CS 2429 - Foundations of Communication Complexity

Lecture #5: 13 October 2006

Lecturer: Toniann Pitassi and Avner Magen

Scribe Notes by: Daniel Fryer

1 Discrepancy and Duality of Sign Degree

Theorem 1 (Duality of sign degree) *Let* $f : \{-1,1\}^n$ $d \geq 0$

Then sign-deg(f) d iff $\exists$ *a distribution* μ *over* $\{-1,1\}^n$ *s.t.*

$E_{x\sim\mu}\left[f(x) \cdot \chi_S(x)\right] = 0 \ \forall S,\ |S| < d$

That is to say, "f is orthogonal to χ_S for small s", where χ_S is the parity function over the indices in S

Theorem 2 (Duality of approximation degree) *(Sherstov, Shi-Zhu)*
Fix $\varepsilon \geq 0$. *Let* $f : \{-1,1\}^n \to \{-1,1\}$, $deg_\varepsilon(f) = d \geq 1$.
Then $\exists g : \{-1,1\}^n \to \{-1,1\}$ *and a distribution* μ *over* $\{-1,1\}^n$ *such that:*

$$(1) \quad E_{x\sim\mu}\left[g(x)\chi_S(x)\right] = 0 \quad \forall S \quad |S| \leq d$$

$$(2) \quad corr_\mu(f,g) > \varepsilon \qquad (corr_\mu(f,g) = E_{x\sim\mu}[f(x)g(x)])$$

Proof (Duality of sign degree) This is an instance of the "Gordon Transposition Lemma"
Let A be a matrix of dimension $m \times n$. Then $\exists \vec{u}$ s.t. $\vec{u}^T A > 0$ iff $\exists \vec{v} > 0$ s.t. $A\vec{v} = 0$

We want a polynomial f' which sign-approximates f. We look for coefficients α_s, $|S| < d$ to produce $f' = \sum_S \alpha_s \chi_s$

Fix ρ. If $f(\rho) = 1$ $\sum_S \alpha_s \chi_s > 0$, and if $f(\rho) = -1$ $\sum_S \alpha_s \chi_s < 0$. So, $\sum \alpha_s \chi_s f(\rho) > 0$, that is to say, they match in sign.

We construct a matrix with columns representing values for rho and rows representing values for s, that is, subsets of $1..n$ of size $\leq d$. For each value we fill in $\chi_s(\rho)f(\rho)$. Then the rows of our matrix are the values for α_s, which is $\vec{u}^T$ in the above lemma, and $\vec{v}$ is a distribution over our columns.

Using duality of sign degree we can prove 2-party communication complexity lower bounds.

(1) We start with a base function $f : \{-1,1\}^n$ with large sign degree d. For example, $f(x) = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} x_{ij}$ has sign-degree m, or the parity function, with sign degree n.

(2) Use the pattern matrix method to construct a 2-player CC problem $F(\bar{x}, \bar{y})$ $|\bar{x}| = N$ and $|\bar{y}| = \log \binom{N}{n}$, $N = O(n^k)$. $F(\bar{x}, \bar{y}) = f(\bar{(x)}|_{\bar{y}})$, which is read "f of $\bar{x}$, restricted to the bits specified by $\bar{y}$"

(3) Use duality and BNS upper bound for discrepancy to show that there exists a distribution λ such that $F(\bar{x}, \bar{y})$ has 2^{-d} discrepancy w.r.t λ, for appropriate N.

Theorem 3 *Let f be boolean over $x_1..x_n$ with sign degree $\geq d$.*
Then $disc(F) \leq (\frac{4en^2}{Nd})^{\frac{d}{2}}$ where e has its usual meaning as the base of the natural logarithm.

We set $N = \frac{16en^2}{d}$ so that $disc \leq 2^{-d}$. See Sherstov, Seperating AC^0 from depth-2 majority circuits, and Sherstov, Pattern Matrix Method.

Proof BNS Lemma: $F(X \times Y) \to \{-1,1\}$ $|X| = 2^N$ $|Y| = 2^N$

$$disc_\lambda(F)^2 \leq 4^N \sum_{y,y' \in Y} \left| \sum_{x \in X} \lambda(x,y)\lambda(x,y')F(x,y)F(x,y') \right|$$

We rename y, y' S and T. λ is a distribution on $X \times Y$ induced by μ. To obtain λ we pick $y \in Y$ uniformly at random. We choose $x|_S$ according to μ. Then we set the rest of the bits of x uniformly at random.

By the above lemma,

$$disc_\Pi(\mu)^2 \leq (*) 4^n E_{S,T \sim U} |\Gamma(S,T)|$$

where

$$\Gamma(S,T) = E_{x \sim U} \left[\mu(x|_S)\mu(x|_T)f(x|_S)f(x|_T) \right]$$

Claim 1 When $|S \cap T| \leq d-1$ then $\Gamma(S,T) = 0$.

Claim 2 When $|S \cap T| = i$, $|\Gamma(S,T)| \leq 2^{i-2}$.

By these claims,

$$(*) \leq \sum_{k=d}^{n} 2^k Pr\left[|S \cap T| = k\right]$$

$$Pr\left[|S \cap T| = k\right] = \frac{\binom{n}{k}\binom{N-n}{n-k}}{\binom{N}{n}} \leq \left(\frac{en^2}{Nk}\right)^k$$

$$disc_\lambda(F)^2 \leq \sum_{k=d}^{n} 2^k \left(\frac{en^2}{Nk}\right)^k = \sum_{k=d}^{n} \left(\frac{2en^2}{Nk}\right)^k \leq \left(\frac{4en^2}{Nd}\right)^k$$

by magic.

Proof of Claim 1 Proving that when $|S \cap T| \leq d-1$ then $\Gamma(S,T) = 0$. Let S be $x_1 ... x_n$

$$\Gamma(S,T) = E_x \left[\mu(x_1...x_n) f(x_1...x_n) \mu(x|_T) f(x|_T)\right]$$

$$\Gamma(S,T) = 2^{\frac{1}{N}} \sum_{x_1..x_n} \mu(x_1..x_n) f(x_1..x_n) \sum_{x_{n+1}..x_N} \mu(x|_T) f(x|_T)$$

$$\Gamma(S,T) = 2^{\frac{1}{N}} E_{x_1..x_n \sim \mu} f(x|_{x_1..x_n}) \left[\sum_{x_{n+1}..x_N} \mu(x|_T) f(x|_T)\right]$$

$\sum_{x_{n+1}..x_N} \mu(x|_T) f(x|_T)$ depends on $\leq d$ bits, so

$$\Gamma(S,T) = 0$$

Proof of Claim 2 When $|S \cap T| = i$, $|\Gamma(S,T)| \leq 2^{i-2}$

$$|\Gamma(S,T)| = E_{x_1..x_n} \left[\mu(x_1..x_n)\right] \cdot \max_{x_1..x_n} E_{x_{n+1}..x_{2n-i}} \left[\mu(x_1..x_i x_{n+1}..x_{2n-i})\right]$$

where we assume that $f(x_1..x_i x_{n+1}..x_{2n-i}) = 1$ because we're searching for a maximal value. $E_{x_1..x_n} \left[\mu(x_1..x_n)\right] = 2^{-n}$ and $E_{x_{n+1}..x_{2n-i}} \left[\mu(x_1..x_i x_{n+1}..x_{2n-i})\right] \leq 2^{-n-i}$ so

$$|\Gamma(S,T)| = 2^{i-2n}$$

2 Application to Circuits

Allender '89 Any AC^0 function can be computed by a depth-3 majority circuit of quasipolynomial ($O(n^{polylog(n)})$ size.

(Formerly) open question - Can this be improved? Can every function in AC^0 be computed by depth-2 majority-of-threshold circuits of quasipolynomial size?

Theorem 4 (Sherstov) *$\exists F \in AC^0_3$ (depth 3) whose computation requires majority of exponentially many threshold gates.*

It suffices to show an AC^0 function with exponentially small discrepancy. We start with the AC^0_2 function:

$$f = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} e_{ij}$$

We construct F(x,y) where $F(x,y) = f(x|_y)$, that is, f of the bits of x specified by y. F(x,y) is in AC^0_3:

$$F(x,y) = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} \bigvee_{\alpha} (y_{ij\alpha_1} \wedge y_{ij\alpha_2} \wedge ... \wedge y_{ij\alpha_q} \wedge x_{ij\alpha})$$

because we can swap the order of the $\wedge$'s within the brackets with the last $\bigvee$ and then merge them with the middle $\bigwedge$.

By the degree/discrepancy theorem we know that because f requires a high degree polynomial to compute, F(x,y) has low discrepancy. Each threshold gate can be computed by a $O(\log n)$ bit probabilistic CC protocol with $R_{\epsilon}^{pub}(f) = O(\log n + \log \frac{1}{\epsilon})$.

Suppose F has (low) discrepancy $e^{-N^{\varepsilon}}$. Then any randomized protocol requires N^{ε} bits. Also let $F = MAJ(h_1..h_S)$ where each h_i is a threshold circuit.

The players pick a random $i \in [S]$. They evaluate h_i, using $O(\log n)$ bits and output the result.

The probability of correctness of the threshold-computing protocol is $1 - \frac{1}{4S}$ if we set $\varepsilon' \sim \frac{1}{S}$.

The total cost is $O(\log n) + \log S$ bits. The probability of correctness is $(\frac{1}{2} + \frac{1}{2S}) - \frac{1}{4S} = \frac{1}{2} + \frac{1}{4S}$ on every input.

Since we know that F requires $O(N^{\varepsilon})$ bits to compute, S must be exponentially large! And so there is no polynomially-sized majority-of-threshold circuit to compute $F \in AC_3^0$.