1 Introduction

Communication complexity is a very rich area within complexity which hits the sweet spot of simplicity vs. application. The complexity of a lot of problems are well understood, and has been the source of very many elegant applications, such as in circuit complexity, data structure lower bounds and many more!

Note: An extensive reference on communication complexity is the book by Kushilevitz-Nisan. Chapter 13 in Arora-Barak also provides a good introduction.

The model

There is a joint function \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \) where Alice gets input \( x \in \{0, 1\}^n \) and Bob gets input \( y \in \{0, 1\}^n \), and they wish to compute \( f(x, y) \). They exchange messages with one another according to some pre-coordinated protocol.\(^\dagger\) The last bit of the protocol should be \( f(x, y) \). We want to minimize the number of bits exchanged. We don’t care about the computational complexity of Alice and Bob (they have unbounded computation power!).

Definition 1. \( D(f) \) is the number of bits sent by the smallest (deterministic) protocol that computes \( f \).

Note: For any \( f \), \( D(f) \leq n + 1 \), as Alice can simply send her entire input \( x \) over to Bob who can then compute \( f(x, y) \) and declare the answer.

Examples

(a) XOR : \( \text{XOR}(x, y) = \bigoplus_{i=1}^{n} x_i \oplus y_i \)

(b) EQUALITY : \( \text{EQ}_n(x, y) = 1 \) iff \( x = y \).

(c) INNER PRODUCT : \( \text{IP}_n(x, y) = \bigoplus_{i=1}^{n} (x_i \cdot y_i) \)

(d) DISJOINTNESS : \( \text{DISJ}_n(x, y) = \neg \bigvee_{i=1}^{n} (x_i \land y_i) = \begin{cases} 1 & x \land y = 0 \\ 0 & \text{otherwise} \end{cases} \)

\(^\dagger\)if we don’t care about constant factors, without loss of generality the protocol can be that they have \( k(n) \) rounds, where in odd rounds Alice sends a bit to Bob and in even rounds Bob sends a bit to Alice
2 Lower bounds on deterministic CC

Theorem 1. \( D(\text{EQ}_n) \geq n \)

Proof. Suppose \( D(\text{EQ}_n) < n \) and let \( \pi \) be a protocol computing this function with communication at most \( n - 1 \) bits. Then there are inputs \( x \neq x' \) such that \( \pi(x, x) = \pi(x', x') \). We claim that \( \pi(x, x') = \pi(x, x) \). This would yield a contradiction because \( \text{EQ}_n(x, x) = 1 \) but \( \text{EQ}_n(x, x') = 0 \).

Indeed, let \( m = (m_1, \ldots, m_k) \) be the messages \( \pi(x, x) = \pi(x', x') \). Clearly \( m_1 \) is also the first message in \( \pi(x, x') \) since it's only based on Alice’s input \( x \). Now \( m = \pi(x', x') \) means that on input \( x' \) and after seeing message \( m_1 \), Bob sends the message \( m_2 \), and hence \( m_2 \) is the second message of \( \pi(x, x') \). We can continue this way for all messages. \( \square \)

\( \text{EQ}_n \) was easy, but how do we prove lower bounds for other functions?

Tree view of protocols

We can think of protocols as a tree. If the first bit in the protocol is to be sent by Alice, we think of the root node as \( A(x) \). If Alice sends a 0, then the next bit is sent by Bob given by \( B_0(y) \). If Bob sends a 0, then the next bit is sent by Alice given by \( A_{00}(x) \) and so on.

\[
\begin{array}{c}
A(x) \\
B_0(y) \quad A_1(x) \\
A_{00}(x) \quad 1 \quad B_{10}(y) \quad A_{11}(x) \\
0 \quad 1 \quad 0 \quad 1 \\
1 \quad 1 \quad 1 \quad 1 \\
0 \quad 0 \quad 0 \quad 0 \\
1 \quad 0 \quad 0 \quad 0
\end{array}
\]

Monochromatic Partition

Another way to view a protocol is by looking at what it does to the matrix \( M_f \) corresponding to the function.

Definition 2. For any function \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \), define \( M_f \) to be a \( 2^n \times 2^n \) matrix where the rows and columns are indexed by \( \{0, 1\}^n \) and \( M_f(x, y) = f(x, y) \). When the context is clear, we will denote \( M_f \) by \( M \).

Any protocol partitions the matrix \( M_f \) into monochromatic rectangles as shown in the following figure.

\[
\begin{pmatrix}
0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix}
\]
Moreover, if the protocol \( \pi \) has \( c \) bits of communication, then the matrix \( M_f \) is partitioned into at most \( 2^c \) monochromatic rectangles. Thus, \( D(f) \) is at least the log of the smallest number of monochromatic rectangles needed to partition \( M_f \). Also, note that since each monochromatic rectangle is a rank-1 matrix, the matrix \( M_f \) can be written as the sum of \( 2^{c-1} \) rank-1 matrices (because only one out of two leaves at the end can output 1). Thus, we get that the smallest number of monochromatic rectangles needed to partition \( M_f \) is at least \( 2 \cdot \text{rank}(M_f) \). Thus, we have our first lower bound result,

**Theorem 2.** For all \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} : D(f) \geq \log(\text{rank}(M_f)) + 1 \)

**Applications**

(a) \( D(\text{EQ}_n) \geq n + 1 \): since \( M_{\text{EQ}_n} = I_{2^n \times 2^n} \), and hence \( \log(\text{rank}(M_{\text{EQ}_n})) = n \).

(b) \( D(\text{IP}_n) \geq n + 1 \) (easy exercise)

(c) \( D(\text{DISJ}_n) \geq n + 1 \) (difficult exercise)

The log-rank lower bound is conjectured to be tight upto polynomials factors. This remains a tantalizing open conjecture till date!

**Conjecture 1** (Log-rank conjecture). For all \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} : D(f) \leq (\log(\text{rank}(M_f)))^{O(1)} \)

### 3 Randomized Communication Complexity

**Definition 3.** For \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \), define \( R_\varepsilon(f) = \min \text{CC}(\pi) \) where \( \pi \) is a randomized protocol such that for all \( x, y \in \{0, 1\}^n \), \( \Pr[\pi(x, y) \neq f(x, y)] \leq \varepsilon \). We think of the randomness as coming from the “sky”, and so Alice and Bob use the same random string without communicating at all. For ease of notation, we denote \( R_{1/3}(f) \) by simply \( R(f) \).

Even though computing \( \text{EQ}_n \) requires \( n + 1 \) communication deterministically, it can be computed by a randomized protocol with only \( O(1) \) communication! This is achieved as follows:

- Choose a random string \( r \in \{0, 1\}^n \)
- Alice computes \( r \cdot x \) and sends it to Bob.
- Bob checks whether \( r \cdot x = r \cdot y \). If yes, then he outputs 1 and 0 otherwise.

If \( x = y \), then this protocol will always output 1. If \( x \neq y \), then the protocol will output 0 with probability 1/2 (exercise). By repeating this protocol \( \log(1/\varepsilon) \) times, we can reduce the error to \( \varepsilon \), and hence \( R_\varepsilon(\text{EQ}_n) = O(\log(1/\varepsilon)) \).

But what about \( R(\text{DISJ}_n) \) and \( R(\text{IP}_n) \)? Can they be solved efficiently? If not, how do we prove lower bounds on randomized communication complexity?
4 Distributional Complexity

Definition 4. For \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \), define \( D^\mu_\varepsilon(f) = \min CC(\pi) \) where \( \pi \) is a deterministic protocol such that, \( \Pr_{(x,y)\sim\mu}[\pi(x, y) \neq f(x, y)] \leq \varepsilon \).

For ease of notation, we denote \( D^\mu_0(f) \) to be deterministic; in particular, if it makes an error on input \( x \), then it works correctly on input \( y \).

The key difference between \( R(f) \) and \( D^\mu_\varepsilon(f) \) is as follows: In \( R(f) \) the probability is over the choice of randomness in the protocol \( \pi \) and the protocol has to work with low error on every input. In \( D^\mu_\varepsilon(f) \) the probability is over the choice of input drawn from \( \mu \), and the protocol \( \pi \) has to be deterministic; in particular, if it makes an error on input \( (x, y) \), then it will make an error on that input with probability 1. The only thing we require is that the total probability of inputs on which \( \pi \) makes an error is small.

Example. Let \( \mu \) being the distribution over \( \{0, 1\}^n \times \{0, 1\}^n \) generated as follows: First we generate a random bit \( b \). If \( b = 1 \), then we sample \( (x, x) \) uniformly. If \( b = 0 \), then we sample \( (x, y) \) independently and uniformly. For this distribution, there is a very trivial protocol for \( EQ_n \). Namely, Alice sends the first bit of \( x \), and Bob accepts if it matches the first bit of \( y \).

Exercise. The probability that this protocol succeeds is at least 3/4.

4.1 Yao’s Min-Max principle

Yao introduced the following simple but powerful lemma as an approach to prove lower bounds on randomized communication complexity.

Lemma 1. For any function \( f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\} \) and any distribution \( \mu \):

\[
R_\varepsilon(f) = \max_{\mu} D^\mu_\varepsilon(f)
\]

Proof. \([R_\varepsilon(f) \geq D^\mu_\varepsilon(f)]\) Suppose we have a randomized protocol \( \pi \) that makes error with probability at most \( \varepsilon \). We can think of the protocol as a “convex” combination of several deterministic protocols. Namely, Alice and Bob first sample a random string \( R \) from the ‘sky’, and then use the deterministic protocol \( \pi_R \), where \( CC(\pi_R) \leq CC(\pi) \).

Consider the matrix \( A \) where the rows are the inputs \( (x, y) \) and the columns are all deterministic protocols of length \( CC(\pi) \). The entries of the matrix are given by \( A((x, y), R) = \text{1}(f(x, y) = \pi_R(x, y)) \) that is, it is 1 if \( \pi_R \) works correctly on input \( (x, y) \) and 0 otherwise. We think of \( \pi \) as a distribution, say \( \psi \), over these deterministic protocols. Clearly, \( \pi \) makes error with probability at most \( \varepsilon \) even when the inputs are drawn from a distribution \( \mu \). But observe that the correctness probability of the protocol \( \pi \) over distribution \( \mu \) is exactly \( \mu^T A\psi \).

Thus, we have that,

\[
1 - \varepsilon \leq \mu^T A\psi = \sum_R (\mu^T A_R) \cdot \psi(R)
\]

Thus, there exists an \( R_0 \) such that \( \mu^T A_{R_0} \geq 1 - \varepsilon \), which means that the deterministic protocol \( \pi_{R_0} \) makes error with probability less than \( \varepsilon \) when the input is coming from distribution \( \mu \).

\([R_\varepsilon(f) \leq \max_{\mu} D^\mu_\varepsilon(f)]\) Suppose we know that for any \( \mu \), there exists a protocol \( \pi_R \) with \( CC(\pi_R) \leq c \), such that \( \Pr_{(x,y)\sim\mu}[\pi_R(x, y) = f(x, y)] \geq 1 - \varepsilon \). This can equivalently be written as,

\[
\forall \mu \exists \psi \mu^T A\psi \geq 1 - \varepsilon \quad \iff \quad \min_{\psi} \max_{\mu} \mu^T A\psi \geq 1 - \varepsilon
\]
From LP duality, we know that \( \max_a \min_b a^T A b = \min_b \max_a a^T A b \), and hence we have that,

\[
\max_{\psi} \min_{\mu} \mu^T A \psi \geq 1 - \varepsilon \quad \equiv \quad \exists \psi \forall \mu \mu^T A \psi \geq 1 - \varepsilon
\]

That is, there exists a randomized protocol \( \pi \), given by the distribution \( \psi \), which makes error with probability less than \( \varepsilon \) when the inputs are coming from any distribution \( \mu \). In particular, when the \( \mu \) is the single-support distribution \( \{(x, y)\} \) we get that \( \pi \) makes error with probability at most \( \varepsilon \) on input \( (x, y) \). Thus, we have the desired randomized protocol.

5 Discrepancy method

With Yao’s min-max principle, we obtain that in order to prove a lower bound on \( R(f) \), it suffices to prove a lower bound on \( D \) for any \( \mu \) of our choice!

Suppose \( \pi \) is a deterministic protocol that makes error less than \( \varepsilon \) when inputs are drawn from distribution \( \mu \). We know that any deterministic protocol partitions the rectangle \( M_f \) into rectangles, but now these rectangles need not be monochromatic. But since the protocol makes error at most \( \varepsilon \), each of these rectangles must be almost monochromatic.

**Definition 5** (Discrepancy). Let \( A_0 = f^{-1}(0) \) and \( A_1 = f^{-1}(1) \). Define \( \text{disc}_\mu(f) \), the discrepancy of \( f \) under distribution \( \mu \) as,

\[
\text{disc}_\mu(f) = \max_{R = S \times T} |\mu(R \cap A_0) - \mu(R \cap A_1)|
\]

Now suppose protocol \( \pi \) makes error at most \( \varepsilon \) on distribution \( \mu \). Let \( C \subseteq \{0, 1\}^n \times \{0, 1\}^n \) be the set of inputs on which \( \pi \) works correctly. Thus, we have that,

\[
1 - 2\varepsilon \leq \Pr_\mu[(x, y) \in C] - \Pr_\mu[(x, y) \notin C]
= \sum_R \Pr_\mu[(x, y) \in C \cap R] - \Pr_\mu[(x, y) \notin C \cap R]
\leq \sum_R |\Pr_\mu[(x, y) \in C \cap R] - \Pr_\mu[(x, y) \notin C \cap R]|
\leq \sum_R \text{disc}_\mu(f)
\leq 2^{\text{CC}(\pi)} \text{disc}_\mu(f)
\]

where the summation is over the rectangles in the partition induced by \( \pi \). Thus, we obtain the following theorem.

**Theorem 3** (Discrepancy lower bound). For any function \( f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\} \) and any distribution \( \mu \):

\[
D^\mu(f) \geq \log_2 \left( \frac{1 - 2\varepsilon}{\text{disc}_\mu(f)} \right)
\]
**Lower bound on $\IP_n$**

We will now show a lower bound on $R(\IP_n)$ using the discrepancy method.

For convenience we will look at the $\{1,-1\}$ representation of $M_{IP}$, let’s call it $H$. Namely, $H(x,y) = (-1)^{\sum_i x_i y_i}$. We first show a simple observation,

**Observation 1.** $HH^T = 2^n I_{2^n \times 2^n}$

**Proof.** Let $G = HH^T$. We have that

$$G(x,y) = \sum_z H(x,z) \cdot H(z,y)$$

$$= \sum_z (-1)^{\sum_i x_i z_i + \sum_i z_i y_i}$$

$$= \sum_z (-1)^{\sum_i z_i (x_i + y_i)}$$

$$= 2^n \cdot \mathbb{1}_{x=y}$$

\[ \square \]

**Theorem 4.** $R(\IP_n) \geq \Omega(n)$

**Proof.** Let $\mu$ be the uniform distribution on $\{0,1\}^n \times \{0,1\}^n$. We will show that $D^\mu(\IP_n) \geq \Omega(n)$. For any rectangle $R = S \times T$, we have that,

$$|\mu(R \cap A_0) - \mu(R \cap A_1)| = \frac{\mathbb{1}_S \cdot H \cdot \mathbb{1}_T}{2^{2n}}$$

$$\leq \frac{\|\mathbb{1}_S\|_2 \cdot \|H\|_2 \cdot \|\mathbb{1}_T\|_2}{2^{2n}}$$

$$\leq \frac{\sqrt{2^n} \cdot \sqrt{2^n} \cdot \sqrt{2^n}}{2^{2n}} \cdot \|H\|_2 = 2^{n/2}$$

$$= \frac{1}{2^{n/2}}$$

Thus, we have that $\text{disc}_\mu(\IP) \leq 2^{-n/2}$ and hence $D^\mu(\IP_n) \geq \Omega(n)$. \[ \square \]

**6 Some recent results**

Although the log-rank conjecture remains tantalizingly open, there was a recent breakthrough progress by Shachar Lovett.

**Theorem 5 ([1]).** For all $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\} : D(f) \leq O(\sqrt{r \log(r)})$ where $r = \text{rank}(M_f)$.

The proof uses a discrepancy-based upper bound!

Discrepancy has also been used to obtain some circuit complexity separations! A well-known result (due to Allender) states that every function in $\text{AC}^0$ can be computed by a depth-3 threshold circuit of quasipolynomial size. It was an open problem to determine whether Allender’s
simulation is optimal. Specifically, Krause and Pudlák asked whether every function in $AC^0$ can be computed by a depth-2 threshold circuits of quasipolynomial size. A recent work of Alexander Sherstov has resolved this problem, using discrepancy related techniques from communication complexity.

**Theorem 6 ([2]).** There is a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, explicitly given and computable by an $AC^0$ circuit of depth-3, whose computation requires a majority vote of $\exp(\Omega(n^{1/5}))$ linear threshold gates.

**References**
