CSG399: Gems of Theoretical Computer Science. Lectures 6-7. Jan. 27-30, 2009
Instructor: Emanuele Viola Scribe: Eric Miles

Parity requires large constant-depth circuits (II)

Correlation bounds for parity

1 Computing parity with constant-depth circuits

In this section we will build on the last two lectures and finish the proof that computing PARITY requires large constant-depth circuits.

Recall from last lecture that we say a polynomial p over $\{-1,1\}^n$ *weakly computes* $\prod_i x_i$ if $p \not\equiv 0$ and $\forall x: \ p(x) \neq 0 \Rightarrow \text{SIGN}(p(x)) = \prod_i x_i$. We will now see how to transform a polynomial which computes $\prod_i x_i$ with high probability into one which weakly computes $\prod_i x_i$. Intuitively, this is done by "zeroing out" the points on which the product function is not correctly computed.

Lemma 1. *Let p be a polynomial over $\{-1,1\}^n$ which computes $\prod_i x_i$ correctly on more than half of its inputs. Specifically, let $S \subset \{-1,1\}^n$ be the subset of size $|S| = 2^n \cdot \gamma$ (for a constant $\gamma < 1/2$) such that $x \in S \Leftrightarrow p(x) \neq \prod_i x_i$. Then, there exists a polynomial $\bar{p}$ such that $\text{SIGN}(\bar{p}(x))$ weakly computes $\prod_i x_i$, and degree($\bar{p}$) = degree(p) $+ n - \epsilon'\sqrt{n}$, for a constant ϵ' which depends only on γ.*

Proof. We first define a polynomial q such that $q \not\equiv 0$ and $x \in S \Rightarrow q(x) = 0$. Let M be the set of all monomials of size at most $\frac{1}{2}(n - \epsilon'\sqrt{n})$ (for a parameter ϵ' to be chosen later) over variables $x_1, \ldots, x_n$ each having degree at most 1. Formally,

$$M = \left\{ \prod_{i \in I} x_i : I \subset [n], |I| \leq \frac{1}{2}(n - \epsilon'\sqrt{n}) \right\}.$$

Then, let $q(x) = \sum_{m \in M} (a_m \cdot m(x))$ be the weighted sum of these monomials, for a set of weights $\{a_m\}_{m \in M}$. We want to choose the weights in order to give q the properties already stated, which is equivalent to finding a non-trivial solution to a certain system of equations. Denote $S = \{s_1, \ldots, s_{|S|}\}$ and $M = \{m_1, \ldots, m_{|M|}\}$. Then, the system of equations we would like to solve is

$$\begin{pmatrix} m_1(s_1) & m_2(s_1) & \cdots & m_{|M|}(s_1) \\ m_1(s_2) & m_2(s_2) & \cdots & m_{|M|}(s_2) \\ \vdots & & \ddots & \\ m_1(s_{|S|}) & m_2(s_{|S|}) & \cdots & m_{|M|}(s_{|S|}) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{|M|} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

From linear algebra, we know that this system has a non-trivial solution if there are more variables (the a_is) than equations (rows in the matrix). Here, this is equivalent to $|M| > |S|$, which we will now show.

$$
\begin{aligned}
|M| &= \textstyle\sum_{i=0}^{\frac{1}{2}(n-\epsilon'\sqrt{n})} \binom{n}{i} \\
&= \textstyle\sum_{i=0}^{\frac{n}{2}} \binom{n}{i} - \sum_{i=\frac{1}{2}(n-\epsilon'\sqrt{n})}^{\frac{n}{2}} \binom{n}{i} \\
&= 2^{n-1} - \textstyle\sum_{i=\frac{1}{2}(n-\epsilon'\sqrt{n})}^{\frac{n}{2}} \binom{n}{i} \\
&> 2^{n-1} - \tfrac{\epsilon'}{2}\sqrt{n} \cdot \binom{n}{n/2} && \text{using } \tbinom{n}{n/2} \text{ for all terms of the summation} \\
&= 2^{n-1} - \tfrac{\epsilon'}{2}\sqrt{n} \cdot \Theta\left(\tfrac{2^n}{\sqrt{n}}\right) && \text{Stirling's approximation} \\
&= 2^n \left(\tfrac{1}{2} - \Theta(\epsilon')\right).
\end{aligned}
$$

Because γ is bounded away from $\frac{1}{2}$, we can choose ϵ' small enough so that $|M| > 2^n \left(\frac{1}{2} - \Theta(\epsilon')\right) > 2^n \cdot \gamma = |S|$. We have now shown the existence of the polynomial q.

Finally, let

$$\bar{p} := p \cdot q^2.$$

It can be easily checked that $\text{SIGN}(\bar{p}(x))$ satisfies the requirements for weakly computing $\prod_i x_i$. Furthermore, $\text{degree}(q) = \frac{1}{2}(n - \epsilon'\sqrt{n})$, so $\text{degree}(q^2) = n - \epsilon'\sqrt{n}$, and $\text{degree}(\bar{p}) = \text{degree}(p) + n - \epsilon'\sqrt{n}$. □

With this last piece in place, we are now ready to prove our main theorem.

Theorem 2. *Let C be a circuit with n inputs, depth d and size w which computes* PARITY. *Then, $w \geq 2^{n^{\epsilon/d}}$, for a fixed universal constant ϵ.*

Proof. This proof simply combines the results that we have proved in the last few lectures. Let C be as stated in the theorem. Then we know there exists a polynomial p such that $\Pr_{x\in\{0,1\}^n}\left[p(x) = C(x) \stackrel{\text{def}}{=} \text{PARITY}(x)\right] \geq \frac{2}{3}$ and $\text{degree}(p) = \log^{O(d)} w$. By normalizing appropriately, as we saw in the previous lecture, we can get another polynomial p' such that $\Pr_{x\in\{-1,1\}^n}\left[p'(x) = \prod_i x_i\right] \geq \frac{2}{3}$ and $\text{degree}(p') = \text{degree}(p)$. This polynomial satisfies the requirements for Lemma 1, and so we obtain a polynomial $\bar{p}$ which weakly computes $\prod_i x_i$ and has degree $\log^{O(d)} w + n - \epsilon'\sqrt{n}$. From the lemma at the end of the last lecture, we know that the degree of $\bar{p}$ must be at least n, and thus we can see the following:

$$
\begin{aligned}
\log^{O(d)} w + n - \epsilon'\sqrt{n} \;\geq\; n \quad &\Longrightarrow \quad \log^{O(d)} w \;\geq\; \epsilon'\sqrt{n} \\
&\Longrightarrow \quad \log w \;\geq\; (\epsilon'\sqrt{n})^{1/O(d)} \\
&\Longrightarrow \quad w \;\geq\; 2^{(\epsilon'\sqrt{n})^{1/O(d)}} \\
&\Longrightarrow \quad w \;\geq\; 2^{n^{\epsilon/d}}
\end{aligned}
$$

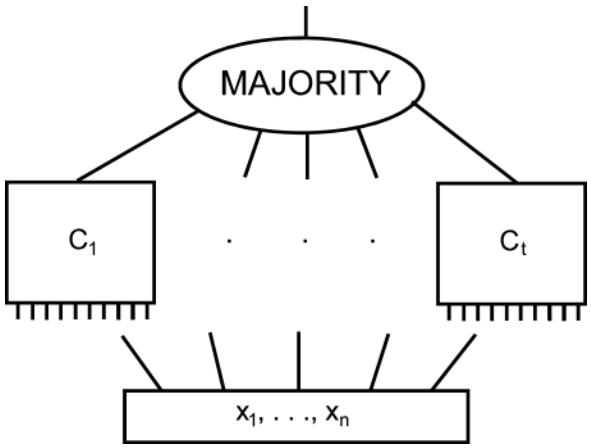
for some absolute constant ϵ. □

2 Correlating with parity using constant-depth circuits

In this section, we will prove a theorem, analogous to Theorem 2, which gives a lower bound on the size of constant-depth circuits that can only *correlate* with PARITY. This is done by showing that a circuit which correlates with PARITY implies the existence of a certain type of circuit that computes PARITY, enabling us to make use of the lower bound obtained in the last section.

To get an idea of the intuition behind the proof, imagine that you are given a biased coin. That is, for some $p \in \left(\frac{1}{2}, 1\right]$, the coin either has probability p of landing heads or probability p of landing tails. Your job is to determine which is the case. The naïve algorithm for this problem simply flips the coin a large number of times and takes the majority, and in fact, this is essentially what we will do with our circuit that correlates with PARITY: evaluate it many times, and output 1 iff the majority of the evaluations were 1. We will see that, for a large enough choice of the number of evaluations, this will allow us to compute PARITY exactly. This type of algorithm is captured with the following definition.

Definition 3. *A* majority-on-top *circuit is a circuit of the following form:*



where $x_1, \ldots, x_n$ are the inputs, the output gate is a MAJORITY *gate, and each intermediate $C_1, \ldots, C_t$ is a circuit on n inputs with only* AND, OR *and* NOT *gates.*

We now proceed by proving the following two results:

1. Any majority-on-top circuit computing PARITY exactly has large size. (Theorem 5)
2. A circuit which correlates with PARITY implies the existence of a majority-on-top circuit which computes PARITY and whose size is only polynomially larger. (Theorem 6)

The following lemma is very similar to one proved in Lecture 5.

Lemma 4. *Let C be a majority-on-top circuit with n inputs, depth d and size w. Then, there exists a polynomial $p : \{0,1\}^n \to \mathbb{R}$ of degree $\log^{O(d)}(w/\epsilon)$ such that*

$$\Pr_{x \in \{0,1\}^n}\left[\frac{\text{sign}(p(x)) + 1}{2} = C(x)\right] \geq 1 - \epsilon.$$

Proof. Let C be the circuit stated in the lemma, and denote its subcircuits by $C_1, \ldots, C_t$. We apply a lemma from the last lecture with $\epsilon' = \epsilon/t$ to get polynomials $p_1, \ldots, p_t$ such that $\forall i : \Pr_x [p_i(x) = C_i(x)] \geq 1 - \epsilon'$. The degree of each p_i is at most $\log^{O(d)}(w/\epsilon')$, since each circuit C_i has size $\leq w$ and depth $\leq d$. Then, define $p(x) := \sum_i p_i(x) - \frac{t}{2}$, and note that the degree of p is equal to the maximum degree of any p_i, which is $\log^{O(d)}(w/\epsilon') = \log^{O(d)}(w/\epsilon)$ (because $t \leq w$). Notice that whenever we have $p_i(x) = C_i(x)$ for all i, then $\frac{\text{sign}(p(x))+1}{2} = \text{MAJORITY}\,(C_1(x), \ldots, C_t(x)) = C(x)$. By a union bound, this happens with probability at least $1 - t \cdot \epsilon' = 1 - \epsilon$. □

By the same series of steps that led to Theorem 2, we can obtain the following analog.

Theorem 5. *Let C be a majority-on-top circuit with n inputs, depth d and size w which computes* PARITY. *Then, $w \geq 2^{n^{\epsilon/d}}$, for a fixed universal constant ϵ.*

Next, we prove the existence of a majority-on-top circuit computing PARITY, starting from one that only correlates with PARITY.

Theorem 6. *Let C be a circuit of depth d and size w such that* $\text{COR}(\text{PARITY}, C) \geq \frac{1}{w}$. *Then, there exists a majority-on-top circuit $\bar{C}$ of depth $O(d)$ and size* $\text{poly}(w)$ *which computes* PARITY *exactly.*

The following proof of Theorem 6 appears here for the first time. It is due to Adam Klivans and Salil Vadhan who kindly agreed to have it presented in this class.

Proof. From the definition of correlation, we know that $|\mathbb{E}_x e\,[C(x) + \text{PARITY}(x)]| \geq \frac{1}{w}$. W.l.o.g., we drop the absolute value signs. We rewrite the inequality as $\Pr_x [C(x) = \text{PARITY}(x)] \geq \frac{1}{2} + \frac{1}{2w}$. The intuition now is that we have a circuit C that computes PARITY 51% of the time, and we want a circuit $\bar{C}$ that computes it 100% of the time.

We will now make use of the *random self-reducibility* of PARITY. Let $a \in \{0,1\}^n$, and consider the circuit C'_a defined by

$$C'_a(x) := C(a + x) + \text{PARITY}(a).$$

First of all, notice that for any fixed a, C'_a is an $O(d)$-depth $O(w)$-size circuit because the step of adding $\text{PARITY}(a)$, which is just a bit, can be done with a constant number of extra gates and wires. Now observe that for any fixed x,

$$\begin{aligned} \mathbb{E}_a e\,[C'_a(x) + \text{PARITY}(x)] &= \mathbb{E}_a e\,[C(a+x) + \text{PARITY}(a) + \text{PARITY}(x)] \\ &= \mathbb{E}_a e\,[C(a+x) + \text{PARITY}(a+x)] \\ &= \mathbb{E}_a e\,[C(a) + \text{PARITY}(a)] \quad \geq \quad \frac{1}{w} \end{aligned}$$

and therefore $\Pr_a [C'_a(x) = \text{PARITY}(x)] \geq \frac{1}{2} + \frac{1}{2w}$. So, we will let our intermediate circuits be of the form C'_a. That is, for any choice of $a_1, \ldots, a_t \in \{0,1\}^n$, define the majority-on-top circuit $\bar{C}_{a_1,\ldots,a_t}(x) = \text{MAJORITY}\left(C'_{a_1}(x), \ldots, C'_{a_t}(x)\right)$, and note that $\bar{C}_{a_1,\ldots,a_t}$ has depth $O(d)$ and size $O(t \cdot w)$. Remember that the goal for the size of our final circuit was $\text{poly}(w)$; the following lemma shows that this is achievable.

Lemma 7. *We can choose $t = \text{poly}(w)$ such that $\forall x$,*

$$\Pr_{a_1,\ldots,a_t}\left[\bar{C}_{a_1,\ldots,a_t}(x) \neq \textsc{parity}(x)\right] < 2^{-n}.$$

Deferring the proof for a moment, let's finish the theorem assuming this lemma. A simple union bound gives us that $\Pr_{a_1,\ldots,a_t}\left[\exists x : \bar{C}_{a_1,\ldots,a_t}(x) \neq \textsc{parity}(x)\right] < 2^n \cdot 2^{-n} = 1$. Another way of stating this is that $\Pr_{a_1,\ldots,a_t}\left[\forall x : \bar{C}_{a_1,\ldots,a_t}(x) = \textsc{parity}(x)\right] > 0$. Thus, there is some fixed choice of $a_1, \ldots, a_t$ which results in a ($O(d)$-depth, poly(w)-size) majority-on-top circuit that computes PARITY. □

We now turn to the proof of the lemma.

Proof. (of Lemma 7) Fix an input $x \in \{0,1\}^n$. Define indicator random variables $Y_1, \ldots, Y_t$, where $Y_i := 1$ iff $C'_{a_i}(x) = \textsc{parity}(x)$. Let $\epsilon \in \left[\frac{1}{2w}, \frac{1}{2}\right]$ be the "advantage" that each C'_{a_i} has, so that $\Pr_{a_i}[Y_i = 1] = \frac{1}{2} + \epsilon$. Then, we can rewrite the probability of interest as $\Pr\left[\textsc{majority}(Y_1, \ldots, Y_t) \neq 1\right] = \Pr\left[\sum_i Y_i < \frac{t}{2}\right]$. This is equal to the probability that at least $\frac{t}{2}$ of the Y_is are 0, i.e.

$$\begin{aligned}
& \textstyle\sum_{I \subseteq [t], |I| \geq \frac{t}{2}} \Pr\left[Y_i = 0 \ \forall i \in I\right] \cdot \Pr\left[Y_i = 1 \ \forall i \notin I\right] \\
= \ & \textstyle\sum_I \left(\frac{1}{2} - \epsilon\right)^{|I|} \cdot \left(\frac{1}{2} + \epsilon\right)^{t-|I|} \\
\leq \ & \textstyle\sum_I \left(\frac{1}{2} - \epsilon\right)^{t/2} \cdot \left(\frac{1}{2} + \epsilon\right)^{t/2} && \text{multiply by } \left(\frac{1}{2} - \epsilon\right)^{t/2-|I|} \left(\frac{1}{2} + \epsilon\right)^{|I|-t/2} \geq 1 \\
\leq \ & 2^t \cdot \left(\frac{1}{4} - \epsilon^2\right)^{t/2} \leq \left(1 - \frac{1}{w^2}\right)^{t/2} \\
\leq \ & e^{-t/(2 \cdot w^2)}
\end{aligned}.$$

The latter quantity if less than 2^{-n} for $t = \text{poly}(w, n) = \text{poly}(n)$, concluding the proof of the claim. □

Combining Theorems 2, 5 and 6, we have the main result of this section.

Theorem 8. *Let C be a circuit with n inputs, depth d and size w such that $\textsc{cor}(C, \textsc{parity}) \geq 1/w$. Then, $w \geq 2^{n^{\epsilon/d}}$, for a fixed universal constant ϵ.*

3 An Application of Constant-Depth Circuits

In this section, we will solve Problem 3 from the list of problems for this course. Here is the setup:

> Imagine you are handed an algorithm $M : (\{0,1\}^a)^b \to \{0,1\}$ that on input $(x_1, \ldots, x_b)$ evaluates to 1 iff $\forall i : x_i \in A_i$, where $A_1, \ldots, A_b \subseteq \{0,1\}^a$ are arbitrary fixed subsets. Your goal is to compute an approximation ϵ to the volume $V = \prod_{i=0}^b |A_i|/2^a$, so that $|\epsilon - V| \leq \frac{1}{100}$, while only making $2^{\text{poly}(a, \log b)}$ queries to M.

First, notice that by making 2^{ab} queries to M (i.e. one for each element of $(\{0,1\}^a)^b$), we could compute the volume exactly. If $b \gg a$, the number of queries we are allowed to make is much less than this.

The key observation is that, however algorithm M operates, it is computable by a constant-depth circuit of size $\text{poly}(2^a \cdot b)$. This is because, for each set A_i, there is a "brute force" circuit C_i of depth 3 and size $O(2^a)$ which tests for membership in A_i. (This was Fact 2 in the first lecture.) So, by placing circuits $C_1, \ldots, C_b$ in parallel and wiring their outputs to a single AND gate (which is also the output gate), we get a circuit that computes M. For the remainder, say that this circuit has constant depth d and size $w = \text{poly}(2^a \cdot b)$.

We have seen (for example in Problem 2) that there is an explicit pseudorandom generator $G : \{0,1\}^s \to \{0,1\}^w$, where $s = \log^{O(d)} w$, that fools circuits of size w and depth d with error $\frac{1}{w}$. So, our algorithm for computing ϵ is to iterate over all seeds x of length s, and return $\epsilon = 2^{-s} \cdot \sum_x M(G(x))$, where $M(G(x))$ denotes the result of querying M on the first ab bits of $G(x)$. Note that the number of queries is $2^s = 2^{\text{poly}(a, \log b)}$ as required. Assume for contradiction that $|\epsilon - V| > \frac{1}{100}$. Then, because the volume is simply the fraction of inputs that evaluate to 1, we would have $\left|\Pr_{x \in \{0,1\}^{ab}}[M(x) = 1] - \Pr_{x \in \{0,1\}^s}[M(G(x)) = 1]\right| > \frac{1}{100}$, contradicting the fact that G fools constant-depth circuits.

4 Log-Depth Circuits

We now move to considering circuits with logarithmic depth. While with constant-depth circuits we allowed our gates to have unbounded fan-in, we will now restrict ourselves to gates with fan-in 2. We first note that these circuits are, in a sense, at least as powerful as the circuits we previously considered.

Theorem 9. *Let C be a circuit on n inputs of size w and depth d with unbounded fan-in. Then there is a circuit C' on n inputs with size $O(w)$ and depth $O(d \cdot \log w)$ with fan-in 2 such that $C(x) = C'(x)$ $\forall x$.*

Note that in particular, if $w = \text{poly}(n)$, the depth of C' is $O(\log n)$.

Proof. Let g be a gate in C which has fan-in $f_g > 2$. We replace g in C' with a depth-$O(\log f_w)$ "binary tree" of gates (of the same type), each of which has fan-in 2. Because f_g must be less than w, the total depth of the circuit increases by a factor of $O(\log w)$. Also, the number of wires in the gate-tree that replaced g is at most $2f_g$, so the size of C' only increases by a constant factor. □