The Nisan-Wigderson pseudorandom generator (lecture notes)

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

---

1 Pseudorandom generators

Previously, we proved the following theorem, showing that AC^0 circuits do a very poor job of computing or even approximating the parity function.

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

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

Next, as an application of Theorem 1, we will construct a *pseudorandom generator* (PRG) that fools AC^0 circuits. That is, we will show how to use a small number of truly random bits to sample a long sequence of bits that "appear random" to any AC^0 circuit. To make this precise, let U_n denote the uniform distribution over $\{0,1\}^n$. A PRG is defined as follows.

Definition 1 (Distinguishers, fooling, and PRGs)**.** Let X be a distribution over $\{0,1\}^n$, and let $C\colon \{0,1\}^n \to \{0,1\}$. We say that C *distinguishes* X from U_n with advantage ε if

$$|\Pr[C(X) = 1] - \Pr[C(U_n) = 1]| > \varepsilon.$$

Otherwise, we say that X *fools* C with error ε. A *pseudorandom generator* (PRG) is a function $G\colon \{0,1\}^s \to \{0,1\}^n$. We say that G fools C with error ε if $G(U_s)$ fools C with error ε. The parameter s is called the *seed length* of the PRG.

We will use Theorem 1 to prove the following.

Theorem 2 (PRG that fools AC^0)**.** *For every $n, S, d \in \mathbb{N}$ such that $S \ge n$, for every $\varepsilon \in (0,1)$, there exists a PRG $G\colon \{0,1\}^s \to \{0,1\}^n$ such that:*

- *The generator G fools all AC^0_d circuits of size at most S with error ε.*
- *The seed length is $s = (\log(S/\varepsilon))^{O(d)}$.*
- *Given the parameters n, S, d, ε and a seed $x \in \{0,1\}^s$, the output $G(x)$ can be computed in $\mathrm{poly}(n)$ time.*

Other than Theorem 1, the proof of Theorem 2 uses barely any facts about AC^0. That is, the proof technique is a *general framework* for converting correlation bounds into PRGs, called the "Nisan-Wigderson framework." In these lecture notes, we will focus on the case of AC^0 for simplicity's sake, but you can study a more general formulation of the framework in, for example, Hatami and Hoza's survey [HH24].

2 Generating unpredictable bits

Our goal is to sample pseudorandom bits that are *indistinguishable* from uniform random bits by AC^0 circuits. First, we will show how to sample pseudorandom bits such that each bit is *unpredictable* by AC^0 circuits that get to see all the previous bits.

Definition 2 (Next-bit predictors)**.** Let X be a distribution over $\{0,1\}^n$ and let $i \in [n]$. A *next-bit predictor for X with advantage ε* is a function $C \colon \{0,1\}^{i-1} \to \{0,1\}$, for some $i \in [n]$, such that

$$\Pr[C(X_1, X_2, \ldots, X_{i-1}) = X_i] > \frac{1}{2} + \varepsilon.$$

For example, if we define $G \colon \{0,1\}^{n-1} \to \{0,1\}^n$ by $G(x) = (x, x_1 \oplus x_2 \oplus \cdots \oplus x_{n-1})$, then Theorem 1 immediately implies that next-bit predictors for $G(U_{n-1})$ with non-negligible advantage cannot be computed in AC^0. To improve the seed length, our approach will be to apply the XOR operation to n different *subsets* of the seed bits. We will use the following family of subsets.

Lemma 1 (Nearly disjoint sets)**.** *For every $r, n \in \mathbb{N}$ with $r \leq n$, there exist $S_1, S_2, \ldots, S_n \subseteq [s]$, where $s = O(r^2)$, such that:*

- *$|S_1| = |S_2| = \cdots = |S_n| = r$.*
- *For every $i, j \in [n]$ such that $i \neq j$, we have $|S_i \cap S_j| < \log n$.*
- *Given r and n, the sets $S_1, \ldots, S_n$ can be constructed in* $\mathrm{poly}(n)$ *time.*

Proof. In $\mathrm{poly}(r)$ time, we can find a prime number $p \in [r, 2r]$ via naïve brute-force search. (Such a prime always exists.) For each string $a = (a_0, a_1, \ldots, a_{\log n - 1}) \in \{0,1\}^{\log n} \cong [n]$, we define

$$S_a = \{(x, a_0 + a_1 x + a_2 x^2 + \cdots + a_{\log n - 1} x^{\log n - 1} \bmod p) : x \in [r]\}.$$

In other words, if we let $\mathbb{F}_p$ denote the field of integers modulo p and we define $P_a \in \mathbb{F}_p[x]$ by $P_a(x) = a_0 + a_1 x + \cdots + a_{\log n - 1} x^{\log n - 1}$, then $S_a = \{(x, P_a(x)) : x \in [r]\}$.

Clearly, $|S_a| = r$, and $S_a \subseteq \mathbb{F}_p^2 \cong [p^2]$. Constructing these sets just involves some simple arithmetic. Finally, if $a \neq b$, we have $|S_a \cap S_b| = |\{x \in [r] : P_a(x) - P_b(x) = 0\}|$. Over any field, a nonzero degree-D polynomial can have at most D distinct roots.[1] Consequently, $P_a - P_b$ has at most $\log n - 1$ distinct roots. □

We define $G \colon \{0,1\}^s \to \{0,1\}^n$ by the formula

$$G(x) = \left(\bigoplus_{i \in S_1} x_i, \bigoplus_{i \in S_2} x_i, \ldots, \bigoplus_{i \in S_n} x_i \right), \tag{1}$$

where $S_1, \ldots, S_n \subseteq [s]$ are the sets from Lemma 1, using a value r that we will choose later.

Theorem 3 (The NW PRG is unpredictable)**.** *Let $C \colon \{0,1\}^{j-1} \to \{0,1\}$ be a next-bit predictor for $G(U_s)$ with advantage ε, where G is defined above, and assume that C can be computed by an AC^0_d circuit of size S. Then either $S \geq 2^{r^{\Omega(1/d)}} - \mathrm{poly}(n)$ or else $\varepsilon \leq 2^{-r^{\Omega(1/d)}}$.*

Proof. The definition of next-bit predictors says that

$$\Pr_{x \in \{0,1\}^s} \left[C\left(\bigoplus_{i \in S_1} x_i, \ldots, \bigoplus_{i \in S_{j-1}} x_i \right) = \bigoplus_{i \in S_j} x_i \right] > \frac{1}{2} + \varepsilon.$$

The best case is at least as good as the average case, so there is some way of fixing x_i for all $i \notin S_j$ such that we preserve the advantage:

$$\Pr_{x \in \{0,1\}^{S_j}} \left[C\left(\bigoplus_{i \in S_1} x_i, \ldots, \bigoplus_{i \in S_{j-1}} x_i \right) = \bigoplus_{i \in S_j} x_i \right] > \frac{1}{2} + \varepsilon. \tag{2}$$

[1] Proof by induction on D: Let x_* be a root of a degree-D polynomial P. Perform long division to write $P(x) = (x - x_*) \cdot P'(x) + c$ for some constant c. Then $P(x_*) = c$, so $c = 0$, so $P(x) = (x - x_*) \cdot P'(x)$. If y_* is a root of P with $y_* \neq x_*$, then y_* must be a root of P', because a product of nonzero field elements is always nonzero. By induction, P' has at most $D - 1$ distinct roots.

For each $k \in [j-1]$, define $b_k = \bigoplus_{i \in S_k \setminus S_j} x_i$ (a parity involving only the fixed bits). Define $C' \colon \{0,1\}^{S_j} \to \{0,1\}$ by the rule

$$C'(x) = C\left(b_1 \oplus \bigoplus_{i \in S_1 \cap S_j} x_i, \dots, b_{j-1} \oplus \bigoplus_{i \in S_{j-1} \cap S_j} x_i\right).$$

Then Eq. (2) implies that C' correlates with PARITY_r:

$$\Pr_{x \in \{0,1\}^{S_j}}[C'(x) = \mathsf{PARITY}_r(x)] > \frac{1}{2} + \varepsilon.$$

Each XOR operation $\bigoplus_{i \in S_k \cap S_j} x_i$ can be performed by a polynomial-size brute-force DNF formula, because $|S_k \cap S_j| < \log n$. Therefore, C' can be computed by an AC^0_{d+2} circuit of size $S + \mathrm{poly}(n)$. Consequently, by Theorem 1, either $\varepsilon \leq 2^{-r^{\Omega(1/d)}}$, or else $S \geq 2^{r^{\Omega(1/d)}} - \mathrm{poly}(n)$. □

3 Yao's distinguisher-to-predictor lemma

The last ingredient in the proof of Theorem 2 is the following lemma. We specialize to the case of AC^0 circuits only for simplicity's sake.

Lemma 2 (Yao's distinguisher-to-predictor lemma)**.** *Let $n, d, S \in \mathbb{N}$ and let $\varepsilon \in (0,1)$. Let X be a random variable distributed over $\{0,1\}^n$, and assume there exists an AC^0_d circuit C of size S that distinguishes X from U_n with advantage ε. Then there exists a next-bit predictor for X with advantage $\varepsilon/(2n)$ that is computable by an AC^0_d circuit of size S.*

Proof. The first step is a hybrid argument. Sample $R \sim U_n$ independently of X. For each $i \in \{0, 1, \dots, n\}$, define

$$Y^{(i)} = X_1 X_2 \dots X_i R_{i+1} R_{i+2} \dots R_n,$$

so $Y^{(0)} = R$ and $Y^{(n)} = X$. Then by the triangle inequality,

$$\varepsilon < |\Pr[C(R) = 1] - \Pr[C(X) = 1]| \leq \sum_{i=1}^{n} |\Pr[C(Y^{(i-1)}) = 1] - \Pr[C(Y^{(i)}) = 1]|.$$

Consequently, there is some $i \in [n]$ such that

$$|\Pr[C(Y^{(i-1)}) = 1] - \Pr[C(Y^{(i)}) = 1]| > \frac{\varepsilon}{n}.$$

By flipping the output bit if necessary, we can assume without loss of generality that

$$\Pr[C(Y^{(i)}) = 1] > \Pr[C(Y^{(i-1)}) = 1] + \frac{\varepsilon}{n}. \tag{3}$$

Intuitively, Eq. (3) says that "$C(x) = 1$" is a signal suggesting that the first i bits of x were sampled from the distribution X. This intuition suggests the following randomized next-bit predictor: Given $X_1, \dots, X_{i-1}$:

1. Sample $R \sim U_n$.
2. If $C(X_1 X_2 \dots X_{i-1} R_i R_{i+1} \dots R_n) = 1$, then output R_i.
3. Otherwise, sample $Z \in \{0,1\}$ uniformly at random, and output Z.

Let $\mathsf{Success}$ denote the event that the procedure above correctly outputs X_i. Then

$$\begin{aligned}
\Pr[\mathsf{Success}] &= \Pr[C(Y^{(i-1)}) = 1 \text{ and } R_i = X_i] + \Pr[C(Y^{(i-1)}) = 0 \text{ and } Z = X_i] \\
&= \Pr[C(Y^{(i)}) = 1 \text{ and } R_i = X_i] + \Pr[C(Y^{(i-1)}) = 0 \text{ and } Z = X_i] \\
&= \frac{1}{2} \cdot \Pr[C(Y^{(i)}) = 1] + \frac{1}{2} \Pr[C(Y^{(i-1)}) = 0] && \text{(Independence)} \\
&> \frac{1}{2} \cdot \Pr[C(Y^{(i-1)}) = 1] + \frac{\varepsilon}{2n} + \frac{1}{2} \Pr[C(Y^{(i-1)}) = 0] && \text{(Eq. (3))} \\
&= \frac{1}{2} + \frac{\varepsilon}{2n}.
\end{aligned}$$

So far, we have described a randomized next-bit predictor. By averaging, there is some way to fix the internal randomness in the predictor (R and Z) while preserving the advantage. Now we have three cases based on the fixed values of R and Z.

- If $R_i = Z$, then the predictor is a constant function.
- If $R_i = 1$ and $Z = 0$, then the predictor has the form $C'(X_1 \dots X_{i-1}) = C(X_1 \dots X_{i-1} R_i \dots R_n)$.
- If $R_i = 0$ and $Z = 1$, then the predictor has the form $C'(X_1 \dots X_{i-1}) = 1 - C(X_1 \dots X_{i-1} R_i \dots R_n)$.

In all three cases, the predictor is computable by an AC^0_d circuit of size at most S. □

Proof of Theorem 2. Let G be the generator from Eq. (1), and suppose there exists an AC^0_d circuit of size S that distinguishes $G(U_s)$ from U_n with advantage ε. By Lemma 2, there exists a next-bit predictor for $G(U_s)$ with advantage $\varepsilon/(2n)$, computable by an AC^0_d circuit of size S. By Theorem 3, this implies that either $S \geq 2^{r^{\Omega(1/d)}} - \mathrm{poly}(n)$ or else $\varepsilon/(2n) \leq 2^{-r^{\Omega(1/d)}}$. Either way, we get $r \leq r_* = (\log(S/\varepsilon))^{O(d)}$. Taking a contrapositive, we have shown that if construct G using $r = r_* + 1$, then G fools AC^0_d circuits of size S with error ε. This generator is clearly computable in $\mathrm{poly}(n)$ time, and furthermore, it has seed length $s = O(r^2) = (\log(S/\varepsilon))^{O(d)}$. □

References

[HH24] Pooya Hatami and William Hoza. "Paradigms for Unconditional Pseudorandom Generators". In: *Foundations and Trends in Theoretical Computer Science* 16.1-2 (2024), pp. 1–210. ISSN: 1551-305X. DOI: 10.1561/0400000109.