Communication Complexity

Boaz Barak

October 4, 2012

This lecture is loosely based on lectures 8 and 9 from Anup Rao's course (see also some background in Lecture 3).

Disjointness Recall the disjointness function $DISJ(x,y) = 1$ iff $x_i y_i = 0$ for all $i \in \{1..n\}$. Our goal in this lecture is to show $R(DISJ) \geq \Omega(n)$. This was first shown by Kalyanasundaram and Schnitger (1987), Razborov (1990) gave a simpler proof, and our exposition follows the information theoretic viewpoint of Bar-Yossef, Jayram, Kumar and Sivakumar (2002)

The disjointness problem is one of the most important questions in communication complexity, and lower bounds for disjointness have been used to derive lower bounds for streaming algorithms, data structures, barriers in complexity (algebraization), approaches for circuit lower bounds, proof complexity, game theory. Some sources on this include the survey "The Story of Set Disjointness" by Chattopadhyay and Pittasi `http://www.cs.toronto.edu/~arkadev/commSurvey.pdf`, and the paper "Unifying the Landscape of Cell-Probe Lower Bounds" by Patrascu `http://people.csail.mit.edu/mip/papers/structures/paper.pdf`.

We'll first prove that $R(DISJ) \geq \Omega(\sqrt{n})$, which was shown by Babai, Frankl, and Simon (1987), and then strengthen the result to a linear lower bound.

Needed facts on entropy, mutual information We are going to use the following facts about entropy, mutual information, and distances between measures: (see also entropy handout)

- Definitions: $H(X)$ is number of information bits in X, defined as $\sum_x \Pr[X=x]\log(1/\Pr[X=x])$. $I(X;Y)$ is number of bits one can learn about X from Y (or vice versa) defined as $H(X)+H(Y)-H(XY)$.
- If $X_1...X_n$ are independent, and you can learn m_i bits about X_i from Y, then you can learn at least$m_1 + .. + m_n$ bits about $X_1...X_n$ from Y. That is,

$$I(X_1...X_n;Y) \geq \sum_{i=1}^{n} H(X_i;Y)$$

- Conditioning: we define $H(X|Y)$ as the expectation over y of $H(X_y)$ where $X_y = X|Y=y$. Similarly define $I(X;Y|W)$ as the expectation of $I(X_w;Y_w)$ over w where $X_w = X|W=w$ and $Y_w = Y|W=w$.
- Distances between distribution: we define

$$\Delta_{\mathsf{TV}}(X,Y) = \tfrac{1}{2}\sum_w |\Pr[X=w]-\Pr[Y=w]| = \tfrac{1}{2}\max_{f:\mathrm{Domain}(X,Y)\to\{0,1\}} |\mathbb{E}f(X)-\mathbb{E}f(Y)|$$

and

$$\Delta^2_{\mathsf{Hel}}(X,Y) = \tfrac{1}{2}\sum_w \left|\sqrt{\Pr[X=w]} - \sqrt{\Pr[Y=w]}\right|^2 = 1 - \sum_w \sqrt{\Pr[X=w]\Pr[Y=w]}$$

It turns out we can move between these two distances freely via the following relation

$$\tfrac{1}{2}\Delta^2_{\mathsf{Hel}}(X,Y) \leq \Delta_{\mathsf{TV}}(X,Y) \leq \Delta_{\mathsf{Hel}}(X,Y)$$

- Mutual information and distances: if $I(X;Y)$ is small then the distribution Y is close to the distribution $Y|X$. That is, we have the following lemma:

  Lemma: Let $Y_x = Y|X=x$, then $\mathbb{E}_{x\in X}\Delta^2_{\mathsf{Hel}}(Y,Y_x) \leq I(X;Y)$.

  Proof: We use the fact that for every two distributions X,Y, $\Delta^2_{\mathsf{Hel}}(XY, X\times Y) \leq \Delta_{\mathsf{KL}}(XY\|X\times Y) = I(X;Y)$. So, letting $p(x) = \Pr[X=x]$, $p(y) = \Pr[Y=y]$ and $p(y|x) = \Pr[Y=y|X=x]$, we need to prove that

$$\mathbb{E}_x \sum_y \sqrt{p(y)p(y|x)} \geq \sum_{x,y} \sqrt{p(x)p(y)p(x)p(y|x)}$$

  but in fact the LHS is equal to the RHS as it can be written as

$$\sum_x p(x) \sum_y \sqrt{p(y)p(y|x)} = \sum_{x,y} \sqrt{p(x)^2 p(y)p(y|x)}$$

□

The $\sqrt{n}$ lower bound We now prove the following:

Theorem: $R(DISJ) \geq \Omega(\sqrt{n})$.

The distribution D: Using Yao's Min-Max principle, it's enough to come up with a distribution D on inputs, so that for every protocol π of $o(\sqrt{n})$ communication, $\Pr_{(x,y)\in D}[\pi(x,y) = DISJ(x,y)] < 0.999$. We're going to use the following distribution D: chooses $x_1....x_n, y_1..., y_n$ independently so that each is equal to 1 with probability $1/\sqrt{n}$ and equal to 0 with probability $1 - 1/\sqrt{n}$. We denote by $X_1...X_n, Y_1...., Y_n$ the random variables distributed according to D and note the following properties of D:

Balance: $\Pr[DISJ(X_1,...,X_n,Y_1,...,Y_n) = 1] \in (0.1, 0.9)$

Party independence: if we choose $(x,y) \in D$ then x is independent from y. This is also known as D being a *product distribution* of the form $D = D' \times D''$.

Coordinate independence: for $i \neq j$, the distribution X_iY_i is independent from X_jY_j. (Here X_iY_i denotes concatenation, not multiplication).

We will assume towards a contradiction that for some tiny $\epsilon > 0$ there is a protocol π with $k =\leq \epsilon\sqrt{n}$ communication satisfying $\Pr_{(x,y)\in D}[\pi(x,y) = DISJ(x,y)] \geq 0.999$. We will use Π to denote the random variable that is π's transcript on $X_1...X_nY_1...Y_n$.

Proof of the theorem We can now use **coordinate independence** of D to argue that

$$I(X_1Y_1;\Pi)+\cdots+I(X_nY_n;\Pi)\leq I(X_1Y_1...X_nY_n;\Pi)\leq k\leq \epsilon/\sqrt{n} \quad (1)$$

and so we get that for some typical j, $I(X_jY_j;\Pi)\leq k/n\leq \epsilon/\sqrt{n}$.

Fix this j and write for every $x,y\in\{0,1\}$, $\Pi_{x,y}$ the random variable Π obtained when fixing $X_j=x$ and $Y_j=y$. We use the fact that

$$\mathbb{E}_{x\in X}\Delta^2_{\mathsf{Hel}}(W,W_x)\leq I(X;W)$$

where $W_x=W|X=x$.

So, we can write for $p=1/\sqrt{n}$

$$(1-p)^2\Delta^2_{\mathsf{Hel}}(\Pi_{0,0},\Pi)+p(1-p)\Delta^2_{\mathsf{Hel}}(\Pi_{1,0},\Pi)+(1-p)p\Delta^2_{\mathsf{Hel}}(\Pi_{0,1},\Pi)+p^2\Delta^2_{\mathsf{Hel}}(\Pi_{1,1},P)\leq\epsilon/\sqrt{n} \quad (2)$$

In particular, this implies that

$$\Delta^2_{\mathsf{Hel}}(\Pi_{1,0},\Pi)\leq\epsilon$$
$$\Delta^2_{\mathsf{Hel}}(\Pi_{0,1},\Pi)\leq\epsilon$$

and so by triangle inequality

$$\Delta_{\mathsf{Hel}}(\Pi_{0,1},\Pi_{1,0})\leq\epsilon'=2\sqrt{\epsilon} \quad (3)$$

We now make the following claim

Claim: $\Delta_{\mathsf{Hel}}(\Pi_{0,0},\Pi_{1,1})\leq\epsilon'$

Assuming the claim we're done, since we know using the **balance** condition that:

- In $\Pi_{1,1}$ the value of $DISJ(X,Y)$ is always equal to 0, and moreover if we choose j at random then $\Pi^j_{1,1}$ covers a constant fraction of the probability space, and so the protocol outputs 1 with probability at most 0.01.
- In $\Pi_{0,0}$ the value of $DISJ(X,Y)$ is equal to 1 with probability at least $1/2$, and so the protocol outputs 1 with probability at least 0.1.
- This implies that $\Delta_{\mathsf{TV}}(\Pi_{1,1},\Pi_{0,0})\geq 0.05$, contradicting the fact that for every Z,W, $\Delta_{\mathsf{TV}}(Z,W)\leq O(\sqrt{\Delta_{\mathsf{Hel}}(Z,W)})$.

□

Proof of claim Let say that a set of 4 distributions $P_{0,0},P_{0,1},P_{1,0},P_{1,1}$ over some domain $\mathcal{D}$ is *separable* if there are some non-negative A_0,A_1,B_0,B_1 such that $P_{xy}(\alpha)=A_x(\alpha)B_y(\alpha)$ for all $x,y\in\{0,1\}$ and $\alpha\in\mathcal{D}$.

The claim immediately follows from (3) and the following two lemmas:

Lemma 1: The set $\{\Pi_{0,0},\Pi_{0,1},\Pi_{1,0},\Pi_{1,1}\}$ is separable.

Lemma 2: If $P_{0,0},P_{0,1},P_{1,0},P_{1,1}$ is separable then $\Delta_{\mathsf{Hel}}(P_{0,1},P_{1,0})=\Delta_{\mathsf{Hel}}(P_{0,0},P_{1,1})$.

Proof of Lemma 1: Let $\vec{m}=(m_1...m_k)$ be a transcript in the domain of Π, and let $x,y\in\{0,1\}$. $\Pi_{x,y}(\vec{m})$ can be computed as follows:

1. Pick $\{X_i\}_{i\neq j}$, $\{Y_i\}_{i\neq j}$ at random from D.
2. Now let p_1 be the probability that Alice sends m_1 on inputs $(X_{-j}, x) = X_1..X_{j-1}, x, X_j, .., X_n$.
   Let p_2 be the probability that Bob sends m_2 after receiving m_1 on inputs Y_{-j}, y.
   And similarly define $p_3, p_4, ..., p_k$

We get that $\Pi_{x,y}(\vec{m}) = p_1 p_2 \cdots p_k$, but the odd terms in this product can be computed by knowing X_{-j}, x and the even terms in this product can be computed by knowing Y_{-j}, y, and moreover using the **coordinate independence** and (most importantly) **party independence** property of D, X_{-j} and Y_{-j} are independent. Hence, we can define functions A_x and B_y such that $\Pi_{x,y}(\vec{m}) = A_x(\vec{m})B_x(\vec{m})$. □

Proof of Lemma 2: We know that for every two distributions P, P', $\Delta^2_{\mathsf{Hel}}(P, P') = 1 - \sum_\alpha \sqrt{P(\alpha)P'(\alpha)}$, but if $\{P_{x,y}\}$ is separable then for every α,

$$\sqrt{P_{0,1}(\alpha)P_{1,0}(\alpha)} = \sqrt{A_0(\alpha)}\sqrt{B_1(\alpha)}\sqrt{A_1(\alpha)}\sqrt{B_0(\alpha)} = \sqrt{P_{0,0}(\alpha)P_{1,1}(\alpha)}$$

□

A linear lower bound We now want to prove the stronger theorem

Theorem: $R(DISJ) \geq \Omega(n)$.

We will use the same general approach via the min-max principle. However, we cannot use the same distribution D (can you see why?)

In fact, it turns out that you can solve on average disjointness on *any* distribution D satisfying **player independence** (also known as being a *product distribution*) with communication $\tilde{O}(\sqrt{n})$.

The new distribution D We define the new input random variables $X_1...X_n, Y_1...., Y_n$ as follows: we choose for every i the pair X_iY_i to equal one of $\{00, 01, 10, 11\}$ with probabilities $p_{00}, p_{01}, p_{10}, p_{11}$ respectively defined as follows: $p_{00} = p_{01} = p_{10} = \frac{1}{3}(1 - 1/n)$ and $P_{11} = 1/n$. Note that this distribution satisfies **balance** and **coordinate independence** but not **party independence**.

The auxiliary random variables W We also define the following random variables $W_1...W_n$ that are correlated with X, Y as follows: for every i, with probability $1/2$, $W_i = X_i?$ and with probability $1/2$, $W_i = ?Y_i$, with these choices made independently. That is, for every coordinate i, W_i reveals either X_i or Y_i. The important properties we'll use about the relation between W, X, Y are the following:

Conditional balance: For a typical $w \in W$, the probability of $DISJ(X, Y) = 1$ conditioned on $W = w$ is in $(0.1, 0.9)$. The reason is that with very high probability, about a third of the the coordinates revealed in W will equal to 1, and then for each such coordinate there is probability $3/n$ that it will make sets not disjoint.

Conditional coordinate independence: Conditioned on $W = w$, we still have that X_iY_i is independent from X_jY_j for $i \neq j$. This is because the W_i's themselves are chosen independently for each coordinate.

Conditional party independence: Conditioned on $W = w$, X is independent from Y. This is because in each coordinate w will fix either X_i or Y_i to be some constant. This property turns out to be crucial for our analysis.

We are now going to carry out the analysis as before but always conditioning on $W = w$. In some sense this may seem very similar to us simply picking w from W and then using the distribution D_w instead of D where $D_w = D|W = w$. But using such a distribution D_w would never work, since there is a trivial $O(1)$ communication protocol to solve the problem on D_w: the protocol π can depend on w and for every coordinate where $w_i = ?y_i$, Alice can check if $x_i y_i = 1$ and for every coordinate where $w_i = x_i?$ Bob can check if $x_i y_i = 1$, and so each of them can send one bit to let the other know whether they found a 11 coordinate. Still, a lot of the analysis from above can be done by working with D_w, though at some point we'll use the fact that w is actually chosen at random and not known to the parties.

Bounding mutual information The first step of the analysis is as before. We show the following analog to (1):

$$\sum_j I(X_j Y_j; \Pi | W) \leq I(X_1 Y_1 .. X_n Y_n; \Pi | W) \leq \epsilon n \tag{4}$$

and so we get that for some typical j,

$$I(X_j Y_j; \Pi | W) \leq \epsilon \tag{5}$$

This means that for a random choice of $w = (w_{-j}, w_j) \in W$, we have that $I(X_j^w Y_k^w; \Pi^w) \leq \epsilon$, where X_j^w, Y_j^w, Π^w are the distributions of X_j, Y_j, Π respectively conditioned on $W = w$. Note that X_j and Y_j only depend on the j^{th} coordinate of w, and so we'll drop all other coordinates from the superscripts of X_j^w and Y_j^w.

Let $\Pi_{x,y}^{w_{-j}}$ denote the distribution of Π conditioned not only on $W = w$ but also on $X_j = x$ and $Y_j = y$ (because this condition implies the information in the j^{th} coordinate of w we dropped it from the superscript). In this notation (and using the relation between mutual entropy and squared Hellinger distance), **5** becomes:

$$\mathbb{E}_{w_{-j},x,y} \Delta^2_{\mathsf{Hel}}(\Pi_{x,y}^{w_{-j}}, \Pi^w) \leq \epsilon \tag{6}$$

Lets fix a typical choice for w_{-j} and for simplicity of notation drop w_{-j} from the superscripts and write Π^{w_j} for Π^w and $\Pi_{x,y}$ for $\Pi_{x,y}^{w_{-j}}$.

So we can write (6) as

$$\begin{aligned}
\epsilon \geq \quad & \tfrac{1}{2} p_{00} \Delta^2_{\mathsf{Hel}}(\Pi_{0,0}; \Pi^{0?}) \quad + \quad \tfrac{1}{2} p_{00} \Delta^2_{\mathsf{Hel}}(\Pi_{0,0}; \Pi^{?0}) + \\
& \tfrac{1}{2} p_{01} \Delta^2_{\mathsf{Hel}}(\Pi_{0,1}; \Pi^{0?}) \quad + \quad \tfrac{1}{2} p_{01} \Delta^2_{\mathsf{Hel}}(\Pi_{0,1}; \Pi^{?1}) + \\
& \tfrac{1}{2} p_{10} \Delta^2_{\mathsf{Hel}}(\Pi_{1,0}; \Pi^{1?}) \quad + \quad \tfrac{1}{2} p_{10} \Delta^2_{\mathsf{Hel}}(\Pi_{1,0}; \Pi^{?0}) + \\
& \tfrac{1}{2} p_{11} \Delta^2_{\mathsf{Hel}}(\Pi_{1,1}; \Pi^{1?}) \quad + \quad \tfrac{1}{2} p_{11} \Delta^2_{\mathsf{Hel}}(\Pi_{1,1}; \Pi^{?1})
\end{aligned}$$

Since $p_{00} = p_{01} = p_{10} \sim 1/3$, we get that (letting $\approx$ denotes closeness up to, say, $10\sqrt{\epsilon}$ in Hellinger distance)

$$\Pi_{0,1} \approx \Pi^{0?} \approx \Pi_{1,0}$$

But for the same reasons as above, the distributions $\{\Pi_{x,y}\}_{x,y \in \{0,1\}}$ are separable (this is because after fixing w_{-j}, the remaining distributions of X_{-j} and Y_{-j} satisfy both **party independence** and **coordinate independence** and so the proof of Lemma 1 works as is).

So we get that (up to $20\sqrt{\epsilon}$ in Hellinger distance, and hence also in total variation distance)

$$\Pi_{0,0} \approx \Pi_{1,1} \tag{7}$$

But because of our **conditional balance** property, for a typical w_{-j}, while $DISJ(\Pi_{1,1}) = 0$ with probability 1, $DISJ(\Pi_{0,0}) = 0$ with probability at most 0.9. Now because by choosing a random j and w_{-j}, we cover a constant fraction of the measure space with $\Pi_{1,1}^{w_{-j}}$ and with $\Pi^{w_{-j}}, 00$ then the protocol would have to answer 0 on $\Pi_{1,1}$ with probability roughly 0.99 while answering 0 on $\Pi_{0,0}$ with probability at most 0.01 thus contradicting (7). □

Note on the log rank conjecture I mentioned last lecture that there are known functions f such that $C(f) \geq (\log \mathsf{rank}(f))^{1.01}$. This was first shown by Nisan and Wigderson in 1994 using the following example. Let $g_1(z_1, z_2, z_3) = 1$ iff $z_1+z_2+z_3 \in \{1, 2\}$, and we define g_k that takes 3^k inputs,splits them to three blocks Z_1, Z_2, Z_3 and outputs $g(g_{k-1}(Z_1), g_{k-1}(Z_2), g_{k-1}(Z_3))$. For $n = 3^k$, let $f_k(x_1...x_n, y_1..., y_n) = g(x_1y_1, ..., x_ny_n)$. We can show the following:

- $C(f_k)$ (and in fact $R(f_k)$) is at least $\Omega(n) = \Omega(3^k)$. This is by reduction to the unique disjointness problem (distinguishing between the case that $x \cap y = \emptyset$ and the case that $|x \cap y| = 1$). The same proofs as above establish an $\Omega(n)$ lower bound for this problem.
- $\mathsf{rank}(f_k) \leq c^{k \cdot 2^k}$ for some constant c, meaning that $\log \mathsf{rank}(f) = \tilde{O}(2^k) = O(n^{0.99})$. This is a good exercise, see footnote for hint.[1]

[1] **Hint:** Prove that the degree of g_k as a polynomial in n variables is at most 2^k, and then argue this implies that $\mathsf{rank}(f_k)$ is at most the number of monomials in g_k.