MATHEMATICAL METHODS IN THEORETICAL CS LECTURE 6: RAZBOROV DISJOINTNESS LOWER BOUND, FORSTER'S THEOREM

LECTURER: MORITZ HARDT, SCRIBE: ARAVINDAN VIJAYARAGHAVAN

Summary: In this lecture, we show two results dealing with lower bounds in communication complexity. The first lower bound is an $\Omega(n)$ lower bound on the distributional complexity of *Disjointness* due to [3, 8]. Here we will present the simplified proof presented in [8]. In the second part, we will show how to obtain lower bounds on the unbounded error probabilistic communication complexity by Forster's method [2] of lower-bounding the sign rank of the corresponding matrix by showing that it has a small spectral norm.

1. Disjointness lower bound

In the first part of the lecture, we will try to lower bound the ϵ-error probabilistic communication complexity of a predicate A, $C_\epsilon(A)$ by using the concept of the Distributional complexity.

The ϵ-error distributional communication complexity $D^\mu_\epsilon(A)$ is the minimum cost (w.r.t. number of communications, measured in say, bits) deterministic protocol P such that

$$Pr_\mu[P(X,Y) = A(X,Y)] \geq 1-\epsilon \tag{1}$$

where μ is a given probability distribution over the inputs X and Y (each n bits long) to the two parties. Yao showed that $D^\mu_\epsilon(A)$ can be used to lower bound $C_\epsilon(A)$ as $C_\epsilon(A) \geq \frac{1}{2}D^\mu_{2\epsilon}(A)$ [5].

We will now prove a $\Omega(n)$ lower bound on probabilistic communication complexity of the Disjointness function by constructing a μ for which $D^\mu_\epsilon(DISJ_n) = \Omega(n)$ where $DISJ_n$ is the *Disjointness* predicate, where the inputs X and Y are $\in \{0,1\}^n$ each representing a subset of $\{1, 2, \ldots, n\}$ (represented as $[n]$).

Theorem 1.1. *$\exists \mu$ such that $D^\mu_\epsilon(DISJ_n) \geq \Omega(n)$ where $\epsilon < \frac{1}{100}$, where*

$$DISJ(X,Y) = \begin{cases} 1 & \text{if } X \cap Y = \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Proof. We first give the probability distribution μ over the inputs. Let $n = 4m-1$ and let $T = (T_x, T_y, \{i\})$ be an arbitrary partition of $[n]$ such that $|T_x| = |T_y| = 2m-1$. Now, the input to the first party, an m-element set $X(|X| = m)$ is chosen uniformly at random from $T_x \cup \{i\}$, and the m-element set Y is similarly chosen uniformly at random from $T_y \cup \{i\}$. Let $X_0(Y_0)$ correspond to an input X(resp. Y) such that $i \notin X$(resp. Y), and X_1(resp Y_1) correspond to an input such that $i \in X$(resp. Y). Hence, in the distribution defined, we get inputs of the types $(X_0, Y_0), (X_0, Y_1), (X_1, Y_0)$ and (X_1, Y_1) with an equal probability of $\frac{1}{4}$ each.

Let A be the set of inputs of the type (X_1, Y_1) ($DISJ$ value 1) and B be the other inputs with non-zero weight ($DISJ$ value 0). The proof of the theorem follows from Lemma 1.5 as follows.
Let $D^{\mu}_{\epsilon} = k$. Let $R_1, R_2, \ldots, R_t$ (with $t \leq 2^k$) be the (almost) monochromatic rectangles which have the function value 1.

$$\begin{aligned}
\mu(B \cap \bigcup_{i=1}^{t} R_i) &\leq \epsilon \\
\mu(B \cap \bigcup_{i=1}^{t} R_i) &= \sum_{i=1}^{t} \mu(R_i \cap B) \\
&\geq \sum_{i=1}^{t} \alpha\mu(A \cap R_i) - 2^{-\delta n} \text{ (from Lemma 1.5)} \\
&\geq \alpha(\frac{3}{4} - \epsilon) - t2^{-\delta n}
\end{aligned}$$

Hence by choosing a small enough ϵ, we get $k = \Omega(n)$. □

The lemma 1.5 shows that in any rectangle (of inputs) which is not too small (weight more than $\frac{2^{-\delta n}}{\alpha}$), $R = \mathbb{X} \times \mathbb{Y}$, the number of inputs which return a (disjointness) function value 0 is not too small, hence showing that no rectangle(which is not too small) is close to be being monochromatic. However, we show another lemma with a similar flavour first.

Lemma 1.2. *For any* $R = \mathbb{X} \times \mathbb{Y}$ *, where* $\mathbb{X}, \mathbb{Y} \subseteq 2^{[n]}$,

$$P[(x_1, y_1) \in R] \geq \alpha Pr[(x_0, y_0) \in R] - 2^{-\Omega(n)} \tag{3}$$

Proof. Let $t = (t_X, t_Y, \{i\})$ be a partition of $[n]$. For ease of notation, we define the following terms:

$$p(t) = Pr(X \in \mathbb{X} | T = t) \tag{4}$$

$$p_0(t) = Pr(X_0 \in \mathbb{X} | T = t) \tag{5}$$

$$p_!(t) = Pr(X_1 \in \mathbb{X} | T = t) \tag{6}$$

(7)

Further, we call a partition $t = (t_X, t_Y, \{i\})$ *X-bad* if

$$p_1(t) < \frac{1}{3} p_0(t) - 2^{-\epsilon n} \tag{8}$$

Similarly we define the terms $q(t), q_0(t), q_1(t)$ and *Y-bad* for the inputs to the second party Y. We also call a partition t bad iff it is either *X-bad* or *Y-bad*. To prove the lemma, we would like to show that most partitions are not bad (claim 1.3).

We first observe that by fixing the partition $T = (T_X, T_Y, \{i\})$, the two quantities $Pr(X_1 \in R)$ and $Pr(Y_1 \in R)$ becomes independent [1]. Hence

$$Pr[(X_\lambda, Y_\lambda) \in R] = \mathbb{E}_t[p_\lambda(t) q_\lambda(t)] \text{ for } \lambda \in \{0, 1\} \tag{9}$$

[1] This method of fixing t to make the two events independent in order to get a convex combination has also been used subsequently in the proof of the Parallel Repetition theorem by Raz [7].

We also observe that

$$p(t) = \frac{1}{2}(p_0(t) + p_1(t)) \tag{10}$$

Finally, we note that fixing t_Y, fixes $p(t)$ and $q_0(t)$ and fixing t_X, also fixes $q(t)$ and $p_0(t)$.
Now, we will proceed to show the lemma component-wise.

Claim 1.3. *For every set $t_Y \subseteq [n]$ such that $|t_Y| = 2m - 1$,*

$$Pr(T \text{ is } X\text{-bad } | T_Y = t_Y) < \frac{1}{5} \tag{11}$$

Proof. Given t_Y, $X \leftarrow_{\mathrm{R}} [n] \backslash t_Y$ (with $|X| = m$), and hence $Pr(T$ is X-bad$|T_Y = t_Y)$ becomes fixed too.
If $p(t) < 2^{-\epsilon n}$, from equation 10, $p_0(t) \leq 2p(t)$. Hence, if t is X-bad,
$p_1(t) < \frac{1}{3}p_0(t) - 2^{-\epsilon n} < -\frac{1}{3}2^{-\epsilon n} < 0$, which is impossible.
Consider the case when $p(t) \geq 2^{-\epsilon n}$. Let $\Gamma = \mathbb{X} \cap \{X | X \subseteq [n] \text{ s.t } |X| = m\}$.
$p(t) = \frac{|\Gamma|}{\binom{2m}{m}}$. Also, if $s \leftarrow_{\mathrm{R}} \Gamma$, then,

$$p_0(t) = 2p(t)Pr(i \in s)$$

This follows because,

$$\begin{aligned} p_0(t) &= Pr(X \in \mathbb{X} | t = t, i \notin X) \\ &= 2Pr(X \in \mathbb{X}, i \notin X | T = t) \\ &= 2Pr(i \notin X | X \in \mathbb{X}, T = t)Pr(X \in \mathbb{X} | T = t) \\ &= 2p(t)Pr[i \notin s] \end{aligned}$$

Similarly,

$$p_1(t) = 2p(t)Pr(i \notin s)$$

Now, if partition t is *X-bad*, we have from equations 8,1,1

$$Pr[i \in s] < \frac{1}{3}Pr[i \notin s] - \frac{2^{-\epsilon n}}{2p(t)} \tag{12}$$

Since $p(t) \geq 2^{-\epsilon n}$, $Pr[i \in s] < \frac{1}{3}Pr[i \notin s]$. Hence,

$$Pr(i \in s) < \frac{1}{4} \tag{13}$$

Let $\{i_1, i_2, \ldots, i_{2m}\} = [n] - T_y$ and let $\vec{s} = (s_1, s_2, \ldots, s_{2m})$ where the s_j indicates whether $i_j \in s$. We now show that claim by using an entropy argument on the possible choice of vectors $\vec{s}$. Let the claim 1.3 not hold, in which case
$Pr[T = (T_X, T_Y, \{i\})$ is X-bad$|T_Y = T_y] \geq \frac{1}{5}$. Then, calculating the entropy we get

$$\begin{aligned} H(s) &\geq m(2 - 4\epsilon - o(1)) \\ H(s) &\leq \sum i = 1^{2m} H(s_i) \\ &\leq \frac{8m}{5} + \frac{2m}{5}H(\frac{1}{4}) \\ &\leq 1.93m \end{aligned}$$

which is a contradiction if we choose a small enough ϵ. Hence the claim is true. $\square$

Let $Bad(T)$ denote the indicator of the event $T = (T_X, T_Y, \{i\})$ being Bad, and let $Bad_X(T), Bad_Y(T)$ be the indicator of the events T being X-Bad and Y-Bad respectively. In the next claim, we prove that contribution of non-Bad partitions T to the RHS of Lemma 1.2 is insignificant.

Claim 1.4.

$$\mathbb{E}_T[p_0(T)q_0(T)Bad(T)] \leq \frac{4}{5}\mathbb{E}_T[p_0(T)q_0(T)]$$

Proof. Since $Bad(T) \leq Bad_X(T) + Bad_Y(T)$, and by symmetry, it suffices to show that

$$\mathbb{E}_T[p_0(T)q_0(T)Bad_X(T)] \leq \frac{2}{5}\mathbb{E}_T[p_0(T)q_0(T)]$$

Fixing T_Y also fixes $p(T), q_0(T)$. Hence,

$$\begin{aligned}
\mathbb{E}[p_0(T)q_0(T)Bad_X(T)|T_Y = t_Y] &= 2pq_0\mathbb{E}_T[Bad_X|T_Y = t_Y] \\
&\leq \frac{2}{5}pq_0 \text{ from claim 1.3} \\
&\leq \frac{2}{5}\mathbb{E}_T[q_0(T)|T_Y = t_Y] \\
&\leq \frac{2}{5}\mathbb{E}_T[p_0(T)q_0(T)|T_Y = t_Y]
\end{aligned}$$

□

Now, to complete the proof of the lemma (1.2),

$$\begin{aligned}
P[(X_1, Y_1) \in R] &= \mathbb{E}_T[p_1(T)q_1(T)] \\
&\geq \mathbb{E}_T[p_1(T)q_1(T)(1 - Bad(T))] \\
&\geq \mathbb{E}_T[(\frac{1}{3}p_0(T) - 2^{-\epsilon n})(\frac{1}{3}q_0(T) - 2^{-\epsilon n})(1 - Bad(T))] \text{ (from eq 8)} \\
&\geq \alpha\mathbb{E}_T[p_0(T)q_0(T)] - 2^{-\Omega(n)}] \\
&\geq \alpha Pr[(X_0, Y_0) \in R] - 2^{-\Omega(n)}
\end{aligned}$$

□

Now, we prove the lemma required in the proof of the theorem.

Lemma 1.5. *For any $R = \mathbb{X} \times \mathbb{Y}$, where $\mathbb{X}, \mathbb{Y} \subseteq 2^{[n]}$,*

$$\mu(B \cap R) \geq \alpha\mu(A \cap R) - 2^{-\delta n} \tag{14}$$

for some constants $k, \delta > 0$.

Proof. This lemma follows from the previous lemma (1.2) by just observing that

$$\begin{aligned}
\mu(B \cap R) &= \frac{1}{4}\mathbb{E}_T[p_1(T)q_1(T)] \\
\mu(A \cap R) &= \frac{3}{4}\mathbb{E}_T[p_0(T)q_0(T)]
\end{aligned}$$

We prove just the first equation. The second equation follows along similar lines.

$$\begin{aligned}\mu(B \cap R) &= \mu(B)\mu(R|B) \\ &= \frac{1}{4}\sum T Pr[T]Pr[X \in \mathbb{X}|T = t, i \in X]Pr[Y \in \mathbb{Y}|T = t, i \in Y] \\ &= \frac{1}{4}\mathbb{E}_T[p_1(T)q_1(T)]\end{aligned}$$

□

2. Forster's lower bound

In this section, we present Forster's lower bound for the *Unbounded Error Probabilistic Communication Complexity* by showing a lower bound on *the sign rank* of the corresponding communication matrix, if it has a low spectral norm. In [6], it was shown that the the communication complexity (C_f) of a distributed function f is closely related to *sign rank* (say k) (2.1) of the communication matrix as

$$\lceil \log_2 k \rceil \leq C_f \leq \lceil \log_2 k \rceil + 1 \tag{15}$$

For a distributed function $f : \{1, 2, \ldots, n\}^2 \to \{-1, 1\}$ represented by the matrix $M(f) \in \{-1, 1\}^{n \times n}$, we define the sign rank of the corresponding matrix.

Definition 2.1. *For a matrix $M \in \{-1, 1\}^{n \times n}$, we say that the $signrank(M) \leq k$ iff there exists $A \in \mathbb{R}^{n \times n}$ of rank $\leq k$ such that $M_{i,j} = sign(A_{i,j})$ (where $sign(x)$ is the usual sign function defined on $\mathbb{R}$). Equivalently, there exists $X = \{x_1, x_2, \ldots, x_n\}, Y = \{y_1, y_2, \ldots, y_n\} \subseteq \mathbb{R}^k$ such that $M_{i,j} = sign(< x_i, y_j >)$, where $< a, b >$ represents the inner product of vectors a and b.*

We now present and prove Forster's theorem [2]

Theorem 2.2. *$signrank(M) \geq \frac{n}{||M||}$ where $||A||$ represents the spectral norm of the matrix A.*

Note that this theorem along with theorem 15 implies that for any distributed function $f : \{0, 1\}^m \times \{0, 1\}^m \to \{0, 1\}$, the communication complexity

$$C_f \geq m - \log_2 ||M_f||$$

where M_f represents the corresponding communication matrix. For the sake of notation, for vectors $x \in X$ and $y \in Y$, we refer to the corresponding entry in M by $M_{x,y}$.

Proof. The proof follows in two steps. The first is in establishing a connection between a relaxation of the Discrepancy($disc(M)$) and the spectral norm $||M||$. The second part of the proof lower bounds $disc$ by using Lemma 2.3, which forms the crux of the proof. We now state the lemma (proved in the special note by David Steurer)

Lemma 2.3. *For every $X \subseteq \mathbb{R}^k$ such that $|X| = n$, such that all subsets of X with at most k elements are linearly independent, there exists a linear map $A \in GL(k)$ such that*

$$\sum_{x \in X} \frac{1}{||Ax||^2}(Ax)(Ax)^T = \frac{n}{k} I_k$$

We first define $disc(M)$.

$$disc(M) = \max_{\substack{X=\{x \mid ||x||=1\} \\ Y=\{y \mid ||y||=1\} \\ |Y|=|X|=n}} \sum_{x \in X, y \in Y} M_{x,y} < x, y >$$

This is related to within a constant factor of the earlier definition of discrepancy by the Groethendieck's inequality ([1]). Now, we establish the following relation $disc \leq n||M||$.

$$||M|| = \max_{||u||=1, ||v||=1} < u, Mv > \tag{16}$$

We now show that

$$||M|| = \max_{\substack{\sum_x ||x||^2=1 \\ \sum_y ||y||^2=1}} M_{x,y} < x, y > \tag{17}$$

Clearly, $LHS \leq RHS$. We now show the other direction by apply Cauchy-Schwartz inequality twice.

$$\begin{aligned}
\max_{\substack{\sum_x ||x||^2=1 \\ \sum_y ||y||^2=1}} M_{x,y} < x, y > &\leq \sum_i \sum_{\substack{\sum_x ||x||^2=1 \\ \sum_y ||y||^2=1}} M_{x,y} x_i y_i \\
&\leq \sum_i ||M|| \sqrt{(\sum_x x_i^2)(\sum_y y_i^2)} \\
&\leq ||M|| \sum_i \sqrt{\sum_x x_i^2} \sqrt{\sum_y y_i^2} \\
&\leq ||M|| \sqrt{\sum_i \sum_x x_i^2} \sqrt{\sum_i \sum_y y_i^2} \\
&\leq ||M||
\end{aligned}$$

Finally, we proceed to prove the theorem. M has sign rank k, and assume that the corresponding unit vectors are $x \in X$ and $y \in Y$.

$$\begin{aligned}
\sum_{x,y} M_{x,y} < x, y > &= \sum_{x,y} | < x, y > | \\
&\geq \sum_{x,y} < x, y >^2 \\
&= \sum_y Y^T (\sum_x x x^T) Y \\
&= \frac{n^2}{k} \text{ from Lemma2.3} \\
\text{Hence } n||M|| &\geq disc \\
&\geq \frac{n^2}{k}
\end{aligned}$$

Thus, $||M|| \geq \frac{n}{k}$. □

In particular, this result by Forster also resolved a long-standing conjecture of [6, 4] in showing that the unbounding error probabilistic communication complexity of the distributed function given by the Hadamard matrix is linear ($\geq \frac{n}{2}$).

References

[1] N. Alon and A. Naor. Approximating the cut-norm via grothendieck's inequality. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, New York, NY, USA, 2004. ACM.

[2] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[3] B. Kalyanasundaram and G. Schintger. The probabilistic communication complexity of set intersection. *SIAM J. Discret. Math.*, 5(4):545–557, 1992.

[4] M. Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theor. Comput. Sci.*, 156(1-2):99–117, 1996.

[5] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, New York, NY, USA, 2006.

[6] R. Paturi and J. Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[7] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[8] A. A. Razborov. On the distributional complexity of disjointness. In *Proceedings of the seventeenth international colloquium on Automata, languages and programming*, pages 249–253, New York, NY, USA, 1990. Springer-Verlag New York, Inc.