15-859: Information Theory and Applications in TCS **CMU: Spring 2013**

Lecture 20: Lower Bounds for Inner Product & Indexing

April 9, 2013

Lecturer: Venkatesan Guruswami *Scribe: Albert Gu*

1 Recap

- **Last class**
  - *Randomized Communication Complexity*
  - *Distributional CC*: $D_\delta^\mu(f)$ is the best communication complexity of a deterministic protocol Π such that $\Pr_\mu[\Pi(x,y) \neq f(x,y)] \leq \delta$.
  - Lemma: $R_\delta^{pub}(f) = \max_\mu D_\delta^\mu(f)$ – can be used to lower bound $R(f)$ by choosing an adverse distribution μ.
- **Today**
  - Lower bounds on Distributional CC
  - *Discrepancy*
  - *Indexing problem* via Information Theory

2 Discrepancy technique

We would like to develop a method to lower bound D_δ^μ, which in turn will lower bound $R(f)$. Every deterministic protocol induces a partition of $X \times Y$ into rectangles; previously, for zero-error protocols, these rectangles had to be monochromatic, but now we allow some to not be monochromatic.

The discrepancy technique aims to show that, for a specific function f, every large rectangle has nearly equal numbers of 0s and 1s. This forces an accurate protocol to use only small rectangles and hence require many rectangles.

Definition 1 (Discrepancy)**.** *Let $f : X \times Y \to \{0,1\}$, $R = S \times T : S \subseteq X, T \subseteq Y$, and μ be a distribution on $X \times Y$.*

Denote

$$\begin{aligned} Disc_\mu(R,f) &= \left| \Pr_{(x,y)\sim\mu}[f(x,y) = 0 \wedge (x,y) \in R] - \Pr_\mu[f(x,y) = 1 \wedge (x,y) \in R] \right| \\ &= \left| \sum_{(x,y)\in R} (-1)^{f(x,y)} \mu(x,y) \right| \end{aligned}$$

and

$$Disc_\mu(f) = \max_{R \in X \times Y} Disc_\mu(R, f)$$

(Note: If R is monochromatic, $Disc_\mu(R, f) = \mu(R)$.)

Note that large discrepancy (monochromatic and large rectangle) is good for a protocol. The following is a generalization of the deterministic bound $D(f) \geq \log_2 \left(\frac{1}{\max_{R monochr} \mu(R)} \right)$:

Proposition 2 (Discrepancy lower bound)**.** $D^\mu_{\frac{1}{2}-\gamma}(f) \geq \log_2 \left(\frac{2\gamma}{Disc_\mu(f)} \right)$

Proof. Let Π be a protocol using c bits of communication with error probability at most $\frac{1}{2} - \gamma$. Since it is deterministic, the matrix is split into at most 2^c rectangles.

By the maximum error allowed from this protocol, we have

$$\Pr_{(x,y)\sim\mu} [\Pi(x, y) = f(x, y)] - \Pr_{(x,y)\sim\mu} [\Pi(x, y) \neq f(x, y)] \geq 2\gamma$$

We can bound the LHS by breaking it into rectangles according to the protocol and noting that Π is constant on each rectangle:

$$\begin{aligned}
\Pr_\mu [\Pi(x, y) = f(x, y)] - \Pr_\mu [\Pi(x, y) \neq f(x, y)] &= \sum_{R_\ell \in protocol} \Pr_\mu [\Pi(x, y) = f(x, y) \wedge (x, y) \in R_\ell] \\
&\quad - \Pr_\mu [\Pi(x, y) \neq f(x, y) \wedge (x, y) \in R_\ell] \\
&\leq \sum_{R_\ell} \left| \Pr_\mu [f(x, y) = 0 \wedge (x, y) \in R_\ell] - [f(x, y) = 1 \wedge (x, y) \in R_\ell] \right| \\
&= \sum_{R_\ell} Disc_\mu(R_\ell, f) \\
&\leq 2^c Disc_\mu(f)
\end{aligned}$$

This implies $2^c Disc_\mu(f) \geq 2\gamma \implies c \geq \log_2 \left(\frac{2\gamma}{Disc_\mu(f)} \right)$, as desired. ∎

2.1 Dot product function

We will now apply this technique to bound the randomized CC of the dot product function, defined as

$$IP(x, y) = x \cdot y = \sum x_i y_i \pmod 2$$

In the deterministic case, we showed in a previous lecture that $n + 1$ is the best we could do.

Theorem 3. $R_{\frac{1}{3}}(IP) \geq \Omega(n) = \frac{n}{2} - O(1)$

It suffices to show that $D^{\mu}_{\frac{1}{3}}(IP) \geq \frac{n}{2} - O(1)$ for some distribution μ. This has two parts: we need to come up with a clever μ, and then need to bound it. Since dot product is pretty evenly distributed for random inputs, we take μ to be uniform.

Goal: Prove $Disc_{uniform}(IP) \leq \frac{1}{2^{n/2}}$ (Note that this implies the claimed bound by the above Proposition).

Proof. Let $R = S \times T$ be any rectangle. Then

$$Disc_{\mu}(R, IP) = \left| \sum_{x \in S, y \in T} (-1)^{x \cdot y} \frac{1}{2^{2n}} \right|$$

Let $\mathbf{H_n} \in \{1, -1\}^{2^n \times 2^n}$ be the matrix indexed by X and Y where the (x, y)th entry is $(-1)^{x \cdot y}$. First we show the following fact.

Exercise: $\mathbf{H_n}$ is an orthogonal matrix ($\mathbf{H_n^t H_n} = 2^n \mathbf{I}$).

Now we can bound $Disc_{\mu}(R, IP)$:

$$\begin{aligned}
Disc_{\mu}(R, IP) &= \frac{1}{2^{2n}} \mathbf{1_S}^t \mathbf{H_n} \mathbf{1_T} \\
&= \frac{1}{2^{2n}} (\mathbf{1_S}^t) \cdot (\mathbf{H_n} \mathbf{1_T}) \\
&\leq \frac{1}{2^{2n}} \|\mathbf{1_S}\| \|\mathbf{H_n} \mathbf{1_T}\| \\
&= \frac{1}{2^{2n}} \sqrt{|S|} \sqrt{(\mathbf{H_n} \mathbf{1_T}) \cdot (\mathbf{H_n} \mathbf{1_T})} \\
&= \frac{1}{2^{2n}} \sqrt{|S|} \sqrt{\mathbf{1_T^t} \mathbf{H_n^t} \mathbf{H_n} \mathbf{1_T}} \\
&= \frac{1}{2^{2n}} \sqrt{|S|} \sqrt{2^n |T|} \\
&\leq \frac{1}{2^{2n}} \sqrt{2^n} \sqrt{2^n 2^n} \\
&= \frac{1}{2^{n/2}}
\end{aligned}$$

□

(Note: It is possible to improve the bound for $R(IP)$ to $n - O(1)$, which appears on Problem Set 4)

In summary, we have shown that $R(EQ) = \theta(\log n)$ and $R(IP) = \theta(n)$. In upcoming lectures, we will tackle $R(DISJ)$, which is in some sense the poster child of this whole field.

3 Indexing Problem

Alice and Bob are again communicating, but the setup is slightly asymmetrical this time. As before, Alice has a string $x \in \{0,1\}^n$, but now Bob has an index $i \in \{1, 2, \cdots, n\}$, and the goal is for Bob to learn x_i. There is a trivial $\lceil \log n \rceil$ protocol by just sending the index.

Now, suppose we only allow Alice to send a single message so that Bob can figure out x_i. Can we do better than the trivial n bit solution?

3.1 Deterministic

Suppose Alice and Bob use a deterministic protocol and Alice sends less than n bits. Then there exists $a \neq b$ such that Alice sends the same message for a, b and Bob cannot distinguish if Alice sent A or B. Let j be such that $a_j \neq b_j$, then the protocol is wrong on either (a, j) or (b, j).

3.2 Randomized

The above proof does not give us enough for a randomized lower bound: we need Bob to be wrong on a lot of inputs. However, it turns out that even a randomized protocol requires $\Omega(n)$ bits to be sent, which can be shown in several ways.

Exercise: Come up with a μ on $\{0,1\}^n \times \{1, \cdots, n\}$ such that $D^{\mu}_{1/3}(Index) \geq \Omega(n)$.
Hint: If a and b in the deterministic proof differ in only 1 bit, Bob has low chance of error. We would like them to differ in more. Try finding a distribution on X supported on a code of distance $n/3$, and also supported on $2^{\Omega(n)}$ elements.

In contrast to the coding theory proof hinted at above, we will present an information theoretical proof of this fact.

Proof. We will bound the distributional complexity. We take the distribution $X = X_1 X_2 \cdots X_n$ uniform on $\{0,1\}^n$ and i uniform on $\{1, \cdots, n\}$. Let Π be a deterministic protocol with error at most $\frac{1}{3}$. Alice will send $M = M(x)$, also a random variable. We can bound

$$CC(\Pi) \geq \log(\text{supp}(M)) \geq H(M) = I(M; X) = I(X_1 X_2 \cdots X_n; M) \geq \sum \mathcal{I}(X_i; M)$$

The goal is now to show that M has a lot of information, since Bob can tell a lot about X from M. We would like to show that each $\mathcal{I}(X_i; M)$ is about a constant, which makes sense since Bob can figure out any bit with high probability.

Continuing the chain of inequalities,

$$CC(\Pi) = \sum \mathcal{I}(X_i; M) = \sum H(X_i) - H(X_i|M) = n - \sum H(X_i|M)$$

For notation, let $P_e^{m,i}$ be the probability of error given that Alice sent m and Bob has i. By the error guarantee of the protocol, we have

$$\mathbb{E}_{m,i}[P_e^i] \leq \frac{1}{3}$$

By Fano's Inequality, $h(P_e^{m,i}) \geq H(X_i|M = m)$. Therefore

$$\begin{aligned}
\mathbb{E}_{m,i}[h(P_e^i)] &= \mathbb{E}_i[\mathbb{E}_m[h(P_e^{m,i})]] \\
&\geq \mathbb{E}_i[\mathbb{E}_m[H(X_i|M = m)]] \\
&\geq \mathbb{E}_i[H(X_i|M)] \\
&= \frac{\sum_i H(X_i|M)}{n}
\end{aligned}$$

Finally, by the concavity of h, this gives

$$\sum_i H(X_i|M) \leq \mathbb{E}[h(P_e^{m,i})]n \leq h\left(\mathbb{E}[P_e^{m,i}]\right) n \leq h\left(\frac{1}{3}\right) n$$

Wrapping it all up, we have

$$CC(\Pi) \geq n - \sum_i H(X_i|M) \geq n - h\left(\frac{1}{3}\right) n \geq \Omega(n)\ . \qquad \square$$