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

1 Lectures 2-3, Scribe: Tanay Mehta

In these lectures we conclude the proof that bounded independence fools the AND function, and look at the more recent result that bounded independence fools the circuit class AC^0.

1.1 Bounded Independence Fools AND

We state again the theorem from last time.

Theorem 1. Let $(X_1, \ldots, X_n)$ be a distribution over $\{0,1\}^n$ such that any k X_i are independent (but not necessarily uniform). Then, we have that

$$\left| \Pr\left[\bigwedge_{i=1}^{n} X_i = 1\right] - \prod_{i=1}^{n} \Pr[X_i = 1] \right| \leq 2^{-\Omega(k)}$$

Proof. Let D be the distribution of $(X_1, \ldots, X_n)$. Let B be the n-wise independent distribution $(Y_1, \ldots, Y_n)$ such that $\Pr[Y_i = 1] = \Pr[X_i = 1]$ for all $i \in [n]$ and the Y_i are independent. The theorem is equivalent to the following statement.

$$| \Pr_{X \leftarrow D}\left[\bigwedge_{i=1}^{n} X_i = 1\right] - \Pr_{X \leftarrow B}\left[\bigwedge_{i=1}^{n} X_i = 1\right] | \leq 2^{-\Omega(k)}$$

We will prove the above statement by the following version of the Inclusion-Exclusion principle.

1.1.1 Inclusion-Exclusion Principle

Let V be any distribution over $\{0,1\}^n$. Note that by De Morgan's laws, we have

$$\Pr\left[\bigwedge V_i = 1\right] = 1 - \Pr\left[\bigvee V_i = 0\right]$$

Let E_i be the event that $V_i = 0$. We want to bound the quantity $\Pr\left[\bigcup E_i\right]$. By looking at the Venn diagram of the events E_i, we can see that

$$\Pr\left[\bigcup E_i\right] \leq \Pr[E_1] + \cdots + \Pr[E_n] = \sum_i \Pr[E_i]$$

$$\Pr\left[\bigcup E_i\right] \geq \sum_i \Pr[E_i] - \sum_{i,j} \Pr[E_i \cap E_j]$$

$$\Pr\left[\bigcup E_i\right] \leq \sum_i \Pr[E_i] \quad - \sum_{S\subseteq[n],|S|=2} \Pr\left[\bigcap_{i\in S} E_i\right] \quad + \sum_{S\subseteq[n],|S|=3} \Pr\left[\bigcap_{i\in S} E_i\right],$$

and so on. In general, we have the following. Define

$$T_j := \sum_{S\subseteq[n],|S|=j} \Pr\left[\bigcap_{i\in S} E_i\right]$$

$$S_h := \sum_{i=1}^{h} (-1)^{i+1} T_i$$

Then, we have the bounds $\Pr\left[\bigcup E_i\right] \leq S_j$ for odd j, and $\Pr\left[\bigcup E_i\right] \geq S_j$ for even j. This fact holds for *any* distribution.

Let us return to the proof. Note that the S_h are the same for D and B up to $h = k$ because they only involve sums of ANDs of at most k events. Hence, we have that

$$\left|\Pr_D\left[\bigwedge X_i = 1\right] - \Pr_B\left[\bigwedge X_i = 1\right]\right| \leq |S_k - S_{k-1}| = |T_k|$$

where the last equality comes from the definition of S_k. Therefore, we are done if $|T_k| \leq 2^{-\Omega(k)}$. We have that

$$T_k = \sum_{S\subseteq[n],|S|=k} \Pr\left[\bigcap_{i\in S} E_i\right] = \binom{n}{k} \ \mathbb{E}_{S\subseteq[n],|S|=k}\left[\prod_{i\in S} P_i\right]$$

where $P_i := \Pr[E_i] = 1 - \Pr[X_i = 1]$. To bound the expectation we recall a useful inequality.

1.1.2 A Useful Inequality

Let $Q_1, \ldots, Q_n$ be non-negative real numbers. Then, by the AM-GM inequality, we have that

$$\frac{\sum_i Q_i}{n} \geq \left(\prod_i Q_i\right)^{1/n}.$$

Consider the following more general statement,

$$\mathbb{E}_{S\subseteq[n],|S|=1}\left[\prod_{i\in S} Q_i\right] \geq \mathbb{E}_{S\subseteq[n],|S|=2}\left[\prod_{i\in S} Q_i\right]^{1/2} \geq \cdots$$
$$\cdots \geq \mathbb{E}_{S\subseteq[n],|S|=k}\left[\prod_{i\in S} Q_i\right]^{1/k} \geq \cdots \geq \mathbb{E}_{S\subseteq[n],|S|=n}\left[\prod_{i\in S} Q_i\right]^{1/n}$$

and note that the left most term is equal to $\frac{\sum_i Q_i}{n}$, while the right most term is equal to $(\prod_i Q_i)^{1/n}$

Applying the above inequality to T_k and a common approximation for the binomial coefficient, we have that

$$T_k = \binom{n}{k}\ \mathbb{E}_{S\subseteq[n],|S|=k}\left[\prod_{i\in S} P_i\right] \leq \binom{n}{k}\sum_{i=1}^{n}\left(\frac{P_i}{n}\right)^k \leq \left(\frac{en}{k}\right)^k\left(\frac{\sum P_i}{n}\right)^k = \left(\frac{e\sum P_i}{k}\right)^k.$$

Therefore, we are done if $\sum P_i \leq \frac{k}{2e}$. Recall that $P_i = \Pr[E_i] = 1 - \Pr[X_i = 1]$. So if P_i is small then $\Pr[X_i = 1]$ is close to 1.

It remains to handle the case that $\sum P_i \geq \frac{k}{2e}$. Pick n' such that

$$\sum_{i=1}^{n'} P_i = \frac{k}{2e} \pm 1.$$

By the previous argument, the AND of the first n' is the same up to $2^{-\Omega(k)}$ for D and B. Also, for every distribution the probability of that the And of n bits is 1 is at most the probability that the And of n' bits is 1. And also,

for the n-wise independent distribution B we have

$$\begin{aligned}\Pr_B\left[\bigwedge_{i=1}^{n'} X_i = 1\right] &= \prod_{i=1}^{n'} \Pr[X_i = 1] \\ &= \prod_{i=1}^{n'} (1 - P_i) \\ &\leq \left(\frac{\sum_{i=1}^{n'}(1-P_i)}{n'}\right)^{n'} \text{ by the AM-GM inequality} \\ &\leq \left(\frac{n' - k/2e}{n'}\right)^{n'} \leq (1 - k/2en')^{n'} \leq e^{-\Omega(k)}.\end{aligned}$$

The combination of these facts concludes this case. To summarize, in this case we showed that

$$\Pr_D[\bigwedge_{i=1}^{n} X_i = 1] \leq \Pr_D[\bigwedge_{i=1}^{n'} X_i = 1].$$

as well as

$$\Pr_B[\bigwedge_{i=1}^{n} X_i = 1] \leq \Pr_B[\bigwedge_{i=1}^{n'} X_i = 1] \leq 2^{-\Omega(k)}.$$

By the choice of n' and the previous argument, we also know that $|\Pr_D[\bigwedge_{i=1}^{n'} X_i = 1] - \Pr_B[\bigwedge_{i=1}^{n'} X_i = 1]| \leq 2^{-\Omega(k)}$ and so we are done, as all quantities above are at most $2^{-\Omega(k)}$ (and at least 0). □

Remark 2. The bound is tight up to $\Omega(.)$

Proof. Let D be the distribution over $\{0,1\}^{k+1}$ as follows: $D_{1,\ldots,k} = U_k$ and $D_{k+1} = D_1 + \cdots + D_k \mod 2$. Then, D is k-wise independent. However, if k is even, then

$$\Pr[\bigwedge_{i=1}^{k+1} D_i = 1] = 0.$$

Yet, we have that

$$\Pr[\bigwedge_{i=1}^{k+1} U_i = 1] = 2^{-(k+1)}.$$

□

1.2 Bounded Independence Fools AC^0

Acknowledgement. This section is based on Amnon Ta-Shma's notes for the class 0368.4159 Expanders, Pseudorandomness and Derandomization CS dept, Tel-Aviv University, Fall 2016.

Note that a DNF on n bits can be modeled as a depth two circuit where the top layer is an OR-gate whose inputs are AND-gates, which take inputs $X_1, \ldots, X_n$ and their negations. The circuit class AC^0 can be viewed as a generalization of this to higher (but constant) depth circuits. That is, AC^0 consists of circuits using AND-gates, OR-gates, NOT-gates, and input registers. Each of the gates have unbounded fan-in (i.e. the number of input wires). The size of the circuit is defined to be the number of gates.

AC^0 is one of the most studied classes in complexity theory. AC^0 circuits of polynomial size can do many things, including adding and subtracting n-bit integers.

Conjecture 3.[Linial-Nisan[LN90]] $\log^{O(d)} s$-wise independence fools AC^0 circuits of depth d and size s.

The conjecture was open for a long time, even for in the special case $d = 2$. In 2007 a breakthrough work by Bazzi [Baz09] proved it for $d = 2$. Shortly afterwards, Razborov presented a simpler proof of Bazzi's result [Raz09], and Braverman proved the conjecture for any d with $\log^{d^2} s$-wise independence [Bra10]. Tal improved the result to $\log^{O(d)} s$ [Tal17].

Interestingly, the progress on the conjecture does not use ideas that were not around since the time of its formulation. Bottom line: if a problem is open for a long time, you should immediately attack it with existing tools.

The high-level intuition why such a result should be true is the following:

1. AC^0 is approximated by polylog degree polynomials.

2. k-wise independence fools degree-k polynomials.

Proof of (2). Let $x = (x_1, \ldots, x_n) \in \{0,1\}^n$. Let $p(x_1, \ldots, x_n)$ be a degree k polynomial over $\mathbb{R}$. Write p as

$$p(x_1, \ldots, x_n) = \sum_{M \subseteq [n], |M| \leq k} c_M \cdot x_M.$$

If D is a k-wise independent distribution on $\{0,1\}^n$, then by linearity of expectation

$$\mathbb{E}_D[P] = \sum_{M \subseteq [n], |M| \leq k} c_M \mathbb{E}_D[x_M] = \sum_{M \subseteq [n], |M| \leq k} c_M \mathbb{E}_U[x_M] = \mathbb{E}_U[P].$$

□

There are several notions of approximating AC^0 by low-degree polynomials. We now review two of them, explaining why neither of them is sufficient. Braverman showed how to cleverly combine the two methods to prove a version of (1) that's strong enough.

1.2.1 Approximation 1

Theorem 4. For all AC^0 circuits $C(x_1, \ldots, x_n)$ of size s and depth d, for all distributions D over $\{0,1\}^n$, for all ϵ, there exists a polynomial $p(x_1, \ldots, x_n)$ of degree $\log^{O(d)} s/\epsilon$ such that

$$\Pr_{x \leftarrow D}[p(x) = C(x)] \geq 1 - \epsilon.$$

The important features of this approximation are that it works under any distribution, and when the polynomial is correct it outputs a boolean value.

Similar approximations appear in many papers, going back to Razborov's paper [Raz87] (who considers polynomials modulo 2) which uses ideas from earlier still work.

Note that the polynomial p depends on the circuit C chosen, and on the distribution. This theorem is not a good enough approximation because on the ϵ fraction of inputs where the polynomial and circuit are unequal, the value of the polynomial can (and does) explode to be much greater than $1/\epsilon$. This prevents us from bounding the average of the polynomial.

Nevertheless, let us prove the above theorem.

Proof. Consider one OR-gate of fan-in s. We construct a distribution of polynomials that compute any input with high probability. This implies that there is a fixed polynomial that computes the circuit on a large fraction of the inputs by an averaging argument.

For $i = 1, 2, \ldots, \log s$. let S_i be a random subset of $[s]$ where every element is included with probability $1/2^i$, independently.

Suppose x has Hamming weight 2^j. Then, $\mathbb{E}[\sum_{n \in S_j} x_n] = 1$. And the sum can be shown to equal 1 with constant probability.

Define the approximation polynomial p to be

$$p(x) := 1 - \prod_{i=1}^{\log s} (1 - \sum_{h \in S_i} x_h)$$

Note that if x has weight $w > 0$, then $p(x) = 0$ with constant probability. If $w = 0$, then $p(x) = 1$ with probability 1. We can adjust the error probability to ϵ by repeating each term in the product $\log(1/\epsilon)$ times.

Thus, we can approximate one gate with the above polynomial of degree $O(\log(s) \cdot \log(1/\epsilon))$. Construct polynomials as p above for each gate, with error parameter ϵ/s. The probability that any of them is wrong is at most ϵ by a union bound. To obtain the approximating polynomial for the whole circuit compose all the polynomials together. Since the circuit is of depth d, the final degree of the approximating polynomial is $(\log(s) \cdot \log(s/\epsilon))^d$, as desired.

As mentioned at the beginning, this is a distribution on polynomials that computes correctly any input with probability at least $1 - \epsilon$. By averaging, there exists a fixed polynomial that computes correctly a $1 - \epsilon$ fraction of inputs. □

It can be verified that the value of the polynomial can be larger than $1/\epsilon$. The polynomial for the gates closest to the input can be as large as s. Then at the next level it can be as large as $s^{\log s/\epsilon}$, which is already much larger than $1/\epsilon$.

1.3 Approximation 2

Theorem 5. For all circuits C of size s and depth d, for all error values ϵ, there exists a polynomial $p(x_1, \ldots, x_n)$ of degree $O(\log(s)^{d-1} \log(1/\epsilon))$ such that

$$\mathbb{E}_{x \leftarrow U_n}[(C(x) - p(x))^2] \leq \epsilon.$$

The important feature of this approximation is that it bounds the average, but only under the uniform distribution. Because it does not provide any guarantee on other distributions, including k-wise independent distributions, it cannot be used directly for our aims.

Remark 6. Approximation 2 is proved via the *switching lemma*, an influential lemma first proved in the early 80's by Ajtai [Ajt83] and by Furst, Saxe, and Sipser [FSS84]. The idea is to randomly set a subset of the variables to simplify the circuit. You can do this repeatedly to simplify the circuit even further, but it only works on the uniform distribution. Hastad [Hås87] gave a much tighter analysis of the switching lemma, and the paper [LMN93] used it to prove a version of Approximation 2 with a slightly worse dependence on the error. Recently, a refinement of the switching lemma was proved in [Hås14, IMP12]. Based on that, Tal [Tal17] obtained the corresponding refinement of Approximation 2 where the parameters are as stated above. (The polynomial is simply obtained from the Fourier expansion of the function computed by the circuit by removing all Fourier coefficients larger than a certain threshold. The bound on the Fourier decay in [Tal17] implies the desired approximation.)

1.4 Bounded Independence Fools $\mathbf{AC}^0$

Theorem 7. For all circuits C with unbounded fan-in of size s and depth d, for all error values ϵ, for all k-wise independent distributions D on $\{0,1\}^n$, we have that

$$|\mathbb{E}[C(D)] - \mathbb{E}[C(U_n)]| \leq \epsilon$$

for $k = \log(s/\epsilon)^{O(d)}$.

Corollary 8. In particular, if $s = \text{poly}(n)$, $d = O(1)$, $s = 1/\text{poly}(n)$, then $k = \log^{O(1)}(n)$ suffices.

The next claim is the ultimate polynomial approximation used to prove the theorem.

Claim 9. For all circuits C with unbounded fan-in of size s and depth d, for all error values ϵ, for all k-wise independent distributions D on $\{0,1\}^n$, there is a set E of inputs, and a degree-k polynomial p such that:

1. E is 'rare' under both D and U_n:

$\Pr_{x \leftarrow U_n}[E(x) = 1] \leq \epsilon$, and $\Pr_{x \leftarrow D}[E(x) = 1] \leq \epsilon$. Here we write $E(x)$ for the indicator function of the event $x \in E$.

2. For all x, $p(x) \leq C(x) \vee E(x)$. Here $\vee$ is the logical Or.

3. $\mathbb{E}[p(U_n)] = \mathbb{E}[C(U_n)] \pm \epsilon$.

We only need (1) under D, but (1) under U is used to prove (3).

Proof of Theorem 7 from Claim 9.

$$\begin{aligned}
\mathbb{E}[C(D)] &= \mathbb{E}[C(D) \vee E(D)] \pm \epsilon, \text{ by Claim.(1)} \\
&\geq \mathbb{E}[p(D)] \pm \epsilon, \text{ by Claim.(2)} \\
&= \mathbb{E}[p(U_n)] \pm \epsilon, \text{ because } p \text{ has degree } k \text{ and } D \text{ is } k\text{-wise independent} \\
&= \mathbb{E}[C(U_n)] \pm \epsilon, \text{ by Claim.(3)}
\end{aligned}$$

For the other direction, repeat the argument for 'not C'. □

We can construct the polynomial approximation from Claim 9 by using a combination of Approximation 1 and 2. First we need a little more information about Approximation 1.

Claim 10. Two properties of approximation 1:

1. For all x, $p(x) \leq 2^{\log(s/\epsilon)^{O(d)}}$.

2. The 'bad' set E is computable by a circuit of size poly(s), and depth $d + O(1)$.

Proof of Claim 10 part 2. Consider a single OR gate with input gates $g_1, \ldots, g_s$. This is represented in the approximating polynomial by the term

$$1 - \prod_{i=1}^{\text{polylog}(s/\epsilon)} (1 - \sum_{j \in S_i} g_j).$$

Note that the term is incorrect exactly when the input $g_1, \ldots, g_s$ has weight > 0 but all the sets S_i intersect 0 or ≥ 2 ones. This can be checked in AC^0, in parallel for all gates in the circuit. □

Proof of Claim 9. Run approximation 1 for the distribution $\frac{D+U}{2}$, yielding the polynomial p_c and the set E. This already proves the first part of the claim for both D and U, because if E has probability ϵ under D it has probability $\geq \epsilon/2$ under $(D+U)/2$, and the same for U . Use Claim 10 part 2, and run approximation 2 on E. Call the resulting polynomial p_E, which has degree $\log(s/\delta)^{O(d)}$ with error bound δ.

The idea in the ultimate approximating polynomial is to "check if there is a mistake, and if so, output 0. Otherwise, output C". Formally:

$$p(x) := 1 - (1 - p_c(1 - p_E))^2$$

Claim 9 part 2 can be shown as follows. $p(x) \leq 1$ by definition. So, if $C(x) \vee E(x) = 1$, then we are done. Otherwise, $C(x) \vee E(x) = 0$. So there is no mistake, and $C = 0$. Hence, by the properties of Approximation 1, $p_c(x) = 0$. This implies $p(x) = 0$.

It only remains to show Claim 9 part 3:

$$\mathbb{E}_U[p(x)] = \mathbb{E}_U[C(x)] \pm \epsilon.$$

By part 1 of Claim 9,

$$\mathbb{E}_U[C(x) - p(x)] = \mathbb{E}_U[C(x) \vee E(x) - p(x)] \pm \epsilon.$$

We can show that this equals

$$\mathbb{E}_U\left[(C(x) \vee E(x) - p_c(x)(1 - p_E(x)))^2\right] \pm \epsilon$$

by the following argument: If $C(x) \vee E(x) = 1$ then $1 - p(x) = (1 - p_c(x)(1 - p_E(x)))^2$ by definition. If $C(x) \vee E(x) = 0$, then there is no mistake, and $C(x) = 0$. This implies that $p_c(x)(1 - p_E(x)) = p(x) = 0$.

Let us rewrite the above expression in terms of the expectation ℓ_2 norm.

$$||C \vee E - p_c(1 - p_E)||_2^2.$$

Recall the triangle inequality, which states: $||u - v||_2 \leq ||u - w||_2 + ||w - v||_2$. Therefore, letting $w = p_c(1 - E)$ we have that the above quantity is

$$\leq (||p_c(1 - E) - p_c(1 - p_E)||_2 \quad + \quad ||p_c(1 - E) - C \vee E||_2)^2$$

$$\leq O(||p_c(1 - E) - p_c(1 - p_E)||_2^2 + ||p_c(1 - E) - C \vee E||_2^2).$$

To conclude, we will show that each of the above terms are $\leq \epsilon$. Note that

$$||p_c(1-E) - p_c(1-p_E)||_2^2 \leq \max_x |p_c(x)|^2 ||(1-E) - (1-p_E)||_2^2.$$

By Claim 10 part 1 and Approximation 2, this is at most

$$2^{\log(s/\epsilon)^{O(d)}} \cdot ||E - p_E||_2^2 \leq 2^{\log(s/\epsilon)^{O(d)}} \cdot \delta.$$

For this quantity to be at most ϵ we set $\delta = \epsilon \cdot 2^{-\log(s/\epsilon)^{O(d)}}$. Here we critically set the error in Approximation 2 much lower, to cancel the large values arising from Approximation 1. By Theorem 5, the polynomial arising from approximation 2 has degree $O(\log(s)^{d-1}\log(1/\delta)) = \log(s/\epsilon)^{O(d)}$.

Finally, let us bound the other term, $||p_c(1-E) - C \vee E||_2^2$. If $E(x) = 0$, then the distance is 0. If $E(x) = 1$, then the distance is ≤ 1. Therefore, this term is at most $\Pr_U[E(x) = 1]$, which we know to be at most ϵ. □

References

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. of the ACM*, 57(5), 2010.

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

[Hås87] Johan Håstad. *Computational limitations of small-depth circuits.* MIT Press, 1987.

[Hås14] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. on Computing*, 43(5):1699–1708, 2014.

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 961–972, 2012.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. of the ACM*, 40(3):607–620, 1993.

[LN90] Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[Raz87] Alexander Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Akademiya Nauk SSSR. Matematicheskie Zametki*, 41(4):598–607, 1987. English translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333-338, 1987.

[Raz09] Alexander A. Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1), 2009.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of AC0. In *Conf. on Computational Complexity (CCC)*, pages 15:1–15:31, 2017.