Natural proofs (lecture notes)

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

---

1 Sipser's program

How can we prove $\mathsf{P} \neq \mathsf{NP}$? "Sipser's program" is the following strategy: Prove $\mathsf{NP} \not\subseteq \mathcal{C}$ for stronger and stronger circuit classes $\mathcal{C}$, until eventually we prove $\mathsf{NP} \not\subseteq \mathsf{P/poly}$, which implies $\mathsf{P} \neq \mathsf{NP}$. For example, in Homework Exercise 1, you proved $\mathsf{NP} \not\subseteq \mathsf{AC}^0_2$; in class, we proved $\mathsf{NP} \not\subseteq \mathsf{AC}^0$ and $\mathsf{NP} \not\subseteq \mathsf{AC}^0[\oplus]$; and in Homework Exercise 6, you proved $\mathsf{NP} \not\subseteq \mathsf{AC}^0[p]$ for every prime p.

Unfortunately, despite many decades of intense effort, Sipser's program has not gone much further than $\mathsf{AC}^0[p]$. For example, it remains an open problem to prove $\mathsf{NP} \not\subseteq \mathsf{TC}^0$. In these notes, we will take a step back and try to reason abstractly about the process of proving circuit lower bounds.

- *Why* haven't we managed to prove $\mathsf{NP} \not\subseteq \mathsf{TC}^0$? What makes AC^0 and TC^0 so different?
- *What will it take* to prove $\mathsf{NP} \not\subseteq \mathsf{TC}^0$? What types of techniques should we explore?

2 Natural proofs

For a function $f\colon \{0,1\}^* \to \{0,1\}$, let us use the notation $f_n\colon \{0,1\}^n \to \{0,1\}$ to denote the restriction of f to the domain $\{0,1\}^n$. Let $\mathcal{C}$ be a class of Boolean functions $f\colon \{0,1\}^* \to \{0,1\}$, such as $\mathcal{C} = \mathsf{AC}^0$ or $\mathcal{C} = \mathsf{TC}^0$. In general, how can one prove $\mathsf{NP} \not\subseteq \mathcal{C}$? It is natural to try the following two-step approach.

1. Prove that functions in $\mathcal{C}$ have some "special property." For example, maybe we can show that functions in $\mathcal{C}$ drastically simplify under random restrictions, or maybe we can show that they can be computed by low-degree probabilistic polynomials.
2. Prove that some function $h \in \mathsf{NP}$ does not have that special property. For example, maybe a good choice is the parity function, or the majority function, or Andreev's function, or 3-SAT.

Actually, it is more standard to reason about the complement property, i.e., we will identify a property that the hard function h *does* have and functions in $\mathcal{C}$ do *not* have. We use the letter H to denote this property (H for "Hard.")

Mathematically, we can model H as a function $H\colon \{0,1\}^* \to \{0,1\}$. Given a function $f_n\colon \{0,1\}^n \to \{0,1\}$, described as an N-bit truth table where $N = 2^n$, the value $H(f_n) \in \{0,1\}$ indicates whether the function f_n has the property H. We say that H is *useful against* $\mathcal{C}$ if $H(f_n) = 0$ for all $f \in \mathcal{C}$ and all sufficiently large $n \in \mathbb{N}$. Clearly, if H is useful against $\mathcal{C}$ and $H(h_n) = 1$ for all n, then $h \notin \mathcal{C}$.

Experience shows that when we can prove a circuit lower bound, we can often construct a closely related property H that is "mathematically nice" in addition to being useful, in the following sense.

Definition 1 (Natural property)**.** Let $H\colon \{0,1\}^* \to \{0,1\}$ and let $\mathcal{H}$ be a complexity class. We say that H is *$\mathcal{H}$-natural* if $H \in \mathcal{H}$ and for every $n \in \mathbb{N}$, when we pick $f_n\colon \{0,1\}^n \to \{0,1\}$ uniformly at random, we have $\Pr[H(f_n) = 1] \geq 2^{-O(n)}$.

The first condition ($H \in \mathcal{H}$) is called *constructivity*. We emphasize that the input to H is an N-bit truth table where $N = 2^n$. So, for example, H is P-natural if $H(f_n)$ can be computed in $2^{O(n)}$ time. Constructivity captures the idea that H is a relatively "concrete" property that we can feasibly reason about.

The second condition ($\Pr[H(f_n) = 1] \geq 2^{-O(n)}$) is called *denseness*. The threshold $2^{-O(n)}$ is just one possible choice; it would also be perfectly reasonable to insist that $\Pr[H(f_n) = 1] \geq 0.99$. This condition captures the idea that the property H represents something truly *special* about the functions in $\mathcal{C}$, i.e., something that distinguishes functions in $\mathcal{C}$ from random functions.

Informally, a *natural proof* of a circuit lower bound is a proof based on a natural (and useful) property.

2.1 Example: Naturalness of the random-restrictions proof that $\mathsf{PARITY} \notin \mathsf{AC}^0$

Theorem 1. *There exists an AC^0_2-natural property H such that H is useful against AC^0 and $H(\mathsf{PARITY}_n) = 1$.*

Proof. Let $n \in \mathbb{N}$ and let $N = 2^n$. For a function $f \colon \{0,1\}^n \to \{0,1\}$, define

$$H(f) = 0 \iff \text{there exists } \rho \in \{0,1,\star\}^n \text{ such that } |\rho^{-1}(\star)| \geq \sqrt{n} \text{ and } f|_\rho \text{ is constant.}$$

Clearly, $H(\mathsf{PARITY}_n) = 1$. To show that H is useful against AC^0, let C is an AC^0_d circuit of size S. By the AC^0 Criticality Theorem, there is a value $p = 1/O(\log S)^{d-1}$ such that if we sample $\rho \sim \mathcal{R}_p$, then with probability at least 0.9, the function $C|_\rho$ is constant. Furthermore, by the Chernoff bound, except with probability $2^{-\Omega(n/O(\log S)^{d-1})}$, we have $|\rho^{-1}(\star)| \geq pn/2$. Consequently, if $H(C) = 1$, then C must have size $2^{n^{\Omega(1/d)}}$.

Next, let us show that H is dense. For any fixed $\rho \in \{0,1,\star\}^n$ with at least $\sqrt{n}$ stars, if we pick $f \in \{0,1\}^N$ uniformly at random, the function $f|_\rho$ is a random Boolean function on at least $\sqrt{n}$ many variables. The probability that it is a constant function is at most $2 \cdot 2^{-2^{\sqrt{n}}}$. There are at most 3^n restrictions ρ, so by the union bound, the probability that $H(f) = 0$ is at most $3^n \cdot 2^{-2^{\sqrt{n}}} = \exp(-\Omega(2^{\sqrt{n}}))$.

Finally, let us show that $H \in \mathsf{AC}^0_2$. For each restriction $\rho \in \{0,1,\star\}^n$ and each $b \in \{0,1\}$, there is a circuit $C_{\rho,b}$ consisting of simply a conjunction of literals such that

$$C_{\rho,b}(f) = 1 \iff f|_\rho \equiv b.$$

We can compute H using the formula

$$\neg H(f) = \bigvee_{\substack{\rho \in \{0,1,\star\}^n \\ |\rho^{-1}(\star)| \geq \sqrt{n}}} \bigvee_{b \in \{0,1\}} C_{\rho,b}(f).$$

This is an AC^0_2 circuit of size $2^{O(n)} = \mathrm{poly}(N)$. □

3 Limitations of AC^0-natural proofs

The following theorem should be contrasted with Theorem 1.

Theorem 2. *Let H be an AC^0-natural property. Then H is not useful against $\mathsf{AC}^0_4[\oplus]$.*

Theorem 2 can be interpreted to mean that any proof showing $\mathsf{NP} \not\subseteq \mathsf{AC}^0[\oplus]$, including the Razborov-Smolensky proofs that we studied in this course, must be at least a little bit "unnatural." The proof of Theorem 2 is based on the Nisan-Wigderson PRG, which we studied earlier in this course. Each output bit of the generator is the XOR of a subset of the seed bits, so the following lemma is hopefully not surprising.

Lemma 1 (Implementing the Nisan-Wigderson PRG to run in $\mathsf{AC}^0[\oplus]$)**.** *Let $n, d, S \in \mathbb{N}$, let $\varepsilon \in (0,1)$, let $N = 2^n$, and assume $S \geq N$. There exists a PRG $G \colon \{0,1\}^s \to \{0,1\}^N$ with the following properties.*

1. *The PRG G fools AC^0_d circuits of size S with error ε.*
2. *For each fixed seed $x \in \{0,1\}^s$, there is an $\mathsf{AC}^0_4[\oplus]$ circuit $C_x \colon \{0,1\}^n \to \{0,1\}$ of size $\mathrm{polylog}(S/\varepsilon)$ such that for every $i \in [N]$, we have $C_x(i) = G(x)_i$.*

The proof of Lemma 1 is almost the same as the Nisan-Wigderson construction and analysis that we did in class. The only real difference is that we should use a finite field of characteristic two instead of using a prime field $\mathbb{F}_p$ to construct nearly-disjoint sets. The details are omitted.

Proof of Theorem 2 using Lemma 1. Let $N \in \mathbb{N}$. By constructivity, H_N can be computed by an AC^0_d circuit of size $S = \mathrm{poly}(N)$, where $d = O(1)$. Let $\varepsilon = \Pr_f[H(f) = 1] = 1/\mathrm{poly}(N)$, where $f\colon \{0,1\}^n \to \{0,1\}$ is chosen uniformly at random. Let $G\colon \{0,1\}^s \to \{0,1\}^N$ be the PRG from Lemma 1 that fools AC^0_d circuits of size $S + N$ with error $\varepsilon/2$. Then

$$\Pr[H(G(U_s)) = 1] \geq \Pr[H(U_N) = 1] - \varepsilon/2 > 0.$$

Therefore, there is some seed $x \in \{0,1\}^s$ such that $H(G(x)) = 1$. By Lemma 1, $G(x)$ is the truth table of an $\mathsf{AC}^0_4[\oplus]$ circuit C_x of size $|C_x| = (\log(SN/\varepsilon))^{O(d)} = \mathrm{poly}(n)$. Therefore, H is not useful against $\mathsf{AC}^0_4[\oplus]$. ☐

4 Limitations of P-natural proofs

In the previous section, we showed that there is no AC^0-natural property that is useful against $\mathsf{AC}^0[\oplus]$. Of course, AC^0 is a relatively weak circuit class, so perhaps it is not very surprising to find that AC^0-natural proofs are limited. Traditionally, we model efficient algorithms using the complexity class P. How powerful are P-natural proofs?

Using the Razborov-Smolensky technique, one can construct P-natural properties that are useful against $\mathsf{AC}^0[\oplus]$. On the other hand, it turns out that P-natural proofs are probably too weak to prove $\mathsf{NP} \not\subseteq \mathsf{TC}^0$. The evidence comes from cryptography. A *pseudorandom function* (PRF) is a distribution $\mathcal{F}$ over functions $f\colon \{0,1\}^m \to \{0,1\}$ that fools every efficient adversary A that only has query access to f, i.e., if we sample $f \sim \mathcal{F}$ and we sample $f'\colon \{0,1\}^m \to \{0,1\}$ uniformly at random, then $\Pr[A^f = 1] \approx \Pr[A^{f'} = 1]$. Naor and Reingold [NR04] constructed a candidate PRF such that:

- The PRF is extremely efficient. In particular, Krause and Lucks showed that $\mathsf{Supp}(\mathcal{F}) \subseteq \mathsf{TC}^0_4$ [KL01].[1]
- The PRF is (seemingly) extremely secure. In particular, it is conjectured that there is some constant $\alpha > 0$ such that the PRF fools adversaries that run in time 2^{m^α} with error 2^{-m^α}.[2]

Proposition 1. *Assume PRFs exist with the parameters described above. Then there does not exist a* P*-natural property that is useful against* TC^0_4.

Proof. We will show the contrapositive. Let $H\colon \{0,1\}^* \to \{0,1\}$ be a P-natural property that is useful against TC^0_4. By P-naturalness, there exists a constant $c > 1$ such that:

- (Constructivity) Given the truth table of a function $f\colon \{0,1\}^n \to \{0,1\}$, the value $H(f)$ can be computed in 2^{cn} time.
- (Density) If $f\colon \{0,1\}^n \to \{0,1\}$ is chosen uniformly at random, then $\Pr[H(f) = 1] \geq 2^{-cn}$.

Now let $\alpha > 0$ be any constant. Let $n \in \mathbb{N}$, let $m = (2cn)^{1/\alpha}$, and let $\mathcal{F}$ be a distribution over functions $f\colon \{0,1\}^m \to \{0,1\}$ such that $\mathsf{Supp}(\mathcal{F}) \subseteq \mathsf{TC}^0_4$. We will describe an attack on the security of $\mathcal{F}$ as a candidate PRF. Given oracle access to $f\colon \{0,1\}^m \to \{0,1\}$:

1. Let g be the first 2^n bits of the truth table of f.
2. Compute g by making 2^n queries.
3. Output $H(g)$.

[1] To be clear about what this means, Naor and Reingold constructed a family of distributions $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \dots$, where $\mathcal{F}_m$ is a distribution over functions $f_m\colon \{0,1\}^m \to \{0,1\}$. Krause and Lucks showed that there is a constant $c \in \mathbb{N}$ such that for all sufficiently large $m \in \mathbb{N}$, every $f_m \in \mathsf{Supp}(\mathcal{F}_m)$ can be computed by a depth-4 majority circuit of size m^c.

[2] Naor and Reingold prove that their PRF is secure under the so-called "decisional Diffie-Hellman assumption."

The running time of the attack described above is $2^n \cdot \mathrm{poly}(n) + 2^{cn} < 2^{m^\alpha}$. When f is chosen uniformly at random, g is also uniform random, and hence the attack accepts with probability at least $2^{-cn} > 2^{-m^\alpha}$. On the other hand, if we choose $f \sim \mathcal{F}$, then $f \in \mathsf{TC}^0_4$, which implies $g \in \mathsf{TC}^0_4$ as well, since $\mathrm{poly}(m) = \mathrm{poly}(n)$. Since H is useful against TC^0_4, we have $H(g) = 0$, assuming n is sufficiently large, so the attack rejects. Therefore, $\mathcal{F}$ is not secure as a PRF. □

The conventional interpretation of Proposition 1 is that we ought to develop more non-natural proof techniques, so that one day we can prove $\mathsf{NP} \not\subseteq \mathsf{TC}^0$. Of course, there are other possibilities: maybe the Naor-Reingold PRF and other candidate PRFs are not actually secure, or maybe $\mathsf{NP} \subseteq \mathsf{TC}^0$.

References

[KL01] Matthias Krause and Stefan Lucks. “On the minimal hardware complexity of pseudorandom function generators”. In: *Proceedings of the 18th Symposium on Theoretical Aspects of Computer Science (STACS)*. 2001, pp. 419–430. DOI: 10.1007/3-540-44693-1_37.

[NR04] Moni Naor and Omer Reingold. “Number-theoretic constructions of efficient pseudo-random functions”. In: *J. ACM* 51.2 (2004), pp. 231–262. ISSN: 0004-5411. DOI: 10.1145/972639.972643.