Lecture 4: AC^0 lower bounds and pseudorandomness

Topics in Complexity Theory and Pseudorandomness (Spring 2013)
Rutgers University
Swastik Kopparty
Scribes: Jason Perry and Brian Garnett

In this lecture, we will finish the proof that Approximate Majority has polynomial-size AC^0 circuits. Then we will show that majority is hard for AC^0 even with parity gates. Lastly, we will see that parity is hard-on-average for AC^0, and that this can be used to give us pseudorandom generators against AC^0.

1 Approximate Majority and AC^0

Previously we saw how PARITY requires AC^0 circuits of size $\geq 2^{\Omega(n^{1/(d-1)})}$, where d is the circuit depth. The key to the proof was the switching lemma, which intuitively says that random restrictions drastically simplify AC^0 circuits. This immediately gives us that MAJORITY requires AC^0 circuits of size $2^{\Omega(n^{1/d})}$, since we can produce a parity circuit out of a majority circuit.

At the end of the last lecture, we were attempting to show that randomized AC^0 circuits were no more powerful than deterministic. Recall that for general circuits and formulas, we were able to do so by means of probability amplification. However, this required a circuit computing Majority, which we don't have for polynomial-size AC^0. Ajtai and Ben-Or showed that instead, we can use Approximate majority, which does have polynomial AC^0 circuits, to get the same result.

Now we will wrap up the proof that Approximate Majority has polynomial-size AC^0 circuits. Let's recall the definition of Approximate majority:

Definition 1. *For $x \in \{0,1\}^n$, let $wt(x)$ be the number of ones in x. Then*

$$AM(x) = \begin{cases} 0 & \textit{if } wt(x) \leq n/4 \\ 1 & \textit{if } wt(x) \geq 3n/4 \\ \textit{don't care} & \textit{otherwise} \end{cases}$$

Theorem 2. *There exist polynomial-size AC^0 circuits computing AM.*

Proof. Note that Approximate Majority as we have defined it is not a single function; its definition may be satisfied by many circuits computing different functions. To show that the desired circuits exist, we construct a probability distribution C over small AC^0 circuits.

We want our distribution to satisfy the following, for all x:

$$\text{if } wt(x) < n/4, \ \Pr_C[C(x) \neq 0] \ll 2^{-n^2}$$
$$\text{if } wt(x) > 3n/4, \ \Pr_C[C(x) \neq 1] \ll 2^{-n^2}$$

Then for a fixed x, with high probability a C chosen from this distribution will compute AM:

$$\Pr_C [\exists x \; wt(x) < n/4, C(x) \neq 0 \vee \\ \exists x \; wt(x) > 3n/4, C(x) \neq 1] \ll 2^n \cdot 2^{-n^2} \sim 2^{-\Omega(n^2)}$$

It follows that there exists a single circuit that computes AM exactly, i.e.

$$\forall x \; wt(x) < n/4, C(x) = 0 \quad \wedge \quad \forall x \; wt(x) > 3n/4, C(x) = 1$$

To obtain the desired probabilities, we define a sequence of distributions. Let $p = wt(x)/n$, which is the probability that a random bit in x is 1. The following table describes the distributions and gives bounds on $\Pr_C[C(x) = 1]$ for the two weight conditions:

Distribution	$p < 1/4$	$p > 3/4$
$C_0(x) =$ uniformly chosen circuit to select one bit of x	p	p
$C_1(x) = \bigwedge$ ($10 \log n$ independent random ckts from $C_0(x)$)	$< 1/n^{20}$	$> 1/n^{10}$
$C_2(x) = \bigvee$ (n^{15} independent ckts from $C_1(x)$)	$< 1/n^5$	$> 1 - e^{-n^5}$
$C_3(x) = \bigwedge$ (n^2 independent ckts from $C_2(x)$)	$\ll 2^{-n^2}$	$\gg 1 - e^{-n^4}$

The C_0 circuits are clearly in AC^0. The number of gates in C_{i+1} is polynomial in the size of C_i, and each new level adds 1 to the depth.

The probabilities for C_1 come from $p^{10 \log n}$. For the $p < 1/4$ probability of C_2, the union bound is good enough. The more accurate bound for 3/4 is from $1 - (1 - \frac{1}{n^{10}})^{n^{15}}$. For C_3, the bound on $p > 3/4$ comes from the expansion of $(1 - e^{-n^5})^{n^2}$. Since we have the desired probabilities, the proof is complete. □

Remark: This proof method is interesting because it uses the probabilistic method not over inputs, but over computations.

2 MAJORITY $\notin \mathsf{AC}^0(\oplus)$

The result given in the section title implies that majority is strictly harder than parity for AC^0, since there is no polynomial-size AC^0 circuit to compute majority even if we are given parity gates. The result is Razborov's, and the proof technique uses ideas due to both Razborov and Smolensky.

Consider the class of circuits $\mathsf{AC}^0(\oplus)$ with $\wedge$, $\vee$, $\neg$, and $\oplus$ gates of unbounded fan-in. The parity gate $\oplus$ outputs 1 if an odd number of its inputs are 1.

We first state and prove a lemma showing that every AC^0 circuit can be approximated by a low-degree polynomial.

Lemma 3. *(Razborov's Lemma) For all AC^0 circuits C of size s, depth d, there exists a distribution P of polynomials $p(x_1, ..., x_n) \in \mathbb{F}_2[X_1, ..., X_n]$ such that, for all x,*

$$\Pr_{p \in P}[p(x) \neq C(x)] \leq \epsilon,$$

and $\deg(p) \leq (\log(\frac{s}{\epsilon}))^d$ always.

A corollary to this lemma is that there exists a single low-degree polynomial which computes C on all but an ϵ fraction of the $x \in \{0,1\}^n$. This corollary suffices for our lower bound on Majority; but it is the form stated in the lemma that is the easier one to prove, because it is strong enough to lend itself to a proof by induction.

Before proving the lemma, let's try to develop some intuition about representing $\mathsf{AC}^0(\oplus)$ circuits with low-degree polynomials. Is it possible to encode all $\mathsf{AC}^0(\oplus)$ circuits *exactly* with low-degree polynomials?

The single-gate circuit taking the parity of all inputs has the exact degree-1 representation $\sum_{i=1}^n x_i$, since parity corresponds to addition in $\mathbb{F}_2$. So far so good. But the single-gate circuit computing the AND of all inputs is equivalent to the polynomial $\prod_{i=1}^n x_i$, whose degree cannot be reduced.

We need to find a polynomial approximating AND of degree $\leq (\log(\frac{1}{\epsilon}))^d$. We can't approximate AND by always outputting zero, since the desired correctness probability must hold for *all* inputs x. Multiplying a random constant-size subset of the bits of x will not work either, for the same reason. However, we can find desirable properties by *summing* random subsets.

Take a uniform random subset s of $[n]$ and consider the polynomial $\sum_{i \in s} x_i$. Observe that

$$\Pr_s[\sum_{i \in s} x_i = 1] = 0, \quad \text{if } x = (0, ..., 0)$$
$$\Pr_s[\sum_{i \in s} x_i = 1] = \frac{1}{2}, \quad \text{if } x \neq (0, ..., 0).$$

This is an approximation of OR with a constant error probability, of degree 1. Of course, if we have OR and negation, we can also produce AND. So we have the idea for the proof.

Proof. Choose subsets $s_1, s_2, ..., s_k \subseteq [n]$ uniformly at random. We define the following degree-k polynomial to approximate an OR gate:

$$p_\vee(x_1, ..., x_n) = 1 - \prod_{j=1}^{k}(1 - \sum_{i \in s_j} x_i)$$

Since we use k subsets, we have:

$$\text{if } x = (0, ..., 0), \ \Pr_p[p_\vee(x) = 0] = 1$$
$$x \neq (0, ..., 0), \ \Pr_p[p_\vee(x) = 0] = \frac{1}{2^k}$$

To obtain error probability ϵ, let $k = \log(\frac{1}{\epsilon})$. This shows there exists a random $p(x_1, ..., x_n)$ of degree $(\log(\frac{1}{\epsilon}))$ that randomly computes $\bigvee$. We obtain the same bound for AND with $p_\wedge(x) = \prod_{j=1}^k(1 - \sum_{i \in s_j}(1 - x_i))$.

To construct an approximating polynomial for any circuit, replace each AND and OR gate with $p_\wedge(x)$ and $p_\vee(x)$, respectively, and each $\oplus$ gate with the deterministic sum. The total polynomial is constructed using composition of functions in the natural way, with each gate's s_j's sampled from its input wires.

Take each $p_\wedge$ or $p_\vee$ gate to have error probability ϵ/s (s is the size of the circuit.) This makes the degree of the gate $\mathcal{O}(\log(\frac{s}{\epsilon}))$. By the use of constant-size subsets, fan-in cannot increase the degree; only depth causes products to be nested. So the "tree of polynomials" has degree $\leq (\mathcal{O}(\log(\frac{s}{\epsilon})))^d$.

We bound the circuit's probability of error by the probability that any gate produces a wrong output. A union bound is sufficient:

$$\Pr[\text{any gate polynomial outputs a different answer from circuit}] \leq \frac{\epsilon}{s} \cdot s \leq \epsilon.$$

□

The next piece of the puzzle is to show that MAJORITY does *not* have approximating polynomials—that no low-degree polynomial can agree with MAJORITY on greater than a $(1-\epsilon)$ fraction of inputs.

Claim 4. *For all polynomials* $p(x_1, ..., x_n)$*,* $\deg(p) = t$*,*

$$\Pr_{x \in \{0,1\}^n}[p(x) = \text{Maj}(x)] \leq \frac{1}{2} + \mathcal{O}(t/\sqrt{n}).$$

First we show how this claim and the lemma give the main result. Suppose circuit C has size s, depth d. Then by the above lemma, there is a distribution P of polynomials of degree $(\mathcal{O}(\log(\frac{s}{\epsilon})))^d$ with error probability $\leq \epsilon$. This implies that there exists a fixed polynomial p such that $\Pr_x[p(x) = \text{Maj}(x)] \geq 1-\epsilon$. Plugging in the value from the claim, we have:

$$\frac{1}{2} + \mathcal{O}\left(\frac{\log(\frac{s}{\epsilon})^d}{\sqrt{n}}\right) \geq 1-\epsilon$$

To get a concrete bound, we can set $1-\epsilon = 0.9$, and

$$(\log(10s))^d \geq \sqrt{n}$$
$$\Rightarrow \; s \geq 2^{\Omega(n^{1/2d})}.$$

To prove the claim, we first make the observation that every polynomial can be made *multilinear* without changing its evaluation on $\{0,1\}^n$, and without increasing its degree. A multilinear polynomial is one in which no single variable appears with degree > 1. To make a polynomial multilinear, simply replace each x_i^k with x_i. Clearly this gives exactly the same values on $\{0,1\}^n$.

Now we want to show that if Maj had an approximating polynomial of low degree then every $f : \{0,1\}^n \to \{0,1\}$ has an approximating polynomial of low degree.

Lemma 5 (Versatility)**.** $\forall f : \{0,1\}^n \to \{0,1\}$*,* $\exists g, h \in \mathbb{F}_2[x_1, \ldots, x_n]$ *such that*

$$\forall x, f(x) = g(x) \cdot Maj(x) + h(x), \text{ where } \deg(g), \deg(h) \leq n/2.$$

Proof. Let $S_0 = Maj^{-1}(0)$ and $S_1 = Maj^{-1}(1)$. We want to show that these are interpolating sets for polynomials of degree at most $n/2$, that is, for $i = 0, 1$:

$$\forall f : \{0,1\}^n \to \{0,1\}, \exists f_i \text{ such that } f_i|_{S_i} = f|_{S_i} \text{ and } \deg(f_i) \leq n/2.$$

We'll show the argument for $i = 0$; the other case is similar. Consider a square matrix M with rows and columns both indexed by subsets $I \subset [n]$ where $|I| \leq n/2$. Order them so that the sizes are nondecreasing and use this same ordering for both the rows and columns. Associate a row indexed by I with the incidence vector x_I, and for a column indexed by J the monomial $\prod_{j \in J} x_j$. For $I, J \subset [n]$, let

$$M(I, J) = \prod_{j \in J} x_{Ij} \text{ (where } x_{Ij} \text{ denotes the jth component of the incidence vector } x_I\text{).}$$

The key observation to make is that

$$M(I, J) = \begin{cases} 1 & \text{if } I = J \\ 0 & \text{if } |J| > |I| \text{ or } |I| = |J|, I \neq J \end{cases}.$$

Thus M is a lower triangular matrix with 1's along the diagonal. This implies that the collection of monomials

$$\mathcal{J} = \{\prod_{j \in J} x_j : J \subseteq [n], |J| \leq n/2\}$$

is linearly independent over $\mathbb{F}_2^{S_0}$. (See the "independence criterion" in Jukna, p. 188.) Also note that if $\mathcal{F}_0$ denotes the vector space of all functions $f : S_0 \to \{0, 1\}$, then $|\mathcal{J}| = |S_0| = \dim(\mathcal{F}_0)$. The elements of $|\mathcal{J}|$ thus form a linear basis over $\mathcal{F}_0$.

Now for $i = 0, 1$, take f_i of degree at most $n/2$ with $f_i|_{S_i} = f|_{S_i}$. Then

$$\begin{aligned} f &= f_1 \cdot Maj + f_0 \cdot (1 - Maj) \\ &= Maj \cdot (f_1 - f_0) + f_0. \end{aligned}$$

This proves our lemma. □

Let $p(x_1, \ldots, x_n)$ be a polynomial of degree t and let $S = \{x : p(x) = Maj(x)\}$.

Letting $\mathcal{F}$ be the vector space of all functions $f : S \to \{0, 1\}$ and $\mathcal{P}$ be the vector space of all polynomials (in $\mathbb{F}_2[x_1, \ldots, x_n]$) of degree at most $n/2 + t$, the above lemma gives $\mathcal{F} \subseteq \mathcal{P}$ (For $f \in \mathcal{F}$, $f = g \cdot Maj + h = g \cdot p + h$). Since $\mathcal{P}$ is generated by multilinear monomials of degree up to $n/2 + t$,

$$\dim(\mathcal{P}) = \binom{n}{0} + \binom{n}{1} + \ldots \binom{n}{n/2 + t} \leq (1/2 + o(t/\sqrt{(}n))) \cdot 2^n.$$

Combining this with

$$|S| = \dim(\mathcal{F}) \leq \dim(\mathcal{P})$$

proves Claim 4.

3 Average case hardness

We have actually shown that an $\mathsf{AC}^0(\oplus)$ circuit C of size $poly(n)$ cannot have

$$\Pr_{x \in \{0,1\}^n}[C(x) = Maj(x)] > \frac{1}{2} + \frac{poly(\log n)}{\sqrt{n}}.$$

This is an "average case" lower bound.

In a similar vein, Håstad showed that PARITY cannot be approximated by a polynomial-sized AC^0 circuit, giving a much stronger bound for this case.

Theorem 6 (Håstad)**.** *$\forall$ circuits C with AC^0 size $s = 2^{O(n^{1/d})}$ and depth d,*

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

Proof. (SKETCH) We know by the switching lemma that a random restriction ρ makes $C|\rho$ a t-DNF w.p. $\geq 1 - \delta$, where ρ sets some Z variables to z.

$$\Pr_x[C(x) = PARITY(x)] = \Pr_{y,\rho}[C|\rho(y) = PARITY(z, y)]$$

$$\leq (1 - \delta) \Pr_y[\text{some } t\text{-}DNF(y) = PARITY(y)] + \delta.$$

In the homework you will show that t-DNFs have very small correlation with PARITY. □

4 Pseudo-random Generators

We can use the hardness of PARITY to construct a pseudorandom generator against AC^0.

We want a $S \subset \{0,1\}^n$ of small size such that $\forall\, C$, $|C| \leq n^c$,

$$|\Pr_{x \in S}[C(x) = 1] - \Pr_{x \in \{0,1\}^n}[C(x) = 1]| \leq \epsilon. \qquad (1)$$

Note that we've shown by the probabilistic method the existence of such an S (with carefully chosen parameters) but now we'll do one better by providing an explicit one. This was done by Nisan and Wigderson.

First we'll take n subsets $S_1, \ldots, S_n \subseteq [k]$ such that $|S_i| = a$ $\forall i$ and $\forall i \neq j$, $|S_i \cap S_j| \leq b$, with a and b to be determined. Define $G : \{0,1\}^k \to \{0,1\}^n$ as $G(y) = (G_1(y), \ldots, G_n(y))$, where $G_i(y) = \bigoplus_{j \in S_i} y_j$. Let $S = \mathrm{Im}(G)$ (so $|S| \leq 2^k$).

Claim 7. *If $\exists\, C$ such that*

$$|\Pr_{y \in \{0,1\}^k}[C(G(y)) = 1] - \Pr_{x \in \{0,1\}^n}[C(x) = 1]| > \epsilon,$$

then there exists a circuit of size at most $|C| + n * 2^b$ for computing $PARITY(z)$ for a random z w.p. $\geq 1/2 + \epsilon/2n$.

Proof. (By a "hybrid argument")

Let $D_0 = (x_1, \ldots, x_n)$, $D_n = (G_1(y), \ldots, G_n(y))$ and for $1 \leq k < n$, $D_k = (G_1(y), \ldots, G_k(y), x_{k+1}, \ldots, x_n)$. By the hypothesis, there must be an i such that

$$|\Pr[C(D_i) = 1] - \Pr[C(D_{i-1}) = 1]| > \epsilon/n.$$

Without loss of generality, we can assume we have

$$\Pr_{y,x_{i+1},\ldots,x_n}[C(G_1(y),\ldots,G_i(y),x_{i+1},\ldots,x_n)=1]$$
$$-\Pr_{y,x_i,\ldots,x_n}[C(G_1(y),\ldots,G_{i-1}(y),x_i,x_{i+1},\ldots,x_n)=1] > \epsilon/n.$$

So there exists a fixing of the variables $x_{i+1},\ldots,x_n$ that makes this hold. Now

$$\Pr_y[C(G_1(y),G_2(y),\ldots,G_i(y))=1]$$
$$-\Pr_{y,x_i}[C(G_1(y),\ldots,G_{i-1}(y),x_i)=1] > \epsilon/n.$$

Let $y=(w,z)$ where w are the bits outside S_i and z the bits inside. Then

$$\Pr_{w,z}[C(G_1(w,z|_{S_1}),G_2(w,z|_{S_2}),\ldots,G_i(z)=1]$$
$$-\Pr_{w,z,x_i}[C(G_1(w,z|_{S_1}),\ldots,x_i)=1] > \epsilon/n$$

means there exists a w that makes this hold. Fix this w and integrate it into C. Let $f_i = G_i|_w$.

Now

$$\Pr_z[C(f_1(z|_{S_1}),f_2(z|_{S_2}),\ldots,f_{i-1}(z|_{S_{i-1}}),G_i(z))=1]$$
$$-\Pr_z[C(f_1(z|_{S_1}),f_2(z|_{S_2}),\ldots,f_{i-1}(z|_{S_{i-1}}),x_i)=1] > \epsilon/n.$$

Define $C^*(z,r)=C(f_1(z|_{S_1}),f_2(z|_{S_2}),\ldots,f_{i-1}(z|_{S_{i-1}}),r)$.

This says that

$$\Pr_z[C^*(z,PARITY(z))=1]-\Pr_{z,x_i}[C^*(z,x_i)=1] > \epsilon/n.$$

Since we've replaced inputs of C with circuits computing PARITY on at most b variables, $|C^*| \leq |C| + n * 2^b$.

By a theorem of Yao (which we'll see later), Distinguishability ⇒ Predictability. This implies that

$$\exists\ C^{**} \text{ such that } |C^{**}| = |C^*| \text{ and}$$

$$\Pr_z[C^{**}(z)=PARITY(z)] \geq \frac{1}{2}+\frac{\epsilon}{2n}.$$

□

Lemma 8 (Yao: distinguishability ⇒ predictability)**.** *Suppose $C:\{0,1\}^n\times\{0,1\}\to\{0,1\}$ is a circuit such that:*

$$\Pr_{x\in\{0,1\}^n}[C(x,f(x))=1]-\Pr_{x\in\{0,1\}^n,b\in\{0,1\}}[C(x,b)=1] > \delta.$$

Then there exists another circuit $C':\{0,1\}^n\to\{0,1\}$ such that

$$\Pr_{x\in\{0,1\}^n}[C'(x)=f(x)] > \frac{1}{2}+\delta.$$

Proof. This is a tricky statement, and it is a great exercise to try to prove this.

Here is a non-magical (but slightly longer than necessary) way of arriving at the proof.

We are trying to understand exactly what the hypothesis says. Define the following 3 subsets of $\{0,1\}^n$.

$$S = \{x \in \{0,1\}^n \mid f(x) = 0\}.$$
$$T_0 = \{x \in \{0,1\}^n \mid C(x,0) = 1\}.$$
$$T_1 = \{x \in \{0,1\}^n \mid C(x,1) = 1\}.$$

Now draw a Venn diagram of these sets in $\{0,1\}^n$ (they are in general position).

We have:

$$\Pr[C(x,b) = 1] = \frac{1}{2} \cdot \Pr[T_0] + \frac{1}{2} \cdot \Pr[T_1].$$
$$\Pr[C(x,f(x)) = 1] = \Pr[T_0 \cap S] + \Pr[T_1 \cap S^c].$$

Thus if $\Pr[C(x,f(x)) = 1] - \Pr[C(x,b) = 1] > \delta$, then we must have that one of the two possibilities occurs:

- Case 1: $\Pr[T_0 \cap S] > \frac{1}{2}\Pr[T_0] + \frac{\delta}{2}$. In this case, we see that for a random $x \in \{0,1\}^n$, the event $x \in T_0$ occurring indicates that $f(x) = 0$ is slightly more likely than $f(x) = 1$.

  So our circuit $C'(x)$ does the following: If $C(x,0) = 1$ (this means $x \in T_0$), output 0, otherwise output a random bit.

- Case 2: $\Pr[T_1 \cap S^c] > \frac{1}{2}\Pr[T_1] + \frac{\delta}{2}$. In this case, we see that for a random $x \in \{0,1\}^n$, the event $x \in T_1$ occurring indicates that $f(x) = 1$ is slightly more likely than $f(x) = 0$.

  So our circuit $C'(x)$ does the following: If $C(x,1) = 1$ (this means $x \in T_1$), output 1, otherwise output a random bit.

□

If we let $b = \log(n)$, $a = \log^{2d}(n)$, and $k = \log^{4d}(n)$, then we can produce n sets with these parameters (Exercise. Hint: probabilistic method. Later we will construct such sets deterministically). Then a circuit of size $|C| + n * 2^b = |C| + n^2$ approximates PARITY on $\log^{2d}(n)$ bits. Håstad's theorem implies that the size is at least $2^{\log^2(n)}$ if $\epsilon/2n > 2^{-\log^2(n)}$, so take $\epsilon > n * 2^{-\log^2(n)+1}$. Thus we've shown that a circuit of polynomial size must satisfy inequality (1) above for our constructed S.