12. Introduction to communication complexity

Lecturer: Sagnik Mukhopadhyay *Scribe: Sagnik Mukhopadhyay*

12.1 Yao's two-party communication model

The model [Yao79] consists of two parties, Alice and Bob, each holding inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. They wish to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Each of them have unbounded computational power, unlike as in other models of computation. However their goal is to compute the value of $f(x, y)$ with the least amount of interaction with each other. The measure of communication between the parties will usually be the number of bits exchanged between them. Typically the inputs are boolean, namely $\mathcal{X} = \{0, 1\}^n$ and $\mathcal{Y} = \{0, 1\}^m$, and the output is generally 1 bit, namely $\mathcal{Z} = \{0, 1\}$. We assume the communication will be carried out according to some fixed protocol π, which Alice and Bob have mutually decided beforehand. At each stage, the protocol must decide whether the run terminates. If it does, then it must specify the answer $f(x, y)$, else it must specify which player communicates next. Moreover, this information must be decided by only the bits communicated between them so far in this run of the protocol. Additionally, if it is Alice's (Bob's) turn, the protocol must decide what they send depending on the communication so far and Alice's (Bob's) input. The following figure describes a protocol.

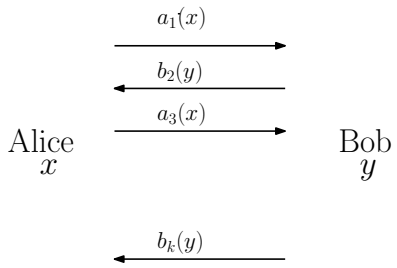

Figure 1: Communication protocol

Here, $b_i(y)$ is a message which depends on Bob's input y and the communication so far. Similarly, $a_j(x)$ is a message which depends on Alice's input x and the communication so far. At each round, the protocol determines whether the run terminates. If at the k-th round the protocol terminates, then $b_k(y)$ is the output of the protocol, whose value is $f(x, y)$. The cost of the protocol π on input (x, y) is the number of bits communicated by π over the input (x, y). The cost of a protocol is the worst case cost of π over all inputs (x, y). Finally, the deterministic communication complexity of a function f is the minimum cost of a protocol computing f. It will be denoted by $\mathsf{D}^{\mathrm{cc}}(f)$.

$$\mathsf{D}^{\mathrm{cc}}(f) = \min_{\pi} \max_{(x,y)} |\pi(x, y)|,$$

where π ranges over all protocols computing f, and (x, y) over $\mathcal{X} \times \mathcal{Y}$.

12.1.1 Some examples

- The parity function $\mathsf{PAR}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ function over n variables are defined as follows:

$$\mathsf{PAR}_n(x, y) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i + y_i \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

  where the summation is on $\mathbb{F}_2$. We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{PAR}_n) \leq 2$.

- The majority function $\mathsf{MAJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ function over n variables are defined as follows:

$$\mathsf{MAJ}_n(x, y) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i + y_i \geq n + 1, \\ 0 & \text{otherwise.} \end{cases}$$

  where the summation is over natural numbers. We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{MAJ}_n) \leq \log n + 1$.

- The $\mathsf{EQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ function over n variables are defined as follows:

$$\mathsf{EQ}_n(x, y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

  We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{EQ}_n) \leq n + 1$.

- The $\mathsf{DISJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ function over n variables are defined as follows:

$$\mathsf{DISJ}_n(x, y) = \begin{cases} 1 & \text{if } x \cap y = \varnothing, \\ 0 & \text{otherwise.} \end{cases}$$

  Think of $x, y \in \{0,1\}^n$ to be characteristic vectors of sets $x, y \subseteq [n]$. We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{DISJ}_n) \leq n + 1$.

12.2 Protocol tree

To give protocols a more combinatorial structure, consider the following *canonicalization* of a protocol:

> Alice and Bob communicate only through messages of length one bit.

Remark 12.1. *Canonicalization does not increase the communication complexity by much.*

Canonicalization gives rise to a representation of a protocol known as the protocol tree: A protocol tree is a rooted binary tree which specifies at any point during the execution of the protocol which party should communicate and what bit they should send as a function of their input.

- The nodes of the tree correspond to states of the protocol execution. (One to one correspondence with the partial transcripts.)
- Each node (apart from the leaves) is labeled with either Alice or Bob, denoting that it is their turn to communicate.
- Each node (apart from the leaves) has a function associated with it (either $A_v : \mathcal{X} \to \{0,1\}$ or $B_v : \mathcal{Y} \to \{0,1\}$) that maps the communicating party's input to the bit that the party should send.
- The root node corresponds to the beginning of the protocol execution.
- On communicating a bit, the state of the protocol execution changes to the left or right child of the current state, depending on whether the bit sent was 0 or 1.
- The leaves of the protocol tree are labeled with a 0 or a 1, denoting the output of the execution.

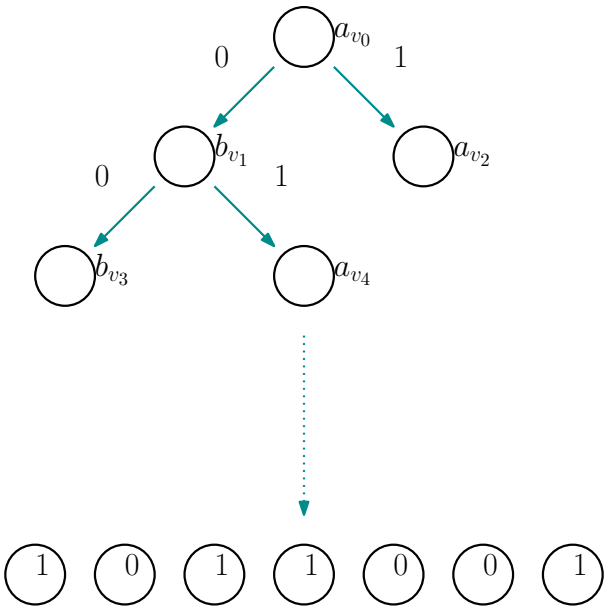

Figure 2: Protocol tree

Given input (x, y), the output of every function A_v and B_v is fixed. This fixes a path from the root to a leaf, the label of the leaf being the output of the protocol tree. A protocol tree is correct for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ iff for every input (x, y), the label of the leaf node reached is $f(x, y)$. The cost of a protocol is the length of the longest path in the protocol tree.

12.2.1 Combinatorial rectangles

Definition 12.2 (Combinatorial rectangles)**.** *$R \subseteq \mathcal{X} \times \mathcal{Y}$ is a combinatorial rectangle (or, simply, rectangle) if $R \equiv A \times B$ for some $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$.*

Claim 12.3. *R is a rectangle iff $\forall x, x', y, y'$; $(x, y), (x', y') \in R \Rightarrow (x', y), (x, y') \in R$.*

Proof. It is easy to see that every rectangle satisfies the given condition.

Now consider any set S satisfying the condition. Let A be the set of all the first members of elements of S and B be the set of all the second members of element of S. Now for any elements $x \in A$, $y \in B$ there must be some element of S having first member x and some element of S having second member y. Therefore, by the condition, $(x, y) \in S$. Furthermore if some $(x, y) \notin A \times B$, then it is clearly not in S. Therefore $S \equiv A \times B$ is a rectangle. □

Lemma 12.4. *A protocol of cost c partitions $\mathcal{X} \times \mathcal{Y}$ into atmost 2^c rectangles.*

Monochromatic rectangle. Consider a rectangle R is the input space of a function f. R is said to be c-monochromatic if for every $(x, y) \in R$, $f(x, y) = c$.

Lemma 12.5. *A protocol of cost c partitions $\mathcal{X} \times \mathcal{Y}$ into atmost 2^c monochromatic rectangles.*

12.3 Composed functions & decision trees

Definition 12.6. *A decision tree is a binary tree which computes a function $f : \{0,1\}^n \to \{0,1\}$ in the following way.*

- *Leaves are labeled by 0 or 1.*
- *Non leaf nodes are labeled by variables.*
- *Each node has a '0-child' and a '1-child'.*

It is said to compute a function f if for all inputs $z \in \{0,1\}^n$, the value of f on z is the label of the leaf on the path induced by z as follows: Suppose we are currently at an internal node with label z_i . If $z_i = 0$, we go to the 0-child and continue, else to the 1-child.

The *query complexity* or *decision tree complexity* of f, denoted as $\mathsf{D}^{\mathsf{q}}(f)$, is the depth of the best protocol computing f.

12.3.1 One example

Consider the fork relation FORK which takes input any string z that starts with 0 and ends with 1, and outputs a position i of the string z such that $z_i = 0, z_{i+1} = 1$. We can show $\mathsf{D}^{\mathsf{q}}(\mathsf{FORK}) \leq \log n$.

12.3.2 Function composition

Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$. Define the composition $F \equiv f \circ g$ as follows: There are n blocks, each with m inputs to Alice and m inputs to Bob. Apply g on each of the blocks, and apply f on the resulting n bit string. Thus, F is a function from $\{0,1\}^{nm} \times \{0,1\}^{nm} \to \{0,1\}$, where $N = nm$.

$$F\Big(\langle x_1, \cdots, x_n\rangle, \langle y_1, \cdots, y_n\rangle\Big) = f\Big(g(x_1, y_1), \cdots, g(x_n, y_n)\Big).$$

For a composed function, g is sometimes referred to as a *gadget*.

Examples. Two functions we have seen before can be expressed as composed functions: EQ_n and DISJ_n.

$$\mathsf{EQ}_n(x,y) = \bigwedge_{i\in[n]} \overline{(x_i \oplus y_i)}; \qquad \overline{\mathsf{DISJ}}_n(x,y) = \bigvee_{i\in[n]} (x_i \wedge y_i).$$

12.3.3 Gadgets of interest

- The inner-product function on m-bits, denoted IP_m in defined on $\{0,1\}^m \times \{0,1\}^m$ to be:

$$\mathsf{IP}_m(x,y) = \sum_{i\in[m]} x_i \cdot y_i \bmod 2.$$

  We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{IP}_m) = \Theta(m)$.

- The indexing function on m-bits, denoted IND_m in defined on $\{0,1\}^m \times [m]$ to be:

$$\mathsf{IND}_m(x,y) = x_y.$$

  We can show that $\mathsf{D}^{\mathrm{cc}}(\mathsf{IND}_m) = \Theta(\log m)$.

12.3.4 Complexity for composed functions

Lemma 12.7 (Naïve upper bound)**.** *For any composed function $F \equiv f \circ g$, $\mathsf{D}^{\mathrm{cc}}(F) \leq \mathsf{D}^{\mathrm{q}}(f) \times \mathsf{D}^{\mathrm{cc}}(g)$.*

Proof sketch. Alice and Bob try to solve f using a decision tree algorithm. Such an algorithm queries the input bits of f frugally. Whenever there is a query, Alice and Bob solve the relevant instance of g by using the best communication protocol for g. □

Theorem 12.8 (Simulation theorem [RM99, GPW15, dRNV16, CKLM17])**.** *For any function f, $\mathsf{D}^{\mathrm{cc}}(f \circ \mathsf{IND}_m) \geq \mathsf{D}^{\mathrm{q}}(f) \times \log m$ for $n = m^{\Omega(1)}$.*

Theorem 12.9 (Simulation theorem [CKLM17])**.** *For any function f, $\mathsf{D}^{\mathrm{cc}}(f \circ \mathsf{IP}_m) \geq \mathsf{D}^{\mathrm{q}}(f) \times m$ for $n = \Omega(\log m)$.*

We show a generalized version of these simulation theorems.

Definition 12.10 (Hitting rectangle-distributions)**.** *Let $0 \leq \delta < 1$ be a real, $h \geq 1$ be an integer, and $\mathcal{X}, \mathcal{Y}$ be some sets. A distribution σ over rectangles within $\mathcal{X} \times \mathcal{Y}$ is called a* (δ, h)-hitting rectangle-distribution *if, for any rectangle $A \times B$ with $|A|/|\mathcal{X}|, |B|/|\mathcal{Y}| \geq 2^{-h}$,*

$$\Pr_{R\sim\sigma}[R \cap (A \times B) \neq \varnothing] \geq 1 - \delta.$$

Let $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function. A rectangle $A \times B$ is c-monochromatic with respect to g if $g(x,y) = c$ for every $(x,y) \in A \times B$.

Definition 12.11. *For a real $\delta \geq 0$ and an integer $h \geq 1$, we say that a (possibly partial) function $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$* has (δ, h)-hitting monochromatic rectangle-distributions *if there are two (δ, h)-hitting rectangle-distributions σ_0 and σ_1, where each σ_c is a distribution over rectangles within $\mathcal{X} \times \mathcal{Y}$ that are c-monochromatic with respect to g.*

Now we show the following:

Theorem 12.12 (Generalized simulation [CKLM17])**.** *Let $\varepsilon \in (0,1)$ and $\delta \in (0, \frac{1}{100})$ be real numbers, and let $h \geq 6/\varepsilon$ and $1 \leq n \leq 2^{h(1-\varepsilon)}$ be integers. Let $f : \{0,1\}^n \to \mathcal{Z}$ be a function and $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function. If g has (δ, h)-hitting monochromatic rectangle-distributions then*

$$\mathsf{D}^{\mathrm{q}}(f) \leq \frac{4}{\varepsilon \cdot h} \cdot \mathsf{D}^{\mathrm{cc}}(f \circ g^n).$$

Now Theorem 12.8 and 12.9 follows from Theorem 12.12 by showing the following:

Theorem 12.13. *IP_m has $(o(1), m(\frac{1}{2} - \varepsilon))$-hitting monochromatic rectangle-distributions; and IND_m has $(\frac{1}{150}, \frac{3}{20} \log m)$-hitting rectangle-distribution.*

12.4 Hitting rectangle distribution for IP_m

We use the following two linear-algebra results which we state without proof. They can be proved by using second-moment method. For detailed proof, see [CKLM17].

Lemma 12.14. *Let $\varepsilon < \frac{1}{2}$ be a positive real number, and consider a set $B \subseteq \{0,1\}^m$ of density $\beta = \frac{|B|}{2^m} \geq 2^{-(\frac{1}{2}-\varepsilon)m}$. Pick V to be a random linear subspace of $\{0,1\}^m$ of dimension d, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})m + 6$. Then*

$$\Pr_V \left[\frac{|B \cap V|}{|V|} \in (1 \pm 2^{-\frac{\varepsilon}{4}m}) \cdot \beta \right] \geq 1 - \frac{1}{4} \cdot 2^{-\frac{\varepsilon}{4}m}.$$

We will show a similar result when we pick the set V in the following manner: First we pick a uniformly random odd-Hamming weight vector $a \in \{0,1\}^m$, and then we pick W to be a random subspace of dimension d within $a^\perp$, where $d \geq (\frac{1}{2} - \frac{\varepsilon}{4})m + 6$; then $V = a + W$.

Lemma 12.15. *Consider a set $B \subseteq \{0,1\}^m$ of density $\beta = \frac{|B|}{2^m} \geq 2^{-(\frac{1}{2}-\varepsilon)m}$. Pick V as described above. Then*

$$\Pr_V \left[\frac{|B \cap V|}{|V|} \in \beta(1 \pm 2^{-\frac{\varepsilon}{4}m}) \right] \geq 1 - 2^{-\frac{\varepsilon}{4}m}.$$

We define the distributions σ_0 and σ_1 by the following sampling methods:

Sampling from σ_0: We choose a uniformly-random $\frac{n}{2}$-dimensional subspaces V of $\mathbb{F}_2^m$, and let $V^\perp$ be its orthogonal complement; output $V \times V^\perp$.

Sampling from σ_1: First we pick $a \in \{0,1\}^m$ uniformly at random conditioned on the fact that a has odd Hamming weight; then we pick random subspace W of dimension $(m-1)/2$ from $a^\perp$, and let $W^\perp$ be the orthogonal complement of W *inside* $a^\perp$. We output $V \times V^\parallel$, where $V = a + W$ and $V^\parallel = a + W^\perp$.

The rectangles produced above are monochromatic as required. Also, V and $V^\perp$ of σ_0 are both random subspaces of dimension $\geq (\frac{1}{2} - \frac{\varepsilon}{4})m + 6$ — as required by Lemma 12.14 — and V and $V^\parallel$ of σ_1 are both obtained by the the kind of procedure required in Lemma 12.15. It then follows by a union bound that if R is chosen by either σ_0 or σ_1 that, if A, B are subsets of $\{0,1\}^m$ of densities $\alpha, \beta \geq 2^{-(\frac{1}{2}-\varepsilon)m}$, then

$$\Pr_R\left[\frac{|A \times B \cap R|}{|R|} = (1 \pm 9 \cdot 2^{-\frac{\varepsilon}{4}m}) \cdot \alpha\beta\right] \geq 1 - 2 \cdot 2^{-\frac{\varepsilon}{4}m}.$$

Hence the same probability lower-bounds the event that $A \times B \cap R \neq \varnothing$.

References

[CKLM17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017.

[dRNV16] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How limited interaction hinders real communication. In *Proceedings of the 56th FOCS*, 2016.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th FOCS*, 2015.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the 11h STOC*, pages 209–213, 1979.