Communication Complexity 21 Sep, 2011 (@ TIFR)

11. Disjointness

Lecturer: Jaikumar Radhakrishnan *Scribe: Swagato Sanyal*

In this lecture we will see lower bounds on communication complexity of two problems, using ideas from information theory. The first one is a bound on the deterministic communication complexity of the Pointer Chasing problem introduced in the last lecture (the proof presented is from [**?**]). The next one is a bound on randomized communication complexity of the Disjointness problem using private coins.

11.1 Pointer Chasing Problem

Here we consider the k-level pointer chasing problem. Alice has $L_0 = \{V_0\}$, and the goal is to find the V_k. We denote the vertex pointed to by v as $f(v)$. Assume that there is a $(k-1)$ round deterministic protocol solving this problem, with Bob starting the communication, such that each message contains at most cn bits. We want to bound c from below. To do that, we analyze the performance of this protocol on certain distribution of inputs. At each step we will start with inputs whose entrypy is high. We will fix V_i and the message sent at that round such that we can get down to a reduced problem by incurring little loss to the entropy of our inputs. Finally when no more message is to be exchanged, the lower bound on c will come from the requirement that the entropy of V_k must be zero bit (as the protocol is deterministic and correct). Choose $V_0, L_1, V_1, \dots, V_k, L_k$ uniformly at random from all possible inputs. Call the distribution of Alice's inputs (Bob's inputs) at step t, $X^{(t)}$ ($Y^{(t)}$). Thus $X^{(0)}$ and $Y^{(0)}$ are uniform distribution. Now let us look at step $t (t \geq 0)$. We prove the following by induction on t.

Claim 11.1. *There exists messages $m_1, m_2, \dots, m_t$ and vertices $v_1 \in L_1, v_2 \in L_2, \dots, v_t \in L_t$ such that under $X^{(t)}$ ($Y^{(t)}$) $= X^{(0)}$ ($Y^{(0)}$) conditioned on $M_1 = m_1, \dots, M_t = m_t$,$V_1 = v_1, \dots, V_t = v_t$, the following holds:*

1. $H[f(V_t)] \geq \log n - \delta(n)$
2. $H[X^{(t)}] \geq kn \log n - (2^t - 1)cn$
3. $H[X^{(t)}] \geq kn \log n - (2^t - 1)cn$

where $\delta()$ is the function given by the relations $\delta(0) = 0$ and for $j \geq 1$, $\delta(j) = (2^j - 1)2^{2(\delta(j-1)+1)}).c.$

Proof. The base case ($t = 0$) is clearly true, as we start with uniform distribution over all inputs. For the industion step, we assume it to be true for $t = i - 1$ and show it to be true for $t = i$. We assume that i is odd so that after choosing the m_j's and v_j's, Alice holds the starting vertex. The proof of the other possibility is identical. According to the inductive hypothesis, we have mesages $m_1, m_2, \dots, m_{i-1}$ and vertices $v_1 \in L_1, v_2 \in L_2, \dots, v_i \in L_{i-1}$ such that under $X^{(i-1)}$ and $Y^{(i-1)}$,

1. $H[f(V_{i-1})] \geq \log n - \delta(i-1)$
2. $H[X^{(i-1)}] \geq kn \log n - (2^{i-1} - 1)cn$
3. $H[Y^{(i-1)}] \geq kn \log n - (2^{i-1} - 1)cn$

We first describe how we choose m_i. If we fix an m_i, knowledge of that m_i does not affect the entropy of Alice's distribution. According to our hypothesis, $H[Y^{(i-1)}|M_i]$ is at least $kn \log n - (2^{i-1} - 1)cn - cn$. So there is one m_i for which $H[Y^{(i-1)}|M_i = m_i]$ is at least $kn \log n - (2^{i-1} - 1)cn - cn \geq kn \log n - (2^i - 1)cn$. Since fixing V_i which is part of Alice's input does reveal anything about Bob's input, $H[Y^{(i)}] = H[Y^{(i-1)}|M_i = m_i] \geq kn \log n - (2^i - 1)cn$. Thus 3 holds after i-th round.
Now we describe how we fix v_i. From inductive hypothesis, $H[X^{(i-1)}|V_i] \geq kn \log n - (2^{i-1} - 1)cn - \log n$. Again from inductive hypothesis, $H[V_i] \geq \log n - \delta(i-1)$. Let the set of V_i's for which $H[X^{(i-1)}|V_i] \geq n \log n - 2((2^{i-1} - 1)cn - \log n)$ be $\mathcal{N}$. Let $|\mathcal{N}| = N$. By Markov's inequality we have that the total probability p of vertices in $\mathcal{N}$ is at least $1/2$. Also the sum of entropies of vertices of $\mathcal{N}$ is at most $\log N$. Thus $H[V_i]$ is at most $p \log N + (1-p) \log n$. As $p \geq 1/2$, we have $N \geq kn/2^{2(\delta(i-1)+1)}$. Choose one vertex v_i from the set $\mathcal{N}$ for which $H[f(v_i)]$ is maximum. $H[X^{(i)}]$ is at least $kn \log n - 2((2^{i-1} - 1)cn - \log n) \geq kn \log n - (2^i - 1)cn$. Thus 2 holds after i-th round. Finally, to show that 1 holds after i-th round, note that $H[Y^{(i)}] \leq \Sigma_{v \notin \mathcal{N}} \log n + \Sigma_{v \in \mathcal{N}} f(v)$. Using 3, the fact that $N \geq kn/2^{2(\delta(i-1)+1)}$, and the fact that v_i is the vertex in $\mathcal{N}$ for which $H[X^{(i-1)}|V_i = v_i]$ is maximum, we have $H[V_i] \geq \log n - \delta(i)$. □

After k-th round we thus have $H[f(V_{i-1})] \geq \log n - \delta(k-1)$. But since we are not allowed any further message, and the protocol is deterministic and correct, we must have $H[f(V_{i-1})] = 0$. This gives us $\log n \leq \delta(k-1)$ which gives us $c = \Omega(\log^{(k-1)} n)$. If we denote the k-level pointer chasing function by g_k and for any function f, the cost of the best deterministic protocol to compute f where Alice(Bob) sends the first message by $C^{A,k}(f)(C^{B,k}(f))$, then we have proved the following theorem:

Theorem 11.2 (PRV)**.** *$C^{B,k}(g_k) = \Omega(n \log^{(k-1)} n)$ for any fixed k.*

11.2 Disjointness Problem

In this problem, Alice and Bob are both given subsets X and Y of $[n]$. They are to decide whether the sets are disjoint or not.

$$\mathsf{DISJ}_n(X,Y) \quad = \quad \begin{cases} 1 & \text{if } X \cap Y = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

The parties are allowed to use private randomness. The sets X and Y can also be thought of as bit vectors with n components (which are the characteristic vectors of the sets). Here onwards we will mean vectors in $\{0,1\}^n$ whenever we talk of inputs. DISJ_n under this representation can be written as follows

$$\mathsf{DISJ}_n(X,Y) = \bigwedge_{i=1}^{n} \left(\overline{X_i} \vee \overline{Y_i}\right) = \bigwedge_{i=1}^{n} \mathsf{NAND}(X_i, Y_i)$$

We prove the following result by Kalyanasundaram, Schnitger, the proof of which was simplied by Razborov and further by Bar-Yossef, Jayram, Kumar and Sivakumar. We will present the proof along the lines of Bar-Yossef et al.

Theorem 11.3 ([?, ?, ?]).

$$\mathrm{R}_{\frac{1}{2}-\varepsilon}(\mathsf{DISJ}_n) = \Omega(\varepsilon^2 n).$$

The proof has two broad parts. In the first part it is shown that existense of a efficient randomized private coin protocol for disjointness implies existense of a randmomized private coin protocol for 2-input NAND function with the property that for a particular input distribution, the information about its inputs that its transcript contains is very small. The second part derives a contradiction to it by showing that for any protocol for NAND, for that particular input distribution, the information about its inputs in its transcript has to be large.

Let us fix a $(1/2-\varepsilon)$-error private coin randmomized protocol Π for computing the Disjointness problem of length at most δn for some $\delta > 0$. Let $\Pi = \Pi(X,Y) = \Pi(X,Y;R_X,R_Y)$ stand for the transcript of the protocol, and let $|\Pi| = m$.

We first define two distribution η_A and η_B on $\{0,1\}^2$ as follows.

$$\Pr_{(X,Y)\sim\eta_A}[(X,Y)=(1,0)] = \Pr_{(X,Y)\sim\eta_A}[(X,Y)=(0,0)] = \frac{1}{2}$$
$$\Pr_{(X,Y)\sim\eta_B}[(X,Y)=(0,1)] = \Pr_{(X,Y)\sim\eta_B}[(X,Y)=(0,0)] = \frac{1}{2}$$

Both η_A and η_B are supported on the YES instances of NAND. For η_A, Y is always set to 0 while X is 0 or 1 with equal probability and the other way round for η_B. Thus, "Alice is active" in η_A while "Bob is active" in η_B. Observe that in both η_A and η_B, X and Y are independent of each other.

For every $\sigma \in \{A,B\}^n$, we define joint random variables (X^σ, Y^σ) with distribution μ_σ on $(\{0,1\}^n)^2$ as follows: for each $i \in [n]$ independently do the following, if $\sigma_i = A$, set $(X_i^\sigma, Y_i^\sigma) \sim \eta_A$. In this case, we say "Alice is active" in coordinate i. if $\sigma_i = B$, set $(X_i^\sigma, Y_i^\sigma) \sim \eta_B$. In this case, we say "Bob is active" in coordinate i. By construction, $(X_i^\sigma, Y_i^\sigma)_{i=1}^n$ is independent across the coordinates for any σ.

Remark 11.4. *It might appear strange that we are choosing a distribution which is supported on the YES instances of disjointness (and* NAND*). Clearly, there exists a 1-bit protocol that obtains the correct answer on this distribution and furthermore this transcript carries zero information about the inputs. How do we manage to get a lower bound then? The reasoning goes as follows. We have assumed the protocol is correct on every input, not just on the inputs supported by the distribution. We will use this fact to show that the transcript must carry some information about the inputs even over this seemingly strange distribution supported on the YES instances. Note that this is similar in spirit to the fooling set argument used to bound deterministic communication complexity.*

We will consider the quantity $I[XY : \Pi]$, which is the amount of information the transcript contains about the pair of inputs. This quantity is at most the entropy of Π which in turn is at most the length of Π. We thus, have that $I[XY : \Pi] \leq \delta n$. Since for each σ, the pair (X_i^σ, Y_i^σ) is independent over different i, it follows from the sub-additivity of mutual information that

$$\delta n \geq I[X^\sigma Y^\sigma : \Pi(X^\sigma, Y^\sigma)] \geq \sum_{i=1}^{n} (I[X_i^\sigma Y_i^\sigma : \Pi]).$$

If we choose a coordinate $k \in [n]$ uniformly at random, then

$$\mathbb{E}_k(I[X^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

The above expectation is over choice of k and for every fixed σ. Now we choose a σ uniformly at random from $\{A, B\}^n$ and take expectation over σ to obtain

$$\mathbb{E}_\sigma \mathbb{E}_k(I[X_k^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

Interchanging the order expectations, we have

$$\mathbb{E}_k \mathbb{E}_\sigma(I[X_k^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

Thus, there exists a coordinate k, such that,

$$\mathbb{E}_\sigma(I[X_k^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

We fix such a k. We now split the choice of σ into two parts: choosing active party for k-th coordinate (denoted by $\sigma_k \in \{A, B\}$) and choosing active parties for the remaining coordinates (denoted by $\sigma_{-k} \in \{A, B\}^{n-1}$). Thus we have,

$$\mathbb{E}_{\sigma_{-k}} \mathbb{E}_{\sigma_k}(I[X_k^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

Thus, there exists a σ_{-k}, such that

$$\mathbb{E}_{\sigma_k}(I[X_k^\sigma Y_k^\sigma : \Pi]) \leq \delta.$$

Fix such a σ_{-k} for the rest of the argument. Expanding the above expectation over the two possible choices for σ_k, we get

$$\frac{1}{2}\left(I[X_k^\sigma Y_k^\sigma : \Pi | \text{``Alice is active''}] + I[X_k^\sigma Y_k^\sigma : \Pi | \text{``Bob is active''}]\right) \leq \delta. \tag{11.2.1}$$

The above equations tells us that the transcript hardly contains any information about the inputs in the k-th coordinate. We will use this to show that the protocol Π cannot compute the NAND of the k bits. For this purpose, we describe a protocol π for NAND based on the protocol Π. Suppose Alice and Bob are given input bits x and y respectively. First, Alice and Bob construct n bit inputs X and Y for DISJ_n function from x and y respectively as follows: Alice and Bob set the k-bit of X and Y to be x and y respectively (i.e., $X_k = x$ and $Y_k = y$). For the remaining $n-1$ bits, Alice and Bob behave as follows. For each $i \neq k$, $\sigma_{-k}|_i$ tells if Alice or Bob is active. If Alice is active (ie., $\sigma_{-k}(i) = A$), then Alice sets X_i with equal probability to 0 or 1, while Bob sets Y_i to be 0. Similarly, if Bob is active, then Bob sets Y_i with equal probability to 0 or 1, while Alice sets X_i to be 0. Observe, that all of this can be done by Alice and Bob independently using their private randomness and the knowledge of k and σ_{-k}. They, then run the protocol Π on this input (X, Y). Since

$\mathsf{NAND}(X_i, Y_i) = 1$ for all $i \neq k$, we have that $\mathsf{DISJ}_n(X, Y) = \mathsf{NAND}(x, y)$. Hence, if Π is a protocol that computes DISJ_n correctly on every input with error at most $1/2 - \varepsilon$, then π is a protocol that computes NAND correctly on every input with error at most $1/2 - \varepsilon$. It is to obtain this conclusion, that we chose distributions η_A and η_B which were entirely supported on the YES instances of NAND. Let us know rewrite (11.2.1) which was written for protocol Π in terms of the protocol π.

$$I[XY : \pi(X,Y) | (X,Y) \sim \eta_A] + I[Y : \pi(X,Y) | (X,Y) \sim \eta_B] \leq 2\delta.$$

Recall that when $(X, Y) \sim \eta_A$, we have X is a random bit and Y is 0 (and the other way round for η_B). We can thus, rewrite the above inequality as follows.

$$I[Z : \pi(Z, 0)] + I[Z : \pi(0, Z)] \leq 2\delta, \tag{11.2.2}$$

where Z is a random bit that takes 0 and 1 with equal probability.

What does the above inequality imply? The fact that $I[Z : \pi(Z, 0)]$ is small reveals that conditioned on Bob's bit being 0, the transcript does not contain much information about Alice's random bit Z. In other words, the transcript distributions $\pi(0, 0)$ and $\pi(1, 0)$ look alike. Similarly, from the fact that $I[Z : \pi(0, Z)]$ is small, we conclude that the transcript distributions $\pi(0, 0)$ and $\pi(0, 1)$ look alike. But then the transcript distributions $\pi(0, 1)$ and $\pi(1, 0)$ must also look alike. This by itself is not surprising since $\mathsf{NAND}(0, 1) = \mathsf{NAND}(1, 0) = 1$. Let us pretend that π is a determinstic protocol and "look alike" means the transcripts are identical. Now, if the transcripts $\pi(0, 1)$ and $\pi(1, 0)$ are identical, then so do the transcripts $\pi(0, 0)$ and $\pi(1, 1)$ by the rectangle property of deterministic protocols. However this cannot be true, since $\mathsf{NAND}(0, 0) = 1$ while $\mathsf{NAND}(1, 1) = 0$ and the protocol π distinguishes between YES and NO instances of NAND. This leads us to our contradiction. We will get around the fact that π is randomized and the transcript distributions only look alike and are not identical, using a formalization via *Hellinger distance*.

Definition 11.5 (Hellinger distance)**.** *Let $P = (p_i)_i$ and $Q = (q_i)_i$ be two probability distributions on some universe. The* Hellinger Distance *$h(P, Q)$ between them is defined as*

$$h(P, Q) = \sqrt{\frac{1}{2} \sum_i (\sqrt{p_i} - \sqrt{q_i})^2}$$

We will study Hellinger Distance in greater detail in the next lecture. In this lecture, we state without proof a few properties of Hellinger Distance, which we will help us complete the disjointness lower bound. The first one relates Hellinger distance to the standard total variation distance between two distributions.

Lemma 11.6 (Hellinger vs. total variation distance)**.**

$$\Delta(p - q) = \frac{1}{2} \cdot \|p - q\|_1 \leq \sqrt{2} \cdot h(p, q).$$

The next one relates Hellinger Distance to information.

Lemma 11.7 (Hellinger vs. Information)**.** *Let a random variable Z be distributed uniformly over $\{z_1, z_2\}$ and $\pi(Z)$ be a function (possibly randomized) of Z. Then*

$$I[Z : \pi(z)] \geq h^2(\pi(z_1), \pi(z_2))$$

The third lemma is the reason why Hellinger is useful while studying randomized communication protocols. It is the extension of the rectangle property for deterministic protocols to private-coins randomized protocols.

Lemma 11.8 (cut-and-paste lemma)**.** *Let $\pi(x,y)$ denote the transcript of a randomized protocol of some communcation problem on input (x,y). Then for all x, x', y, y',*

$$h^2(\pi(x,y'),\pi(x',y)) = h^2(\pi(x,y),\pi(x',y'))$$

We can now complete the lower bound for disjointness formalizing the above intuition using Hellinger distance.

$$\begin{aligned}
2\delta &\geq I[Z:\pi(Z,0)] + I[Z:\pi(0,Z)] && \text{[from (11.2.2)]}\\
&\geq h^2(\pi(0,0),\pi(1,0)) + h^2(\pi(0,0),\pi(0,1)) && \text{[from Lemma 11.7]}\\
&\geq \frac{1}{2}\cdot\left(h(\pi(0,0),\pi(1,0)) + h(\pi(0,0),\pi(0,1))\right)^2 && \text{[Cauchy-Schwarz inequality]}\\
&\geq \frac{1}{2}\cdot h^2(\pi(1,0),\pi(0,1)) && \text{[triangle inequality, since } h \text{ is a metric]}\\
&= \frac{1}{2}\cdot h^2(\pi(0,0),\pi(1,1)) && \text{[by cut-and-past lemma 11.8]}\\
&\geq \frac{1}{4}\cdot \Delta^2(\pi(0,0),\pi(1,1)) && \text{[by Lemma 11.6]}\\
&\geq \varepsilon^2
\end{aligned}$$

The last inequality follows since the protocol π outputs NAND with error at most $1/2-\varepsilon$. Hence, the total variation distance between the transcript distributions $\pi(0,0)$ and $\pi(1,1)$ is at least 2ε. Hence, $\delta \geq \varepsilon^2/2$. We thus, have $\mathrm{R}_{1/2-\varepsilon}(\mathsf{DISJ}_n) \geq \varepsilon^2 n/2$.