CS 2429 - Foundations of Communication Complexity

Lecture #3: 29 September 2009

Lecturer: Toniann Pitassi and Avner Magen

Scribe Notes by: Avery Miller

Randomized Communication Complexity

Definitions

A *(private coin) randomized protocol* is a protocol where Alice and Bob have access to random strings r_A and r_B, respectively. These two strings are chosen independently, according to some probability distribution. We can classify randomized protocols by considering different types of error:

- *zero-error protocol* $\mathcal{P}$:
$$\forall x, y \Pr_{r_A, r_B}[\mathcal{P}(x, r_A, y, r_B) = f(x, y)] = 1$$
- *ϵ-error protocol* $\mathcal{P}$:
$$\forall x, y \Pr_{r_A, r_B}[\mathcal{P}(x, r_A, y, r_B) = f(x, y)] \geq 1 - \epsilon$$
- *one-sided ϵ-error protocol* $\mathcal{P}$:
$$\begin{aligned} \forall x, y: \quad & f(x, y) = 0 \Rightarrow \Pr_{r_A, r_B}[\mathcal{P}(x, r_A, y, r_B) = 0] = 1 \\ & f(x, y) = 1 \Rightarrow \Pr_{r_A, r_B}[\mathcal{P}(x, r_A, y, r_B) = 1] \geq 1 - \epsilon \end{aligned}$$

Due to randomization, the number of bits exchanged may differ in different executions of the protocol on the same input (x, y). So, there are two natural choices for measuring the running time of a randomized protocol:

- The *worst case running time $\mathcal{P}$ on input* (x, y) is the maximum number of bits communicated over all choices of the random strings r_A and r_B. The *worst case cost of* $\mathcal{P}$ is the maximum, over all inputs (x, y), of the worst case running time of $\mathcal{P}$ on (x, y).
- The *average case running time $\mathcal{P}$ on input* (x, y) is the expected number of bits communicated over all choices of the random strings r_A and r_B. The *average case cost of* $\mathcal{P}$ is the maximum, over all inputs (x, y), of the average case running time of $\mathcal{P}$ on (x, y).

So, for a function $f : X \times Y \to \{0, 1\}$, we define the following complexity measures:

- $R_0(f)$ is the minimum average case cost of a randomized protocol that computes f with zero error.

- For $0 < \epsilon < \frac{1}{2}$, $R_\epsilon(f)$ is the minimum worst case cost of a randomized protocol that computes f with error ϵ.
- For $0 < \epsilon < 1$, $R^1_\epsilon(f)$ is the minimum worst case cost of a randomized protocol that computes f with one-sided error ϵ.

These lead naturally to the following complexity classes:

- $ZPP^{cc} = \{f \mid R_0(f) \in O(\text{polylog}(n))\}$
- $BPP^{cc} = \{f \mid R_\epsilon(f) \in O(\text{polylog}(n))\}$
- $RP^{cc} = \{f \mid R^1_\epsilon(f) \in O(\text{polylog}(n))\}$

Analogous definitions hold in a *public coin* model, that is, a model where both Alice and Bob see the results of a single series of random coin flips. A randomized protocol in the public coin model can be viewed as a distribution of deterministic protocols, that is, Alice and Bob choose together a string r (according to a probability distribution Π, and independently of x and y) and then follow the deterministic protocol P_r. The *success probability* of a public coin protocol on input (x, y) is the probability of choosing a deterministic protocol, according to the probability distribution Π, that computes $f(x, y)$ correctly. We use the same complexity measures as in the private coin model, but add a superscript 'pub', i.e., $R^{pub}_0(f)$, $R^{pub}_\epsilon(f)$, $R^{1\ pub}_\epsilon(f)$. We have previously seen the following facts:

- $R^{pub}_\epsilon(f) \leq R_\epsilon(f)$
- for every $\delta > 0$ and every $\epsilon > 0$, $R_{\epsilon+\delta}(f) \leq R^{pub}_\epsilon(f) + O(\log n + \log \delta^{-1})$

Distributional Complexity

Let μ be a probability distribution over $X \times Y$, $X = \{0,1\}^n$, $Y = \{0,1\}^n$. The (μ, ϵ)-*distributional communication complexity* of f, $D^\mu_\epsilon(f)$, is the cost of the best deterministic protocol that gives the correct answer for f on at least a $(1 - \epsilon)$ fraction of all inputs in $X \times Y$, weighted by μ.

Theorem 1 $R^{pub}_\epsilon(f) = \max_\mu D^\mu_\epsilon(f)$

Proof First, we show that $R^{pub}_\epsilon(f) \geq \max_\mu D^\mu_\epsilon(f)$. Let $\mathcal{P}$ be a randomized public coin protocol with worst-case cost $R^{pub}_\epsilon(f)$ that computes f with success probability at least $1 - \epsilon$ for every input (x, y). Therefore, if Π is the probability distribution of $\mathcal{P}$'s public coin flips,

$$\Pr_{r \in \Pi, (x,y) \in (X \times Y)_\mu} (\mathcal{P}_r(x, y) = f(x, y)) \geq 1 - \epsilon$$

By a counting argument, there exists a fixed choice of public coin flips r' such that

$$\Pr_{(x,y) \in (X \times Y)_\mu} (\mathcal{P}_{r'}(x, y) = f(x, y)) \geq 1 - \epsilon$$

Thus, $\mathcal{P}_{r'}$ is a deterministic protocol that gives the correct answer for f on at least a $1 - \epsilon$ fraction of all inputs in $X \times Y$, weighted by μ. So, $R^{pub}_\epsilon(f) \geq \text{cost}(\mathcal{P}_{r'}) \geq \max_\mu D^\mu_\epsilon(f)$.

Next, we show that $R^{pub}_\epsilon(f) \leq \max_\mu D^\mu_\epsilon(f)$. Let $c = \max_\mu D^\mu_\epsilon(f)$. We define a two-player zero-sum game:

- Player $P1$ has all deterministic c-bit communication protocols. Player $P2$ has all inputs $X \times Y$.
- $P1$ chooses a protocol $\mathcal{P}$, $P2$ chooses an input (x, y) (independently of one another).
- $P1$ wins if $\mathcal{P}(x, y) = f(x, y)$, otherwise $P2$ wins.

Each mixed strategy of $P2$ can be viewed as a distribution μ' on the inputs. Since $D^{\mu'}_\epsilon(f) \leq c$, there is a protocol that $P1$ can pick that ensures that the expected payment is at least $1 - \epsilon$. By John von Neumann's Minimax theorem, $P1$ has a randomized strategy that guarantees payoff $1-\epsilon$ for every choice (x, y) of $P2$. This randomized strategy is a distribution Π over c-bit deterministic protocols, so it is a randomized public coin protocol $\mathcal{P}$ for f with cost at most c and error at most ϵ. Therefore, $c \geq \text{cost}(\mathcal{P}) \geq R^{pub}_\epsilon(f)$. □

Theorem 1 is useful because, for any choice of μ, a lower bound for D^μ_ϵ gives a lower bound on $R^{pub}_\epsilon(f)$.

Definition A distribution μ over $X \times Y$ is a *product distribution* if $\mu(x, y) = \mu_X(x) \cdot \mu_Y(y)$ for some distributions μ_X over X and μ_Y over Y. Let $R^{[\,]}(f) = \max_\mu D^\mu(f)$, where the maximum is taken over all product distributions μ.

Exercise: Prove that $R^{[\,]}_\epsilon(DISJ) = O(\sqrt{n} \log n)$. On the other hand, show that $R_\epsilon(DISJ) = \Theta(n)$.

Sherstov showed a separation between product and non-product distributional complexity by proving the existence of a function f such that $R^{[\,]}(f) = \Theta(1)$ but $R_\epsilon(f) = \Theta(n)$.

Discrepancy

We now consider a technique for proving lower bounds for D^μ_ϵ. It consists of finding an upper bound for the size of rectangles in M_f that are "almost" monochromatic. If we can prove that all such rectangles are small, then we need a lot of rectangles to "cover" the function.

Definition Let $f : X \times Y \to \{0, 1\}$ be a function, R be any rectangle, and μ be a probability distribution on $X \times Y$. Denote

$$\text{Disc}_\mu(R, f) = \left| \Pr_\mu[f(x, y) = 0 \text{ and } (x, y) \in R] - \Pr_\mu[f(x, y) = 1 \text{ and } (x, y) \in R] \right|$$

The *discrepancy of f according to μ* is

$$\text{Disc}_\mu(f) = \max_R \{\text{Disc}_\mu(R, f)\},$$

where the maximum is taken over all rectangles R.

Proposition 2 *For every function $f : X \times Y \to \{0, 1\}$, every probability distribution μ on $X \times Y$, and every $\epsilon \geq 0$,*

$$D^\mu_{\frac{1}{2}-\epsilon} \geq \log_2\left(\frac{2\epsilon}{Disc_\mu(f)}\right)$$

Proof Let $\mathcal{P}$ be a c-bit deterministic protocol for f which is correct with probability at least $\frac{1}{2}+\epsilon$, where the inputs are weighted by μ. Then,

$$\begin{aligned}(\tfrac{1}{2}+\epsilon)-(\tfrac{1}{2}-\epsilon) &\leq \Pr_{\mu}[\mathcal{P}(x,y)=f(x,y)]-\Pr_{\mu}[\mathcal{P}(x,y)\neq f(x,y)] \\ 2\epsilon &= \sum_{\ell}\left(\Pr_{\mu}[\mathcal{P}(x,y)=f(x,y) \text{ and } (x,y)\in R_\ell]-\Pr_{\mu}[\mathcal{P}(x,y)\neq f(x,y) \text{ and } (x,y)\in R_\ell]\right)\end{aligned}$$

where the summation is over all leaves ℓ of the protocol. Since each leaf designates either a 0 or a 1, we can bound this expression from above by

$$\sum_{\ell}\left|\Pr_{\mu}[f(x,y)=0 \text{ and } (x,y)\in R_\ell]-\Pr_{\mu}[f(x,y)=1 \text{ and } (x,y)\in R_\ell]\right|$$

Each R_ℓ is a rectangle, so each of the terms in this sum is bounded from above by $\mathrm{Disc}_\mu(f)$. Since there at most 2^c leaves, we get $2\epsilon \leq 2^c \cdot \mathrm{Disc}_\mu(f)$, which implies the result. □

We now demonstrate how to prove a lower bound for the inner product (IP) function by calculating the discrepancy of IP according to the uniform distribution. First, we prove the following result, known as the Lindsey Lemma. We will use the fact that, for any two rows r_i and r_ℓ of a Hadamard matrix, $< r_i, r_\ell >= 0$ whenever $i \neq \ell$.

Lemma 3 *Let H be an $N \times N$ Hadamard matrix. Let $K = S \times T$, where $|S| = a$ and $|T| = b$, be an $(a \times b)$ submatrix of H. The absolute value of the sum of all entries in K is bounded above by $\sqrt{abN}$.*

Proof Let $\alpha = \sum_{i\in S}\sum_{j\in T} K_{ij}$. Let K_i denote the i^{th} row of K and define $\overline{y} = \sum_{i\in S} K_i$. Denote by $\overline{x}$ the vector such that

$$\overline{x}_j = \begin{cases} 1 & \text{if } j \in T \\ 0 & \text{if } j \notin T \end{cases}$$

It follows that $\alpha =< \overline{x}, \overline{y} >$. But,

$$\begin{aligned} < \overline{x}, \overline{y} >^2 &\leq |\overline{x}|_2^2 \cdot |\overline{y}|_2^2 \\ &= b \cdot |\overline{y}|_2^2 \\ &= b \cdot < \overline{y}, \overline{y} > \\ &= b \cdot < \textstyle\sum_{i\in S} K_i \ , \ \sum_{i\in S} K_i > \\ &= b \cdot \textstyle\sum_{i\in S}\sum_{\ell\in S} < K_i, K_\ell > \\ &= b \cdot \textstyle\sum_{i\in S} < K_i, K_i > \quad (\text{as } < K_i, K_\ell >= 0 \text{ when } i \neq \ell) \\ &= b \cdot (aN) \end{aligned}$$

Thus, $\alpha = \sqrt{abN}$. □

Now, we calculate an upper bound on the discrepancy of IP according to the uniform distribution. We define $f = IP$ as:

$$f(x,y) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i y_i \bmod 2 = 0 \\ -1 & \text{if } \sum_{i=1}^n x_i y_i \bmod 2 = 1 \end{cases}$$

Then, any matrix M_f is a $2^n \times 2^n$ Hadamard matrix. So, for any rectangle $K = S \times T$ with $|S| = a$ and $|T| = b$,

$$\begin{aligned} \text{Disc}_{uniform}(K, f) &= \textstyle\sum_{i \in S} \sum_{j \in T} K_{ij} \\ &\leq \frac{\sqrt{ab}2^n}{2^{2n}} \quad \text{by Lemma 3} \end{aligned}$$

As $a, b \leq 2^n$,

$$\begin{aligned} \text{Disc}_{uniform}(f) &\leq \frac{\sqrt{2^n 2^n 2^n}}{2^{2n}} \\ &= 2^{-\frac{n}{2}} \end{aligned}$$

So, by Proposition 2, $D^{\mu}_{\frac{1}{2}-\epsilon}(IP) \geq \log_2\left(\frac{2\epsilon}{2^{-\frac{n}{2}}}\right) = \frac{n}{2} + 1 + \log_2(\epsilon)$.