Special Topics in Complexity Theory, Fall 2017. Instructor: Emanuele Viola

1 Lecture 8-9, Scribe: Xuangui Huang

In these lectures, we finish the proof of the approximate degree lower bound for AND-OR function, then we move to the surjectivity function SURJ. Finally we discuss quasirandom groups.

1.1 Lower Bound of $d_{1/3}$(AND-OR)

Recall from the last lecture that AND-OR: $\{0,1\}^{R\times N} \to \{0,1\}$ is the composition of the AND function on R bits and the OR function on N bits. We also proved the following lemma.

Lemma 1. Suppose that distributions A^0, A^1 over $\{0,1\}^{n_A}$ are k_A-wise indistinguishable distributions; and distributions B^0, B^1 over $\{0,1\}^{n_B}$ are k_B-wise indistinguishable distributions. Define C^0, C^1 over $\{0,1\}^{n_A \cdot n_B}$ as follows:

C^b: draw a sample $x \in \{0,1\}^{n_A}$ from A^b, and replace each bit x_i by a sample of B^{x_i} (independently).

Then C^0 and C^1 are $k_A \cdot k_B$-wise indistinguishable.

To finish the proof of the lower bound on the approximate degree of the AND-OR function, it remains to see that AND-OR can distinguish well the distributions C^0 and C^1. For this, we begin with observing that we can assume without loss of generality that the distributions have disjoint supports.

Claim 2. For any function f, and for any k-wise indistinguishable distributions A^0 and A^1, if f can distinguish A^0 and A^1 with probability ϵ then there are distributions B^0 and B^1 with the same properties (k-wise indistinguishability yet distinguishable by f) and also with disjoint supports. (By disjoint support we mean for any x either $\Pr[B^0 = x] = 0$ or $\Pr[B^1 = x] = 0$.)

Proof. Let distribution C be the "common part" of A^0 and A^1. That is to say, we define C such that $\Pr[C = x] := \min\{\Pr[A^0 = x], \Pr[A^1 = x]\}$ multiplied by some constant that normalize C into a distribution.

Then we can write A^0 and A^1 as

$$A^0 = pC + (1-p)B^0\,,$$
$$A^1 = pC + (1-p)B^1\,,$$

where $p \in [0,1]$, B^0 and B^1 are two distributions. Clearly B^0 and B^1 have disjoint supports.

Then we have

$$\begin{aligned}\mathbb{E}[f(A^0)] - \mathbb{E}[f(A^1)] &= p\mathbb{E}[f(C)] + (1-p)\mathbb{E}[f(B^0)] \\ &\quad - p\mathbb{E}[f(C)] - (1-p)\mathbb{E}[f(B^1)] \\ &= (1-p)\big(\mathbb{E}[f(B^0)] - \mathbb{E}[f(B^1)]\big) \\ &\le \mathbb{E}[f(B^0)] - \mathbb{E}[f(B^1)]\,.\end{aligned}$$

Therefore if f can distinguish A^0 and A^1 with probability ϵ then it can also distinguish B^0 and B^1 with such probability.

Similarly, for all $S \neq \varnothing$ such that $|S| \le k$, we have

$$0 = \mathbb{E}[\chi_S(A^0)] - \mathbb{E}[\chi_S(A^1)] = (1-p)\big(\mathbb{E}[\chi_S(B^0)] - \mathbb{E}[\chi_S(B^1)]\big) = 0\,.$$

Hence, B^0 and B^1 are k-wise indistinguishable. □

Equipped with the above lemma and claim, we can finally prove the following lower bound on the approximate degree of AND-OR.

Theorem 3. $d_{1/3}(\text{AND-OR}) = \Omega(\sqrt{RN})$.

Proof. Let A^0, A^1 be $\Omega(\sqrt{R})$-wise indistinguishable distributions for AND with advantage 0.99, i.e. $\Pr[\text{AND}(A^1) = 1] > \Pr[\text{AND}(A^0) = 1] + 0.99$. Let B^0, B^1 be $\Omega(\sqrt{N})$-wise indistinguishable distributions for OR with advantage 0.99. By the above claim, we can assume that A^0, A^1 have disjoint supports, and the same for B^0, B^1. Compose them by the lemma, getting $\Omega(\sqrt{RN})$-wise indistinguishable distributions C^0, C^1. We now show that AND-OR can distinguish C^0, C^1:

- C_0: First sample A^0. As there exists a unique $x = 1^R$ such that $\text{AND}(x) = 1$, $\Pr[A^1 = 1^R] > 0$. Thus by disjointness of support $\Pr[A^0 = 1^R] = 0$. Therefore when sampling A^0 we always get a string with at least one "0". But then "0" is replaced with sample from B^0. We have $\Pr[B^0 = 0^N] \ge 0.99$, and when $B^0 = 0^N$, AND-OR$= 0$.

- C_1: First sample A^1, and we know that $A^1 = 1^R$ with probability at least 0.99. Each bit "1" is replaced by a sample from B^1, and we know that $\Pr[B^1 = 0^N] = 0$ by disjointness of support. Then AND-OR$= 1$.

Therefore we have $d_{1/3}(\text{AND-OR}) = \Omega(\sqrt{RN})$. □

1.2 Lower Bound of $d_{1/3}$(SURJ)

In this subsection we discuss the approximate degree of the surjectivity function. This function is defined as follows.

Definition 4. The surjectivity function SURJ: $\left(\{0,1\}^{\log R}\right)^N \to \{0,1\}$, which takes input $(x_1, \ldots, x_N)$ where $x_i \in [R]$ for all i, has value 1 if and only if $\forall j \in [R], \exists i \colon x_i = j$.

First, some history. Aaronson first proved that the approximate degree of SURJ and other functions on n bits including "the collision problem" is $n^{\Omega(1)}$. This was motivated by an application in quantum computing. Before this result, even a lower bound of $\omega(1)$ had not been known. Later Shi improved the lower bound to $n^{2/3}$, see [AS04]. The instructor believes that the quantum framework may have blocked some people from studying this problem, though it may have very well attracted others. Recently Bun and Thaler [BT17] reproved the $n^{2/3}$ lower bound, but in a quantum-free paper, and introducing some different intuition. Soon after, together with Kothari, they proved [BKT17] that the approximate degree of SURJ is $\Theta(n^{3/4})$.

We shall now prove the $\Omega(n^{3/4})$ lower bound, though one piece is only sketched. Again we present some things in a different way from the papers.

For the proof, we consider the AND-OR function under the promise that the Hamming weight of the RN input bits is at most N. Call the approximate degree of AND-OR under this promise $d_{1/3}^{\leq N}(\text{AND-OR})$. Then we can prove the following theorems.

Theorem 5. $d_{1/3}(\text{SURJ}) \geq d_{1/3}^{\leq N}(\text{AND-OR})$.

Theorem 6. $d_{1/3}^{\leq N}(\text{AND-OR}) \geq \Omega(N^{3/4})$ for some suitable $R = \Theta(N)$.

In our settings, we consider $R = \Theta(N)$. Theorem 5 shows surprisingly that we can somehow "shrink" $\Theta(N^2)$ bits of input into $N \log N$ bits while maintaining the approximate degree of the function, under some promise. Without this promise, we just showed in the last subsection that the ap-

proximate degree of AND-OR is $\Omega(N)$ instead of $\Omega(N^{3/4})$ as in Theorem 6.

Proof of Theorem 5. Define an $N \times R$ matrix Y s.t. the 0/1 variable y_{ij} is the entry in the i-th row j-th column, and $y_{ij} = 1$ iff $x_i = j$. We can prove this theorem in following steps:

1. $d_{1/3}(\text{SURJ}(\overline{x})) \geq d_{1/3}(\text{AND-OR}(\overline{y}))$ under the promise that each row has weight 1;
2. let z_j be the sum of the j-th column, then $d_{1/3}(\text{AND-OR}(\overline{y}))$ under the promise that each row has weight 1, is at least $d_{1/3}(\text{AND-OR}(\overline{z}))$ under the promise that $\sum_j z_j = N$;
3. $d_{1/3}(\text{AND-OR}(\overline{z}))$ under the promise that $\sum_j z_j = N$, is at least $d_{1/3}^{=N}(\text{AND-OR}(\overline{y}))$;
4. we can change "$= N$" into "$\leq N$".

Now we prove this theorem step by step.

1. Let $P(x_1, \ldots, x_N)$ be a polynomial for SURJ, where $x_i = (x_i)_1, \ldots, (x_i)_{\log R}$. Then we have
$$(x_i)_k = \sum_{j:k\text{-th bit of } j \text{ is } 1} y_{ij}.$$
Then the polynomial $P'(\overline{y})$ for AND-OR$(\overline{y})$ is the polynomial $P(\overline{x})$ with $(x_i)_k$ replaced as above, thus the degree won't increase. Correctness follows by the promise.

2. This is the most extraordinary step, due to Ambainis [Amb05]. In this notation, AND-OR becomes the indicator function of $\forall j, z_j \neq 0$. Define
$$Q(z_1, \ldots, z_R) := \mathop{\mathbb{E}}_{\substack{\overline{y}\text{: his rows have weight 1} \\ \text{and is consistent with } \overline{z}}} P(\overline{y}).$$
Clearly it is a good approximation of AND-OR$(\overline{z})$. It remains to show that it's a polynomial of degree k in z's if P is a polynomial of degree k in y's.

Let's look at one monomial of degree k in P: $y_{i_1 j_1} y_{i_2 j_2} \cdots y_{i_k j_k}$. Observe that all i_ℓ's are distinct by the promise, and by $u^2 = u$ over $\{0, 1\}$. By chain rule we have

$$\mathbb{E}[y_{i_1 j_1} \cdots y_{i_k j_k}] = \mathbb{E}[y_{i_1 j_1}]\mathbb{E}[y_{i_2 j_2} | y_{i_1 j_1} = 1] \cdots \mathbb{E}[y_{i_k j_k} | y_{i_1 j_1} = \cdots = y_{i_{k-1} j_{k-1}} = 1].$$

By symmetry we have $\mathbb{E}[y_{i_1 j_1}] = \frac{z_{j_1}}{N}$, which is linear in z's. To get $\mathbb{E}[y_{i_2 j_2} | y_{i_1 j_1} = 1]$, we know that every other entry in row i_1 is 0, so we give away row i_1, average over y's such that $\begin{cases} y_{i_1 j_1} = 1 \\ y_{ij} = 0 \quad j \neq j_1 \end{cases}$ under the promise and consistent with z's. Therefore

$$\mathbb{E}[y_{i_2 j_2} | y_{i_1 j_1} = 1] = \begin{cases} \frac{z_{j_2}}{N-1} & j_1 \neq j_2, \\ \frac{z_{j_2}-1}{N-1} & j_1 = j_2. \end{cases}$$

In general we have

$$\mathbb{E}[y_{i_k j_k} | y_{i_1 j_1} = \cdots = y_{i_{k-1} j_{k-1}} = 1] = \frac{z_{j_k} - \#\ell < k \colon j_\ell = j_k}{N - k + 1},$$

which has degree 1 in z's. Therefore the degree of Q is not larger than that of P.

3. Note that $\forall j$, $z_j = \sum_i y_{ij}$. Hence by replacing z's by y's, the degree won't increase.

4. We can add a "slack" variable z_0, or equivalently $y_{01}, \ldots, y_{0N}$; then the condition $\sum_{j=0}^{R} z_j = N$ actually means $\sum_{j=1}^{R} z_j \leq N$.

□

Proof idea for Theorem 6. First, by the duality argument we can verify that $d_{1/3}^{\leq N}(f) \geq d$ if and only if there exists d-wise indistinguishable distributions A, B such that:

- f can distinguish A, B;
- A and B are supported on strings of weight $\leq N$.

Claim 7. $d_{1/3}^{\leq\sqrt{N}}(\mathrm{OR}_N) = \Omega(N^{1/4})$.

The proof needs a little more information about the weight distribution of the indistinguishable distributions corresponding to this claim. Basically, their expected weight is very small.

Now we combine these distributions with the usual ones for And using the lemma mentioned at the beginning.

What remains to show is that the final distribution is supported on Hamming weight $\leq N$. Because by construction the R copies of the distributions for Or are sampled independently, we can use concentration of measure to prove a tail bound. This gives that all but an exponentially small measure of the distribution is supported on strings of weight $\leq N$. The final step of the proof consists of slightly tweaking the distributions to make that measure 0. □

1.3 Groups

Groups have many applications in theoretical computer science. Barrington [Bar89] used the permutation group S_5 to prove a very surprising result, which states that the majority function can be computed efficiently using only constant bits of memory (something which was conjectured to be false). More recently, catalytic computation [BCK+14] shows that if we have a lot of memory, but it's full with junk that cannot be erased, we can still compute more than if we had little memory. We will see some interesting properties of groups in the following.

Some famous groups used in computer science are:

- $\{0,1\}^n$ with bit-wise addition;
- $\mathbb{Z}_m$ with addition mod m ;
- S_n, which are permutations of n elements;
- Wreath product $G := (\mathbb{Z}_m \times \mathbb{Z}_m) \wr \mathbb{Z}_2$, whose elements are of the form $(a,b)z$ where z is a "flip bit", with the following multiplication rules:
  - $(a,b)1 = 1(b,a)$;
  - $z \cdot z' := z + z'$ in $\mathbb{Z}_2$;
  - $(a,b) \cdot (a',b') := (a+a', b+b')$ is the $\mathbb{Z}_m \times \mathbb{Z}_m$ operation;

An example is $(5,7)1 \cdot (2,1)1 = (5,7)1 \cdot 1(1,2) = (6,9)0$. Generally we have

$$(a,b)z \cdot (a',b')z' = \begin{cases} (a+a', b+b')z + z' & z = 1\,, \\ (a+b', b+a')z + z' & z = 0\,; \end{cases}$$

- $SL_2(q) := \{2 \times 2 \text{ matrices over } \mathbb{F}_q \text{ with determinant } 1\}$, in other words, group of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $ad - bc = 1$.

The group $SL_2(q)$ was invented by Galois. (If you haven't, read his biography on wikipedia.)

Quiz. Among these groups, which is the "least abelian"? The latter can be defined in several ways. We focus on this: If we have two high-entropy distributions X, Y over G, does $X \cdot Y$ has more entropy? For example, if X and Y are uniform over some $\Omega(|G|)$ elements, is $X \cdot Y$ close to uniform over G? By "close to" we mean that the statistical distance is less that a small constant from the uniform distribution. For $G = (\{0,1\}^n, +)$, if $Y = X$ uniform over $\{0\} \times \{0,1\}^{n-1}$, then $X \cdot Y$ is the same, so there is not entropy increase even though X and Y are uniform on half the elements.

Definition 8.[Measure of Entropy] For $\|A\|_2 = (\sum_x A(x)^2)^{\frac{1}{2}}$, we think of $\|A\|_2^2 = 100\frac{1}{|G|}$ for "high entropy".

Note that $\|A\|_2^2$ is exactly the "collision probability", i.e. $\Pr[A = A']$. We will consider the entropy of the uniform distribution U as very small, i.e. $\|U\|_2^2 = \frac{1}{|G|} \approx \|\overline{0}\|_2^2$. Then we have

$$\begin{aligned} \|A - U\|_2^2 &= \sum_x \left(A(x) - \frac{1}{|G|} \right)^2 \\ &= \sum_x A(x)^2 - 2A(x)\frac{1}{|G|} + \frac{1}{|G|^2} \\ &= \|A\|_2^2 - \frac{1}{|G|} \\ &= \|A\|_2^2 - \|U\|_2^2 \\ &\approx \|A\|_2^2\,. \end{aligned}$$

Theorem 9.[[Gow08], [BNP08]] If X, Y are independent over G, then

$$\|X \cdot Y - U\|_2 \le \|X\|_2 \|Y\|_2 \sqrt{\frac{|G|}{d}},$$

where d is the minimum dimension of irreducible representation of G.

By this theorem, for high entropy distributions X and Y, we get $\|X \cdot Y - U\|_2 \leq \frac{O(1)}{\sqrt{|G|d}}$, thus we have

$$\|X \cdot Y - U\|_1 \leq \sqrt{|G|}\|X \cdot Y - U\|_2 \leq \frac{O(1)}{\sqrt{d}}. \tag{1}$$

If d is large, then $X \cdot Y$ is very close to uniform. The following table shows the d's for the groups we've introduced.

G	$\{0,1\}^n$	$\mathbb{Z}_m$	$(\mathbb{Z}_m \times \mathbb{Z}_m) \wr \mathbb{Z}_2$	A_n	$SL_2(q)$
d	1	1	should be very small	$\frac{\log \lvert G\rvert}{\log\log \lvert G\rvert}$	$\lvert G\rvert^{1/3}$

Here A_n is the alternating group of even permutations. We can see that for the first groups, Equation (1) doesn't give non-trivial bounds.

But for A_n we get a non-trivial bound, and for $SL_2(q)$ we get a strong bound: we have $\|X \cdot Y - U\|_2 \leq \frac{1}{|G|^{\Omega(1)}}$.

References

[Amb05] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. of the ACM*, 51(4):595–605, 2004.

[Bar89] David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. of Computer and System Sciences*, 38(1):150–164, 1989.

[BCK+14] Harry Buhrman, Richard Cleve, Michal Koucký, Bruno Loff, and Florian Speelman. Computing with a full memory: catalytic space. In *ACM Symp. on the Theory of Computing (STOC)*, pages 857–866, 2014.

[BKT17] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *CoRR*, arXiv:1710.09079, 2017.

[BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 248–257, 2008.

[BT17] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC0. *CoRR*, abs/1703.05784, 2017.

[Gow08] W. T. Gowers. Quasirandom groups. *Combinatorics, Probability & Computing*, 17(3):363–387, 2008.