Lecture 06

Preliminaries

☞ $f : \{1,-1\}^n \to \{1,-1\}$, $\rho \sim \{1,-1\}^n$, $F : \mathcal{X} \times \mathcal{Y} \to \{1,-1\}$, $\mu \sim \mathcal{X} \times \mathcal{Y}$.

1 *Communication complexity.* $R_\varepsilon(F)$, $D^\mu_\varepsilon(F)$.

2 *Yao's Principle.* $R_\varepsilon(F) = \max_\mu D^\mu_\varepsilon(F)$

☞ A rectangle R is a product set $X \times Y$ where $X \subseteq \mathcal{X}, Y \subseteq \mathcal{Y}$.

3 *Discrepancy.* Discrepancy of F w.r.t. μ is the maximum μ-weighted F-bias over any rectangle:

$$\mathsf{Disc}_\mu(F) = \max_R \left| \sum_{(x,y)\in R} \mu(x,y)F(x,y) \right|$$

4 Theorem (Small Discrepancy $\implies$ Big Communication). For all μ, F and $\gamma > 0$:

$$R_{1/2-\gamma/2}(F) \geq D^\mu_{1/2-\gamma/2}(F) \geq \log \frac{\gamma}{\mathsf{Disc}_\mu(F)}$$

Proof. Let π be some deterministic protocol to compute F on μ with error $\varepsilon \leq 1/2 - \gamma/2$ using C bits of communication. Then

$$\gamma \leq \Pr[\pi \text{ correct}] - \Pr[\pi \text{ wrong}] = \sum_{x,y} \mu(x,y)\pi(x,y)F(x,y)$$

Let R range over the rectangles that fix the transcript of π. In absolute value, the right-hand side becomes

$$\left| \sum_R \pi(R) \sum_{(x,y)\in R} \mu(x,y)F(x,y) \right| \leq 2^C \mathsf{Disc}_\mu(F). \qquad \blacksquare$$

Relating discrepancy and MAJ ∘ THR circuits

5 Theorem. Let $g : \{1,-1\}^n \to \{1,-1\}$ be a linear threshold function. Then $R_\varepsilon(f) = O(\log n \cdot \log \frac{n}{\varepsilon})$, for any partition of the variables and any $\varepsilon = \varepsilon(n)$.

Proof. $R_\varepsilon(\mathsf{EQ_n}) \leq 2 \log \frac{1}{\varepsilon}$. Why? Alice and Bob can use shared randomness to pick a subset of the input coordinates, and send each other the XOR of the bits at those coordinates. If $x = y$, then the two XORs will be the same with probability 1, and if $x \neq y$, then the two XORs will be different with probability 1/2 (the probability that the common subset picks an odd number of different coordinates is 1/2). Do this $\log \frac{1}{\varepsilon}$ times, and the success probability becomes $1 - \varepsilon$.

This implies that $R_\varepsilon(\mathsf{GEQ}_n) = O(\log n \cdot \log \frac{n}{\varepsilon})$. To know the minimum of their two n-bit positive integers x and y, Alice and Bob use the equality protocol on prefixes of x and y to find, via binary search, the first index where x and y differ. They need to do $\log n$ rounds, and they make sure that the error in each round is at most $\varepsilon/\log n$, so that the total error union-bounds to at most ε. We're being careless, so we use error ε/n instead.

To compute $g(x,y) = sign(\sum_i a_i x_i + \sum_i b_i y_i - \theta)$, where a_i, b_i are n-bit coefficients, Alice first computes $x' = \sum_i a_i x_i - \theta + 2^{n+1}$, and Bob computes $y' = \sum_i b_i y_i + 2^{n+1}$; x' and y' have $\leq 2n$ bits. Then they use the previous protocol for GEQ_{2n}, and accept iff $x' \geq y$. ■

☞ If F is the majority of s linear threshold functions, then for any input x, y, the probability that a random threshold function (among the s) correctly computes F is at least $1/2 + 1/2s$.[1]. A two-player protocol that picks a random g among the s-many, and uses the previous protocol to output $g(x,y)$ with $\varepsilon = 1/4s$ error.

6 Corollary. This implies that

$$R_{1/2-1/4s}(F) = O(\log n \cdot \log(sn)).$$

7 Theorem. If $\mathsf{Disc}_\mu(F) \leq 2^V$, for $v \gg (\log n)^2$, then any majority of s linear threshold functions to compute F must have $s = \frac{2^{\Omega(V)}}{n}$.

[1] This is easy to see if s is odd; if s is even, and the condition does not hold, we may replace the least predictive threshold function with 0 or 1 (effectively making s odd) while still having the majority compute the same function.

Proof. If $\mathsf{Disc}_\mu(F) \leq 2^{-V}$, then by §4, §6 and §2:

$$V - \log 2s \leq D^\mu_{1/2-1/4s}(F) = O(\log n \cdot \log sn),$$

and so $s = 2^{\Omega(V)}/n$, as stated. ■

Our aim for this part of the course is to prove the following

8 Theorem (Degree/Discrepancy theorem, roughly speaking). We can take an AC^0 function f having high $\mathsf{deg}_\pm(f)$, and produce from it an AC^0 function F having very small discrepancy (something like $2^{-d/2}$).

9 Theorem (Degree/Discrepancy theorem). Let $\mathsf{deg}_\pm(f) = d \geq 1$. Let $N \geq n$, and define $F(x, y) = f(x|_y)$ with $x \in \{1, -1\}^N$ and $y \in \binom{N}{n}$. Then

$$\mathsf{Disc}_\mu(F) \leq \left(\frac{4en^2}{Nd}\right)^{d/2} \qquad (\leq 2^{-d/2} \text{ for } N \geq n^2/d).$$

Gordan's Transposition Theorem

10 *Threshold Degree.* The threshold degree of f, denotd $\mathsf{deg}_\pm(f)$, is the least degree of a real polynomial $p(x_1, \ldots, x_n)$ such that $f(x) = sign(p(x))$.

If we let $\chi_S = \sum_{i \in S} x_i$, then $\mathsf{deg}_\pm(f)$ is the smallest d such that

$$f(x) = \sum_{|S| \leq d} a_S \chi_S(x)$$

for some real coefficients a_S.

11 *Gordan's Transposition Theorem.* Let $M \in \mathbb{R}^{m \times n}$. Then exactly one of the following statements must hold:

(i) For some vector u, $Mu \geq \varepsilon$ (all coordinates of Mu are positive);

(ii) $M^T v = 0$ for some non-zero vector $v \geq 0$.

11.1. We will use linear program duality (Schrijver, p. 91). Provided both sets are non-empty, it holds that:

$$\min\{\langle b, y\rangle \mid Ay \geq c\} = \max\{\langle c, x\rangle \mid x \geq 0, A^T x = b\}$$

Proof. Let e, e' be all-1 vectors of suitable dimension. Consider the following dual pair of linear programs:

$$\mathrm{P}\left\{\begin{array}{lrcl} \min & \langle e, z\rangle & & \\ \text{s.t.} & Mu + z & \geq & e' \\ & z & \geq & 0 \end{array}\right. \qquad \mathrm{Q}\left\{\begin{array}{lrcl} \max & \langle e', v\rangle & & \\ \text{s.t.} & M^T v & = & 0 \\ & v & \leq & e \\ & v & \geq & 0 \end{array}\right.$$

Clearly both P and Q are feasible; $z = e'$ satisfies P, and $v = 0$ satisfies Q. Hence both programs have solutions, regardless of M.

Now suppose that for any u, some coordinate of Mu is ≤ 0. Then any z satisfying the constraints of P must have some coordinate > 0. This is also true for the optimum solution u^*, z^* of P, and hence the optimum value of P is > 0. But this equals the optimum value of Q, and so the optimum solution $v = v^*$ of Q must be non-zero (and have $M^T v = 0$).

Now suppose instead that some u has $Mu \geq \varepsilon$; then we may assume $\varepsilon = 2$ by multiplying u with a sufficiently large number. Then $z = 0$ is a solution to P, and so 0 is the optimum value of both P and Q. But that implies that $v = 0$ is the only solution to Q. ■

12 Corollary. For any f, d, exactly one of the following holds:

(i) $\mathsf{deg}_{\pm}(f) \leq d$;

(ii) there exists ρ such that $\sum_x \rho(x) f(x) \chi_S(x) = 0$ whenever $|S| \leq d$.

☞ I.e., either f can be computed by low degree polynomials or, with respect to some ρ, f has zero correlation with low-degree polynomials!

Proof. Let $M[x, S] = f(x)\chi_S(x)$ for any S with $|S| \leq d$, and any $x \in \{1, -1\}^n$. Refer to §11 with respect to this M. If case (i) holds there, then $(Mu)_x > 0$ is positive for every x, for some vector $u = (u_S)$; meaning, the polynomial $p_u(x) = \sum_S u_S \chi_S(x)$ has the same sign as $f(x)$ for every x, and so $\mathsf{deg}_{\pm}(f) \leq d$.

If case (ii) holds, then some non-zero $v = (v_x) \geq 0$ gives $(M^T v)_S = 0$ for every S, which means $\sum_x v_x f(x) \chi_S(x) = 0$. Then $\rho(x) = v_x / \|v\|_1$. ■