

Lecture 9: Proof of the switching lemma and lower bounds for PARITY

Lecturer: Kristoffer Arnsfelt Hansen

Scribe: Thomas Dueholm Hansen

**Definition 1** Denote by  $R_n^l$  the set of restrictions to  $n$  variables leaving  $l$  variables free. That is,  $n - l$  variables are fixed to 0 or 1.

**Definition 2** Denote by  $T(F)$  the canonical decision tree for a DNF  $F = C_1 \vee C_2 \vee \dots \vee C_k$ , defined in the following recursive way. Assume that the literals in the terms are ordered. Construct a binary decision tree over  $C_1$  such that at the  $i$ 'th level we query the  $i$ 'th literal of  $C_1$ , and proceed left if it is 0 and right if it is 1. At every leaf attach  $T(F_\rho)$ , where  $\rho$  is the restriction corresponding to the path to the leaf. If  $F_\rho$  is the constant function 0 or 1 attach a leaf with that value.

**Definition 3** Define  $Stars(r, d)$  to be the set of all sequences  $(\beta_1, \dots, \beta_k)$  such that for every  $j$ ,  $\beta_j \in \{*, -\}^r \setminus \{-\}^r$  and the total number of  $*$ 's over all the  $\beta_j$ 's is  $d$ .

**Lemma 4**  $|Stars(r, d)| < \left(\frac{r}{\ln 2}\right)^d$ .

**Proof** Define  $\gamma$  by  $\left(1 + \frac{1}{\gamma}\right)^r = 2$ . Using  $1 + x < e^x$  for  $x \neq 0$ , we get:

$$2 < e^{\frac{r}{\gamma}} \Rightarrow \ln 2 < \frac{r}{\gamma} \Rightarrow \gamma < \frac{r}{\ln 2}$$

Next we will prove that  $|Stars(r, d)| \leq \gamma^d$ . We will do this by induction. The base case  $d = 0$  follows trivially. For the induction step assume that for all  $n < d$ ,  $|Stars(r, n)| \leq \gamma^n$ . We then get:

$$\begin{aligned} |Stars(r, d)| &= \sum_{i=1}^{\min(r, d)} \binom{r}{i} |Stars(r, d-i)| \leq \sum_{i=1}^r \binom{r}{i} \gamma^{d-i} \\ &= \gamma^d \sum_{i=1}^r \binom{r}{i} \left(\frac{1}{\gamma}\right)^i = \gamma^d \left( \left(1 + \frac{1}{\gamma}\right)^r - 1 \right) = \gamma^d (2 - 1) = \gamma^d \end{aligned}$$

In the first step we count the number of ways to append an extra element to a sequence such that the total number of  $*$ 's becomes  $d$ . In the second step we can change  $\min(r, d)$  to  $r$  because if  $\min(r, d) = r$  the sum is unchanged, but if  $\min(r, d) = d$  then  $d \leq r$  and the sum is just larger. In the fourth step we use the identity  $\sum_{i=0}^k \binom{k}{i} a^i b^{k-i} = (a + b)^k$ . □

**Lemma 5 (Switching Lemma)** Let  $F = C_1 \vee C_2 \vee \dots \vee C_k$  be a DNF with terms of size  $\leq r$ . Let  $l = pn$ , for  $0 < p \leq \frac{1}{7}$ . Pick  $\rho \in R_n^l$  at random, then  $Pr[F_\rho$  does not have a decision tree of depth at most  $d] < (7pr)^d$ .

**Proof** Let  $S \subseteq R_n^l$  be the set of restrictions, such that for  $\rho \in S$  the depth of  $T(F_\rho)$  is  $\geq d$ . We will define a 1-1 map  $S \rightarrow T$ , where  $T = R_n^{l-d} \times Stars(r, d) \times \{0, 1\}^d$ . Let some  $\rho \in S$  be given. Let  $\pi$  be the restriction corresponding to the first  $d$  variables of the lexicographically first path in  $T(F_\rho)$  of length  $\geq d$ .

Let  $C_{v_1}$  be the first term of  $F_\rho$ , and let  $\pi_1$  be the part of  $\pi$  in  $C_{v_1}$ . Also, let  $\sigma_1$  be the unique restriction making  $C_{v_1} = 1$ . For  $i > 1$  let  $C_{v_i}$  be the first term of  $F_{\rho\pi_1\dots\pi_{i-1}}$ , and let  $\pi_i$  be the part of  $\pi$  in  $C_{v_i}$ . Also, let  $\sigma_i$  be the unique restriction making  $C_{v_i} = 1$ .

In the following we define the 1-1 map  $S \rightarrow T$ , but first let us introduce some notation. For every  $i = 1, \dots, k$ , let the  $j$ 'th component of  $\beta_i$  be  $*$  if and only if the  $j$ 'th variable in  $C_{v_i}$  is set by  $\sigma_i$ . Also, define  $\delta \in \{0, 1\}^d$  to be the bit-string for which the  $i$ 'th bit is 1 if and only if  $\pi$  and  $\sigma_1 \dots \sigma_k$  agree on the  $i$ 'th variable.

Let the 1-1 map be such that for every  $\rho \in S$ , we get  $\rho \mapsto (\rho\sigma_1 \dots \sigma_k, (\beta_1, \dots, \beta_k), \delta)$ . This is called the deconstruction of  $\rho$ . For the reconstruction, i.e. the mapping from  $T$  to  $S$ , we are given  $\rho\sigma_1 \dots \sigma_k$  (a composition of restrictions),  $(\beta_1, \dots, \beta_k)$ ,  $\delta$  and  $F$ , and we want to find  $\rho$ .

Let  $C_{v_1}$  be the first term set to 1 by  $\rho\sigma_1 \dots \sigma_k$ . Using  $\beta_1$  we know which variables in  $C_{v_1}$  are set by  $\sigma_1$ , but these are the only variables set by  $\sigma_1$ , and hence we know  $\sigma_1$ . Knowing  $\sigma_1$  and the variables of  $C_{v_1}$ , we can use  $\delta$  to reconstruct  $\pi_1$ . We can then consider  $\rho\pi_1\sigma_2 \dots \sigma_k$ , and repeat the procedure until we find  $\rho\pi_1 \dots \pi_k$ . Knowing  $\pi_i$  for  $i = 1..k$  we can then reconstruct  $\rho$ , and the reconstruction is complete.

With this we have shown that  $|S| \leq |T|$ , and we know that  $|T| \leq |R_n^{l-d}| \cdot |\text{Stars}(r, d)| \cdot 2^d$ . We are now ready to prove the statement of the lemma.

$$\begin{aligned}
\Pr[F_\rho \text{ does not have a decision tree of depth at most } d] &= \frac{|S|}{|R_n^l|} \\
&\leq \frac{|R_n^{l-d}| \cdot |\text{Stars}(r, d)| \cdot 2^d}{|R_n^l|} \leq \frac{|R_n^{l-d}| \cdot \left(\frac{r}{\ln 2}\right)^d \cdot 2^d}{|R_n^l|} = \frac{|R_n^{l-d}|}{|R_n^l|} \cdot \left(\frac{2r}{\ln 2}\right)^d \\
&= \frac{\binom{n}{l-d} 2^{n-l+d}}{\binom{n}{l} 2^{n-l}} \cdot \left(\frac{2r}{\ln 2}\right)^d = \frac{\binom{n}{l-d}}{\binom{n}{l}} \cdot \left(\frac{4r}{\ln 2}\right)^d = \frac{n!}{(n-l+d)!(l-d)!} \cdot \left(\frac{4r}{\ln 2}\right)^d \\
&\leq \frac{l^d}{(n-l)^d} \cdot \left(\frac{4r}{\ln 2}\right)^d = \left(\frac{4lr}{(n-l)\ln 2}\right)^d = \left(\frac{4\frac{l}{n}r}{\left(1-\frac{l}{n}\right)\ln 2}\right)^d \\
&= \left(\frac{4pr}{(1-p)\ln 2}\right)^d \leq \left(\frac{4pr}{\frac{6}{7}\ln 2}\right)^d < (7pr)^d
\end{aligned}$$

□

**Proposition 6** *Let  $C$  be an AND/OR circuit of depth  $h$  and size  $S$ . Let  $d$  be given and define  $n_h = \frac{n}{14(14d)^{h-1}}$ . Choose  $\rho \in R_n^{n_h}$  at random, then with probability  $1 - S2^{-d}$  every function computed at every gate of  $C$  has a decision tree of depth at most  $d$  after using  $\rho$ .*

**Proof** First, we will describe a useful alternative way of choosing a  $\rho \in R_n^{n_h}$  at random. Define for  $0 \leq i \leq n-1$ ,  $n_{i+1} = \frac{n}{14(14d)^i}$ . Choose  $\rho$  by choosing  $\rho_1\rho_2 \dots \rho_h$ , where  $\rho_1 \in R_n^{n_1}$  and  $\rho_{i+1} \in R_{n_{i+1}}^{n_{i+1}}$ , for  $1 \leq i \leq h-1$ . Note that this is in fact the same as choosing  $\rho$ , we just restrict the number of variables in steps ending up with the correct number of fixed variables.

We will for each gate show that the probability that the corresponding decision tree has depth greater than  $d$ , given that its input gates have decision trees of depth at most  $d$ , is less than  $2^{-d}$ , and the statement then follows by summing over all gates.

For a given gate, we will do a proof by induction in the depth of the gate. As the base case consider an OR gate at level 1, with level 1 being the lowest level. This can be viewed as a DNF

with terms of size 1, meaning that we can apply the switching lemma. Hence, when we pick a restriction  $\rho_1 \in R_n^{n_1}$  at random, we get from  $n_1 = \frac{n}{14}$  and  $p = \frac{1}{14}$  that:

$$\Pr[F_{\rho_1} \text{ does not have a decision tree of depth at most } d] < (7 \cdot \frac{1}{14} \cdot 1)^d = 2^{-d}$$

For an AND gate at level 1, we simply use the decision tree of the negation, having switched 0 and 1.

For the induction step assume that after using  $\rho_1 \dots \rho_i$ , all gates at levels 1 to  $i$  have decision trees of depth  $\leq d$ .

Consider an OR gate at level  $i + 1$ . All its inputs have decision trees of depth  $\leq d$ . We can rewrite these to DNF's with terms of size  $\leq d$  in the following way. Every root-to-leaf path can be expressed as a term in which the literals are required to have the values observed in the decision tree. Note that there can at most be  $d$  literals in each term. The decision tree can then be expressed as an OR gate having as input every term constructed from a path leading to a leaf of value 1. Since the OR gate at level  $i + 1$  now only has OR gates as inputs, we can collapse all these into one OR gate, achieving a DNF with terms of size at most  $d$ .

We can now apply the switching lemma, picking  $\rho_{i+1} \in R_{n_i}^{n_{i+1}}$  at random. We now have:

$$p = \frac{n_{i+1}}{n_i} = \frac{\frac{n}{14(14d)^{i+1}}}{\frac{n}{14(14d)^i}} = \frac{1}{14d}$$

Giving us that:

$$\Pr[F_{\rho_1 \dots \rho_{i+1}} \text{ does not have a decision tree of depth at most } d] < (7 \cdot \frac{1}{14d} \cdot d)^d = 2^{-d}$$

For the case of an AND gate at level  $i + 1$ , we again make use of negation. □

**Theorem 7** *Circuits of depth  $h$  and size  $S$  computing PARITY must satisfy  $S \geq 2^{\frac{1}{14}n^{\frac{1}{h-1}}}$ .*

**Proof** Given a circuit  $C$  of depth  $h$  and size  $S$  computing PARITY, choose  $d = \log S$ . Assume that the topmost gate is an OR gate, otherwise negate the circuit such that it computes non-PARITY. As in the proof of Proposition 6 we can pick a  $\rho \in R_n^{n_{h-1}}$  at random such that there is positive probability that the input gates of the topmost OR gate have decision trees of depth at most  $d$  after applying  $\rho$ . That is, there must exist such decision trees for some  $\rho$ , otherwise the probability would be zero. Let  $\rho$  be fixed such that this is the case, then as in the proof of Proposition 6 we can express the circuit after having applied  $\rho$  as a DNF formula  $F$  with terms of size at most  $d$ .

The key observation is that  $F$  itself decides either PARITY or non-PARITY for  $n_{h-1}$  variables. In both cases we claim that PARITY or non-PARITY for  $n$  variables being decided by a DNF requires terms of size  $n$ . Indeed, if a term can be satisfied using  $n - 1$  or fewer variables, we will get the same result both when setting the last variable to 0 and 1, which cannot be the case.

It follows that we must have

$$d \geq n_{h-1} = \frac{n}{14(14d)^{h-2}} \Rightarrow (14d)^{h-1} \geq n \Rightarrow d \geq \frac{1}{14}n^{\frac{1}{h-1}} \Rightarrow S \geq 2^{\frac{1}{14}n^{\frac{1}{h-1}}}.$$

□

So we see that a circuit that computes parity for  $n$  input variables and is of constant height  $h$  must have size  $S \geq 2^{\frac{1}{14}n^{\frac{1}{h-1}}} = 2^{n^{\Omega(1)}}$ , and therefore not polynomial in size so  $PARITY \notin AC^0$ .