CS 2429 - Foundations of Communication Complexity

Lecturer: Toniann Pitassi

1 Randomized Communication Complexity

1.1 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. All of these definitions are for private coin protocols.

- $R_0(f)$ 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., $R_0^{pub}(f)$, $R_\epsilon^{pub}(f)$, $R_\epsilon^{1\ pub}(f)$. We have previously seen the following facts:

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

1.2 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 $R_\epsilon^{pub}(f) = \max_\mu D^\mu_\epsilon(f)$

Proof First, we show that $R_\epsilon^{pub}(f) \geq \max_\mu D^\mu_\epsilon(f)$. Let $\mathcal{P}$ be a randomized public coin protocol with worst-case cost $R_\epsilon^{pub}(f)$ 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, $R_\epsilon^{pub}(f) \geq \text{cost}(\mathcal{P}_{r'}) \geq \max_\mu D^\mu_\epsilon(f)$.

Next, we show that $R_\epsilon^{pub}(f) \leq \max_\mu D^\mu_\epsilon(f)$. Let $c = \max_\mu D^\mu_\epsilon(f)$.

1.2.1 Minimax Theorem

We will show this direction of the theorem by an application of Von Neumann's Minimax Theorem. In a two-player, zero-sum game, there are two players, $P1$ and $P2$. $P1$ has a finite set $A = \{a_1, \ldots, a_m\}$ of pure strategies, and $P2$ has a finite set of pure strategies, $B = \{b_1, \ldots, b_n\}$. Each player has a utility for each pair (a_i, b_j) of actions. The utility for $P1$ is denoted by $U_1(a_i, b_j)$ and the utility for $P2$ is denoted by $U_2(a_i, b_j)$. It is a zero-sum game if for all i, j $U_1(a_i, b_j) = -U_2(a_i, b_j)$. In our case, for each (a_i, b_j), one of the players will win and the other one will lose.

Each player can use a mixed strategy by creating a probability mass function and playing each pure strategy with a fixed probability. Let p_i denote the probability that $P1$ plays action a_i and let q_j denote the probability that $P2$ plays action b_j. Since p and q are probabilities, we have that each $p_i, q_j \geq 0$, and the sum of the p_i's is 1, and the sum of the q_j's is 1. A mixed strategy for $P1$ will be denoted by p, and similarly q denotes a mixed strategy for $P2$. For each mixed strategy pair (p, q), the payoff $M(p, q)$ is defined to be

$$\sum_{i=1}^{m} \sum_{j=1}^{n} p_i M(a_i, b_j) q_j.$$

When $P1$ uses pure strategy a_i and $P2$ uses mixed strategy q, then $M(a_i, q) = \sum_{j=1}^{n} M(a_i, b_j) q_j$, and analogously for $M(p, b_j)$. We let P and Q denote the set of all mixed strategies available to player 1 and 2 respectively. Player $P1$'s objective is to select a mixed strategy $p \in P$ soas to maximize $min_q M(p, q)$, and at the same time $P2$'s objective is to select a mixed strategy $q \in Q$ soas to minimize $max_p M(p, q)$.

The Minimax theorem states that for every two-person zero-sum game, there exists an equilibrium strategy. That is there exists a value v such that

$$max_p min_q M(p, q) = min_q max_p M(p, q)$$

In other words, in every two-person zero-sum game with finite strategies, there exists a value v and a mixed strategy for each player such that: (a) given Player 2's strategy, the best payoff for Player 1 is v, and (b) given Player 1's strategy, the best payoff for Player 2 is $-v$.

In our context, we define a two-player zero-sum game as follows:

- $P1$ (the protocol designer): his pure strategies are all c-bit deterministic protocols $\mathcal{P}_\nabla$, one for each choice of coin flips. His mixed strategies are all randomized protocols, P, (each of which is a distribution over the deterministic protocols).
- $P2$ (the adversary): her pure strategies are all inputs (x, y). Her mixed strategies are all distributions μ over the inputs.
- $P1$ has payoff 1 if $\mathcal{P}_r(x, y) = f(x, y)$ and -1 otherwise. That is, the designer ($P1$) wins the game iff this protocol is correct on (x, y), and otherwise $P2$ wins.

We are given as our assumption that for all distributions μ over inputs (x, y), there exists a pure strategy (a protocol) P such that the probability of a win is at least $1 - \epsilon$. This means that $Min_\mu Max_P M(\mu, P) \geq 1 - \epsilon$. (Since for each choice of μ, there is a fixed strategy P_r that achieves payoff $1-\epsilon$, so no matter what μ we choose, the designer will be able to come up with a protocol that wins $1-\epsilon$ of the time. Now by the Minimax theorem, this means that $Max_P Min_\mu M(\mu, P) \geq 1-\epsilon$.

From this it follows that there is a randomized strategy P such that for all fixed (x, y), the payoff is at least $1 - \epsilon$.

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 $R_{\epsilon}(f) = \Theta(n)$.

2 Lower Bounds for Randomized Protocols: 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 for a given function f 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 $X \times Y$.

$$Disc_{\mu}(R) = |\mu(R \cap f^{-1}(1)) - \mu(R \cap f^{-1}(0)|.$$

The discrepancy of f under μ is the maximum over all possible rectangles:

$$Disc_{\mu}(f) = max_R Disc_{\mu}(R).$$

If f has small discrepancy it means (informally) that all large rectangles are roughly balanced.

Consider a deterministic protocol that partitions the input space into rectangles $R_1, \ldots, R_{2^c}$. And suppose it has success probability $2/3$ with respect to μ. The best thing that the protocol can do if it has to give one output a_i for all inputs in the rectangle R_i is to set a_i to the bit value with the highest weight in that rectangle. This contributes $\mu(R_i \cap f^{-1}(a_i))$ to the success probability and $\mu(R_i \cap f^{-1}(1 - a_i))$ to the failure probability. Thus the overall success probability is $\sum_i \mu(R_i \cap f^{-1}(a_i))$ and the overall error probability is $\sum_i \mu(R_i \cap f^{-1}(1 - a_i))$. Since the difference between these two has to be at least $2/3 - 1/3 = 1/3$, we have

$$
\begin{aligned}
1/3 &\leq \sum_{i=1}^{2^c} \mu(R_i \cap f^{-1}(a_i)) - \sum_{i=1}^{2^c} \mu(R_i \cap f^{-1}(1-a_i)) && (1)\\
&\leq \sum_{i=1}^{2^c} |\mu(R_i \cap f^{-1}(a_i)) - \mu(R_i \cap f^{-1}(1-a_i))| && (2)\\
&= \sum_{i=1}^{2^c} Disc_\mu(R_i) && (3)\\
&\leq 2^c Disc_\mu(f). && (4)
\end{aligned}
$$

This gives a lower bound on communication: $c \geq \log(1/3Disc_\mu(f))$. To get a lower bound for randomized protocols, it suffices to find a distribution μ such that $Disc_\mu(f)$ is small.

We have proved

Theorem 2 *For every distribution μ, $R_\mu(f) \geq \log(1/3Disc_\mu(f))$.*

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. Before we prove this result, we will study the communication matrix for the IP function for $n = 3$ to get some intuition. We will actually switch things a little bit and analyze the matrix whose (x, y) entry is $(-1)^{x \cdot y}$. This is just the communication matrix for IP, with 0's replaced by 1's and 1's replaced by -1's. With this switch of basis, The associated IP matrices are the Hadamard matrices. Hadamard matrices are defined to be square matrices where each entry is either $+1$ or -1 and such that all pairs of rows are mutually orthogonal.

The IP matrix, H_n, for $n = 3$ looks like this:

1	1	1	1	1	1	1	1
1	-1	1	-1	1	-1	1	-1
1	1	-1	-1	1	1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	1	1	-1	-1	-1	-1
1	-1	1	-1	-1	1	-1	1
1	1	-1	-1	-1	-1	1	1
1	-1	-1	1	-1	1	1	-1

More generally $H_0 = [1]$ and H_n is built from H_{n-1} as follows: the lower right quadrant of H_n is equal to $-H_{n-1}$ and the other three quadrants are equal to H_{n-1}.

The following facts are easy to prove about H_n:

- Every pair of rows is orthogonal, and therefore $H_n^2 = N \cdot I$.
- We can interpret the rows as parity functions
- The matrix is symmetric about the diagonal
- The eigenvectors form an orthonormal basis. (That is $< v_i, v_j >= 0$ for all $i \neq j$, and $v_i^2 = 1$ for all i.)

- The only eigenvalues of H_n are $+/-\sqrt{N}$.

We want to find the eigenvalues of the Hadamard matrices, as claimed in the last bullet point above. Recall these are defined by the following recursive construction:

$$H_0 = [1], \qquad H_{n+1} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}.$$

Lemma 3 *For each n, $H_n^2 = HH^T = 2^n I_{2^n}$.*

Proof The proof is by induction. Since $H_0 = I_1$, the lemma is correct for $n = 0$.

Given that $H_n^2 = 2^n I$, we can calculate H_{n+1}^2 explicitly:

$$\begin{aligned} H_{n+1}^2 &= \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}^2 \\ &= \begin{bmatrix} H_n^2 + H_n^2 & H_n^2 - H_n^2 \\ H_n^2 - H_n^2 & H_n^2 + H_n^2 \end{bmatrix} = \begin{bmatrix} 2^{n+1} I_{2^n} & 0 \\ 0 & 2^{n+1} I_{2^n} \end{bmatrix} = 2^{n+1} I_{2^{n+1}}. \end{aligned}$$

Corollary 4 *The eigenvalues of H_n are all $\pm 2^{n/2}$.*

Proof

By the above lemma, for all v, $vHH^T = 2^n v$ and therefore 2^n is the only eigenvector of HH^T. Thus, the only eigenvectors of H are $\pm 2^{n/2}$.

We denote the discrepancy of f (with respect to the uniform distribution) and a rectangle $A \times B$ by $\mathrm{disc}(f, A \times B)$. All our results can be generalized to arbitrary distributions by multiplying each entry of M_f by the probability of the corresponding cell.

Recall that Boolean functions can be considered as taking values in either $\{0, 1\}$ or $\{+1, -1\}$. In this section, we will use the ± 1 convention when describing the matrices and rectangles.

We use the notation $\mathbf{1}_A$ for the characteristic vector of A, which contains 1 in positions corresponding to the elements of A, and 0's elsewhere.

2.1 The Eigenvalue Method

The eigenvalue method upper bounds the discrepancy using the maximal eigenvalue of M_f.

Lemma 5 (Eigenvalue Bound) *Let f be a symmetric Boolean function, i.e. $f(x, y) = f(y, x)$. Then*

$$\mathrm{disc}(f, A \times B) \leq 2^{-2n} \lambda_{\max} \sqrt{|A| \cdot |B|},$$

where $n = |x| = |y|$ is the input size, and $\lambda_{\max}$ is the largest eigenvalue of the symmetric matrix M_f.

Proof Since M_f is symmetric, its eigenvectors v_i form an orthonormal basis for $\mathbb{R}^n$. Denote by λ_i the eigenvalue corresponding to v_i, so that $M_f v_i = \lambda_i v_i$.

Expand the characteristic vectors of A and B in this basis:

$$\mathbf{1}_A = \sum \alpha_i v_i, \qquad \mathbf{1}_B = \sum \beta_i v_i$$

Putting these expansions into the definition of discrepancy, we are almost done. Since $2^{2n}\text{disc}(f, A\times B)$ is equal to the absolute value of the difference between the number of 1's and the number of 0's in $A \times B$, we have:

$$\begin{aligned} 2^{2n}\text{disc}(f, A \times B) &= \left|\mathbf{1}_A{}^{\mathrm{T}} M_f \mathbf{1}_B\right| \\ &= \left|\left(\sum \alpha_i v_i\right)^{\mathrm{T}} \left(\sum \beta_i \lambda_i v_i\right)\right| \\ &= \left|\sum \alpha_i \beta_i \lambda_i\right| \leq \lambda_{\max} \left|\sum \alpha_i \beta_i\right|. \end{aligned}$$

Note that $\sum \alpha_i^2 = \|\mathbf{1}_A\|^2 = |A|$ by Parseval's identity. (Parseval's identity relates the values of the Fourier coefficients to the values of the function. Namely, it states that for any function $f : \{0,1\}^n \to R$, the sum of the squares of the Fourier coefficients of f is equal to f^2. Note that in our case we have not normalized. If we had normalized – so that the Fourier coefficients were normalized, then the sum of the squares of the Fouerier coefficients of f would be equal to $E[f^2]$.)

and similarly $\sum \beta_i^2 = |B|$. The lemma follows from an application of Cauchy-Schwarz:

$$\begin{aligned} 2^{2n}\text{disc}(f, A \times B) &\leq \lambda_{\max} \left|\sum \alpha_i \beta_i\right| \\ &\leq \lambda_{\max} \sqrt{\sum \alpha_i^2} \sqrt{\sum \beta_i^2} = \lambda_{\max} \sqrt{|A| \cdot |B|}. \end{aligned}$$

We are now ready to prove Lindsey's Lemma which gives a bound on the disrepancy of the inner product function:

Lemma 6 (Lindsey's Lemma) $2^{2n}\text{disc}(\text{IP}_n, A \times B) \leq \sqrt{2^n |A| \cdot |B|}$.

Here $\text{IP}_n(x, y) = \sum x_i y_i \pmod 2$.

Proof The matrix corresponding to IP_n is H_n. We have shown that $\lambda_{\max}(H_n) = 2^{n/2}$, and so the lemma follows by the Eigenvalue Bound.

We are now ready to prove the following theorem.

Theorem 7 $R^{cc}(IP) = \Omega(n)$

By Lindsey's Lemma, discrepancy is maximized when $|A| = |B| = 2^n$, and this gives $disc(IP_n, A\times B) \leq 2^{3n/2} 2^{-2n} = 2^{-n/2}$. Thus $R(IP_n) \geq \log(1/3disc(IP_n)) = log(2^{n/2}/3) = \Omega(n)$.