Impagliazzo's hard-core lemma and Yao's XOR lemma (lecture notes)

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

---

1 Correlation bounds

Previously in this course, we used the Razborov-Smolensky method to prove $\mathsf{PARITY} \notin \mathsf{AC}^0$ and $\mathsf{MAJ} \notin \mathsf{AC}^0[\oplus]$. The proofs actually showed something stronger, namely, that small circuits cannot even *approximately* compute the parity and majority functions. For example, our proof that $\mathsf{MAJ} \notin \mathsf{AC}^0[\oplus]$ actually shows that if C is a size-S $\mathsf{AC}^0_d[\oplus]$ circuit, then

$$\Pr_{x \in \{0,1\}^n}[C(x) = \mathsf{MAJ}_n(x)] \leq \frac{1}{2} + \frac{(\log S)^{O(d)}}{\sqrt{n}}.$$

This type of statement is called a *correlation bound*. In general, if $\Pr_x[C(x) = f(x)] = \frac{1+\varepsilon}{2}$, we say that ε is the "correlation" between C and f.

We will now develop a method for *amplifying* correlation bounds. That is, starting from a "hard function" h that satisfies a mild correlation bound, we will show how to construct a "harder function" h' that satisfies a much stronger correlation bound. Looking ahead, this will eventually enable us to prove that the correlation between the parity function and AC^0 circuits is *exponentially* small, which is much stronger than what the Razborov-Smolensky method gives us. The first step is "Impagliazzo's hard-core lemma," which we discuss in the next section.

2 Impagliazzo's Hard-Core Lemma

Impagliazzo's hard-core lemma can be informally stated as follows. Let $h \colon \{0,1\}^n \to \{0,1\}$, and assume that for every "low-complexity" circuit C, we have

$$\Pr_{x \in \{0,1\}^n}[C(x) = h(x)] \leq 1 - \Omega(1).$$

Then the lemma says there is a set $H \subseteq \{0,1\}^n$ (the "hard core") such that $|H| \geq \Omega(2^n)$ and for every "low-complexity" circuit C, we have

$$\Pr_{x \in H}[C(x) = h(x)] \approx \frac{1}{2}.$$

Thus, the lemma partitions the inputs into the "hard inputs" (H) and the "easy inputs" ($\{0,1\}^n \setminus H$). The existence of the hard core H "explains why" low-complexity circuits attempting to compute h cannot achieve success probability $1 - o(1)$.

Now let us rigorously state and prove the lemma. Instead of a hard-core *set* of inputs, we will actually construct a hard-core *distribution* over inputs. The condition $|H| \geq \Omega(2^n)$ is replaced with the following.

Definition 1 (Dense distributions)**.** Let $\delta \in (0,1]$. A distribution H over $\{0,1\}^n$ is δ-*dense* if for every $y \in \{0,1\}^n$, we have[1]

$$\Pr_{x \sim H}[x = y] \leq \frac{1}{\delta \cdot 2^n}.$$

---

[1] If you're familiar with the concept of "min-entropy," a δ-dense distribution is a distribution with at least $n - \log(1/\delta)$ bits of min-entropy.

Lemma 1 (Impagliazzo's Hard-Core lemma)**.** *For every $\varepsilon, \delta > 0$, there is a value $t = O(\frac{\log(1/\delta)}{\varepsilon^2})$ such that the following holds. Let $\mathcal{C}$ be a class of functions $C\colon \{0,1\}^n \to \{0,1\}$. Let $h\colon \{0,1\}^n \to \{0,1\}$, and assume that for every $C \in \mathsf{MAJ}_t \circ \mathcal{C}$, we have*

$$\Pr_x[C(x) = h(x)] \leq 1 - 2\delta.$$

Then there is a δ-dense distribution H over $\{0,1\}^n$ such that for every $C \in \mathcal{C}$, we have

$$\Pr_{x \sim H}[C(x) = h(x)] \leq 1/2 + \varepsilon.$$

The proof uses von Neumann's minimax theorem from the theory of zero-sum games, stated below.

Theorem 1 (Von Neumann's Minimax Theorem)**.** *Let $\mathcal{S}, \mathcal{C}$ be finite nonempty sets and let $\phi\colon \mathcal{S} \times \mathcal{C} \to \mathbb{R}$. [Interpretation: Alice picks $S \in \mathcal{S}$, Bob picks $C \in \mathcal{C}$, and Bob receives payoff $\phi(S, C)$.] Let $c \in \mathbb{R}$, and assume that for every distribution $\mu_{\mathcal{S}}$ over $\mathcal{S}$, there exists $C \in \mathcal{C}$ such that*

$$\mathop{\mathbb{E}}_{S \sim \mu_{\mathcal{S}}}[\phi(S, C)] > c.$$

Then there exists a distribution $\mu_{\mathcal{C}}$ over $\mathcal{C}$ such that for every $S \in \mathcal{S}$, we have

$$\mathop{\mathbb{E}}_{C \sim \mu_{\mathcal{C}}}[\phi(S, C)] > c.$$

We omit the proof of Theorem 1. Let us now use Theorem 1 to prove Lemma 1.

Proof of Impagliazzo's Hard-Core Lemma (Lemma 1). We will prove the contrapositive. Assume that for every δ-dense distribution H over $\{0,1\}^n$, there exists $C \in \mathcal{C}$ such that

$$\Pr_{x \sim H}[C(x) = h(x)] > 1/2 + \varepsilon.$$

Consider the following two-player game.

- Alice chooses a set $S \subseteq \{0,1\}^n$ with $|S| \geq \delta \cdot 2^n$. Let $\mathcal{S}$ be the collection of all such sets.
- Bob chooses a circuit $C \in \mathcal{C}$.
- Bob receives payoff $\phi(S, C) := \Pr_{x \in S}[C(x) = h(x)]$.

To show that the hypothesis of Theorem 1 is satisfied, let μ_S be any distribution over $\mathcal{S}$. Let H be the distribution over $\{0,1\}^n$ that is sampled by first sampling $S \sim \mu_S$, and then sampling $x \in S$ uniformly at random. Then H is δ-dense, because every S in the support of μ_S has size at least $\delta \cdot 2^n$. Therefore, there exists $C \in \mathcal{C}$ such that

$$\mathop{\mathbb{E}}_{S \sim \mu_{\mathcal{S}}}[\phi(S, C)] = \Pr_{x \sim H}[C(x) = h(x)] > 1/2 + \varepsilon.$$

This shows that the hypothesis of Theorem 1 is satisfied. Therefore, by Theorem 1, there exists a distribution $\mu_{\mathcal{C}}$ over $\mathcal{C}$ such that for every $S \in \mathcal{S}$, we have

$$\mathop{\mathbb{E}}_{C \sim \mu_{\mathcal{C}}}\left[\Pr_{x \in S}[C(x) = h(x)]\right] = \mathop{\mathbb{E}}_{x \in S}\left[\Pr_{C \sim \mu_{\mathcal{C}}}[C(x) = h(x)]\right] > 1/2 + \varepsilon.$$

Define

$$\mathsf{BAD} = \left\{x \in \{0,1\}^n : \Pr_{C \sim \mu_{\mathcal{C}}}[C(x) = h(x)] \leq 1/2 + \varepsilon\right\}.$$

Then evidently $\mathsf{BAD} \notin \mathcal{S}$, i.e., $|\mathsf{BAD}| < \delta \cdot 2^n$.

Now sample t circuits $C_1, \ldots, C_t \sim \mu_{\mathcal{C}}$ independently and let $C(x) = \mathsf{MAJ}_t(C_1(x), \ldots, C_t(x))$. For each $x \notin \mathsf{BAD}$, by Hoeffding's inequality, we have

$$\Pr_{C_1,\ldots,C_t \sim \mu_{\mathcal{C}}}[C(x) \neq h(x)] \leq \exp(-2\varepsilon^2 t).$$

Therefore, if we choose $x \in \{0,1\}^n$ uniformly at random, then

$$\Pr_{\substack{x \in \{0,1\}^n \\ C_1,\ldots,C_t \sim \mu_{\mathcal{C}}}}[C(x) \neq h(x)] \leq \exp(-2\varepsilon^2 t) + \frac{|\mathsf{BAD}|}{2^n} < 2\delta,$$

provided we choose a suitable value $t = O(\log(1/\delta)/\varepsilon^2)$. There is some fixing of $C_1, \ldots, C_t$ that preserves the success probability (the best case is at least as good as the average case). Therefore, there exists $C \in \mathsf{MAJ}_t \circ \mathcal{C}$ such that $\Pr_x[C(x) = h(x)] > 1 - 2\delta$, completing the proof. □

3 Yao's XOR Lemma

For a function $h \colon \{0,1\}^n \to \{0,1\}$ and a number $k \in \mathbb{N}$, we define $h^{\oplus k} \colon \{0,1\}^{nk} \to \{0,1\}$ by the rule

$$h^{\oplus k}(x^{(1)}, \ldots, x^{(k)}) = \bigoplus_{i=1}^{k} h(x^{(i)}).$$

Yao's XOR lemma can be informally stated as follows. If every "low-complexity" circuit C satisfies

$$\Pr_{x \in \{0,1\}^n}[C(x) = h(x)] \leq 1 - \Omega(1),$$

then every "low-complexity" circuit C satisfies

$$\Pr_{x \in \{0,1\}^{nk}}[C(x) = h^{\oplus k}(x)] \leq \frac{1}{2} + 2^{-\Omega(k)}.$$

To make this precise, we introduce the following definition.

Definition 2 (Projections)**.** Let PROJ_n denote the class of functions $f \colon \{0,1\}^n \to \{0,1\}^m$ that can be computed by "circuits consisting only of wires." That is, each output bit is either a literal or a constant.

Lemma 2 (Yao's XOR Lemma)**.** *For every $\varepsilon, \delta > 0$, there is a value $t = O(\frac{\log(1/\delta)}{\varepsilon^2})$ such that the following holds. Let $n, k \in \mathbb{N}$, let $\mathcal{C}$ be a class of functions $C \colon \{0,1\}^{nk} \to \{0,1\}$ that is closed under complementation,*[2] *let $h \colon \{0,1\}^n \to \{0,1\}$, and assume that for every $C \in \mathsf{MAJ}_t \circ \mathcal{C} \circ \mathsf{PROJ}_n$, we have*

$$\Pr_x[C(x) = h(x)] \leq 1 - 2\delta.$$

Then for every $C \in \mathcal{C}$, we have

$$\Pr_x[C(x) = h^{\oplus k}(x)] \leq \frac{1}{2} + \varepsilon + (1-\delta)^k.$$

We will use Impagliazz's Hard-Core Lemma to prove Yao's XOR Lemma. The first step of the proof is an alternative characterization of δ-dense distributions.

Lemma 3 (Dense distributions vs. the uniform distribution)**.** *Let H be a δ-dense distribution over $\{0,1\}^n$. There exists a distribution E over $\{0,1\}^n$ such that the following two distributions are identical:*

1. *Sample $x \in \{0,1\}^n$ uniformly at random.*
2. *With probability δ, sample $x \sim H$, and with probability $1-\delta$, sample $x \sim E$.*

[2] I.e., if $C \in \mathcal{C}$, then $\neg C \in \mathcal{C}$.

Proof. Let us identify probability distributions with their probability mass functions. Let

$$E(x) = \frac{2^{-n} - \delta \cdot H(x)}{1-\delta}.$$

Then $\sum_x E(x) = 1$ because H is a distribution, and $E(x) \geq 0$ for all x because H is δ-dense. Therefore, E is a valid probability distribution, and for every $x \in \{0,1\}^n$, we have

$$2^{-n} = \delta \cdot H(x) + (1-\delta) \cdot E(x). \qquad \square$$

Proof of Yao's XOR Lemma (Lemma 2). By Impagliazzo's Hard-Core Lemma, there is a δ-dense distribution H such that for every $C \in \mathcal{C} \circ \mathsf{PROJ}_n$ and every $b \in \{0,1\}$, we have

$$\Pr_{x \sim H}[C(x) = h(x) \oplus b] \leq \frac{1}{2} + \varepsilon.$$

(Recall that $\mathcal{C}$ is closed under complementation.) Let E be the corresponding distribution from Lemma 3. Then sampling $x = (x^{(1)}, \ldots, x^{(k)}) \in \{0,1\}^{nk}$ uniformly at random is equivalent to the following:

1. Sample $S \subseteq [k]$ by including each index independently with probability δ.
2. For each $i \in S$, sample $x^{(i)} \sim H$.
3. For each $i \notin S$, sample $x^{(i)} \sim E$.

For any $C \in \mathcal{C}$, we have

$$\Pr_x[C(x) = h^{\oplus k}(x)] \leq \Pr[S = \varnothing] + \Pr_x[C(x) = h^{\oplus k}(x) \mid S \neq \varnothing].$$

The first term is $(1-\delta)^k$. To bound the second term, fix any $S \neq \varnothing$, and assume for simplicity that $S = [k']$ for some $k' \in [k]$. Then

$$\Pr_{\substack{x^{(1)},\ldots,x^{(k')} \sim H \\ x^{(k'+1)},\ldots,x^{(k)} \sim E}}[C(x) = h^{\oplus k}(x)] = \mathop{\mathbb{E}}_{\substack{x^{(2)},\ldots,x^{(k')} \sim H \\ x^{(k'+1)},\ldots,x^{(k)} \sim E}}\left[\Pr_{x^{(1)} \sim H}\left[C(x) = h(x^{(1)}) \oplus h(x^{(2)}) \oplus \cdots \oplus h(x^{(k)})\right]\right].$$

The inner probability is always at most $1/2 + \varepsilon$, because for any fixing of $x^{(2)}, \ldots, x^{(k)}$, the function $C'(x^{(1)}) = C(x^{(1)}, \ldots, x^{(k)})$ is in $\mathcal{C} \circ \mathsf{PROJ}_n$. $\square$