Probabilistic polynomials for AC^0 (lecture notes)

Course: Circuit Complexity, Autumn 2024, University of Chicago
Instructor: William Hoza (williamhoza@uchicago.edu)

---

1 Polynomials as a computational model

Let $\mathbb{F}$ be a field.[1] The cases we will care about are $\mathbb{F} \in \{\mathbb{R}, \mathbb{F}_2, \mathbb{F}_3\}$. Recall that $\mathbb{F}[x_1, \ldots, x_n]$ denotes the set of n-variate polynomials with coefficients in $\mathbb{F}$. We will think of *polynomials as algorithms*. Over finite fields, all functions are computable in this model:

Proposition 1 (Lagrange interpolator)**.** *Let $\mathbb{F}$ be a finite field, let $n \in \mathbb{N}$, and let $f \colon \mathbb{F}^n \to \mathbb{F}$ be an arbitrary function. There exists a polynomial $p \in \mathbb{F}[x_1, \ldots, x_n]$ such that for every $x \in \mathbb{F}^n$, we have $p(x) = f(x)$.*

Proof.

$$f(x) = \sum_{z \in \mathbb{F}^n} f(z) \cdot \prod_{i=1}^{n} \prod_{a \in \mathbb{F} \setminus \{z_i\}} \frac{x_i - a}{z_i - a}. \qquad \square$$

The *degree* of a polynomial is a measure of its complexity. In the rest of these notes, we will prove that $\mathsf{PARITY} \notin \mathsf{AC}^0$. Our strategy will be to first show that functions in AC^0 can be computed by low-degree *probabilistic polynomials*, and then show that the parity function cannot. This strategy is often called the *Razborov-Smolenksy method*.

2 Probabilistic polynomials

Definition 1 (Probabilistic polynomials)**.** A *probabilistic polynomial* over a field $\mathbb{F}$ is a distribution P over $\mathbb{F}[x_1, \ldots, x_n]$. We say that P computes $f \colon \{0,1\}^n \to \{0,1\}$ with error ε if for every $x \in \{0,1\}^n$, we have

$$\Pr_{p \sim P}[p(x) = f(x)] \geq 1 - \varepsilon.$$

Since we only plug in $x \in \{0,1\}^n$, we can assume without loss of generality that every $p \in \mathsf{Supp}(P)$ is multilinear. We write $\deg(P)$ to denote $\max_{p \in \mathsf{Supp}(P)} \deg(p)$.

Example 1. The function MAJ_3 can be computed with error $1/3$ by a probabilistic polynomial of degree 1:

$$P(x) = \begin{cases} x_1 & \text{with probability } 1/3 \\ x_2 & \text{with probability } 1/3 \\ x_3 & \text{with probability } 1/3. \end{cases}$$

A probabilistic polynomial is a type of randomized algorithm. As usual with randomized algorithms, we require that the algorithm gives the right answer with high probability on *every* input. This is stronger than having a deterministic algorithm that works on most inputs:

Proposition 2 (Deterministic polynomials that are correct on most inputs)**.** *Let $\mathbb{F}$ be a field, let $f \colon \{0,1\}^n \to \{0,1\}$, and suppose f can be computed by a degree-D probabilistic polynomial P with error ε over $\mathbb{F}$. Then for every distribution μ over $\{0,1\}^n$, there exists a (deterministic) degree-D polynomial p over $\mathbb{F}$ such that*

$$\Pr_{x \sim \mu}[p(x) = f(x)] \geq 1 - \varepsilon.$$

[1] More generally, we could work over any *ring*.

Proof. Pick $p \sim P$ and $x \sim \mu$ independently. Then

$$1 - \varepsilon \leq \mathop{\mathbb{E}}_{x \sim \mu} \left[\Pr_{p \sim P}[p(x) = f(x)] \right] = \mathop{\mathbb{E}}_{p \sim P} \left[\Pr_{x \sim \mu}[p(x) = f(x)] \right].$$

The best case is at least as good as the average case, so there is some fixing of p such that $\Pr_{x \sim \mu}[p(x) = f(x)] \geq 1 - \varepsilon$. □

Proposition 2 is actually an "if and only if" condition; this is a special case of Yao's principle.

3 Simulating AC^0 circuits using probabilistic polynomials

We will show that AC^0 circuits can be computed by probabilistic polynomials of polylogarithmic degree, over any field. As a warm-up, let us begin by designing probabilistic polynomials for the NOR function over the field $\mathbb{F}_2$.

Proposition 3 (Warm-up)**.** *For every $n \in \mathbb{N}$ and $\varepsilon > 0$, there exists a probabilistic polynomial over $\mathbb{F}_2$ of degree $\lceil \log(1/\varepsilon) \rceil$ that computes NOR_n with error ε.*

Proof. Pick $S \subseteq [n]$ uniformly at random and let $P(x) = 1 - \sum_{i \in S} x_i$. Then $\Pr[P(0^n) = 1] = 1$, and if $x \neq 0^n$, then $\Pr[P(x) = 0] = 1/2$. Now we amplify: Sample $t = \lceil \log(1/\varepsilon) \rceil$ polynomials from P independently, say $P_1, \ldots, P_t$, and let $P'(x) = P_1(x) \cdot P_2(x) \cdots P_t(x)$. That way, $\Pr[P'(0^n) = 1] = 1$, and if $x \neq 0^n$, then $\Pr[P'(x) \neq 0] = 2^{-t} \leq \varepsilon$. □

Now let's generalize to any field.

Lemma 1 (Probabilistic polynomials for NOR over any field)**.** *For every field $\mathbb{F}$, for every $n \in \mathbb{N}$, and for every $\varepsilon > 0$, there exists a probabilistic polynomial over $\mathbb{F}$ of degree $O(\log n \cdot \log(1/\varepsilon))$ that computes NOR_n with error ε.*

Proof. Randomly sample sets $S_1, S_2, \ldots, S_{1+\log n} \subseteq [n]$ as follows: Independently for each i and j, with probability 2^{-i} we include $j \in S_i$, and with the remaining probability we exclude it. Now define

$$P(x) = \prod_{i=1}^{1+\log n} \left(1 - \sum_{j \in S_i} x_j \right).$$

Then P has degree $1 + \log n$, and $\Pr[P(0^n) = 1] = 1$ with probability 1.

Now let x be a nonzero string, say with Hamming weight $w > 0$. There is some i such that $2^{-i} \in [0.25/w, 0.5/w]$. For this value of i, we have

$$\begin{aligned}
\Pr\left[\sum_{j \in S_i} x_j = 1 \right] \geq \Pr\left[\sum_{j \in S_i} x_j = 1 \text{ over } \mathbb{Z} \right] &= w \cdot 2^{-i} \cdot (1 - 2^{-i})^{w-1} \\
&\geq 0.25 \cdot (1 - 0.5/w)^w \\
&\geq 0.25 \cdot (1 - 0.5) \qquad \text{(Bernoulli's inequality)} \\
&= 1/8.
\end{aligned}$$

Therefore, with probability at least $1/8$, we have $\sum_{j \in S_i} x_j = 1$, and hence $P(x) = 0$.

Finally, we amplify, just like in the proof of Proposition 3: Sample $t = O(\log(1/\varepsilon))$ polynomials from P independently, say $P_1, P_2, \ldots, P_t$, and let $P'(x) = P_1(x) \cdot P_2(x) \cdots P_t(x)$. That way, $\Pr[P'(0^n) = 1] = 1$, and if $x \neq 0^n$, then $\Pr[P'(x) \neq 0] \leq (7/8)^t \leq \varepsilon$. □

Lemma 1 readily implies that every function in AC^0 can be computed by a low-degree probabilistic polynomial. In the theorem below, we use the following convenient notation: an "AC^0_d circuit" is an AC circuit of depth d.

Theorem 1 (Probabilistic polynomials for AC^0)**.** *For every field $\mathbb{F}$, for every $n, S, d \in \mathbb{N}$ with $S \geq n$, for every size-S AC^0_d circuit $C\colon \{0,1\}^n \to \{0,1\}$, and for every $\varepsilon > 0$, there exists a probabilistic polynomial over $\mathbb{F}$ of degree $O((\log S \cdot \log(S/\varepsilon)))^d$ that computes C with error ε.*

Proof sketch. Lemma 1 implies that we can construct probabilistic polynomials for the AND and OR functions with the same parameters, because $\mathsf{OR}(x) = 1 - \mathsf{NOR}(x)$ and $\mathsf{AND}_n(x) = \mathsf{NOR}_n(1 - x_1, \ldots, 1 - x_n)$.

Replace each gate of C with a probabilistic polynomial that computes that gate's operation (AND or OR) with error ε/S and degree $O(\log S \cdot \log(S/\varepsilon))$. Then compose all these polynomials in the manner dictated by the structure of C. The resulting polynomial has degree $O(\log S \cdot \log(S/\varepsilon))^d$, and by the union bound, it computes C with error ε. □

It is an open problem to determine the optimal degree of probabilistic polynomials for the OR and AND functions over $\mathbb{R}$ [BHMS21].

4 PARITY does not have low-degree approximators

We have shown that functions in AC^0 can be computed by low-degree probabilistic polynomials. Next, we will show that the parity function cannot be computed by low-degree probabilistic polynomials over $\mathbb{F}_3$. In fact, we will show something stronger, namely, every low-degree *deterministic* polynomial attempting to compute the parity function has a significant error rate when the *input* is chosen uniformly at random (see Proposition 2).

Theorem 2 (Parity cannot be approximated by low-degree polynomials over $\mathbb{F}_3$)**.** *Let $p\colon \mathbb{F}_3^n \to \mathbb{F}_3$ be a polynomial of degree D. Then*

$$\Pr_{x \in \{0,1\}^n}[p(x) = \mathsf{PARITY}(x)] \leq \frac{1}{2} + O\left(\frac{D}{\sqrt{n}}\right).$$

The first step of the proof is an encoding trick. We would like to work with $+1$ and -1 instead of 0 and 1. Therefore, define $q\colon \mathbb{F}_3^n \to \mathbb{F}_3$ by

$$q(y) = p(y_1 - 1, \ldots, y_n - 1) + 1.$$

Then $\deg(q) = \deg(p)$, and

$$\Pr_{x \in \{0,1\}^n}[p(x) = \mathsf{PARITY}(x)] = \Pr_{y \in \{\pm 1\}^n}[q(y) = y_1 y_2 \cdots y_n].$$

The next step of the proof is to show there is a low-degree *reduction* from arbitrary functions to the parity function.

Lemma 2 (Low-degree reduction from arbitrary functions to the parity function)**.** *Every function $f\colon \{\pm 1\}^n \to \mathbb{F}_3$ can be written in the form $f(y) = p_0(y) + p_1(y) \cdot (y_1 y_2 \cdots y_n)$, where p_0 and p_1 have degree at most $n/2$.*

Proof. The function f can be computed by some polynomial over $\mathbb{F}_3$ (Proposition 1). Furthermore, we can make this polynomial multilinear, because $y_i^2 = 1$ when $y \in \{\pm 1\}^n$. Hence, this polynomial has the form

$$f(y) = \sum_{S \subseteq [n]} c_S \cdot \prod_{i \in S} y_i.$$

We define

$$p_0(y) = \sum_{\substack{S \subseteq [n] \\ |S| \leq n/2}} c_S \cdot \prod_{i \in S} y_i \qquad p_1(y) = \sum_{\substack{S \subseteq [n] \\ |S| > n/2}} c_S \cdot \prod_{i \notin S} y_i. \qquad \square$$

Proof of Theorem 2. Let $S = \{y \in \{\pm 1\}^n : q(y) = y_1 y_2 \cdots y_n\}$. By Lemma 2, every function $f : S \to \mathbb{F}_3$ can be written as $f(x) = p_0(x) + p_1(x) \cdot q(x)$, a polynomial of degree at most $n/2 + D$. The number of functions $f : S \to \mathbb{F}_3$ is $3^{|S|}$. On the other hand, the number of polynomials of degree at most $n/2 + D$ is $3^{\sum_{i=0}^{n/2+D} \binom{n}{i}}$. Therefore, $|S| \leq \sum_{i=0}^{n/2+D} \binom{n}{i} \leq 2^n \cdot (1/2 + O(D/\sqrt{n}))$.[2] □

Corollary 1 ($\mathsf{PARITY} \notin \mathsf{AC}^0$)**.** *Every* AC^0_d *circuit computing* PARITY_n *has size* $2^{n^{\Omega(1/d)}}$.

Proof. If C is a size-S AC^0_d circuit, then C has a 0.1-error probabilistic polynomial for C over $\mathbb{F}_3$ of degree $D = (\log S)^{O(d)}$ (Theorem 1). Therefore, there is a deterministic degree-D polynomial p that computes C on 90% of inputs (Proposition 2). If C computes the parity function, this implies $D \geq \Omega(\sqrt{n})$ (Theorem 2), and hence $S \geq 2^{n^{\Omega(1/d)}}$. □

References

[BHMS21] Siddharth Bhandari, Prahladh Harsha, Tulasimohan Molli, and Srikanth Srinivasan. "On the probabilistic degree of OR over the reals". In: *Random Structures & Algorithms* 59.1 (2021), pp. 53–67. DOI: https://doi.org/10.1002/rsa.20991.

[2] The last step uses the fact that for every n and k, we have $\binom{n}{k} \leq O(2^n/\sqrt{n})$. This can be proven using Stirling's approximation.