}$$

functions f which can be covered by k monochromatic rectangles. Thus,

$$|\{f : X \times Y \to \{0,1\} \mid d(f) \leq k\}| \leq 2^{2kN+k}.$$

Since there are 2^{N^2} functions boolean functions on $X \times Y$, for $k = N/4$ we estimate

$$\mathbf{P}\left(d(f) \leq \frac{N}{4}\right) \leq \frac{1}{2^{N^2}} 2^{-N^2/2+N/4} = 2^{N^2/2+N/4}.$$

Whence the theorem follows. □

3.2 Equality and disjointness

Example 19 (Equality function)**.** We define the equality function EQ by

$$\mathrm{EQ}(x,y) = \begin{cases} 1 & x = y \\ 0 & x \neq y. \end{cases}$$

Since the matrix representation of EQ, M_{EQ} is the $2^n \times 2^n$ identity matrix, the rank lower bounds immediately implies that $\mathrm{D}(\mathrm{EQ}) \geq n$, hence $\mathrm{D}(\mathbf{E}) = n$ or $n+1$. The fooling set method shows that, in fact, the latter is the correct quantity.

Consider the set $S = \{(x,x) \mid x \in X\}$. Notice that $|S| = |X| = 2^n$, hence theorem 15 gives $\mathrm{D}(\mathrm{EQ}) \geq n$. However, every element in S corresponds $\mathrm{EQ} = 1$. Since there must be at least one 0-rectangle in the partition induced by a protocol Π, there must be at least $2^n + 1$ monochromatic rectangles total. Hence $\mathrm{D}(\mathrm{EQ}) \geq \log(2^n + 1) > n$, we conclude that in fact $\mathrm{D}(\mathrm{EQ}) = n + 1$ since D is always integral. This result should hardly be surprising, for how could Alice and Bob be certain that their inputs are equal without knowing each others' input?

Example 20 (Disjointness function)**.** Suppose Alice and Bob each hold subsets A and B (respectively) of $[n] = \{1, 2, \ldots, n\}$. We can encode those subsets as bit vectors $x, y \in \{0,1\}^n$ where

$$x_i = 1 \iff i \in A, \qquad y_j = 1 \iff j \in B.$$

We define the disjointness function DISJ by

$$\mathrm{DISJ}(x,y) = \begin{cases} 1 & A \cap B = \varnothing \\ 0 & \text{otherwise.} \end{cases}$$

Notice that DISJ has the rather appealing form

$$\mathrm{DISJ}(x,y) = \neg \bigvee_{i=1}^{n} (x_i \wedge y_i).$$

This "direct sum" decomposition of DISJ seems to suggest that we cannot possibly determine that sets A and B are disjoint without essentially verifying that each element in $[n]$ is contained in at most one of the sets. Indeed, this is the case. Consider the fooling set

$$S = \{(A, A^c) \mid A \subset [n]\}.$$

As with EQ, analyzing S and applying theorem 15 gives the desired result.

4 Preview of Things to Come

Thus far, we have only considered deterministic protocols which are guaranteed to produce a correct output for every instance of the communication problem. There are two natural randomized relaxations of the communication problem:

1. Assume that the input (x, y) is chosen according to some probability distribution μ on $X \times Y$. We now only require that a (deterministic) protocol Π output the correct answer with some good probability, say

   $$\mathbf{P}_\mu \left(\Pi(x, y) = f(x, y)\right) \geq 1 - \varepsilon.$$

   The communication complexity resulting from this model is known as the ***distributional complexity***.

2. Suppose we allow the protocol Π to be randomized. That is, we give Alice and Bob access to (private or public) random strings, r_A and r_B, and allow their messages to be functions both their input and the random string. We now require that a protocol Π output the correct for all inputs (x, y) with some good probability over the strings r_A and r_B:

   $$\mathbf{P}_{r_A, r_B} \left(\Pi(x, y) = f(x, y)\right) \geq 1 - \varepsilon, \text{ for all } (x, y) \in X \times Y.$$

   The communication cost of the best randomized protocol for f is the ***randomized communication complexity***.

As we will see in the sequel, distributional and randomized complexity are very closely related. In fact, distribution complexity completely characterizes randomized complexity. These notions may have wildly different complexity than their deterministic counterparts, as the following example shows.

Example 21. Let EQ be the equality function as before. There exists a randomized communication protocol Π for EQ such that $\mathrm{CC}(\Pi) = \mathcal{O}(\log n)$ such that for all $x, y \in \{0, 1\}^n$,

$$\mathbf{P}(\Pi(x, y) = \mathrm{EQ}(x, y)) \geq 1 - \frac{1}{n}.$$

The protocol Π works as follows. Suppose p is a prime with $n^2 < p < 2n^2$ (say p is the smallest prime larger than n^2). Alice and Bob encode their input as polynomials of degree at most n over $\mathbf{F}_p$, the field with p elements in the following manner. Write $x = a_1 a_2 \cdots a_n$ and $y = b_1 b_2 \cdots b_n$ where $a_i, b_j \in \{0, 1\}$ for all i, j. Alice and Bob form

$$P_A(z) = \sum_{i=1}^{n} a_i z^{i-1} \quad \text{and} \quad P_B(z) = \sum_{j=1}^{n} b_j z^{j-1}$$

respectively. Alice chooses $t \in \mathbf{F}_p$ uniformly at random and sends to Bob t and $P_A(t) \in \mathbf{F}_p$. Bob responds with 1 if $P_B(t) = P_A(t)$ and 0 otherwise. Notice that the total number of bits exchanged is $\mathcal{O}(\log n)$.

Clearly, if $x = y$, then $P_A = P_B$ so that Bob responds with 1. On the other hand, if $x \neq y$, then $P_A - P_B$ is a nonzero polynomial of degree at most $n-1$ over $\mathbf{F}_p$ with $p > n^2$. Thus

$$\mathbf{P}_t \left(P_A(t) - P_B(t) = 0\right) \leq \frac{n-1}{n^2} < \frac{1}{n}.$$

Thus, Π correctly computes EQ with probability at least $\frac{1}{n}$ using $\mathcal{O}(\log n)$ communication.

This example shows that while the deterministic communication complexity of EQ is $\Omega(n)$, the randomized communication complexity is $\mathcal{O}(\log n)$. Thus randomization may afford us exponentially more efficient communication protocols. In the next talk, we will give techniques for lower bounds for randomized protocols.

Sources

[1] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[2] Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *CoRR*, abs/1403.8106, 2014.

[3] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213. ACM, 1979.