1 Distinguishing property of the Parity function

Parity\(_n = \oplus_{i=1}^{n} x_i\) cannot be forced to be a constant function (either always 0, or always 1), unless we fix all \(n\) of its variables. In contrast, we show that every function \(f(x_1, \ldots, x_n) \in AC^0\) can always be forced to become a constant function by fixing very few of its \(n\) variables. This will show that Parity\(_n \not\in AC^0\), one of the most famous circuit complexity results!

We’ll be using the Probabilistic Method to argue that for every \(f(x_1, \ldots, x_n) \in AC^0\) there will always exist a partial assignment of constants to variables, leaving some variables free (unfixed), so that the restricted \(f\) is a constant function (even though it still has some unfixed variables).

For a parameter \(p \in [0,1]\), define a random \(p\)-restriction of variables \(x_1, \ldots, x_n\) as follows: For each \(1 \leq i \leq n\), independently at random, leave \(x_i\) free with probability \(p\), and otherwise, assign \(x_i\) either 0, with probability \((1-p)/2\), or 1, with probability \((1-p)/2\). We denote a random \(p\)-restriction by \(\rho\), and use the notation \(\rho \sim R_p\) to mean that \(\rho\) is a random \(p\)-restriction. Given a boolean function \(f(x_1, \ldots, x_n)\), we denote by \(f|_{\rho}\) the result of restricting \(f\) by \(\rho\), i.e., \(f|_{\rho}\) is \(f\) with some variables fixed (according to \(\rho\)) and some variables left free.

Recall that a \(k\)-CNF \(\phi(x_1, \ldots, x_n)\) is a conjunction of clauses, where each clause is a disjunction of at most \(k\) literals (variables or negated variables). A \(k\)-DNF is a disjunction of terms, where each term is a conjunction of at most \(k\) literals. Both \((k-)CNF\) and DNF are examples of depth-2 \(AC^0\) circuits.

**Lemma 1** (Håstad’s Switching Lemma). For any \(p \in [0,1]\), and any positive integers \(k, t\), if a boolean function \(f(x_1, \ldots, x_n)\) is computable by a \(k\)-CNF, then

\[
\Pr_{\rho \sim R_p}[f|_{\rho} \text{ is not computable by } t\text{-DNF}] \leq (5pk)^t.
\]

Note that the same result is also true if we start with a \(k\)-DNF \(f\), then its random restriction is likely to become a \(t\)-CNF, except with probability at most \((5pk)^t\). Also note that the probability in the Switching Lemma does not depend on the size of the \(k\)-CNF \(f\).
2 Strategy: Hit and simplify

Given an $\mathsf{AC}^0$ circuit of depth $d$ and size $s$, computing some boolean function $f(x_1, \ldots, x_n)$, we will apply a number of random $p$-restrictions (for a parameter $p$ to be determined), and argue that, with high enough probability, the resulting restriction of $f$ will become a depth-2 $\mathsf{AC}^0$ circuit (either CNF or DNF).

Consider the bottom two levels of the given $\mathsf{AC}^0$ circuit. Since we assume that the levels of $\mathsf{AC}^0$ circuit always alternate between ANDs and ORs, the depth-2 sub-circuits at these bottom two levels are all either CNFs or DNFs. To apply the Switching Lemma to them, we need to make them into $k$-CNFs (or $k$-DNFs) first, for some small $k$.

Bounding the bottom fan-in. Suppose these are all CNFs. We can view a given CNF $\phi$ on $n$ variables as a depth-3 circuit AND of ORs of ANDs, with the bottom AND gates of fan-in exactly 2, where each bottom AND gate takes as inputs two copies of the same variable. (For example, a CNF $(x \lor y) \land (z \lor w)$ can be viewed as a depth-3 circuit $(((x \land x) \lor (y \land y)) \land ((z \land z) \lor (w \land w)))$.)

Consider the bottom two levels of this depth-3 circuit. It is a 2-DNF. If we apply a random $p_0$-restriction with $p_0 = 1/20$, we get by the Switching Lemma that this 2-DNF will simplify (after the restriction) to a $k$-CNF, with all but at most $2^{-k}$ probability. Pick $k = 2 \cdot \log s$. Then the probability of not switching becomes at most $1/(s^2)$. We can merge these $k$-CNFs with the top AND gate of our original CNF $\phi$, still getting a $k$-CNF after the merge.

Thus, after this $p_0$-restriction, we get each bottom CNF of our original depth-$d$ $\mathsf{AC}^0$ circuit is likely to become a $k$-CNF. By the Union Bound, the probability that some of at most $s$ bottom CNFs does not become a $k$-CNF after a random $p_0$-restriction is at most $s \cdot 1/s^2 = 1/s$.

Reducing the depth by 1. Now that we have depth-2 sub-circuits of our restricted $\mathsf{AC}^0$ to be $k$-CNFs, we can apply another $p$-random restriction and get, by the Switching Lemma, that each such $k$-CNF becomes a $k$-DNF except with probability $(5pk)^k$. Choose $p = 1/(10k)$. Then this probability is at most $2^{-k} = 1/s^2$. By the Union Bound, the probability that there exists some of at most $s$ $k$-CNFs that doesn’t switch to a $k$-DNF is at most $s \cdot (1/s^2) = 1/s$. Thus, with high probability, we get that all $k$-CNFS at the bottom can be replaced with $k$-DNFs, and then merged with the OR level immediately above. This will yield a new restricted $\mathsf{AC}^0$ circuit of depth $d - 1$.

We repeat for $d - 2$ steps (choosing independent $p$-restrictions in each step), until we get (with high probability) to a final depth-2 circuit, which is a $k$-CNF or a $k$-DNF.
Making the circuit into a constant function. Any given $k$-CNF can be forced to be a constant function by fixing at most $k$ of its variables: simply pick one clause and set all literals in that clause to 0, which will make the whole CNF evaluate to 0. The case of a $k$-DNF is analogous. Thus, if we end up with the restricted $\text{AC}^0$ circuit (which is now a $k$-CNF or $k$-DNF), then we can force it to a constant function by fixing at most $k$ more variables. For this to work, we must ensure that we have more than $k$ variables free (unfixed) at this stage, after all of our random restrictions.

For each variable $x_i$, $1 \leq i \leq n$, the probability it survives all our restrictions is

$$q := p_0 \cdot p^{d-2} = \frac{1}{20} \cdot \left( \frac{1}{10k} \right)^{d-2}.$$  

Since these events are independent for $x_1, \ldots, x_n$, we get (by the Chernoff bound — see below) that, with high probability, the actual number of surviving variables will be very close to the expected number, which is $n \cdot q$. Thus, in particular, there will exist a partial assignment that makes our original $\text{AC}^0$ circuit into a $k$-CNF (or $k$-DNF) while leaving at least $nq/2$ variables free. If $nq/2 > k$, we conclude that this circuit can be made into a constant function while still having some free variables left!

Rewriting $nq/2 > k$, we get

$$nq > 2k \iff n > 40 \cdot (10)^{d-2} \cdot k^{d-1} \iff k < O(n^{1/(d-1)}).$$

As $k = 2 \log s$, we conclude that $s < 2^{O(n^{1/(d-1)})}$. That is, if the size $s$ of our original depth-$d$ $\text{AC}^0$ circuit is less than $2^{O(n^{1/(d-1)})}$, then this circuit can be made into a constant function without fixing all of its variables. As the parity function cannot be made into a constant function this way, we conclude the following:

**Theorem 2.** Parity$_n$ requires depth $d$ $\text{AC}^0$ circuits of size at least $2^{O(n^{1/(d-1)})}$.

That is, Parity requires exponential-size $\text{AC}^0$ circuits!

## 3 Optimality

The lower bound of Theorem 2 is tight. Given $d$, we can construct an $\text{AC}^0$ depth-$d$ circuit of size $\text{poly}(n) \cdot 2^{n^{1/(d-1)}}$ that computes Parity$_n$. For $d = 2$, this is obvious! (Every boolean function on $n$ variables can be computed by a CNF (and DNF) of size $n \cdot 2^n$.)

For $d = 3$, partition the $n$-bit input string into $\sqrt{n}$ blocks of size $\sqrt{n}$ each. Compute the parity of each block by a DNF of size $\sqrt{n} \cdot 2^{\sqrt{n}}$. Compute the parity of the resulting $\sqrt{n}$ bits by a CNF of size $\sqrt{n} \cdot 2^{\sqrt{n}}$. Finally, merge the middle two layers of OR gates to get a depth 3 $\text{AC}^0$ circuit of size $\text{poly}(n) \cdot 2^{\sqrt{n}}$. The case of other depth $d > 2$ is analogous. (Exercise!)

### A Probability basics

We review some very basic concepts and facts from probability theory.
Union bound: Given random events $E_1, \ldots, E_n$, we have

$$\Pr[E_1 \lor E_2 \lor \cdots \lor E_n] \leq \sum_{i=1}^{n} \Pr[E_i].$$

That is, the probability of the union of several events is at most the sum of their individual probabilities.

Random variables and expectations: A random variable $X$ is a mapping from the underlying probability space (of elementary random events) to the real numbers $\mathbb{R}$.

For a random event $E$, we can define the indicator random variable $X_E$, which is 1 if event $E$ occurs, and is 0 otherwise. (For example, a random event is “a flipped coin is “heads””. Then the indicator variable is $X$ taking values in $\{0, 1\}$, which is 1 iff a flipped coin is “heads”.)

If a random variable $X$ takes its values from a set $A$, we define the expectation of $X$ as follows:

$$\Exp[X] = \sum_{a \in A} a \cdot \Pr[X = a].$$

Linearity of expectation: For random variables $X_1, \ldots, X_n$ (not necessarily independent), we have

$$\Exp \left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} \Exp[X_i].$$

Markov’s inequality: For nonnegative random variable $X \geq 0$, and for any $k > 0$,

$$\Pr[X \geq k \cdot \Exp[X]] \leq 1/k.$$

Concentration around the expectation: Markov’s inequality above is a simplest bound on how likely a given random variable to assume a value close to the expectation. It says, for example, that a random variable can be more than twice the expectation at most with probability $1/2$.

For a random variable $X = \sum_{i=1}^{n} X_i$ that is a sum of independent random variables (say each $X_i$ takes values in $\{0, 1\}$), a much stronger bound is given by the Chernoff inequality. In particular, it says that the probability that this $X$ is at least twice the expectation is inverse-exponentially small in the expectation.

Theorem 3 (Chernoff bounds). Let $X_1, \ldots, X_n$ be independent identically distributed random variables such that $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$, for some $0 \leq p \leq 1$. Consider the new random variable $X = \sum_{i=1}^{n} X_i$ with expectation $\mu = pn$. Then, for any $0 < \delta < 1$,

$$\Pr[|X - \mu| > \delta \mu] < 2 \cdot e^{-\delta^2 \mu/3}.$$