The switching lemma (lecture notes)

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

---

1 The AC^0 criticality theorem

For a Boolean function f, let $\mathsf{DTDepth}(f)$ denote the minimum depth of a decision tree that computes $f(x)$ by making queries to x. Recall that R_p denotes a random restriction with $\star$-probability p. The following powerful theorem describes the effect of random restrictions on AC^0 circuits.

Theorem 1 (AC^0 Criticality Theorem [Ros17])**.** *Let C be a size-S AC^0_d circuit, let $p \in (0,1)$, and let $D \in \mathbb{N}$. Then*

$$\Pr_{\rho \sim R_p}[\mathsf{DTDepth}(C|_\rho) \geq D] \leq (p \cdot O(\log S)^{d-1})^D.$$

We will not prove Theorem 1 in this course. Instead, we will prove a famous simpler variant called the "switching lemma." Before stating and proving the switching lemma, however, let us illustrate how to use Theorem 1 to prove a very strong bound on the correlation between AC^0 circuits and the parity function.

1.1 Optimal correlation bounds for the parity function

Lemma 1 (Shallow decision trees are uncorrelated with the parity function)**.** *If $T\colon \{0,1\}^n \to \{0,1\}$ is a depth-$(n-1)$ decision tree, then*

$$\Pr_{x \in \{0,1\}^n}[T(x) = \mathsf{PARITY}_n(x)] = \frac{1}{2}.$$

Proof sketch. We can imagine simulating the decision tree and choosing the bits of x "on the fly." That is, whenever the tree tries to query some bit x_i, we toss a coin to decide what x_i is. If the tree makes fewer than n queries, then when the process finishes, the tree outputs an answer b, and at least one of the bits of x is still undetermined. Then, with respect to the random choice of that last bit (or bits), we have $\Pr[\mathsf{PARITY}_n(x) = b] = \frac{1}{2}$. □

Theorem 2 (Correlation between AC^0 and the parity function)**.** *If $C\colon \{0,1\}^n \to \{0,1\}$ is a size-S AC^0_d circuit, then*

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

In particular, if C computes PARITY_n on all inputs, then $S \geq 2^{\Omega(n^{1/(d-1)})}$.

For example, when $d = 3$, the size bound is $S \geq 2^{\Omega(\sqrt{n})}$. It is an open problem to prove that there exists $h \in \mathsf{NP}$ such that every AC^0_3 circuit computing h has size $2^{\omega(\sqrt{n})}$.

Proof. If we sample $\rho \sim R_p$, then

$$\Pr_{x \in \{0,1\}^n}[C(x) = \mathsf{PARITY}_n(x)] = \mathop{\mathbb{E}}_{\rho}\left[\Pr_{x \in \{0,1\}^n}[C|_\rho(x) = (\mathsf{PARITY}_n)|_\rho(x)]\right]$$

Now, $(\mathsf{PARITY}_n)|_\rho$ is the parity function on $|\rho^{-1}(\star)|$ variables (or the negation of that function). Consequently, by Lemma 1, we have

$$\begin{aligned}\Pr_{x \in \{0,1\}^n}[C(x) = \mathsf{PARITY}_n(x)] &\leq \frac{1}{2} + \Pr_\rho[\mathsf{DTDepth}(C|_\rho) \geq |\rho^{-1}(\star)|] \\ &\leq \frac{1}{2} + \Pr_\rho[\mathsf{DTDepth}(C|_\rho) \geq pn/2] + \Pr_\rho[|\rho^{-1}(\star)| \leq pn/2].\end{aligned}$$

If we choose a suitable value $p = 1/O(\log S)^{d-1}$, then the AC^0 Criticality Theorem (Theorem 1) tells us that the second term is at most $2^{-n/O(\log S)^{d-1}}$. Meanwhile, the third term is also at most $2^{-n/O(\log S)^{d-1}}$ by the Chernoff bound.[1] $\square$

2 The switching lemma

Recall that a DNF formula is a disjunction of *terms*, each of which is a conjunction of literals. The *width* of the formula is the maximum number of literals in any term.

Lemma 2 (The Switching Lemma)**.** *Let C be a width-w DNF formula, let $p \in (0,1)$ and let $D \in \mathbb{N}$. Then*

$$\Pr_{\rho \sim R_p}[\mathsf{DTDepth}(C|_\rho) \geq D] \leq O(pw)^D.$$

Several interrelated proofs of the switching lemma are known. The proof we will present is most closely related to the work of Kelley [Kel21].

We begin by presenting a decision tree CDT_ρ[2] that computes $C|_\rho(x)$ by making queries to $x \in \{0,1\}^n$. The algorithm is probably the first thing you would think of: Query all the living variables in the first living term, and then repeat, until either we find a satisfied term, or else we run out of terms. In more detail, let $C_1, \dots, C_S$ be the terms of C, and let $V_i \subseteq [n]$ be the set of variables that appear in C_i. The algorithm CDT_ρ is as follows.

1. Initialize $\pi \leftarrow \rho$. For $t = 1, 2, 3, \dots$:
   (a) If there is a $b \in \{0,1\}$ such that $C|_\pi \equiv b$, then halt and output b. Otherwise, find the first term $i_t \in [S]$ such that $C_{i_t}|_\pi \not\equiv 0$.
   (b) Let Q_t be the set of living variables in C_{i_t}, i.e., $Q_t = V_{i_t} \cap \pi^{-1}(\star)$.
   (c) For every $j \in Q_t$, query x_j and set $\pi_j \leftarrow x_j$.

To analyze CDT_ρ, we will design a strategy for guessing many points in $\rho^{-1}(\star)$ given only a uniform random *completion* of ρ, i.e., a string $y \in \{0,1\}^n$ that agrees with ρ on $\rho^{-1}(\{0,1\})$. On the one hand, y is independent of $\rho^{-1}(\star)$, so all such strategies must be trivial. On the other hand, we will show that our strategy has a good success probability conditioned on CDT_ρ being deep. This will enable us to conclude that CDT_ρ is shallow with high probability.

Let $d \in \mathbb{N}$. Our guessing strategy, denoted $\mathsf{StarGuesser}_d$, is as follows.

1. Pick $x \in \{0,1\}^n$ uniformly at random.
2. Pick a decomposition of d into positive integers, $d = d_1 + d_2 + \cdots + d_r$, uniformly at random.
3. Initialize $z \leftarrow y$. For $t = 1, 2, \dots, r$:
   (a) Find the first term $\widehat{i}_t \in [S]$ such that $C_{\widehat{i}_t}(z) = 1$ (or output "fail" if none exists).
   (b) Pick a size-d_t subset $\widehat{Q}_t \subseteq V_{\widehat{i}_t}$ uniformly at random (or output "fail" if $|V_{\widehat{i}_t}| < d_t$).
   (c) For every $j \in \widehat{Q}_t$, set $z_j \leftarrow x_j$.
4. Output $\widehat{Q}_1 \cup \cdots \cup \widehat{Q}_r$.

[1] One form of the Chernoff bound says that if $X_1, \dots, X_n \in [0,1]$ are independent random variables, and $\mathbb{E}[X_1 + \cdots + X_n] = pn$, then $\Pr[X_1 + \cdots + X_n \leq (1-\varepsilon)pn] \leq e^{-\varepsilon^2 pn/2}$.

[2] CDT stands for "Canonical Decision Tree."

Claim 1 (Correctness of $\mathsf{StarGuesser}_d$)**.** *Let Win_d denote the event that $\mathsf{StarGuesser}_d$ successfully outputs d distinct points, all of which are in $\rho^{-1}(\star)$. Then*

$$\Pr[\mathsf{Win}_d \mid \mathsf{Depth}(\mathsf{CDT}_\rho) = d] \geq \frac{4}{(8w)^d}.$$

(The probability above is with respect to the random choices of ρ and y and the internal randomness of $\mathsf{StarGuesser}_d$.)

Proof. Fix any choice of ρ such that $\mathsf{Depth}(\mathsf{CDT}_\rho) = d$. With probability at least 2^{-d+1}, the strategy $\mathsf{StarGuesser}_d$ picks an input $x \in \{0,1\}^n$ on which CDT_ρ makes d queries. Fix any such x. Let $i_1, \dots, i_r \in [S]$ be the terms visited by CDT_ρ on x, and let $Q_1, \dots, Q_r$ be the sets of variables queried by CDT_ρ on x in those r iterations. With probability 2^{-d+1}, the strategy $\mathsf{StarGuesser}_d$ chooses $d_t = |Q_t|$ for every $t \in [r]$. Assume this occurs.

We can write each term C_i in the form $C_i(x) = \bigwedge_{j \in V_i}(x_j \oplus b_{i,j})$, where $b_{i,j} \in \{0,1\}$. With probability 2^{-d}, we choose a completion y of ρ such that for every $t \in [r]$ and every $j \in Q_t$, we have $y_j \oplus b_{i,j} = 1$. Fix any such y.

Now let us consider the random choices of $\widehat{Q}_1, \dots, \widehat{Q}_r$. For each $t \in [r]$, let E_t be the event that $\widehat{i}_t = i_t$ and $\widehat{Q}_t = Q_t$. Suppose $E_1, \dots, E_{t-1}$ all occur, and consider the beginning of iteration t. At this point, z is a completion of ρ that agrees with x on $Q_1 \cup \cdots \cup Q_{t-1}$. Therefore, based on the way $\mathsf{CDT}_\rho(x)$ chooses i_t, we see that $C_1(z) = C_2(z) = \cdots = C_{i_t - 1}(z) = 0$, and that $z_j \oplus b_{i_t,j} = 1$ for every $j \in V_{i_t} \setminus Q_t$. Furthermore, our assumption on y implies that $z_j \oplus b_{i_t,j} = 1$ for every $j \in Q_t$ as well. Therefore, $C_{i_t}(z) = 1$, hence $\widehat{i}_t = i_t$. Consequently, $\Pr[E_t \mid E_1, \dots, E_{t-1}] = 1/\binom{|V_{i_t}|}{d_t} \geq 1/w^{d_t}$, hence $\Pr[E_1, \dots, E_r] \geq 1/w^{d_1 + \cdots + d_r} = 1/w^d$. Finally, note that if $E_1, \dots, E_r$ occur, then $\mathsf{StarGuesser}_d$ outputs $Q_1 \cup \cdots \cup Q_r$, which is indeed a size-d subset of $\rho^{-1}(\star)$. □

Proof of the Switching Lemma (Lemma 2). For each $d \in \mathbb{N}$, we have

$$\frac{4}{(8w)^d} \leq \Pr[\mathsf{Win}_d \mid \mathsf{Depth}(\mathsf{CDT}_\rho) = d] \leq \frac{\Pr[\mathsf{Win}_d]}{\Pr[\mathsf{Depth}(\mathsf{CDT}_\rho) = d]} \leq \frac{p^d}{\Pr[\mathsf{Depth}(\mathsf{CDT}_\rho) = d]},$$

where the last step uses the fact that the output of $\mathsf{StarGuesser}_d$ is independent of $\rho^{-1}(\star)$. (We could choose $y \in \{0,1\}^n$ uniformly at random first, then run $\mathsf{StarGuesser}_d$, and then choose $\rho^{-1}(\star)$ last.) Rearranging, we get $\Pr[\mathsf{Depth}(\mathsf{CDT}_\rho) = d] \leq 0.25 \cdot (8wp)^d$. We may assume without loss of generality that $16pw \leq 1$, because otherwise the switching lemma is trivial. Therefore,

$$\Pr[\mathsf{DTDepth}(C|_\rho) \geq D] \leq \sum_{d=D}^{\infty} \Pr[\mathsf{Depth}(\mathsf{CDT}_\rho) = d] \leq \frac{1}{4} \cdot \sum_{d=D}^{\infty} (8pw)^d \leq 0.5 \cdot (8pw)^D. \qquad \square$$

References

[Kel21] Zander Kelley. "An improved derandomization of the switching lemma". In: *Proceedings of the 53rd Annual Symposium on Theory of Computing (STOC)*. 2021, 272–282. DOI: 10.1145/3406325.3451054.

[Ros17] Benjamin Rossman. "An entropy proof of the switching lemma and tight bounds on the decision-tree size of AC^0". 2017. URL: https://users.cs.duke.edu/~br148/logsize.pdf.