03683170: Expanders, Pseudorandomness and Derandomization 28/02/16

Lecture 4 – AC^0

Amnon Ta-Shma and Dean Doron

1 Boolean circuits and AC^0

Definition 1. *A boolean circuit is a directed acyclic graph where every input vertex (vertices of in-degree* 0*) is labeled by* 0, 1 *or a variable, every internal vertex is a gate from* $\{\wedge, \vee, \neg\}$ *and the output vertices are those with out-degree* 0 *(the in-degree of the* $\wedge$ *and* $\vee$ *gates is not restricted). A circuit computes a boolean function in the obvious sense.*

The depth of a circuit is the length of the longest path from an input to an output, and the size of a circuit is the number of edges. A family of circuits $\{C_n\}$ solves a language L if for every $x \in \{0,1\}^n$, $x \in L$ iff $C_n(x) = 1$. A family of circuits is uniform if there exists a logspace (or polynomial-time) Turing machine that on input 1^n outputs C_n.

Definition 2. *A language* $L \in \mathsf{NC}^k$ *if it can be solved by a logspace-uniform family of circuits* $\{C_n\}$ *such that* $\text{depth}(C_n) = O(\log^k n)$ *and* $\text{size}(C_n) = \text{poly}(n)$ *and the in-degree of every gate is at most two. A language* $L \in \mathsf{AC}^k$ *if it can be solved by a logspace-uniform family of circuits* $\{C_n\}$ *such that* $\text{depth}(C_n) = O(\log^k n)$*,* $\text{size}(C_n) = \text{poly}(n)$ *and the gates (possibly) have unbounded fan-in.*

For example, parity, addition and multiplication are in NC^1. Addition and boolean matrix multiplication are in AC^0. The following inclusions hold (verify that you understand why):

$$\mathsf{NC}^0 \subseteq \mathsf{AC}^0 \subseteq \mathsf{NC}^1 \subseteq \mathsf{L} \subseteq \mathsf{NL} \subseteq \mathsf{AC}^1 \subseteq \mathsf{NC}^2 \subseteq \mathsf{L}^2.$$

Our first goal is to prove that $\mathsf{AC}^0 \subsetneq \mathsf{NC}^1$ and we will do that by proving Parity $\notin \mathsf{AC}^0$. This is a rare case where we can actually prove a lower bound. In fact, we will also prove that parity is average-case hard for AC^0 circuits. Separation results were first obtained by Ajtai [1], Furst et al. [3] culminating with Håstad's switching lemma (that shows that AC^0 drastically shrink under partial setting of the variables) which leads to almost optimal lower bounds [4]. A different proof of a slightly weaker lower bound (that has its own advantages) was given later by Smolensky [7]. Linial, Mansour and Nisan [5] gave a learning algorithm for AC^0. We will see Smolensky's proof and cite the LMN result.

We then that t-wise independence with $t = \text{polylog}(n)$ fools AC^0. Braverman proved:

Theorem 3 ([2])**.** *t-wise independence ε-fools* AC *circuit of size s and depth d for*

$$t = \left(\log \frac{s}{\varepsilon}\right)^{O(d^2)}.$$

In particular, taking d to be constant and $\varepsilon = \frac{1}{\text{poly}(s)}$*, we see that* $t = \text{poly}(\log s)$ *suffices.*

As it turns out, most results in this section deal with how well we can represent a function as a low-degree multilinear polynomial over the reals. The exact meaning of "representation" (exact or approximate, worst-case or average-case) differs, but the moto is the same: functions computed by AC^0 circuits are "close" to functions that have a low-degree representation, while parity is far from such functions.

2 Smolensky's proof that Parity is not in AC^0

2.1 Exact/approximate, worst-case/average-case representation of a function as a low-degree polynomial

Definition 4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function and $p : \mathbb{R}^n \to \mathbb{R}$ be a multi-linear polynomial. We say that:*

- *p* computes *f on $x \in \{0,1\}^n$ if $p(x) = f(x)$,*
- *p ε–*approximates *f on $x \in \{0,1\}^n$ if $|p(x) - f(x)| \leq \varepsilon$,*
- *p* computes *f on average with δ–error, if $\Pr_{x \in \{0,1\}^n}[p(x) = f(x)] \geq 1 - \delta$,*
- *p ε-*approximates *f on average with δ error if $\Pr_{x \in \{0,1\}^n}[|p(x) - f(x)| \leq \varepsilon] \geq 1 - \delta$.*

Let us check these definitions with the OR and Parity functions. We start with *exact*, *worst-case* computation. If p *computes* f on every $x \in \{0,1\}^n$, then p is the *unique* multi-linear representation of f, and we already saw that the degree of p is the maximal cardinality of a non-zero Fourier coefficient. Therefore, both the OR and Parity functions on n bits require full degree.

Next, we look at *approximate*, *worst-case* computation. Both the OR and Parity functions are symmetric functions. A simple observation is:

Lemma 5. *If $f : \{0,1\}^n \to \{0,1\}$ can be worst-case, ε-approximated by a degree t polynomial $p : \mathbb{R}^n \to \mathbb{R}$, then there exists a degree t uni-variate polynomial $\tilde{p} : \mathbb{R} \to \mathbb{R}$ such that $\tilde{p}(\sum_i x_i)$ worst-case, ε-approximates f.*

The question then reduces to the following problem: Given $\{(i, a_i)\}_{i=1}^n$, where a_i represents the value of f on an input with exactly i ones, what is the smallest degree t such that there is a low degree uni-variate polynomial p of degree t for which $p(i)$ is ε-close to a_i for all i? This is a well studied question in approximation theory. For the OR function there exists a degree $O(\sqrt{n})$ polynomial that approximates OR with at most $\frac{1}{3}$ error, and for this error the degree is tight up to a constant factor. For the parity function a worst-case approximation with $\frac{1}{3}$ error requires $\Theta(n)$ degree. In general, Paturi showed:

Lemma 6 ([6])**.** *Let f be a boolean symmetric function on n variables and let*

$$\Gamma(f) = \min\{|2k - n + 1| : f_k \neq f_{k+1}, 0 \leq k \leq n-1\}$$

where f_i is the value of f on inputs with exactly i 1-s. Then, every polynomial that $\frac{1}{3}$-worst-case approximates f is of degree $\Theta(\sqrt{n(n - \Gamma(f))})$.

This completes the picture (at least for symmetric functions) of worst-case approximation, and we see a small (quadratic) difference between the OR and the Parity functions.

We now turn to average-case, *exact* computation. The OR function is trivially easy, as it can be approximated by the constant function 1. We now show that Parity is hard.

Lemma 7. *If $p : \mathbb{R}^n \to \mathbb{R}$ computes f with average-case success $\frac{1}{2} + \delta$, then $\deg(p) = \Omega(\delta\sqrt{n})$.*

Proof. Suppose there exists a degree t polynomial that exactly computes Parity on average with success $\frac{1}{2}+\delta$. I.e., there exists a set $A \subseteq \{0,1\}^n$ of cardinality at least $\left(\frac{1}{2}+\delta\right)2^n$ such that for every $x \in A$, $p(x) = \text{Parity}(x)$.

Consider the vector space of all functions from A to $\mathbb{R}$. This vector space has dimension $|A|$. We will soon show that this vector space is a subset of a vector space L containing all multi-linear polynomials of total degree at most $\frac{n}{2}+t$. Together this implies that

$$\left(\frac{1}{2}+\delta\right)\cdot 2^n \;\leq\; |A| \;\leq\; \sum_{i=0}^{\frac{n}{2}+t}\binom{n}{i} \;\leq\; \frac{2^n}{2}+\frac{t}{\sqrt{n}}2^n,$$

which implies that $\delta \leq \frac{t}{\sqrt{n}}$.

To see that indeed $A \subseteq L = \text{Span}\left\{\prod_{i\in S} x_i\right\}_{|S|\leq\frac{n}{2}+t}$, expand f in the Fourier basis. It is enough to show that each character $(-1)^{\langle S,a\rangle}$ is contained in L. This is clearly true for sets S of cardinality at most $\frac{n}{2}$. For sets S of cardinality above $\frac{n}{2}$, $\chi_S(x) = \chi_{1,\ldots,1}(x)\cdot\chi_{\bar{S}}(x)$. $\chi_{\bar{S}}(x)$ is represented by a polynomial of degree at most $\frac{n}{2}$. $\chi_{1,\ldots,1}(x) = (-1)^{\text{Parity}(x)} = 1-2\cdot\text{Parity}(x) = 1-2\cdot p(x)$ over A, and therefore can be represented by a degree t polynomial as desired. □

This result is tight, due to [6]. Thus,

Parity	Computation	$\frac{1}{3}$-approximation
Worst-case	n	n
Average-case with success $\frac{2}{3}$	$\Omega(\sqrt{n})$	?

Table 1: The degree needed for the Parity function

Recall that for the OR function we have:

OR	Computation	$\frac{1}{3}$-approximation
Worst-case	n	$\Theta(\sqrt{n})$
Average-case with success $\frac{2}{3}$	1	1

Table 2: The degree needed for the OR function

2.2 Exact, worst-case, probabilistic computation

Next, we would like to extend the positive result, of low-degree average-case approximation of the OR function, to all functions in AC^0 with the intuition that any AC^0 function is a low-depth composition of OR and NOT gates, and NOT gates are trivial. However, while the OR function has a trivial exact, average-case approximation with respect to the uniform distribution, it is not clear how to extend the result to compositions, because the distributions appearing within the circuit are controlled by the circuit and can be far from uniform. To handle this problem, we introduce a new notion of exact, worst-case, *probabilistic* computation.

Definition 8. *Let P be a class of functions. We say P ε-approximates f (worst-case and exact) if for every $x \in \{0,1\}^n$ $\Pr_{p\in P}[p(x) = f(x)] \geq 1-\varepsilon$.*

Let us first consider the OR function.

Lemma 9. *For all n and $\varepsilon > 0$ there exists a set P of polynomials of degree $O(\log n \log \frac{1}{\varepsilon})$ that ε computes OR.*

Proof. Pick T sets $S_1, \ldots, S_T \subseteq [n]$ as follows: The first set is $[n]$. Then, for every $k = 1, \ldots, \log n$, we pick t random subsets of cardinality 2^k. Having $S_1, \ldots, S_T$, define the polynomial

$$p(x) \;=\; 1 - \prod_{j=1}^{T} \left(1 - \sum_{i \in S_j} x_i\right).$$

When $x = 0$ check that $p(x) = 0$ (i.e., we have one-sided error). When $x \neq 0$, the subset $I(x) = \{i \in [n] \mid x_i = 1\}$ is non-empty and has cardinality q for some $2^{k-1} \leq q < 2^k$. We claim that with probability at least $\frac{1}{8}$, a random subset of $[n]$ of cardinality $\frac{n}{2^k}$ has intersection size exactly one with $I(x)$. When that happens the corresponding term $1 - \sum_{i \in S_j} x_i$ is zero. It follows that for every non-zero x, with probability at least $1 - 2^{-\Omega(t)}$, $p(x) = 1$ and p agrees with the or function.

Finally, clearly, $\deg(p) \leq T$. □

In fact, instead of choosing sets (of certain cardinalities) uniformly at random, one can choose them pair-wise independently. Specifically, each time we want to sample a set of size $\frac{n}{2^k}$, instead of choosing a random subset of that size, we pick h uniformly at random from a two-universal family of hash functions $\mathcal{H}_k$, and let $S = h^{-1}(0)$. We will prove in the exercise that this also works.

The same also holds for the AND function:

Lemma 10. *For all n and $\varepsilon > 0$ there exists a set P of polynomials of degree $O(\log n \log \frac{1}{\varepsilon})$ that ε computes AND.*

Computing NOT (exactly) by a polynomial is trivial, so we are ready to prove our main lemma for this section.

Lemma 11. *For every AC circuit C of size s and depth d there exists a set P of polynomials of degree $O((\log s)^{2d})$ that ε computes OR.*

Proof. Set $\varepsilon = \frac{1}{4s}$. For every gate g_i of C, pick its approximating low-degree function p_i independently according to the above distribution and let p be the polynomial resulting from the composition. By the union-bound, for every $x \in \{0,1\}^n$, the probability (over choosing the subsets) that p does not agree with C on x is at most $s \cdot \varepsilon = \frac{1}{4}$. In particular, the expectation (over $x \in \{0,1\}^n$ and the random choices of the subsets) of the agreement between p and C is at least $\frac{3}{4}$. It follows that there exists a choice of subsets for which p agrees with C on at least $\frac{3}{4}$ of the inputs.

Every p_i has degree at most $\log s \cdot \log \frac{1}{\varepsilon} = O(\log^2 s)$ (because the fan-in of every gate is at most s). Also, the composition has the effect of multiplying the degrees (check). Thus, the overall degree is $O((\log s)^{2d})$. □

As a corollary:

Lemma 12. *For every AC circuit C of size s and depth d, distribution π over $\{0,1\}^n$ there exists a function $p : \{0,1\}^n \to \mathbb{R}$ of degree $O((\log s)^{2d})$ such that $\Pr_{x \in \pi}[C(x) = p(x)] \geq \frac{3}{4}$.*

Tracing the error more explicitly,

Lemma 13. *For every* AC *circuit* C *of size* s *and depth* d*, distribution* π *over* $\{0,1\}^n$ *and* $a > 1$ *there exists a function* $p : \{0,1\}^n \to \mathbb{R}$ *of degree* $(a \log s)^d$ *such that* $\Pr_{x\in\pi}[C(x) \neq p(x)] \leq \left(\frac{3}{4}\right)^a \cdot s$.

Here, we have $T \leq a \log s$ and $|S_i| \leq s$ for every $i \in [T]$.

We will need another component: A function E that will tell us whether there is a *mistake* in p (which happens rarely) and will also be computable in AC^0 if C itself is an AC^0 function. Suppose we are given a circuit C and we prepare for some distribution π. We know that there exists a setting of the sT sets (where $T = (a \log s)$) that yields a polynomial p that is ε-good on average with respect to π. Fix these sets. With these fixed sets, a non-zero boolean input to a gate is bad if all sets intersect with the set bits that are 1 in the input with intersection size that is not 1, and there is an easy AC^0 circuit (of depth 3) checking this. Altogether, if C has depth d, we can build a circuit of depth $d + 2$ returning 1 when the input is bad. The blowup in size is at most $O(sT)$.

We then have:

Lemma 14. *Let* π *be any distribution on* $\{0,1\}^n$*. For every* AC *circuit* C *of size* s *and depth* d*, for any* $\varepsilon > 0$*, there is a polynomial* $p : \{0,1\}^n \to \mathbb{R}$ *of degree at most* $r = (4 \log \frac{s}{\varepsilon} \log s)^d$ *and a boolean circuit* E_π *of depth* $d + 2$ *and size at most* $s^2 r$ *such that:*

- $\Pr_{x\sim\pi}[E_\pi(x) = 1] \leq \varepsilon$.
- *Whenever* $E_\pi(x) = 0$*,* $p(x) = C(x)$.

We also note that:

Claim 15. *Using the notations of Lemma 14,* $|p(x)| < (2s)^{r-2}$ *for every* $x \in \{0,1\}^n$.

Finally, we use the results obtained in this section to get a (mild) average-case hardness of Parity for AC^0. We have seen that every function $f \in \mathsf{AC}^0$ can be approximated well by a low-degree polynomial. We also saw that Parity cannot be approximated by a low degree polynomial. The obvious conclusion is that Parity is not in AC^0. Moreover,

Theorem 16. *Let* C *be an* AC^0 *circuit. Then,*

$$\Pr_{x\in\{0,1\}^n}[C(x) = \text{Parity}(x)] \;\leq\; \frac{1}{2} + o(n^{-\frac{1}{2}}).$$

A tighter bound is known, which follows from Håstad's switching lemma:

Theorem 17 ([4])**.** *Let* C *be a circuit of depth* d *and size* $2^{O(n^{1/d})}$*. Then,*

$$\Pr_{x\in\{0,1\}^n}[C(x) = \text{Parity}(x)] \;\leq\; \frac{1}{2} + 2^{-\Omega(n^{1/d})}.$$

3 ℓ_2 approximations

Let C be a small-depth polynomial size circuit computing a function f. $\hat{f}(S)$ measures the correlation f has with the Parity function on the bits in S (no correlation means no bias and zero Fourier

coefficient, while high correlation means high bias and high Fourier coefficient). Since we saw that all small-depth polynomial size circuits have very low advantage computing Parity on many bits, we may conclude that all Fourier coefficients $\hat{f}(S)$ for large cardinality sets S are small. I.e., if we look at the vector $(\hat{f}(S))_{S:|S|\geq k}$ then this vector has small ℓ_∞ norm when k is large enough. The LMN theorem extends this to the ℓ_2 norm, i.e., the sum-of-squares of the high Fourier coefficients is very small.

Definition 18. *We say that p ε-approximates f in the ℓ_2 norm if $\| f - p \|_2^2 = \mathbb{E}_x[|f(x)-p(x)|^2] \leq \varepsilon$.*

Linial, Mansour and Nisan proved:

Theorem 19 ([5])**.** *Suppose $f : \{0,1\}^n \to \{0,1\}$ is computed by an AC circuit of size s and depth d. Then, for every t there is a degree t polynomial p with $\| f - p \|_2^2 = \frac{1}{2^n} \sum_x |f(x) - p(x)|^2 \leq 2s \cdot 2^{-t^{1/d}/20}$.*

Notice that by Parseval $\| f - p \|_2^2 = \sum_S |\hat{f}_S - \hat{p}_S|^2$. Also, because p has degree at most t, $\hat{p}_S = 0$ for all S of cardinality larger than t. On sets of smaller cardinality the best choice for p (minimizing the sum) is choosing $\hat{p}_S = \hat{f}_S$. Thus, for the best p we can choose $f - p$ is f with all small fourier coefficients eliminated, and the theorem claims that their ℓ_2 norm is small.

For the theorem we can again recover that Parity is worst-case hard for AC^0, (though, the LMN proof itself relies on the switching lemma that already implies that). On the one hand,

Claim 20. *Every polynomial $p : \{0,1\}^n \to \mathbb{R}$ of degree at most $n-1$ satisfies $\| p - \mathrm{Parity} \|_2 \geq \frac{1}{2}$.*

Proof. It can be verified that $\widehat{\mathrm{Parity}}(\emptyset) = \frac{1}{2}$ and $\widehat{\mathrm{Parity}}([n]) = -\frac{1}{2}$. Now, write $p(x) = \sum_S \hat{p}(S)\chi_S(x)$. By Parseval,

$$\| p - \mathrm{Parity} \|_2^2 \;=\; \sum_S |p - \widehat{\mathrm{Parity}}(S)|^2 \;=\; \sum_S |\hat{p}(S) - \widehat{\mathrm{Parity}}(S)|^2.$$

p is of degree at most $n-1$, so $\hat{p}([n]) = 0$. Thus,

$$\| p - \mathrm{Parity} \|_2^2 \;=\; \sum_{S \neq [n]} |\hat{p}(S) - \widehat{\mathrm{Parity}}(S)|^2 + \frac{1}{4} \;\geq\; \frac{1}{4},$$

as desired. In fact, this is just a concrete instantiation of the discussion after Theorem 19. □

Therefore,

Corollary 21. *If C is a circuit of depth $d \geq 2$ and size s that computes Parity, then $s = 2^{\Omega(n^{1/(4d)})}$.*

Proof. Immediate by applying Theorem 19 with $t = n - 1$. □

This bound is tight:

Exercise 22. *For every constant $d \geq 2$, there are circuits of size $2^{O(n^{1/(d-1)})}$ that compute the parity of n bits.*

Note that although $\mathbb{E}[|f(x)-p(x)|^2]$ is small, it still might be the case that although $|f(x)-p(x)|$ *is* small in $\frac{2}{3}$-fraction of the inputs, it is extremely large elsewhere, and in fact this is the expected behavior from a low-degree polynomial that is forced to be close to particular values in specific evaluation points.

Another important thing to notice is that unlike the Smolensky's bound, the LMN theorem gives a good approximation of an AC^0 predicate by a low-degree polynomial with respect to the *uniform* distribution, and guarantees nothing with respect to other distributions.

4 Polylog-wise independence fools AC^0

We saw that a distribution $\mathcal{D}$ is k-wise independent iff all its non-trivial Fourier coefficients $\hat{\mathcal{D}}(S)$ for sets of cardinality at most k are zero. Thus $\mathcal{D}$ has no correlation with low-degree polynomials. On the other hand, every small-depth polynomial size circuit computes a boolean function that may be approximated well by a low-degree polynomial (be it using Smolensky's exact, average case or LMN's ℓ_2, average case interpretation). Thus, intuitively, we should expect $\mathcal{D}$ to "fool" any such circuit.

Indeed, recall that a distribution $\mathcal{D}$ ε-fools a function $f : \{0,1\}^n \to \{0,1\}$ if $|\mathbb{E}_{x\sim U_n}[f(x)] - \mathbb{E}_{x\sim\mathcal{D}}[f(x)]| \leq \varepsilon$. Take $\mathcal{D}$ to be any t-wise independent distribution. Braverman proves:

Theorem 23 ([2])**.** *Any t-wise independent distribution ε-fools any AC circuit of size s and depth d for*

$$t = \left(\log \frac{s}{\varepsilon}\right)^{O(d^2)}.$$

In particular, taking d to be constant and $\varepsilon = \frac{1}{\mathrm{poly}(s)}$, we see that $t = \mathrm{poly}(\log s)$ suffices.

A first attempt to prove this is as follows: Suppose f is computed by a depth d size s circuit and $\mathcal{D}$ is some t-wise independent distribution. Then f can be approximated by a polynomial p of some low degree $t = t(d,s)$. But

Claim 24. *If $p : \mathbb{R}^n \to \mathbb{R}$ is a degree t multi-linear polynomial, and $\mathcal{D}$ is some t-wise independent distribution then $\mathbb{E}_{x\sim U_n}[p(x)] = \mathbb{E}_{x\sim\mathcal{D}}[p(x)]$.*

Proof. Write $p(x) = \sum_{I:|I|\leq t} a_I x^I$, where $X^I = x_1^{I_1} \cdot \ldots \cdot x_n^{I_n}$. As $\mathcal{D}$ is t-wise independent we have that $\mathbb{E}_{x\sim U_n}[x^I] = \mathbb{E}_{x\sim\mathcal{D}}[x^I]$ for every I with $|I| \leq t$. By the linearity of expectation, the claim follows. □

Hence, $\mathbb{E}_{x\sim\mathcal{D}}[p(x)] \approx \mathbb{E}_{x\sim U_n}[f(x)]$. To conclude the proof we need $\mathbb{E}_{x\sim\mathcal{D}}[p(x)] \approx \mathbb{E}_{x\sim\mathcal{D}}[f(x)]$. However, here we face a problem:

- If we use the approximation notion of ℓ_2 approximation, then proximity is guaranteed only with respect to the uniform distribution and not with respect to $\mathcal{D}$, that is potentially distributed over a very small set.
- If we use the exact, average-case approximation of Smolensky, then potentially on values on which we err we might make a huge error, that would completely bias the approximation. To see that notice that on an x on which we err we might get $p(x)$ as large as s^{2^d}, and since the probability of making an error might be $\frac{1}{\mathrm{poly}(s)}$, the total bias given on the wrong inputs might be huge.

Braverman overcome this difficulty by showing that:

Lemma 25. *Let f be a boolean function computed by a depth d size s circuit. Let $\varepsilon > 0$ and let $\mathcal{D}$ be a t-wise independent distribution for $t = t(d, s, \varepsilon)$. Then, there exists a set of inputs E and a degree t multi-linear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ for which:*

1. *(E is small with respect to $\mathcal{D}$ and the uniform distribution)* $\Pr_{x \sim U_n}[x \in E] \leq \varepsilon$, $\Pr_{x \sim \mathcal{D}}[x \in E] \leq \varepsilon$.
2. *(Exact answer on No instances) For every $x \in \{0,1\}^n \setminus E$ for which $f(x) = 0$ we have $p(x) = 0$, i.e., we have one-sided error, and,*
3. *(p is upper bounded) For every $x \in \{0,1\}^n$, $p(x) \leq (f \vee E)(x) \leq 1$, where $f \vee E$ is the boolean function that is 1 on x iff $f(x) = 1$ or $x \in E$.*
4. *(Approximate answer in ℓ_1 norm on all instances)* $\mathbb{E}_{x \sim U_n}[(f \vee E)(x)] - \mathbb{E}_{x \sim U_n}[p(x)] \leq \varepsilon$.

Putting it in a diagram, we have:

$$\begin{array}{ccc}
\mathbb{E}_{x \sim U_n}[f(x)] & & \mathbb{E}_{x \sim \mathcal{D}}[f(x)] \\
\uparrow \approx, \text{ item (1)} & & \downarrow \approx, \text{ item (1)} \\
\mathbb{E}_{x \sim U_n}[(f \vee E)(x)] & & \mathbb{E}_{x \sim \mathcal{D}}[(f \vee E)(x)] \\
\uparrow \approx, \text{ item (4)} & & \downarrow \geq, \text{ item (3)} \\
\mathbb{E}_{x \sim U_n}[p(x)] & \xleftarrow[=, \text{ Claim 24}]{} & \mathbb{E}_{x \sim \mathcal{D}}[p(x)]
\end{array}$$

Formally, we claim:

Claim 26. *Let f and p be as above. Then $\mathbb{E}_{x \sim \mathcal{D}}[f(x)] \geq \mathbb{E}_{x \sim U_n}[f(x)] - 3\varepsilon$.*

Proof. As p is of degree t, we know by Claim 24 that $\mathbb{E}_{\mathcal{D}}[p] = \mathbb{E}[p]$. Then:

$$\begin{array}{rcll}
\mathbb{E}_{\mathcal{D}}[f] & \geq & \mathbb{E}_{\mathcal{D}}[f \vee E] - \Pr_{x \sim \mathcal{D}}[x \in E] & \\
& \geq & \mathbb{E}_{\mathcal{D}}[f \vee E] - \varepsilon & \text{By item (1)} \\
& \geq & \mathbb{E}_{\mathcal{D}}[p] - \varepsilon & \text{By item (3)} \\
& = & \mathbb{E}[p] - \varepsilon & \text{By Claim 24} \\
& = & \mathbb{E}[f \vee E] - \mathbb{E}[f \vee E - p] - \varepsilon & \\
& \geq & \mathbb{E}[f \vee E] - 2\varepsilon & \text{By item (4)} \\
& \geq & \mathbb{E}[f] - \Pr_{x \sim U_n}[x \in E] - 2\varepsilon & \\
& \geq & \mathbb{E}[f] - 3\varepsilon. & \text{By item(1)}
\end{array}$$

□

Applying the above claim to both f and the complement of f (that both can be computed by a size s depth d circuit) we get

Claim 27. *Let f and p be as above. Then $|\mathbb{E}_{x \sim \mathcal{D}}[f(x)] - \mathbb{E}_{x \sim U_n}[f(x)]| \leq 3\varepsilon$.*

which concludes the proof.

We are left with proving Lemma 25. The proof uses a combination of Smolensky's result together with the LMN result.

Proof. Fix f computed by a size s and depth d circuit. Fix $\mathcal{D}$, an arbitrary t-wise independent distribution. Set $\pi = \frac{1}{2}(U_n + \mathcal{D})$.

- Let p_f be the polynomial guaranteed by Smolensky (Lemma 14). There exists a set E (of suspected bad inputs) such that whenever $x \notin E$, $p_f(x) = f(x)$ and

$$\Pr_{x\sim\pi}[x \in E] \leq \frac{\varepsilon}{4}.$$

  We recall the parameters guaranteed by the lemma, $\deg(p_f) = d_f = \Omega\left((\log \frac{s}{\varepsilon})^{2d}\right)$.

- We overload notations and also call the AC^0 circuit (of size $s_E = O(s^2 d_f)$ and depth $d+2$) that returns one exactly on elements from E. We let p_E be the degree d_E polynomial guaranteed by LMN (Theorem 19) such that

$$\| E - P_E \|_2^2 \leq \delta = \delta(s_E, d, d_E).$$

  We will chose d_E later.

The way we have chosen π guarantees item (1). To see that, notice $\frac{\varepsilon}{2} \geq \Pr_{x\sim\pi}[x \in E] \geq \frac{1}{2} \cdot \max\{\Pr_{x\sim U_n}[x \in E], \Pr_{x\sim\mathcal{D}}[x \in E]\}$.

Now, denote

$$p \;=\; 1 - (1 - p_f \cdot (1 - p_E))^2.$$

If $(f \vee E)(x) = 0$ then $p_f(x) = 0$ (because $x \notin E$ and $p_f(x) = f(x) = 0$) and therefore $p(x) = 0$ and item (2) follows. Also, since we deduce a non-negative quantity, $p(x) \leq 1$ for every $x \in \{0,1\}^n$. Together, $p \leq f \vee E$ and item (3) follows.

For the last item, write $q = p_f \cdot (1 - p_E)$ so $p = 1 - (1 - q)^2$. We claim that for every $x \in \{0,1\}^n$,

$$(f \vee E)(x) - p(x) = ((f \vee E)(x) - q(x))^2.$$

To see the last equality try it for both x such that $(f \vee E)(x) = 0$ which implies $q(x) = 0$, and x such that $(f \vee E)(x) = 1$. Thus,

$$\begin{aligned}
\mathop{\mathbb{E}}_{x\sim U_n}[(f \vee E)(x) - p(x)] &= \mathop{\mathbb{E}}_{x}((f \vee E)(x) - q(x))^2 \\
&= \| f \vee E - q \|_2^2 \\
&\leq \left(\| f \vee E - p_f(1-E) \|_2 + \| p_f(1-E) - q \|_2\right)^2 \\
&\leq 2 \cdot \| f \vee E - p_f(1-E) \|_2^2 + 2 \cdot \| p_f(1-E) - p_f(1-p_E) \|_2^2 \\
&\leq 2 \cdot \| E \|_2^2 + 2 \cdot \| p_f(E - p_E) \|_2^2 \\
&\leq 2 \cdot \Pr_{x\sim\pi}[x \in E] + 2 \cdot \| p_f \|_\infty^2 \cdot \| E - p_E \|_2^2 \\
&\leq 2 \cdot \Pr_{x\sim\pi}[x \in E] + 2 \cdot \| p_f \|_\infty^2 \cdot \delta.
\end{aligned}$$

Note that $2 \cdot \Pr_{x\sim\pi}[x \in E] \leq \frac{\varepsilon}{2}$. By Claim 15 one can choose $d_E = \left(\log \frac{s}{\varepsilon}\right)^{O(d^2)}$ such that $\| p \|_\infty^2 \leq \frac{\varepsilon}{4\delta}$, so overall item (4) is satisfied. Finally, observe that $t = \deg(p) \leq 2(d_f + d_E)$ and the proof is complete. □

References

[1] Miklós Ajtai. Σ_1-formulae on finite structures. *Annals of pure and applied logic*, 24(1):1–48, 1983.

[2] Mark Braverman. Polylogarithmic independence fools AC0 circuits. *Journal of the ACM (JACM)*, 57(5):28, 2010.

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

[4] Johan Håstad. Computational limitations of small-depth circuits. 1987.

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

[6] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 468–474. ACM, 1992.

[7] Roman Smolensky. On representations by low-degree polynomials. In *Foundations of Computer Science, 1993. Proceedings., 34th Annual Symposium on*, pages 130–138. IEEE, 1993.