CS294-114: Fine Grained Complexity and Algorithms September 10, 2015

Razborov-Smolensky Lower bound for ^{0}

Instructor: Russell Impagliazzo *Scribe: Akshayaram Srinivasan*

1 Introduction

In the previous lectures we have shown that parity is not in AC^0 using switching lemma [1]. We have also seen some scenarios where switching lemma has been used to design better algorithms, namely counting the number of satisfying assignments for a CNF formula and learning functions in AC^0. In this lecture we will prove a lower bound for the class ACC^0. That is, we will show that there exists a function that is not in ACC^0. In subsequent lecture, we will see how this lower bound technique can be used for designing better algorithms.

Let us start with some definitions.

Definition 1 *$ACC_{q,d}$ is defined as the set of all depth d, poly sized boolean circuits consisting of unbounded fan-in $\vee, \wedge$ and $\oplus_q$ gates [2].*

The reason to consider a circuit model where we have unbounded fan-in $\oplus_q$ gates in addition to $\vee, \wedge$ gates is that we want to make sure that lower bound which we prove for this circuit class is something new and not the same lower bound for parity in disguise. We can in fact show that $\oplus_q$ cannot be computed in AC^0 using a similar argument as for parity. That is why, we have added extra power to the circuit (in the form of $\oplus_q$ gates) and still show that there are some functions which are outside this class.

The main technique used to prove the lower bound for this class of circuits is due to Razborov which was later simplified by Smolensky. The high-level overview of the idea is that for all functions f in $ACC_{q,O(1)}$ there exists a low degree polynomial which "approximates" f well. But there are functions for which there exists no low degree polynomial that "approximates" it well enough.

Lets formalize the notion of *Probabilistic Polynomials.* This refers to a distribution $P_r(x)$ of multivariate polynomials over $\mathbb{Z}_q$ in variables $X_1, \cdots, X_n$.

Definition 2 (ϵ-computing a boolean function) *A probabilsitic polynomial $P_r(\cdot)$ is said to ϵ-compute a boolean function f over n variables $X_1, \cdots, X_n$ if for all $\vec{X} = \{X_1 \cdots X_n\}$*

$$\Pr_r[P_r(\vec{X}) \neq f(\vec{X})] \leq \epsilon$$

In this lecture, we will be proving that parity cannot be computed by a polynomial size, constant depth circuits with unbounded fan-in $\vee, \wedge$ and $\oplus_q$ gates where $q > 2$. We will also give the intuition on how to extend this result to show that given unbounded fan-in parity gate it is hard to compute mod q for any $q > 2$.

We will first prove a lemma which states that for all functions $f \in ACC_{q,d}$ there exists a low degree polynomial which "approximates" f well.

[1] AC^0 is the class of polynomial size, constant depth circuits having unbounded fan-in $\vee, \wedge, \neg$ gates.

[2] $\oplus_q(X_1, \cdots, X_n) = 1$ iff $\sum_i X_i \neq 0 \mod q$

Lemma 3 *Let q be a prime > 2. Let f be a function in $ACC_{q,d}$ of size s. For every $\epsilon > 0$ there exists a probabilistic polynomial $P_r(\cdot)$ of degree $D = O((d \log s + \log(1/\epsilon))^d)$ that ϵ-computes f.*

Proof: We will construct $P_r(\cdot)$ in a bottom up way. Firstly, we construct probabilistic polynomials that ϵ-compute the gates at the lowest level and later show how these can be composed to construct polynomials that ϵ-compute a function at a higher level.

The probabilistic polynomials corresponding to the input X_i is just X_i. Now lets see the construction of probabilistic polynomials for the gates in the lowest level. $ACC_{q,d}$ has 3 kinds of gates namely $\oplus_q, \vee, \wedge$.

- $\oplus_q$: $\oplus_q(X_1, \cdots, X_n) = 1$ if $\sum_i X_i \neq 0 \mod q$ and 0 otherwise. Our first attempt is to consider the polynomial $P_r(x) = \sum_i X_i$. $P_r(\vec{X}) = \oplus_q(\vec{X})$ whenever $\sum_i X_i = 0 \mod q$. But whenever it is not the case $P_r(\vec{X})$ and $\oplus_q(\vec{X})$ are different. To overcome this issue, we use Fermat's little theorem: This states that for any $a \neq 0 \mod q$, $a^{q-1} = 1 \mod q$. So we consider the polynomial $P_r(x) = (\sum_i X_i)^{q-1}$. It is easy to see that this polynomial exactly computes $\oplus_q$ gate.
- $\vee$: Lets consider the following probabilistic polynomial. Use the randomness r to pick a subset S from $[n]$ and consider $P_r(\vec{X}) = (\sum_{i \in S} X_i)^{q-1}$. We need to estimate the error that the polynomial could make in computing $\vee$. If all X_i's are 0 then both $P_r(\vec{X})$ and $\vee(\vec{X})$ are 0. Now lets consider the case in which at least one of the X_i is 1. We claim that $\Pr_S[P_r(\vec{X}) = 0 \mod q] \leq 1/2$. This follows from the observation that for every value of $\sum_{i \in S/\{X_i\}} X_i$ including or not including X_i in the subset S could change the value to 0. Since the probability of including X_i in the subset S is $1/2$ (for a random subset S) the claim follows.

  Now we want to amplify the probability of correctly computing $\vee$. This can be achieved using the standard technique of repeating the experiment using independent random coins and outputting 1 even if one of the trials outputs 1 (amplification of a RP algorithm). Consider $t = \log 1/\epsilon$ independent trials of the above experiment where the subsets are sampled uniformly and independently at random. We want to compute $\vee((\sum_{i \in S_1} X_i)^{q-1}, \cdots, (\sum_{i \in S_2} X_i)^{q-1})$. We have reduced the computation to $\vee$ of t values even if the initial $\vee$ has $\Omega(s)$ inputs. This can be computed using a degree $t(q-1)$ polynomial using De-Morgan's law. It is also easy to see that the Probability of making an error in the computation of $\vee$ is at most ϵ.
- $\wedge$: This is computed as $1 - \vee_i(1 - X_i)$ where $\vee$ is approximated as before. Thus, $\wedge$ can be ϵ-computed by a polynomial of degree $t(q-1)$ where $t = \log 1/\epsilon$

We have seen how to approximate functions in the lowest layer. Lets now consider a gate h in a higher level. Let $P_r(\cdot)$ be the probabilistic polynomial of degree D that δ-computes h. Let $g_1, \cdots, g_m$ be the functions which are inputs to h. Inductively, let $P_{r_1,1}, \cdots, P_{r_m,m}$ be the probabilistic polynomials that ϵ-compute $g_1, \cdots, g_m$ respectively. Now consider the polynomial $\tilde{P}_{r,r_1,\cdots,r_m}(\vec{X}) = P_r(P_{r_1,1}(\vec{X}), \cdots, P_{r_m,m}(\vec{X}))$. We claim that this polynomial is of degree $D.D'$ and $m\epsilon + \delta$ computes the function at the subtree with root h (denote by H). The degree condition is easy to verify. We will now estimate the error made by $\tilde{P}$.

$$\begin{aligned} Pr[\tilde{P}(\vec{X}) \neq H(\vec{X})] &= Pr[\vee_{i \in [m]} P_{r_i,i}(\vec{X}) \neq g_i(\vec{X}) \vee P_r(\vec{X}) \neq h(\vec{X})] \\ &\leq m\epsilon + \delta \end{aligned}$$

where the last inequality is obtained through union bound. □

We repeat this procedure inductively until we approximate the function at level d. We will now see how to set the parameters so that the polynomial at the topmost level ϵ computes the function at the root. It can be easily seen that at each level the error grows by an multiplicative factor of the size of the circuit (Since $\delta = \epsilon$ by our construction and $m + 1 \leq s$). Hence, if ϵ_0 is the error at the lowest layer then the error at level d is at most $s^d \epsilon_0$. We want this to be $\epsilon = \epsilon_0 . s^d$. Hence, setting $\epsilon_0 = \epsilon / s^d$ and setting the number of iterations $t = \log(1/\epsilon_0) = d \log s + \log(1/\epsilon)$ we get the overall error to be at most ϵ. We now calculate the degree. The degree at the lowest level is $t(q-1) = O(d \log s + \log(1/\epsilon))$ (since q is a constant). It is easy to see that at level i the polynomial we construct has degree $t(q-1)^i$. Hence in level d we construct a polynomial of degree $O((d \log s + \log(1/\epsilon))^d)$.

We are now ready to prove the main lower bound theorem. We will show that for $q = 3$, parity cannot be computed in $ACC_{q,O(1)}$.

Theorem 4 *If $\oplus(X_1, \cdots, X_n)$ is computed by a circuit of depth d and size s using unbounded fan-in $\vee, \wedge, \oplus_q$ gates then $s \geq 2^{\Omega(n^{1/2d})}$*

Proof: To prove this theorem, lets shift to the Fourier representation of 0 and 1 so that we naturally have a multilinear polynomial computing the function exactly. That is, 0 is represented as 1 and 1 is represented as -1. There is a linear transformation from 0,1 representation to $1, -1$ representation $\mod 3$. The transform is $Y = 1 + X \mod 3$. We have already seen that in order to compute the parity in the 0,1 representation it is enough to compute the product of Y_i's in the 1,-1 representation. Suppose there exists a polynomial p that “exactly computes" $\oplus_{i \in [n]} X_i$ of degree d then there exists a polynomial q that “exactly computes" $\Pi_{i \in [n]} Y_i$ of degree d since we are just doing a linear transform from X to $Y \mod 3$. We also observe that in the 1,-1 representation, $Y^2 = 1$.

Consider any function $F : \{-1, 1\}^n \to \{0, 1, -1\}$. We first observe that there is a multi-linear representation of F as $\sum_S c_s \Pi_{i \in S} Y_i$ where $c_S \in \{0, 1, -1\}$ for every S. This is obtained in an analogous way as described in the previous lectures. First set Y_1 and depending on the set value of Y_1 define F. That is $F(\vec{Y}) = ((1 - Y_1)/2) F_{-1} + ((1 + Y_1)/2) F_1 = (Y_1 - 1) F_{-1} - (1 + Y_1) F_1 \mod 3$ where F_{-1} and F_1 are multilinear representations of the children of the root node.

We will now build up the intuition on how to prove the lower bound. Let us first assume that P' is a polynomial which exactly computes $\Pi_{i \in [n]} Y_i$. We first write F as the following:

$$\begin{aligned} F(Y) &= \sum_{S, |S| \leq n/2} c_S (\Pi_{i \in S} Y_i) + \sum_{S, |S| > n/2} c_S (\Pi_{i \in S} Y_i) \\ &= \sum_{S, |S| \leq n/2} c_S (\Pi_{i \in S} Y_i) + \sum_{S, |S| > n/2} c_S (\Pi_{i \notin S} Y_i)(\Pi_{i \in [n]} Y_i) \\ &= \sum_{S, |S| \leq n/2} c_S (\Pi_{i \in S} Y_i) + \sum_{S, |S| > n/2} c_S (\Pi_{i \notin S} Y_i) P'(\vec{Y}) \end{aligned}$$

The second equation follows from the first since $Y_i^2 = 1$ for all $i \in [n]$.

It now easy to see that $Deg(F) \leq n/2 + Deg(P')$. We will show through a simple counting argument that there are functions which require $n/2 + \Omega(\sqrt{n})$-degree polynomial to represent them. Hence, we can deduce that $Deg(P')$ has to be $\Omega(\sqrt{n})$ which will immediately give a lower bound for the size of the circuit computing the parity. The technical details will take care of the fact that P' need not exactly compute $\Pi_{i \in [n]} Y_i$ but can approximate.

Assume that parity is in $ACC_{q,d}$ with size s. We have to show that $s \geq 2^{\Omega(n^{1/2d})}$. From Lemma **??**, we can deduce that there exists a probabilistic polynomial that ϵ computes parity with degree $D = (d \log s + \log(1/\epsilon))^d$.

$$\begin{aligned}
\forall \vec{Y}, \underset{r}{Pr}[P_r(\vec{Y}) \neq \Pi_{i \in [n]} Y_i] &\leq \epsilon \\
\underset{r,\vec{Y}}{Pr}[P_r(\vec{Y}) \neq \Pi_{i \in [n]} Y_i] &\leq \epsilon \\
\exists r, \underset{\vec{Y}}{Pr}[P_r(\vec{Y}) \neq \Pi_{i \in [n]} Y_i] &\leq \epsilon
\end{aligned}$$

The second equation follows from the first as a direct consequence of the universal quantification over Y and the third equation follows from the second due to the standard averaging argument.

From the third equation we can infer that there exists a polynomial $P_r(\cdot)$ such that it computes $\Pi_{i \in [n]} Y_i$ on all but ϵ-fraction of the inputs. Let E be the set of inputs on which $P_r(\cdot)$ errs. $|E| \leq \epsilon.2^n$.

Consider the restricted Domain $\{-1,1\}^n - E$. We can now write any function over $F : \{-1,1\}^n - E \rightarrow \{0,1,-1\}$ as a multilinear polynomial of degree $n/2 + Deg(P_r) = n/2 + D$ (since P_r computes $\Pi_{i \in [n]} Y_i$ exactly in the domain of F). We now count the total number of such functions. This is equal to $3^{2^n - |E|} = 3^{(1-\epsilon)2^n}$. We now calculate the total number of multilinear polynomials of degree $n/2 + D$. We know that this quantity must be greater than or equal the total number of functions. The total number of multilinear polynomials over mod 3 with degree at most $n/2 + D = 3^{\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n/2+D}}$. We now approximate this quantity. The distribution of $\binom{n}{x}$ is a bell-curve. If $D \leq \sqrt{n}$, then we lose a constant fraction of the total mass which is equal to 2^n while neglecting terms from $n/2 + D + 1 \rightarrow n$. Hence, in that case $3^{(1-\epsilon)2^n} \leq 3^{(1-c)2^n}$ which implies $\epsilon \geq c$. We can always fix the parameter $\epsilon < c$ in our construction. This means that $D > \sqrt{n}$ which immediately gives the lower bound that $s > 2^{\Omega(n^{1/2d})}$ as follows:

$$\begin{aligned}
D &> \sqrt{n} \\
(d \log(1/\epsilon) + \log s)^d &> \sqrt{n} \\
\log s &> \Omega(\sqrt{n}^{1/d}) \\
s &> 2^{\Omega(n^{1/2d})}
\end{aligned}$$

□

2 Intuition for lower bound of $\oplus_3$