CAS CS 591 B: Communication Complexity

Prof. Mark Bun Fall 2019

Lecture Notes 8:

Disjointness Lower Bound

Reading.

- Rao-Yehudayoff Chapter 6

The lower bound of $\Omega(n)$ on the randomized communication complexity of Disjointness is perhaps the most impactful result in the entire area. It has consequences in circuit complexity, property testing, algorithmic game theory, extension complexity, data structures, etc. All known proofs of the tight lower bound use information theory one way or another, but the ideas we've developed in our study of information complexity give a particularly simple and intuitively satisfying proof. The presentation here follows the proof of Bar-Yossef, Jayram, Kumar, and Sivakumar, "An information statistics approach to data stream and communication complexity."

It will be more convenient to work with the complement of the Disjointness function, which we'll call the Set-Intersection function. Define

$$\mathrm{INT}_n(x, y) = \bigvee_{i=1}^{n} (x_i \wedge y_i).$$

The idea of the proof will be to show that for some distribution ζ over $\{0,1\}^2$, a protocol for computing INT_n over inputs drawn from $\zeta^{\otimes n}$ will have to reveal n times as much information as a protocol for computing AND_2 over inputs drawn from ζ. We will then exhibit a distribution ζ over $\{0,1\}^2$ which requires $\Omega(1)$ bits of information to compute AND_2.

1 From INT_n to AND_2

Define the distribution ζ over $\{0,1\}^2 \times \{a,b\}$ as follows. Let D be uniformly random from $\{a,b\}$. If $D = a$, let $A = 0$ and B be uniform over $\{0,1\}$. If $D = b$, let $B = 0$ and A be uniform over $\{0,1\}$. Then A and B are independent conditioned on D.

Lemma 1 (Information Cost Decomposition)**.** *Let ζ be the distribution defined above, and let $((A,B),\vec{D}) \sim \zeta^{\otimes n}$. Then for any protocol Π,*

$$I(AB; \Pi(A,B) | \vec{D}) \geq \sum_{i=1}^{n} I(A_iB_i; \Pi(A,B) | \vec{D}).$$

Proof. Abbreviating $\Pi = \Pi(A, B)$, we have

$$\begin{aligned} I(AB; \Pi\vec{D}) &= H(AB|\vec{D}) - H(AB|\Pi\vec{D}) \\ &= \sum_{i=1}^{n} H(A_iB_i|\vec{D}) - H(AB|\Pi\vec{D}) && \text{since the } A_i, B_i\text{'s are independent given } D \\ &\geq \sum_{i=1}^{n} H(A_iB_i|\vec{D}) - \sum_{i=1}^{n} H(A_iB_i|\Pi\vec{D}) && \text{by subadditivity of entropy} \\ &= \sum_{i=1}^{n} I(A_iB_i; \Pi|\vec{D}). \end{aligned}$$

□

Lemma 2 (Reduction Lemma)**.** *Let Π compute INT_n with probability at least $1-\varepsilon$ on every input. Let ζ be the distribution defined above and let $((A, B), \vec{D}) \sim \zeta^{\otimes n}$ and $((U, V), D) \sim \zeta$. Then for every $i \in [n]$,*

$$I(A_iB_i; \Pi(A, B)|\vec{D}) \geq \inf_P I(UV; P(U, V)|D),$$

where the infimum is taken over protocols P computing AND_2 with probability at least $1-\varepsilon$ on every input.

Proof. Let Π be a protocol computing INT_n with probability at least $1 - \varepsilon$ on every input. Fix an index i. By the definition of conditional mutual information,

$$I(A_iB_i; \Pi(A, B)|\vec{D}) = \mathbb{E}_{\vec{d}\sim\{a,b\}^{n-1}}[I(A_iB_i; \Pi(A, B)|D_i, \vec{D}_{-i} = \vec{d})].$$

So it suffices to show that for every fixed $\vec{d} = (d_1, \ldots, d_{i-1}, d_{i+1}, \ldots, d_n) \in \{a, b\}$, there is a protocol for AND_2 with information cost $I(A_iB_i; \Pi(A, B)|D_i, \vec{D}_{-i} = \vec{d})$.

We now describe such a protocol $P_{i,\vec{d}}$ for AND_2. On input u, v, Alice and Bob set $x_i = u, y_i = v$ and every other (x_j, y_j) to a sample from ζ conditioned on d_j (using private randomness). They then run the protocol $\Pi(x, y)$ and output its result. This protocol computes the AND_2 function since $u = v = 1$ iff x and y are intersecting inputs.

Next, we analyze the conditional information cost of $P_{i,\vec{d}}$. One can check by inspection that the joint distribution of $(U, V, D, P_{i,\vec{d}}(U, V))$ is equal to that of $(A_i, B_i, D_i, \Pi(A, B))$ conditioned on $\vec{D}_{-i} = \vec{d}$.

Hence

$$I(UV; P_{j,\vec{d}}(U, V)|D) = I(A_iB_i; \Pi(A, B)|D_i, \vec{D}_{-i} = \vec{d}).$$

□

Theorem 3. *Let ζ be the distribution defined above and let $((A, B), D) \sim \zeta^{\otimes n}$. Then*

$$\mathbf{BPP}_{\varepsilon}^{pub}(\mathrm{INT}_n) \geq n \cdot \inf_P I(AB; P(A, B)|D)$$

where the infimum is taken over protocols P computing AND_2 with probability at least $1 - \varepsilon$.

Proof. Follows by combining the Reduction Lemma and the Information Cost Decomposition, together with the fact that conditional external information $I(AB; \Pi(A, B)|\vec{D})$ is a lower bound on communication cost. □

2 Why Not a Product Distribution?

Alice and Bob's inputs when sampled from $\zeta^{\otimes n}$ are independent conditioned on D, but they are not fully independent. Why couldn't we use a fully independent distribution to prove a lower bound for Set-Intersection? It turns out that Set-Intersection is easy for product distributions: Babai, Frankl, and Simon showed that for every product distribution μ, there is a distributional protocol for Set-Intersection showing that $D_\mu(\mathrm{INT}_n) = O(\sqrt{n}\log n)$. But this may still not be so convincing, since there is a distributional protocol with respect to $\zeta^{\otimes n}$ as well: all inputs in its support are non-intersecting, so nothing needs to be communicated. The magic in this proof lies in the fact that any protocol for computing INT_n on *every* input with high probability must reveal a lot of information when run on the distribution $\zeta^{\otimes n}$, even though that distribution isn't even supported on yes inputs.

To see why this argument wouldn't work for when Alice and Bob are fully independent, we argue that for any product distribution there is a protocol which succeeds on every input but has low expected communication when run on that distribution. Specifically, let's fix a product distribution on (A, B) and consider the following protocol for Set-Intersection. For some parameter $\varepsilon > 0$, Alice finds a coordinate (if one exists) such that $H(A_i) \geq \varepsilon$ and $H(B_i) \geq \varepsilon$. If one is found, the parties exchange the coordinate, output 1 if both coordinates are 1, condition on the values seen and repeat. When they run out of high-entropy coordinates, then the remaining entropy in A and B must be at most $\varepsilon \cdot n$, so they can just transmit their sets with roughly this many bits of communication.

How many rounds should we expect this protocol to last? If $H(A_i) \geq \varepsilon$, then $\Pr[A_i = 1] \geq \Omega(\varepsilon/\log(1/\varepsilon))$. So the probability of finding an intersection is at least about ε^2. Thus with high probability, the protocol will not last for more than about $1/\varepsilon^2$ rounds when actually run on the distribution (A, B). Setting $\varepsilon \approx n^{-1/3}$ gives expected total communication roughly $n^{2/3}$.

3 Information Complexity of AND

Our goal is now to prove a lower bound on the information complexity of any protocol computing AND_2. In this section, we'll give the proof as a consequence of a sequence of lemmas, and give the proofs of those lemmas in the following section.

Theorem 4. *Let P be a protocol which computes $\mathrm{AND}(u, v)$ with probability at least $1 - \varepsilon$. Let $((U, V), D) \sim \zeta$. Then*

$$I(UV; P(U, V)|D) \geq \frac{1}{4}(1 - 2\sqrt{\varepsilon}).$$

To begin analyzing this, we expand the quantity on the right to make it easier to work with. Let $Z \in \{0, 1\}$ be uniformly random. Then by definition

$$\begin{aligned} I(UV; P(U, V)|D) &= \frac{1}{2}I(UV; P(U, V)|D = a) + \frac{1}{2}I(UV; P(U, V)|D = b) \\ &= \frac{1}{2}I(Z; P(0, Z)) + \frac{1}{2}I(Z; P(Z, 0)). \end{aligned}$$

For $(u, v) \in \{0, 1\}^2$, let p_{uv} denote the distribution of transcripts when running $P(u, v)$. The next step is to relate these mutual information quantities to distances between distributions. The distance which will be convenient for us to use is the Hellinger distance.

Definition 5. The squared Hellinger distance between two probability distributions p, q over domain X is defined by

$$h^2(p,q) = 1 - \sum_{x \in X} \sqrt{p(x)q(x)} = \frac{1}{2} \sum_{x \in X} (\sqrt{p(x)} - \sqrt{q(x)})^2.$$

Lemma 6 (Hellinger Lower Bound). *Let P be a protocol computing* AND *with probability at least $1-\varepsilon$. Then for $Z \in \{0,1\}$ uniform,*

$$I(Z; P(0,Z)) \geq h^2(p_{00}, p_{01}) \qquad \textit{and} \qquad I(Z; P(Z,0)) \geq h^2(p_{00}, p_{10}).$$

Continuing our calculation,

$$\begin{aligned} I(UV; P(U,V)|D) &= \frac{1}{2} I(Z; P(0,Z)) + \frac{1}{2} I(Z; P(Z,0)) \\ &\geq \frac{1}{2} h^2(p_{00}, p_{01}) + \frac{1}{2} h^2(p_{00}, p_{10}) \\ &\geq \frac{1}{4} (h(p_{00}, p_{01}) + h(p_{00}, p_{10}))^2 && \text{Cauchy-Schwarz} \\ &\geq \frac{1}{4} h^2(p_{01}, p_{10}) && \text{Triangle Inequality} \end{aligned}$$

At this point, it is not clear why we should expect the distance between p_{01} and p_{10} to be large, as both are 0-inputs to the AND function. This is the point where we exploit the fact that these are distributions over transcripts, and that the set of inputs resulting in any given transcript is a rectangle:

Lemma 7 (Cut-and-Paste Lemma). *Let P be a randomized protocol over $X \times Y$. Then for every $x, x' \in X$ and every $y, y' \in Y$, we have $h(P_{xy}, P_{x',y'}) = h(P_{x,y'}, P_{x',y})$.*

Applying the Cut-and-Paste Lemma lets us conclude that

$$I(UV; P(U,V)|D) \geq \frac{1}{4} h^2(p_{00}, p_{11}).$$

And now, we should expect to be done, because any accurate protocol for AND must induce very different distributions on $(0,0) \in \text{AND}^{-1}(0)$ and $(1,1) \in \text{AND}^{-1}(1)$. Indeed, we have

Lemma 8 (Distinguishing Lemma). *If P computes a function f with probability $2/3$ on every input, and (x,y) and (x',y') are inputs such that $f(x,y) \neq f(x',y')$, then*

$$h^2(p_{xy}, p_{x'y'}) \geq 1 - 2\sqrt{\varepsilon}.$$

This allows us to conclude the proof of Theorem 4.

4 Deferred Proofs

"Proof" of Hellinger Lower Bound Lemma 6. The statement is true as given, but the proof is more complicated and requires introducing a few more information-theoretic quantities. We will prove

the weaker statement that if $Z \in \{0,1\}$ is uniform and if P is a randomized function on one bit, then

$$I(Z;P(Z)) \geq \frac{\log e}{2} h^2(p,q),$$

where p is the distribution of $P(Z)$ and q_0 and q_1 are the distributions of $P(0)$ and $P(1)$ respectively.

For any pair of distributions q,p, we have

$$\begin{aligned} D(q\|p) &= -\sum_T q(T)\log\frac{p(T)}{q(T)} \\ &\geq \sum_T q(T)\cdot(2\log e)\left(1-\sqrt{\frac{p(T)}{q(T)}}\right) && \text{since} \ln y \leq y-1 \\ &= 2(\log e)h^2(q,p) \end{aligned}$$

For any random variables X,Y with joint distribution p, we may write

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \\ &= \sum_y p(y)\sum_x p(x|y)\log\left(\frac{p(x|y)}{p(x)}\right) \\ &= \mathbb{E}_{y\sim Y} D(p(\cdot|y)\|p). \end{aligned}$$

Hence

$$\begin{aligned} I(Z;P(Z)) &= \mathbb{E}_{z\sim Z} D(q_z\|p) \\ &= \frac{1}{2}(D(q_0\|p)+D(q_1\|p)) \\ &\geq \log e\cdot(h^2(q_0,p)+h^2(q_1,p)) \\ &\geq \frac{\log e}{2}(h(q_0,p)+h(q_1,p))^2 && \text{by Cauchy-Schwarz} \\ &\geq \frac{\log e}{2}h^2(q_0,q_1) && \text{by Triangle Inequality.} \end{aligned}$$

□

Proof of Cut-and-Paste Lemma 7. Let P be a private coin protocol and let p_{xy} denote the distribution on transcripts when P is run on (x,y). Recall that for any P, we can, for every transcript T, decompose $\Pr[P(x,y)=T] = q_A(x,T)\cdot q_B(y,T)$ for some functions q_A, q_B.

Then

$$\begin{aligned} 1-h^2(p_{xy},p_{x'y'}) &= \sum_T \sqrt{\Pr[P(x,y)=T]\cdot\Pr[P(x',y')=T]} \\ &= \sum_T \sqrt{q_A(x,T)q_B(y,T)q_A(x',T)q_B(y',T)} \\ &= \sum_T \sqrt{\Pr[P(x',y)=T]\cdot\Pr[P(x,y')=T]} \\ &= 1-h^2(p_{x'y},p_{x'y}). \end{aligned}$$

□

Proof of Distinguishing Lemma 8. We'll actually begin by showing that p_{xy} and $p_{x',y'}$ are far in total variation distance, where we recall that the total variation distance between two distributions p, q over $\mathcal{T}$ is

$$TV(p,q) = \max_{S \subseteq \mathcal{T}} (p(S) - q(S)) = \frac{1}{2} \sum_{T \in \mathcal{T}} |p(T) - q(T)|.$$

Let S be the set of transcripts on which P outputs $f(x,y)$. Then $p_{xy}(S) \geq 1 - \varepsilon$ and $p_{x'y'}(S) \leq \varepsilon$. Hence $TV(p_{xy}, p_{x'y'}) \geq 1 - 2\varepsilon$.

Now we relate total variation distance to Hellinger distance as follows:

$$\begin{aligned}
TV^2(p,q) &= \frac{1}{4} \left(\sum_T |p(T) - q(T)| \right)^2 \\
&= \frac{1}{4} \left(\sum_T (\sqrt{p(T)} + \sqrt{q(T)})(\sqrt{p(T)} - \sqrt{q(T)}) \right)^2 \\
&\leq \frac{1}{4} \left(\sum_T (\sqrt{p(T)} + \sqrt{q(T)})^2 \right) \left(\sum_T (\sqrt{p(T)} - \sqrt{q(T)})^2 \right) && \text{Cauchy-Schwarz} \\
&\leq \frac{1}{2} h^2(p,q) \cdot \left(2 + 2 \sum_T \sqrt{p(T)}\sqrt{q(T)} \right) \\
&= h^2(p,q)(2 - h^2(p,q)).
\end{aligned}$$

This allows us to conclude that $h^2(p,q) \geq 1 - 2\sqrt{\varepsilon}$. □