CS 229r Information Theory in Computer Science	Feb 23, 2016
Lecture 9	
Lecturer: Madhu Sudan	*Scribe: Ali Vakilian*

In this lecture we explain the Communication Complexity (CC) model introduced by Yao [Yao79] and motivate the model with some well-studied examples. Furthermore, we define the probabilistic variant of CC model and at the end describe some the general techniques such as Rank, Discrepancy and Information theory for proving lower bounds in this model.

1 Basic Concepts and Definitions

What's the goal in Communication Complexity? Intuitively speaking, the goal of this model is to study the problems in which two (or more) players each has some input and they want to compute some pre-specified function f on their inputs. However, each of the players is only aware of its own input and (usually) does not know much about the other player's input. The goal is to give upper/lower bound on the *size of message* (in bits) they have to communicate in order to evaluate the function f over their inputs. More precisely, there are two players: Alice and Bob. Alice has input x and Bob has input y (see Figure 1). Alice and Bob want to compute $f(x, y)$ together with communicating the fewest possible number of bits (which works for all x, y in the input domain).

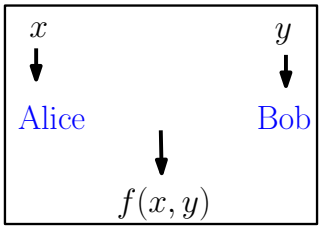


Figure 1: Alice has input x and Bob receives input y and the goal is to compute $f(x, y)$.

A more interesting question arises in the probabilistic variant of CC in which we assume that Alice and Bob are allowed to use *randomized* protocols and at the end they need to output $f(x, y)$ correctly with probability $(1 - \epsilon)$. In general we consider the following variants in CC.

- Deterministic Communication Complexity: As described above, the protocols of Alice and Bob are *deterministic* and at the end they need to compute $f(x, y)$ correctly (no error is allowed). $\mathsf{cc}_0(f)$ denotes the minimum number bits that is required for computing f in this model.
- Private Randomized Communication Complexity: In this model each of the players have some *private* random bits and their protocols are randomized (based on their private random bits) and at the end they must compute $f(x, y)$ correctly with probability $1 - \epsilon$ (probability is over random bits). Similarly, $\mathsf{cc}_\epsilon(f)$ denotes the minimum number bits that is required for computing f in this model.
- Public/Shared Randomized Communication Complexity: In this model, in addition to the private random bits, Alice and Bob are allowed to share some public random bits which are

independent of their input x, y. As before, Alice and Bob are required to compute $f(x, y)$ correctly with probability at least $1 - \epsilon$ and their protocol can be randomized on both their private random bits and shared random bits. Here, $\text{Priv-cc}_\epsilon(f)$ denotes the minimum number of bits that is required for computing f.

Note that it is straightforward to show that for any function f, $\text{cc}_0(f) \geq \text{cc}_\epsilon(f) \geq \text{Priv-cc}_\epsilon(f)$.

Examples: In the following we compute cc_0 of different functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$.

- The trivial upper bound we can give for any arbitrary function f is $\text{cc}_0 \leq n + 1$. Alice sends x to Bob (n bits). Then Bob computes $f(x, y)$ and send it (1 bit) to Alice. In particular, for many functions f, this naive approach is the best we can do.
- $\text{Parity}(x, y) = \oplus_{i=1}^n (x_i \oplus y_i)$: Alice sends $\oplus_{i=1}^n x_i$ to Bob (parity of her bits string which is a single bit) and then Bob computes $\text{Parity}(x, y)$ and return the solution to Alice. Thus $\text{cc}_0(\text{Parity}) \leq 2$.
- $\text{EQ}(x, y) = 1$ if $x = y$ and 0 otherwise: It is known that $\text{cc}_0(\text{EQ}) = n + 1$ and in this lecture we prove that $\text{cc}_0(\text{EQ}) \geq n$.

For proving lower bounds in this model, it is crucial to specify the model CC model carefully. Here, we give a formal description of protocols in Deterministic CC which is due to Yao [Yao79].

Problem) $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$ (In many cases we are only interested in computing f over some parts of the domain and for the rest we do not care about the output of f. Here "?" denotes the output to the parts we do not care. However, in this lecture we do not work with "?".)

Solution) Or **protocol** specifies at each step *who* speaks and *what* says. A k-bit protocol $\Pi = (\Pi_1, \cdots, \Pi_k)$ can be specified as follows:

- $\Pi_i = (z_i, f_i)$ where $z_i \in \{x, y\}$ and $f_i(z_i, b_1, \cdots, b_{i-1}) \in \{0,1\}$ where b_j denote the j^{th} communicated between Alice and Bob. Let g_A and g_B be respectively the output function of Alice and Bob.
  - In randomized CC models, $z_i \in \{(x, R_A, R), (y, R_B, R)\}$ where R_A, R_B and R respectively denote the random bits of Alice, Bob and the shared random bits.

We also need to specify the *correctness* and *efficiency* of a given protocol.

Correctness) If $\forall i, b_i = f_i(z_i, b_i \cdots, b_{i-1})$ then $\forall x, y : g_A(x, b_1, \cdots, b_k) = g_B(y, b_1, \cdots, b_k) = f(x, y)$.

- In randomized CC models, the condition is if $\forall i, b_i = f_i(z_i, b_i \cdots, b_{i-1})$ then $\forall x, y : \Pr[g_A(x, b_1, \cdots, b_k) = g_B(y, b_1, \cdots, b_k) = f(x, y)] \geq 1 - \epsilon$.
- Another CC model which is widely used in the literature is One-way CC. In this model the correctness condition is relaxed and we only require Bob to output $f(x, y)$. More precisely, the correctness in this model is if $\forall i, b_i = f_i(z_i, b_i \cdots, b_{i-1})$ then $\forall x, y : \Pr[g_B(y, b_1, \cdots, b_k) = f(x, y)] \geq 1 - \epsilon$.

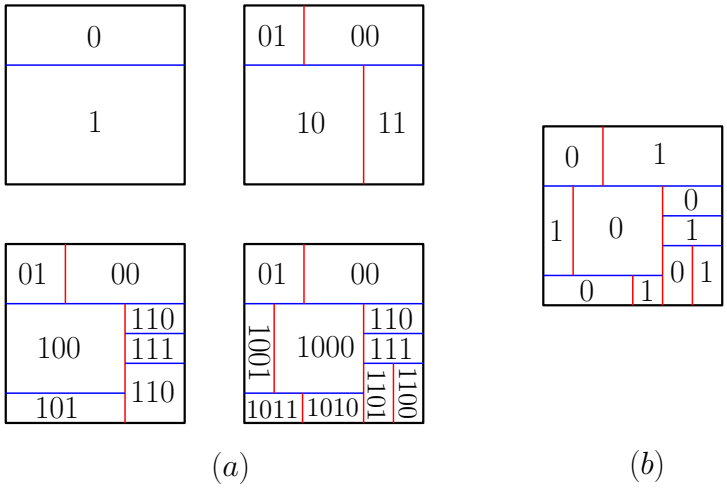
Efficiency) is defined as the length of the protocol. Here it is k.

2 Communication Complexity of EQ

In this section we focus on giving the promised lower bound for $\mathsf{cc}_0(\mathrm{EQ})$ via a *matrix* representation of the problem. Any function f (where $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$) can be shown by a $2^n \times 2^n$ matrix M_f whose rows correspond to the set of all possible inputs of Alice and columns correspond to all possible value of y and $M_f[x_i, y_j] = f(x_i, y_j)$.

Definition 1 (Rectangle) *A (combinatorial) rectangle in* $\{0,1\}^n \times \{0,1\}^n$ *is a subset* $R \subseteq \{0,1\}^n \times \{0,1\}^n$ *such that* $R = A \times B$ *for some* $A \subseteq \{0,1\}^n$ *and* $B \subseteq \{0,1\}^n$.

A useful observation in the matrix view of f in CC model is that at the end of Π_i, M_f is partitioned into a set of rectangles (at the beginning M_f is the only rectangle) and as Alice (Bob) communicates the next bit each existing rectangle may partition into at most two (combinatorial) rectangles (be careful about the definition of rectangles) by horizontal (vertical) lines. Let $\mathcal{R}_i$ be the set of the rectangles at the end of Π_i. Then for all values of x and y that lie in a rectangle $R \in \mathcal{R}_i$, the set of communicated bits is the same and if the protocol has to output at after Π_i, the output would be the same for all such x, y (see Figure 2). The rectangle for which the output is the same are called *monochromatic* rectangles (Figure 2(b)).



(a) (b)

Figure 2: (a) shows the set of communicated bits in each rectangle and (b) is the value of f at each rectangle.

The transcript of a protocol Π on (x, y) is defined as $t(x, y) := (b_1, \cdots, b_k)$, the set of communicated bits given (x, y). Then we have the following result which formalizes what we described above.

Claim 2 *For all* $t \in \{0,1\}^k$, *there exists* $R = S \times T$ *where* $S \subseteq \{0,1\}^n$ *and* $T \subseteq \{0,1\}^n$ *such that* $t(x, y) = t_R \iff (x, y) \in R$.

Proof By induction. ■

Claim 2 implies that if $t(x, y) = t(x', y')$ then $t(x', y) = t(x, y')$. Since g_A (g_B) only depends on $t(x, y)$ and x (y), correctness condition implies that the output all points in a rectangle of R_k have value $r_k \in \{0, 1\}$.

Conclusion) $\Pi = (\Pi_1, \cdots, \Pi_k)$ partitions M_f into 2^k monochromatic rectangles.

Algebraic view of monochromatic rectangles. If the number of monochromatic rectangles in M_f is 2^k $(R_1, \cdots, R_{2^k})$ then $\mathsf{rank}(M_f)$ is at most 2^k. The reason is that we can write M_f as sum of 2^k matrices $M_1, \cdots, M_{2^k}$ such that $M_i[x, y] = 0$ if $(x, y) \notin R_i$ and r_i if $(x, y) \in R_i$ where r_i is the value of a cell in R_i. It is straightforward to verify that $\mathsf{rank}(M_i) \leq 1$; hence, $\mathsf{rank}(M_f)$ is at most 2^k. This implies that $\mathsf{rank}(M_f) \leq 2^{\mathsf{cc}_0(f)} \iff \mathsf{cc}_0(f) \geq \log(\mathsf{rank}(M_f))$. This observation is particularly useful in proving lower bounds for many problems in Deterministic CC. As an application of *rank* method, we show how to bound the number of required bits of communication of EQ.

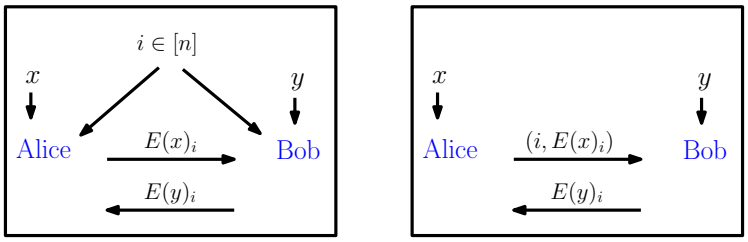
In order to apply rank method, first we need to figure out the matrix representation of EQ. $M_{\mathrm{EQ}} = \mathbb{I}$ and it is straightforward to check that the minimum number of required monochromatic rectangles of any protocol Π_{EQ} is 2^n. Thus any protocol Π_{EQ} requires at least n bits of communications.

Theorem 3 ([Yao79]) $\mathsf{cc}_0(\mathrm{EQ}) \geq n$.

2.1 Randomized Protocol for EQ

Using the following result from *error correcting codes*, we can design an efficient protocol for EQ with randomness.

Theorem 4 *There exists a function* $E : \{0,1\}^n \to \{0,1\}^n$ *such that* $\forall xy$, $E(x)$ *and* $E(y)$ *disagree in at least* $\frac{n}{10}$ *coordinates.*



(a) (b)

Figure 3: (a) shows an efficient protocol for EQ with public randomness and (b) shows the adoption of the same protocol with private randomness.

Consider the following protocol $\mathcal{P}$ for EQ using an error correcting function E.

1. Alice compute $E(x)$ and sends its i^{th} coordinate (which is determined by the public random bits) to Bob
2. Bob also computes $E(y)$ and sends its i^{th} coordinate to Alice.
3. They accept if their i^{th} coordinates agree.

Remark If we want to work with private randomness, Alice has to communicate the random index she selected as well (which requires $O(\log n)$ bits of communication). In fact, the protocol $\mathcal{P}$ gives the following bounds on communication complexity of EQ: $\mathsf{cc}_\epsilon(EQ) = O(1)$ and $\text{Priv-}\mathsf{cc}_\epsilon(O(\log n))$ where $\epsilon = O(1)$.

To show the correctness of $\mathcal{P}$ with constant probability note that,

- If $x = y$, $\quad$ Pr[Alice and Bob accept] $= 1$.
- If $x \neq y$, $\quad$ Pr[Alice and Bob accept]$< \frac{9}{10}$.

We can decrease the failure probability by repeating the protocol several times.

So far we have proved lower bound on the number of required bits of communication for problems in Deterministic CC model. However, it is more interesting to obtain lower bounds that rule out certain bound for randomized protocols as well. The main tool for proving lower bounds of randomized protocols is via Yao's principle which is stated as follows. Proving lower bound for the worst case performance of any randomized protocols solving P is the same as proving lower bound for the performance of any deterministic protocol P over a random instance. This leads to another variant of CC which is called Distributional Communication Complexity (see Section 4).

3 Set Disjointness Problem

Another problem in CC which has lots of applications in other area of TCS such as streaming algorithms is Set Disjointness (DISJ) problem. In DISJ_n, each of Alice and Bob gets a subset of $[n]$ and their goal is to determine whether their sets intersect. More precisely, for $S \subseteq [n]$ and $T \subseteq [n]$, DISJ $(S, T) = 1$ if $S \cap T \neq \emptyset$ and 0 otherwise.

Proving lower bounds for the deterministic protocols of DISJ_n is not so hard, however, a more challenging problem is to give a lower bound for the randomized protocols of the problem. One of the seminal results in CC is the following lower bound on $\mathsf{cc}_\epsilon(\mathsf{DISJ}_n)$ which we will see its proof later in this course.

Theorem 5 ([KS92, Raz92]) $\mathsf{cc}_\epsilon(\mathsf{DISJ}_n) \geq \Omega(n)$.

Theorem 5 has in particular many applications in proving lower bounds for problems in *streaming* setup. As an example, one of the well-known problems in the streaming setup is the Frequent Item problem in which given a data stream of items $\mathcal{E} \subseteq [n]^m$, the goal is to output the most frequent item in the stream using the lowest possible memory space. Suppose there exists an algorithm $\mathcal{A}$ for Frequent Item that returns the correct answer with probability $1 - \epsilon$ and uses $o(n)$ space. Then we can come up with the following protocol $\mathcal{P}$ for DISJ_n:

(a) Alice runs $\mathcal{A}$ on her inputs, S, (as an stream) and then sends the memory state of $\mathcal{A}$ to Bob.

(b) Bob resumes $\mathcal{A}$ on his input, T, from the state sent by Alice and sends to her the final output of $\mathcal{A}$ and a bit stating whether Bob has the item or not.

(c) Alice sends a bit to Bob stating whether she has the most frequent item or not.

It is straightforward to check that the described protocol returns 2 iff S and T intersect and 1 otherwise. Thus $\mathcal{P}$ solves DISJ_n with probability at least $1 - \epsilon$ using $o(n)$ bits of communication which is a contradiction.

Corollary 6 *Any (randomized) streaming algorithm of* Frequent Item *that with probability at least* $1 - \epsilon$ *returns a most frequent item requires* $\Omega(n)$ *space.*

4 Distributional Communication Complexity

In this setting we assume that inputs of both Alice and Bob come from a distribution over $\{0,1\}^n$: $(x,y) \sim \mu(\{0,1\}^n \times \{0,1\}^n)$.

Definition 7 (Distributional Communication Complexity) *The distributional communication complexity of a function $f : [n]^2 \to \{0,1\}$ over the distribution μ and with success probability at least $1-\epsilon$, denoted by $\mathsf{cc}_{\mu,\epsilon}(f)$, is the cost of the best deterministic protocol that gives the correct answer on at least $(1-\epsilon)$ fraction of all the inputs, weighted by μ.*

For any function f, we have the following relation between the cc_ϵ and $\mathsf{cc}_{\mu,\epsilon}$.

Proposition 8 *If $\mathsf{cc}_\epsilon(f) \le k$, then $\mathsf{cc}_{\mu,\epsilon}(f) \le k$.*

Proof If there exists a randomized protocol Π for function f with expected error at most ϵ (over μ and R), then there exists a deterministic protocol Π' whose expected error is at most ϵ (over μ).

$$\mathbb{E}_{R,(x,y)\sim\mu}[\mathsf{Error}(f,\Pi,x,y,R)] \le \epsilon \Rightarrow \exists r, \mathtt{s.t.} \quad \mathbb{E}_{(x,y)\sim\mu}[\mathsf{Error}(f,\Pi,x,y,R=r)] \le \epsilon.$$

■

It is usually useful to work with matrix view of Distributional CC.

- $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$. For convenience, we change 0 to -1 which you see that will help a lot!
- $M_{\mu,f}(x,u) = \mu(x,y)f(x,y)$

Observation 9 $\sum_{x,y} |M_{\mu,f}(x,y)| = 1$.

Consider (combinatorial) rectangles in M_f. By definition, $R_{S\times T}$ denotes the set of (x,y) such that $x \in S$ and $y \in T$ and are all either -1 or 1. Furthermore, we define matrix $M_{S,T}$ as follows: $M_{S,T}[x,y] = 0$ if $(x,y) \notin S \times T$ and r otherwise where $r \in \{-1,1\}$.

Definition 10 (Discrepancy function) *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function and let μ be a probability distribution on $\{0,1\}\times\{0,1\}$. For rectangle R in M_f, $\mathsf{Disc}_{\mu,f}(R) = |\sum_{(x,y)\in R} M_{\mu,f}(x,y)|$. Moreover, the discrepancy of f under μ is defined as the maximum discrepancy over all its rectangles, $\mathsf{Disc}_\mu(f) = \max_{R\in\mathcal{R}_f} \mathsf{Disc}_{\mu,f}(R)$.*

Lemma 11 *Error of protocol Π on f is ϵ iff $|\mathcal{R}| \cdot \mathsf{Disc}_\mu(f) \ge 1-2\epsilon$.*

Proof Let $\epsilon = \mathsf{Error}(\Pi, f)$. Then, $(1-\epsilon) - \epsilon = \mathbb{E}_{x,y}[\mu(x,y) \cdot \Pr(\text{returns correctly on}(x,y))] - \mathbb{E}_{x,y}[\mu(x,y) \cdot \Pr(\text{returns incorrectly on}(x,y))] \le \sum_{R\in\mathcal{R}} \mathsf{Disc}_{\mu,f}(R) \le |\mathcal{R}| \cdot \mathsf{Disc}_\mu(f)$. ■

Lemma 12 *If $\mathsf{cc}_{\mu,\epsilon}(f) \le k$ then there exists a rectangle R such that $\mathsf{Disc}(M_{\mu,f}, R) \ge 2^{-k}(1-2\epsilon)$.*

Proof It follows from Lemma 11 and the fact that $|\mathcal{R}| = 2^k$. ■

4.1 Inner Product Function

In this part, we study the Inner Product(IP) function and prove a lower bound on its communication complexity with (private) randomness. IP function of two vectors x, y is defined as $\mathrm{IP}(x, y) = \oplus_{i=1}^{n}(x_i \wedge y_i)$.

Using Lemma 12 and following lemma which will be shown in next lecture we obtain a lower bound on $\mathrm{cc}_\epsilon(\mathrm{IP})$.

Lemma 13 *For all R,* $\mathrm{Disc}(M_{\mathrm{uniform,IP}}, R) \leq 2^{\frac{-n}{2}}$.

References

[KS92] B. Kalyanasundaram and G. Schintger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[Raz92] A. A. Razborov. On the distributional complexity of disjointness. *Theo. Comp. Sci.*, 106(2):385–390, 1992.

[Yao79] A. C. 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.