CAS CS 591 B: Communication Complexity

Prof. Mark Bun — Fall 2019

Lecture Notes 5:

Discrepancy

Reading.

- Rao-Yehudayoff Chapter 5

Last time we saw how Yao's Principle could be used (in principle) to lower bound the randomized communication complexity of a function by exhibiting a distribution under which it has high distributional complexity. Namely, for any function f, we have $\mathbf{BPP}^{pub}_{\varepsilon}(f) = \max_{\mu} D^{\mu}_{\varepsilon}(f)$.

This lecture will be about a central technique for lower bounding distributional complexity called the discrepancy method. The discrepancy of a collection of sets $S_1, \dots, S_m \subseteq X$ measures how unbalanced any of the sets S_i has to be under a red-blue coloring of the elements of X. For reasonable set systems, a random coloring has low discrepancy with high probability, so discrepancy can be thought of as a measure of how random the coloring looks.

To apply this idea to communication complexity, the collection of sets we will be interested in will (intuitively) be the large combinatorial rectangles in $X \times Y$. If this collection has low discrepancy under the coloring induced by the 0's and 1's of a function f, that means every large rectangle must be close to balanced with respect to that function. Thus, discrepancy upper bounds can be used to establish communication lower bounds.

Definition 1. Let $f : X \times Y \to \{0,1\}$, let μ be a distribution over $X \times Y$, and let $R \subseteq X \times Y$ be a rectangle. Define

$$\mathrm{disc}_{\mu}(f, R) = |\mu(R \cap f^{-1}(0)) - \mu(R \cap f^{-1}(1))|,$$

and let the discrepancy of f under μ be

$$\mathrm{disc}_{\mu}(f) = \max_{R \text{ a rectangle}} \mathrm{disc}_{\mu}(f, R).$$

Suppose $\mathrm{disc}_{\mu}(f, R)$ is small. Then (intuitively) one of two things must be true:

1. R is small with respect to μ, or
2. R is close to balanced, i.e., far from monochromatic.

Either of these things is good for the lower-bound-prover. If a function has small discrepancy, then that means the input cannot be partitioned into large almost-monochromatic rectangles. We make this idea formal as follows.

Theorem 2. *Let $f : X \times Y \to \{0,1\}$ and let μ be a distribution over $X \times Y$. Then*

$$D^{\varepsilon}_{\mu}(f) \geq \log\left(\frac{1-2\varepsilon}{\mathrm{disc}_{\mu}(f)}\right).$$

Proof. Let Π be a deterministic protocol computing f under μ with cost c. We will show that

$$\mathrm{disc}_\mu(f) \geq (1-2\varepsilon)2^{-c},$$

which implies the theorem.

The leaves of Π partition $X \times Y$ into rectangles $R_1, \ldots, R_t$ for some $t \leq 2^c$. For each rectangle i, let a_i be the value that Π outputs on rectangle R_i. Since the protocol has error ε, we have

$$\begin{aligned}
1-\varepsilon &\leq \Pr_{(x,y)\sim\mu}[\Pi(x,y) = f(x,y)] \\
&= \sum_{i=1}^{t} \Pr_{(x,y)\sim\mu}[(x,y) \in R_i \wedge f(x,y) = a_i] \\
&= \sum_{i=1}^{t} \mu(R_i \cap f^{-1}(a_i)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\varepsilon &\geq \Pr_{(x,y)\sim\mu}[\Pi(x,y) \neq f(x,y)] \\
&= \sum_{i=1}^{t} \mu(R_i \cap f^{-1}(1-a_i)).
\end{aligned}$$

Combining the two inequalities,

$$\begin{aligned}
1-2\varepsilon &\leq \sum_{i=1}^{t} \mu(R_i \cap f^{-1}(a_i)) - \mu(R_i \cap f^{-1}(1-a_1)) \\
&\leq \sum_{i=1}^{t} |\mu(R_i \cap f^{-1}(0)) - \mu(R_i \cap f^{-1}(1))| \\
&\leq \sum_{i=1}^{t} \max_R \mathrm{disc}_\mu(f,R) \\
&\leq 2^c \mathrm{disc}_\mu(f).
\end{aligned}$$

□

Now we turn to developing techniques for upper bounding the discrepancy of particular functions.

1 The Spectral Method

Our first technique for upper bounding discrepancy will be in terms of the analytic properties of the communication matrix of f. It will actually be more convenient to work with the "sign matrix" of f defined by

$$S_f[x,y] = (-1)^{f(x,y)} = \begin{cases} -1 & \text{if } f(x,y) = 1 \\ 1 & \text{if } f(x,y) = 0. \end{cases}$$

If f is a symmetric function, i.e., $f(x,y) = f(y,x)$ for every x, y, then the matrix S_f is symmetric. Let $\|S_f\|$ denote the spectral norm of S_f. When S_f is symmetric, it has a few equivalent characterizations:

- $\|S_f\|$ is the absolute value of the largest eigenvalue of S_f.
- $\|S_f\|$ is the largest singular value of S_f. In particular, $\|S_f x\|_2 \le \|x\|_2$ for every vector x.

Theorem 3. *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a symmetric Boolean function, i.e., $f(x,y) = f(y,x)$ for every x, y, and let μ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. Then*

$$\mathrm{disc}_\mu(f,R) \le 2^{-2n}\|S_f\|\sqrt{|R|}.$$

Proof. Let $N = 2^n$. The fact that f is symmetric implies that the real matrix S_f is symmetric, and hence by the spectral theorem, its eigenvectors $v_1, \ldots, v_N$ form an orthonormal basis for $\mathbb{R}^N$. Let λ_i be the associated eigenvalues, i.e., $S_f v_i = \lambda_i v_i$. Let $R = A \times B$ and expand the characteristic vectors for A and B in this basis as

$$\mathbf{1}_A = \sum_{i=1}^N \alpha_i v_i, \qquad \mathbf{1}_B = \sum_{i=1}^N \beta_i v_i.$$

Observe that this expansion satisfies

$$|A| = \mathbf{1}_A^T \mathbf{1}_A = \left(\sum_{i=1}^N \alpha_i v_i\right)^T \left(\sum_{j=1}^N \alpha_j v_j\right) = \sum_{i,j=1}^N \alpha_i \alpha_j v_i^T v_j = \|\alpha\|_2^2,$$

and similarly $|B| = \|\beta\|_2^2$. (This is just Parseval's identity: The sum of squared coefficients of a vector is the same regardless of what orthonormal basis it's expanded in.) Then

$$\begin{aligned}
\mathrm{disc}_\mu(f,R) &= \left|2^{-2n}|R \cap f^{-1}(1)| - 2^{-2n}|R \cap f^{-1}(0)|\right| \\
&= 2^{-2n}\left|\sum_{(x,y)\in R} S_f[x,y]\right| && \text{which is why we work with sign matrices} \\
&= 2^{-2n}\left|\mathbf{1}_A^T S_f \mathbf{1}_B\right| \\
&= 2^{-2n}\left|\left(\sum_{i=1}^N \alpha_i v_i\right)^T \left(\sum_{i=1}^N \beta_i \lambda_i v_i\right)\right| && \text{since the } v_i\text{'s are eigenvectors} \\
&= 2^{-2n}\left|\sum_{i,j} \alpha_i \beta_j \lambda_j v_i^T v_j\right| \\
&= 2^{-2n}\left|\sum_{j=1}^N \alpha_j \beta_j \lambda_j\right| && \text{by orthonormality} \\
&\le 2^{-2n}\|S_f\|\|\alpha\|_2\|\beta\|_2 && \text{by Cauchy-Schwarz} \\
&= 2^{-2n}\|S_f\|\sqrt{|A|}\sqrt{|B|}.
\end{aligned}$$

□

1.1 Inner Product

Recall the Inner Product Mod 2 function is defined by

$$\mathrm{IP}_n(x, y) = \langle x, y \rangle \bmod 2 = \sum_{i=1}^{n} x_i y_i \bmod 2.$$

The sign matrix of IP_n, which we'll denote by H_n, is an example of a Hadamard matrix. A Hadamard matrix is a square $\{\pm 1\}$-valued matrix whose rows are mutually orthogonal. As an example, H_3 takes the form

$$H_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

To get a bound on the spectral norm, we establish the following fact:

Claim 4. *$H_n^2 = 2^n I_{2^n}$, where I_{2^n} is the $2^n \times 2^n$ identity matrix.*

Proof. We calculate each entry $H_n^2[x, x']$ as

$$\begin{aligned} \langle H_{n,x}, H_{n,x'} \rangle &= \sum_{y \in \{0,1\}^n} (-1)^{\langle x,y \rangle} (-1)^{\langle x',y \rangle} \\ &= \sum_{y \in \{0,1\}^n} (-1)^{\langle x \oplus x',y \rangle} \\ &= \begin{cases} 2^n & \text{if } x = x' \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

□

Hence, all of the eigenvalues of H_n have absolute value $2^{n/2}$, so $\|H_n\| = 2^{n/2}$. Combining this with Theorem 3 gives us the following bound on the discrepancy:

Lemma 5 (Lindsey's Lemma)**.** *For every rectangle R, we have*

$$\mathrm{disc}_\mu(\mathrm{IP}_n, R) \leq 2^{-3n/2} \sqrt{|R|}.$$

Since every rectangle has size at most 2^{2n}, this implies that $\mathrm{disc}_\mu(\mathrm{IP}_n) \leq 2^{-n/2}$. Combining this with Theorem 2 establishes that $D^\mu_{1/3}(\mathrm{IP}_n) \geq \Omega(n)$, and hence by Yao's principle, $\mathbf{BPP}^{pub}_{1/3}(\mathrm{IP}_n) \geq \Omega(n)$.

2 The BNS / Cauchy-Schwarz Method

This method of upper bounding discrepancy was introduced by Babai, Nisan, and Szegedy. It replaces discrepancy with an upper bound which is often much easier to compute. (No maximum over rectangles and the dependences on x and y are decoupled, at least for product distributions.)

Theorem 6. *Let μ be the uniform distribution over $X \times Y$. Then*

$$\mathrm{disc}_\mu(f)^2 \leq \mathbb{E}_{y,y'} \left| \mathbb{E}_x S_f[x,y] S_f[x,y'] \right|$$

where $x \sim X$, $y, y' \sim Y$ are chosen independently.

Proof. From the definition of discrepancy,

$$\mathrm{disc}(f, A \times B) = \sum_{x \in A, y \in B} 2^{-2n} S_f[x,y].$$

This can be rewritten as

$$\mathrm{disc}(f, A \times B) = \left| \mathbb{E}_{x,y} \mathbf{1}_A(x) \mathbf{1}_B(y) S_f[x,y] \right|.$$

Squaring,

$$\begin{aligned}
\mathrm{disc}(f, A \times B)^2 &= \left(\mathbb{E}_x \mathbf{1}_A(x) \mathbb{E}_y \mathbf{1}_B(y) S_f[x,y] \right)^2 \\
&\leq \mathbb{E}_x \left(\mathbf{1}_A(x) \mathbb{E}_y \mathbf{1}_B(y) S_f[x,y] \right)^2 && \text{by Jensen: } \mathbb{E}[Z]^2 \leq \mathbb{E}[Z^2] \\
&\leq \mathbb{E}_x \left(\mathbb{E}_y \mathbf{1}_B(y) S_f[x,y] \right)^2 \\
&= \mathbb{E}_x \left(\mathbb{E}_{y,y'} \mathbf{1}_B(y) \mathbf{1}_B(y') S_f[x,y] S_f[x,y'] \right) \\
&\leq \mathbb{E}_{y,y'} \mathbf{1}_B(y) \mathbf{1}_B(y') \left(\mathbb{E}_x S_f[x,y] S_f[x,y'] \right) \\
&\leq \mathbb{E}_{y,y'} \left| \mathbb{E}_x S_f[x,y] S_f[x,y'] \right|.
\end{aligned}$$

□

The BNS approach gives us another way of proving Lindsey's Lemma. Note that for every x, y, y' we have

$$\mathbb{E}_x H_n[x,y] H_n[x,y'] = \begin{cases} 0 & \text{if } y = y' \\ 1 & \text{otherwise} \end{cases}.$$

Hence,

$$\mathrm{disc}(\mathrm{IP}_n, A \times B)^2 \leq \mathbb{E}_{y,y'} \left| \mathbb{E}_x H_n[x,y] H_n[x,y'] \right| = \Pr[y = y'] = 2^{-n}$$

so again we conclude $\mathrm{disc}(\mathrm{IP}_n) \leq 2^{-n/2}$.

3 A Non-Example: Disjointness

Can we use a discrepancy upper bound to prove a lower bound for the set disjointness function? It turns out that disjointness has rather large discrepancy. Here, the discrepancy of a function is the minimum discrepancy over all distributions on the inputs:

Definition 7. The discrepancy of a function $f : X \times Y \to \{0,1\}$ is

$$\mathrm{disc}(f) = \min_{\mu} \mathrm{disc}_{\mu}(f).$$

Note that by combining Yao's principle with Theorem 2, we have the following formulation of the discrepancy lower bound:

Theorem 8. *For every $0 < \delta < 1/2$, we have*

$$\mathbf{BPP}^{pub}_{1/2-\delta}(f) \geq \log\left(\frac{2\delta}{\mathrm{disc}(f)}\right) = \log(1/\mathrm{disc}(f)) - \log(1/2\delta).$$

This formulation reveals that discrepancy not only gives good lower bounds on the bounded error randomized communication complexity of f, but also for protocols where the advantage δ over random guessing is as small as, say, $\sqrt{\mathrm{disc}(f)}/2$. In particular, $\mathbf{BPP}^{pub}_{1/2-\delta}(\mathrm{IP}_n) \geq \Omega(n)$ even for $\delta = 2^{-\Omega(n)}$. But this strength of the discrepancy method can also be its weakness. If a function *does* admit a good large error protocol, then the discrepancy method cannot even prove strong lower bounds even for bounded error.

We can use this connection to prove a lower bound on the discrepancy of Disjointness by exhibiting a good large-error communication protocol.

Theorem 9. $\mathrm{disc}(\mathrm{DISJ}_n) \geq \Omega(1/n)$.

Proof. We exhibit a public coin randomized protocol for disjointness DISJ_n with advantage $\delta = 1/4n$ and communication cost $c = 2$. By Theorem 8 this will imply

$$\mathrm{disc}(f) \geq 2\delta \cdot 2^{-c} \geq \frac{1}{8n}.$$

The protocol is as follows. Using shared randomness, Alice and Bob select a random index $i \in [n]$. Alice sends the bit x_i to Bob. If $x_i = y_i = 1$, he reports $\mathrm{DISJ}_n(x,y) = 0$. Otherwise, he reports 0 with probability $\frac{1}{2} - \frac{1}{4n}$ and 1 with probability $\frac{1}{2} + \frac{1}{4n}$.

To see that this protocol succeeds, first suppose $\mathrm{DISJ}_n(x,y) = 1$. Then Alice and Bob always find themselves in the second case, and their advantage over random guessing is $1/(4n)$. Otherwise, suppose $\mathrm{DISJ}_n(x,y) = 0$. Then with probability at least $1/n$, they select an index i for which $x_i = y_i = 1$, and hence their success probability is at least

$$\frac{1}{n} + \left(1 - \frac{1}{n}\right)\left(\frac{1}{2} - \frac{1}{4n}\right) \geq \frac{1}{2} + \frac{1}{4n}.$$

□

Hence, even for bounded error protocols, the discrepancy method can, at best, only prove an $\Omega(\log n)$ lower bound for Disjointness. (Exercise: Show that $\mathrm{disc}(\mathrm{DISJ}_n) \leq O(1/\sqrt{n})$, so this weak lower bound is actually attainable.) Nevertheless, there are techniques related to the discrepancy method, e.g., the one-sided discrepancy or corruption bound, which are capable of proving a tight lower bound of $\Omega(n)$ for Disjointness. This method is explored in one of your homework problems.

4 Discrepancy and $\mathbf{PP^{cc}}$

Let's take a closer look at the relationship between discrepancy and large-error communication. It turns out that discrepancy actually gives a tight characterization of the cost of large-error protocols when one measures cost in the right way.

Definition 10. Let $f : X \times Y \to \{0,1\}$. A public coin randomized protocol Π computes f with advantage $\delta > 0$ if for every $(x,y) \in X \times Y$,

$$\Pr[\Pi(x,y) = f(x,y)] \geq \frac{1}{2} + \delta.$$

The $\mathbf{PP^{cc}}$ cost of a protocol Π computing f with length c and advantage δ is defined to be

$$\mathbf{PP^{cc}}(\Pi) = c + \log(1/\delta).$$

Moreover, $\mathbf{PP^{cc}}(f)$ is defined to be the least $\mathbf{PP^{cc}}$ cost of a protocol computing f, and $\mathbf{PP^{cc}}$ is the class of sequences of functions with polylogarithmic $\mathbf{PP^{cc}}$ cost.

Theorem 11. *For every function $f : X \times Y \to \{0,1\}$, we have*

$$\mathbf{PP^{cc}}(f) = \Theta(\log(1/\mathrm{disc}(f))).$$

Proof. For the $\geq$ direction, we want to show that any protocol Π computing f has $\mathbf{PP^{cc}}(\Pi) \geq \Omega(\log(1/\mathrm{disc}(f)))$. Let Π be a protocol computing f with length c and advantage δ. If $\delta \leq \sqrt{\mathrm{disc}(f)}/2$, we're done. Otherwise, by Theorem 8, we have

$$c \geq \log(1/\mathrm{disc}(f)) - \log(1/2\delta) \geq \frac{1}{2}\log(1/\mathrm{disc}(f)),$$

and hence $\mathbf{PP^{cc}}(\Pi) \geq \Omega(\log(1/\mathrm{disc}(f)))$.

For the $\leq$ direction, suppose f has large discrepancy $\mathrm{disc}(f) \geq 2^{-c}$. We will exhibit a good distributional protocol for f with respect to any given μ which, by Yao's principle, implies the existence of a good public coin protocol. For any distribution μ we have $\mathrm{disc}_\mu(f) \geq 2^{-c}$, and hence there exists a rectangle $R = A \times B$ such that $\mathrm{disc}_\mu(f,R) \geq 2^{-c}$. Let $a \in \{0,1\}$ such that $\mu(R \cap f^{-1}(a)) \geq 2^{-c} + \mu(R \cap f^{-1}(1-a))$. This implies that $\mu(R \cap f^{-1}(a)) \geq \frac{1}{2}\mu(R) + 2^{-c-1}$. The protocol is as follows. Using 2 bits of communication, Alice and Bob check if $(x,y) \in R$. If so, they output a, and otherwise output the outcome of a coin flip. The success probability of this protocol is

$$\begin{aligned}
\Pr[\Pi(x,y) = f(x,y)] &= \mu(R \cap f^{-1}(a)) + \frac{1}{2}(1 - \mu(R)) \\
&\geq \frac{1}{2}\mu(R) + 2^{-c-1} + \frac{1}{2} - \frac{1}{2}\mu(R) \\
&= \frac{1}{2} + 2^{-c-1}.
\end{aligned}$$

Hence, $\mathbf{PP^{cc}}(f) \leq 2 + (c+1) = O(\log(1/\mathrm{disc}(f)))$. □