$\mathsf{MAJ} \circ \mathsf{AC}^0$ **circuits (lecture notes)**

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

1 Weak polynomial representations

In these notes, we will prove that the parity function cannot be computed, or even approximated, by a small circuit of the form $\mathsf{MAJ} \circ \mathsf{AC}^0$, i.e., a constant-depth circuit with a single majority gate at the top and AND and OR gates elsewhere. The proof builds on the Razborov-Smolensky method: first we will show that $\mathsf{MAJ} \circ \mathsf{AC}^0$ circuits can be simulated by a certain type of low-degree polynomials, and then we will show that the parity function cannot. For this proof, instead of probabilistic polynomials, we will work with so-called *weak polynomial representations.*

Definition 1 (Weak polynomial representation)**.** A *weak polynomial representation* of a function $f\colon \{0,1\}^n \to \{0,1\}$ is a nonzero multilinear polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ such that for every $x \in \{0,1\}^n$, either $p(x) = 0$, or else $\mathrm{sign}(p(x)) = (-1)^{f(x)}$.

In general, when we say that a polynomial is "nonzero," we mean that it has at least one nonzero coefficient. Note that a nonzero polynomial p might satisfy $p(x) = 0$ for every $x \in \{0,1\}^n$! For example, consider $x_1 - x_1^2$. However, such a polynomial could never be multilinear:

Proposition 1 (Nonzero outputs for nonzero polynomials)**.** *If $p \in \mathbb{R}[x_1, \dots, x_n]$ is a nonzero multilinear polynomial, then there exists $x \in \{0,1\}^n$ such that $p(x) \neq 0$.*

Proof. Let x be the indicator for the variables appearing in some minimal nonzero monomial of p. □

Consequently, if p is a weak polynomial representation of f, then there really is at least one point $x \in \{0,1\}^n$ such that $p(x) \neq 0$ and $\mathrm{sign}(p(x)) = (-1)^{f(x)}$.

2 $\mathsf{MAJ} \circ \mathsf{AC}^0$ circuits have low-degree weak polynomial representations

Theorem 1 (Weak polynomial representations for $\mathsf{MAJ} \circ \mathsf{AC}^0$)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$. Assume there exists a size-S $\mathsf{MAJ} \circ \mathsf{AC}^0_d$ circuit C such that $\Pr_x[C(x) = f(x)] \geq 1/2 + \varepsilon$, where $S \geq n$. Then f has a weak polynomial representation of degree at most*

$$n - \Omega(\varepsilon \cdot \sqrt{n}) + (\log S)^{O(d)}.$$

The first step of the proof is the following lemma.

Lemma 1 (Low-degree polynomial that vanishes on a small set)**.** *Let $\varepsilon \in (0,1)$, and let $E \subseteq \mathbb{R}^n$ with $|E| \leq 2^n \cdot (1/2 - \varepsilon)$. There is a nonzero multilinear polynomial $r \in \mathbb{R}[x_1, \dots, x_n]$ of degree at most $n/2 + 1 - \Omega(\varepsilon \cdot \sqrt{n})$ that vanishes on E.*

Proof. Let $E = \{x^{(1)}, \dots, x^{(t)}\}$ and let $D \in \mathbb{N}$ be a parameter that we will choose later. Define a map $\phi\colon \mathbb{R}^{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{D}} \to \mathbb{R}^t$ by the formula

$$\phi(r) = (r(x^{(1)}), \dots, r(x^{(t)})),$$

thinking of r as the list of coefficients of a multilinear polynomial of degree at most D. Then ϕ is a linear transformation. Consequently, provided $\binom{n}{0} + \cdots + \binom{n}{D} > t$, it has a nontrivial kernel, i.e., there is a nonzero real multilinear polynomial r of degree at most D that vanishes on every $x^{(i)}$. If $D = \lceil n/2 - \theta \rceil$ for some $\theta > 0$, then

$$\binom{n}{0} + \cdots + \binom{n}{D} \geq 2^n \cdot \left(\frac{1}{2} - O\left(\frac{\theta}{\sqrt{n}} \right) \right).$$

If we choose a sufficiently small $\theta = \Theta(\varepsilon\sqrt{n})$, we get $\binom{n}{0} + \cdots + \binom{n}{D} > (1/2 - \varepsilon) \cdot 2^n \geq t$. □

Proof of Theorem 1. We may assume without loss of generality that $\varepsilon > 1/\sqrt{n}$, because otherwise the theorem is trivial. Let $C(x) = \mathsf{MAJ}_t(C_1(x), \ldots, C_t(x))$, where $C_1, \ldots, C_t$ are AC^0_d circuits of size at most S and $t \leq S$ and (without loss of generality) t is odd. For each $i \in [t]$, there is a probabilistic polynomial P_i over $\mathbb{R}$ that computes C_i with error $\varepsilon/(2t)$ and degree $(\log S)^{O(d)}$. Consequently, there exist deterministic polynomials $p_1, \ldots, p_t \in \mathbb{R}[x_1, \ldots, x_n]$ such that

$$\Pr_{x \in \{0,1\}^n}[\forall i, p_i(x) = C_i(x)] \geq 1 - \varepsilon.$$

Define

$$E = \{x \in \{0,1\}^n : \exists i, p_i(x) \neq C_i(x)\} \cup \{x \in \{0,1\}^n : C(x) \neq f(x)\},$$

so $|E| \leq 2^n \cdot (1/2 - \varepsilon/2)$. By Lemma 1, there is some nonzero multilinear polynomial $r \in \mathbb{R}[x_1, \ldots, x_n]$ of degree at most $n/2 + 1 - \Omega(\varepsilon \cdot \sqrt{n})$ that vanishes on E. Now define

$$p(x) = \underbrace{(t/2 - p_1(x) - \cdots - p_t(x))}_{(*)} \cdot r(x)^2.$$

Let us show that p is a weak polynomial representation for f.

- We can make p multilinear by replacing each occurrence of x_i^2 with x_i.
- By Proposition 1, there is some point $x \notin E$ such that $r(x) \neq 0$, and consequently $p(x) \neq 0$, hence p is a nonzero polynomial.
- On points $x \in E$, we have $p(x) = r(x) = 0$. Meanwhile, on points $x \notin E$, the expression (*) has the same sign as $(-1)^{f(x)}$ and $r(x)^2 \geq 0$. Thus, on every point $x \in \{0,1\}^n$, either $p(x) = 0$ or else $\operatorname{sign}(p(x)) = (-1)^{f(x)}$.

Finally, note that $\deg(p) = n - \Omega(\varepsilon\sqrt{n}) + (\log S)^{O(d)}$. □

3 Parity does not have a low-degree weak polynomial representation

Theorem 2. *Every weak polynomial representation of* PARITY_n *has degree at least* n.

Proof. Define $\chi \colon \{0,1\}^n \to \{\pm 1\}$ by $\chi(x) = (-1)^{x_1 + x_2 + \cdots + x_n}$. On the one hand, if $p \in \mathbb{R}[x_1, \ldots, x_n]$ is a polynomial of degree less than n, say $p(x) = \sum_{|S|<n} c_S \prod_{i \in S} x_i$, then we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \in \{0,1\}^n}[p(x) \cdot \chi(x)] &= \sum_{|S|<n} c_S \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\left(\prod_{i \in S} x_i \cdot (-1)^{x_i} \right) \cdot \left(\prod_{i \notin S} (-1)^{x_i} \right) \right] \\
&= \sum_{|S|<n} c_S \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\prod_{i \in S} x_i \cdot (-1)^{x_i} \right] \cdot \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[\prod_{i \notin S} (-1)^{x_i} \right] \\
&= 0.
\end{aligned}$$

On the other hand, suppose p is a weak polynomial representation of PARITY_n. By Proposition 1, there is some $x_* \in \{0,1\}^n$ such that $p(x_*) \neq 0$. By the weak representation property, we have $p(x_*) \cdot \chi(x_*) > 0$ and $p(x) \cdot \chi(x) \geq 0$ for all other x. Therefore,

$$\mathop{\mathbb{E}}_{x \in \{0,1\}^n}[p(x) \cdot \chi(x)] > 0.$$

□

Corollary 1 ($\mathsf{PARITY} \notin \mathsf{MAJ} \circ \mathsf{AC}^0$)**.** *If C is a size-S $\mathsf{MAJ} \circ \mathsf{AC}^0_d$ circuit where $S \geq n$, then*

$$\Pr_x[C(x) = \mathsf{PARITY}(x)] \leq 1/2 + \frac{(\log S)^{O(d)}}{\sqrt{n}}.$$

In particular, the success probability is at most 0.8 *for a suitable choice* $S = 2^{n^{\Theta(1/d)}}$.

This proof that $\mathsf{PARITY} \notin \mathsf{MAJ} \circ \mathsf{AC}^0$ is due to Aspnes, Beigel, Furst, and Rudich [ABFR94].

4 Application: The correlation between parity and AC^0

By combining Corollary 1 with Yao's XOR lemma, we can prove that AC^0 circuits do an extremely poor job of approximating the parity function.

Theorem 3 (Non-optimal bound on the correlation between parity and AC^0)**.** *If $C \colon \{0,1\}^n \to \{0,1\}$ is an AC^0_d circuit, then either C has size $2^{n^{\Omega(1/d)}}$, or else*

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

Proof. Let $\mathcal{C}$ be the class of AC^0_d circuits of size S on n bits, for a suitable value $S = 2^{n^{\Theta(1/d)}}$. By Corollary 1, every $C' \in \mathsf{MAJ}_S \circ \mathcal{C} \circ \mathsf{PROJ}_{\sqrt{n}}$ satisfies

$$\Pr_{x \in \{0,1\}^{\sqrt{n}}}[C'(x) = \mathsf{PARITY}_{\sqrt{n}}(x)] \leq \frac{1}{2} + \frac{(\log S)^{O(d)}}{n^{1/4}} \leq 0.8,$$

provided we choose a suitable $S = 2^{n^{\Theta(1/d)}}$. Therefore, by Yao's XOR Lemma, every $C \in \mathcal{C}$ satisfies

$$\Pr_x[C(x) = \mathsf{PARITY}_{\sqrt{n}}^{\oplus \sqrt{n}}(x)] \leq \frac{1}{2} + O\left(\frac{1}{\sqrt{S}}\right) + 0.9^{\sqrt{n}}.$$

But $\mathsf{PARITY}_{\sqrt{n}}^{\oplus \sqrt{n}}$ is simply PARITY_n, and $1/\sqrt{S} = 2^{-n^{\Omega(1/d)}}$, so we are done. ☐

The proof above is due to Klivans [Kli01]. As we will discuss later in this course, the correlation between parity and AC^0 is actually even smaller, namely $2^{-n/O(\log S)^{d-1}}$. Meanwhile, it is an open problem to prove that some function $h \in \mathsf{NP}$ has correlation less than $1/\sqrt{n}$ with $\mathsf{AC}^0[\oplus]$. The function $h = \mathsf{MAJ}_{\sqrt{n}}^{\oplus \sqrt{n}}$ seems like a good candidate.

References

[ABFR94] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. "The expressive power of voting polynomials". In: *Combinatorica* 14.2 (1994), pp. 135–148. ISSN: 0209-9683. DOI: 10.1007/BF01215346.

[Kli01] Adam R. Klivans. "On the derandomization of constant depth circuits". In: *Proceedings of the 5th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2001, pp. 249–260. DOI: 10.1007/3-540-44666-4_28.