Lecture 23: November 19, 1996

Lecturer: Uri Zwick *Scribe: Lisa Hales*

23.1 Another Lower Bound for Parity

23.1.1 Introduction

Last lecture we proved that $PARITY \notin AC^0$, where AC^0 is the class of functions computed by constant depth, polynomial-sized circuits with gates from the set $\{\wedge, \vee, \sim\}$. In fact, we proved an exponential lower bound for the size of such $PARITY$ circuits. Today we get a slightly smaller lower bound for the size of constant depth $PARITY$ circuits with gates from the set $\{\wedge, \vee, \sim, mod_3\}$. The addition of mod_3 to our set of gates is of interest because, like $PARITY$, $MOD_3(n)$ requires exponential-sized constant depth $\{\wedge, \vee, \sim\}$ circuits. Our proof can in fact be generalized to circuits with $\{\wedge, \vee, \sim, mod_{p^k}\}$ gates for any odd prime p. It is unknown whether a similar result holds in general; for example, it is still open whether there are are small constant depth circuits for $PARITY$ which use gates from the set $\{\wedge, \vee, \sim, mod_6\}$.

23.1.2 The Big Picture

The proof, which is due to Razborov and Smolensky, uses techniques similar to the approximation method. We shall prove the following two theorems, from which the result follows almost immediately.

Theorem 1 *If $f : \{0,1\}^n \to \{0,1\}$ can be computed using a depth d $\{\wedge, \vee, \sim, mod_3\}$ circuit with s gates then for every $\epsilon > 0$ there is a multi-linear polynomial $p(x_1, x_2, ..., x_n)$ over GF_3 of degree $(c \log \frac{s}{\epsilon})^d$ such that $f(x_1, ..., x_n) \neq p(x_1, ..., x_n)$ on at most $\epsilon 2^n$ of the values of $\{0,1\}^n$.*

Theorem 2 *If $p(x_1, ..., x_n)$ is a multi-linear polynomial over GF_3 of degree at most $\sqrt{n}$, then $p(x_1, ..., x_n) \neq XOR(x_1, ..., x_n)$ for at least $0.15 \cdot 2^n$ values from $\{0,1\}^n$.*

If we take $\epsilon = 0.15$ then Theorems 1 and 2 imply

$$c(\log \frac{s}{\epsilon})^d \geq \sqrt{n}$$

where s is the number of gates in any $PARITY$ circuit of this form. Hence,

$$s \geq \epsilon(2^{\frac{1}{c}n^{\frac{1}{2d}}}).$$

23.1.3 Proof of Theorem 1

We need the following lemma, which we state and prove in the case of OR_n, and which holds analogously for AND_n.

Lemma 1 *For every input distribution on $\{0,1\}^n$ and for any $\epsilon > 0$ there is a multilinear polynomial $p(x_1, ..., x_n)$ over GF_3 of degree $O(\log(\frac{1}{\epsilon}))$ such that $Pr[OR(x_1, ..., x_n) \neq p(x_1, ..., x_n)] \leq \epsilon$.*

Proof: Let S be a random subset of $\{1, 2, ..., n\}$. Consider $f_S = (\sum_{i \in S} x_i)^2$.

It is clear that if $x_1 = ... = x_n = 0$ then $f_S(x_1, ..., x_n) = 0$. We claim that if $\bigvee_i x_i = 1$ then $Pr[f_S(x_1, ..., x_n) = 1] \geq \frac{1}{2}$. Fix an element of $\{0,1\}^n$ which satisfies $\bigvee_i x_i = 1$ and pick i such that $x_i = 1$. We can decompose our choice of S into two stages: first we chose a random subset S' from $\{1, ...i-1, i+1, ..., n\}$, then we determine whether or not to include i. Thus,

$$\begin{aligned} Pr[f_S = 1] &= \underbrace{Pr[f_S = 1 | \sum_{i \in S'} x_i = 0]}_{(1)} \cdot Pr[\sum_{i \in S'} x_i = 0] \\ &+ \underbrace{Pr[f_S = 1 | \sum_{i \in S'} x_i \neq 0]}_{(2)} \cdot Pr[\sum_{i \in S'} x_i \neq 0] \end{aligned}$$

(1) is at least $\frac{1}{2}$ since we will include i in S with probability $\frac{1}{2}$, and (2) is at least $\frac{1}{2}$ since we do not include i in S with probability $\frac{1}{2}$. Thus we get $Pr[f_S = 1] \geq \frac{1}{2}$.

Thus f_S is an approximation to OR, albeit not a very good one. To get an improved approximation we take l random subsets, S_1,S_2,..., S_l, of $\{1, ..., n\}$ and define $p(x_1, ..., x_n) = 1 - \prod_{i=1}^{l}(1 - f_{S_i})$

The degree of $p(x_1, ..., x_n)$ is $2l$, as the degree of each f_{S_i} is 2. Moreover,

$$\bigvee_i x_i = 0 \Longrightarrow p(x_1, ..., x_n) = 0$$

and

$$\bigvee_i x_i = 1 \Longrightarrow Pr_{S_1,...,S_l}[p(x_1, ..., x_n) = 1] \geq 1 - \frac{1}{2^l}.$$

Thus, letting $l = \log(\frac{1}{\epsilon})$ we get that for all $\vec{x}$,

$$Pr_{S_1,...,S_l}[p(x_1, ..., x_n) \neq OR(x_1, ..., x_n)] \leq \epsilon$$

Finally, let D be any distribution on the set $\{0,1\}^n$, and let $w(\vec{x})$ the weight of $\vec{x}$ induced by D. Then we have

$$\sum_{S_1,\dots,S_l,OR(\vec{x})\neq p(\vec{x})} w(\vec{x}) \leq \epsilon \cdot 2^{nl}$$

Hence there must be at least one choice of $S_1, \dots, S_l$ such that

$$\sum_{OR(\vec{x})\neq p(\vec{x})} w(\vec{x}) \leq \epsilon 2^{nl} \cdot \frac{1}{2^{nl}} = \epsilon,$$

as desired. ∎

To complete the proof of Theorem 1, we give an inductive method of using Lemma 1 to assign polynomials to each gate of the circuit. Suppose we have assigned polynomials $p_1,\dots,p_k$ of degree $(2\log\frac{s}{\epsilon})^{i-1}$ to the top gates of all subcircuits entering gate g at level i. We have three cases.

- Gate g is a mod_3. Then if $p_1,\dots,p_k$ are the assigned polynomials, we use $(\sum_i p_i)^2$, which is of degree at most $2(2\log\frac{s}{\epsilon})^{i-1}$.
- Gate g is an OR. We choose the polynomial defined in the lemma, using the distribution on inputs induced by the polynomials $p_1,p_2,\dots,p_n$ and $l = \log(\frac{\epsilon}{s})$. The polynomial resulting from composition will have degree

$$(2\log\frac{s}{\epsilon})^{i-1} \cdot 2\log\frac{s}{e} = (2\log\frac{s}{\epsilon})^{i}$$

- Gate g is an AND. Similar to OR.

At the conclusion of this process we are left with a polynomial of degree $(2\log\frac{s}{\epsilon})^d$. While this polynomial is not multi-linear, because we are only concerned with its behavior on inputs from the set $\{0,1\}^n$, we can replace x_i^k for by X_i throughout, thus creating a multi-linear polynomial which agrees with the original one on $\{0,1\}^n$ and whose degree is also bounded by $(2\log\frac{s}{\epsilon})^d$. We need only argue that this polynomial errs on at most an ϵ fraction of inputs. But any error must arise from a failure of our assigned polynomials at a particular gate. There are s gates and each contributes at most an $\frac{\epsilon}{s}$ fraction of new failures.

23.1.4 Proof of Theorem 2

For this proof, we represent boolean values by 1 and -1 rather than 0 and 1. In particular, we will let false correspond to 1 and true to -1. With this convention the previous theorem still holds since the degree of a polynomial is preserved under linear transformations, and this change can be accomplished by the mapping $T(x) = 1 - 2x$. We now have

$$XOR(x_1, ..., x_n) = \prod_{i=1}^{n} x_i$$

Suppose that $p(x_1, ..., x_n)$ is a polynomial of degree $\leq \sqrt{n}$. We need to show that this polynomial differs from $\prod_{i=1}^{n} x_i$ on at least a 0.15 fraction of the values of $\{-1, 1\}^n$.

Let $A = \{(x_1, ..., x_n) \in \{-1, 1\}^n | p(x_1, ..., x_n) = \prod_{i=1}^{n} x_i\}$ We wish to show that A is small and we will do this by bounding the number of functions from A to $\{-1, 1\}$.

Let $F = \{-1, 0, 1\}^A$. We claim every function in F can be represented as a multi-linear polynomial over GF_3 of degree at most $\frac{n}{2} + \frac{\sqrt{n}}{2}$. Suppose $f \in F$, and $\prod_{i \in S} x_i$ is a monomial in the representation of f. If $|S| > \frac{n+\sqrt{n}}{2}$ we replace the monomial by $\prod_{i \notin S} x_i \cdot p(x_1, ..., x_n)$. It is easy to see that for inputs in A this monomial agrees with the original one and

$$deg(\prod_{i \notin S} x_i \cdot p(x_1, ...x_n)) \leq \frac{n - \sqrt{n}}{2} + \sqrt{n} = \frac{n + \sqrt{n}}{2}.$$

Now we bound the number N of multi-linear polynomials of degree at most $\frac{n}{2} + \frac{\sqrt{n}}{2}$:

$$N \leq 3^{\sum_{i=0}^{\frac{n+\sqrt{n}}{2}} \binom{n}{i}} \leq 3^{0.85(2^n)}$$

Hence, $|A| \leq 0.85(2^n)$, as desired.

CS277: Concrete Complexity **Fall 1996**

Lecture 24: November 21, 1996

Lecturer: Uri Zwick *Scribe: Lisa Hales*

24.2 $O(\log(n))$-depth Sorting Networks

24.2.1 Introduction

In this lecture we prove the existence of $O(\log n)$-depth sorting networks, commonly known as AKS sorting networks. This result was first proved by Ajtai, Komlos, and Scermeredi (hence the AKS) in 1983, but the version which we present is due to Paterson (*Algorithmica*,'90). The constant coefficient in the original result was 10^{80}. At the present time this constant has been improved to about 1000, but, since there are sorting networks of depth $\frac{1}{2}\log^2(n)$, the result remains of theoretical importance only.

24.2.2 Halvers

Definition 1 *A halver is a comparator network unit which takes as input n numbers and which outputs two sets, L and R, where L=\{$\frac{n}{2}$ smallest inputs\} and R=\{$\frac{n}{2}$ greatest inputs\}.*

Note 1 *In the above definition and throughout this lecture we will be ignoring issues of divisibility and assuming that n is a power of 2. This simplifies considerably the discussion and it is not hard the see that the result generalizes to all n.*

Armed with a halver, we can sort using a binary tree with $\log(n)$ levels. Thus, if we could construct constant-depth halvers we would be done. Unfortunately, we have the following:

Observation 1 *Any halver requires $\Omega(log(n))$ depth*

24.2.3 ϵ-Halvers

Definition 2 *An ϵ-halver is a comparator network unit which receives n inputs and outputs two sets, L and R, such that for all $k \leq \frac{n}{2}$, less than ϵk of the k smallest inputs are in R and likewise less than ϵk of the k largest inputs are in L.*

In order to build our AKS sorting networks we will first need to construct constant depth ϵ-halvers, and then modify our binary tree in figure 2 so that it can accomodate approximate, rather than exact, halvers.

24.2.4 The Construction of Constant-Depth ϵ-Halvers

Definition 3 *A bipartite graph $G = (V, U, E)$ with $|V| = |U| = n$ is an (α, β) expander iff*

- *For all $X \subseteq U$, if $|X| \leq \beta \cdot n$ then $|N(X)| \geq \alpha|X|$.*
- *For all $Y \subseteq V$, if $|Y| \leq \beta \cdot n$ then $|N(Y)| \geq \alpha|Y|$.*

where $N(X)$ is the set of neighbors of X.

Theorem 3 *If there is an $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expander of maximal degree Δ then there is an ϵ-halver of depth Δ.*

Proof: Let G be an $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expander of maximal degree Δ and $|V| = |U| = \frac{n}{2}$. We can assume without loss of generality that G is regular. G can then be divided into Δ perfect matchings. The ϵ-halver will be composed of Δ levels. In each level we do the comparisons that correspond to one matching. In other words, we consider the first $\frac{n}{2}$ registers to be the vertices of V, the second $\frac{n}{2}$ the vertices of U and we compare two registers iff they are connected by an edge.

We need to prove that this construction yields an ϵ-halver. Suppose not. Then there exists an input sequence and a $k \leq \frac{n}{2}$ such that our construction concludes with more than ϵk of the k smallest elements in R. Let X be the set of registers containing these elements at the conclusion of the sort. The values in X are nondecreasing as a function of time (the registers on the right will always receive the larger of the two elements compared). Hence these registers began with inputs from the smallest k and were never compared with any of the $n-k$ largest elements. Let Y be the set of registers which have been compared with the registers in X. Since the elements in X must never have been compared with the $n-k$ largest elements, each register in Y must at some point contain one of the k smallest elements. Since the elements in Y are nonincreasing, at the conclusion of the process the registers in Y will contain only elements from the first k. Finally, since $|X| > \epsilon k$, $|Y| < (1-\epsilon)k$.

But by our assumption that G is an $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expander, we have

$$|N(X)| \geq \alpha\epsilon k = \frac{1-\epsilon}{\epsilon} \cdot \epsilon k = (1-\epsilon)k,$$

a contradiction to $|Y| = |N(X)|$. ∎

Theorem 4 *There exist* $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ *expanders with maximum degree of* $O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})$.

Proof: Construct a random bipartite graph on $2n$ vertices by choosing Δ independent perfect matchings. We need to choose Δ large enough that the probability that the resulting graph is an $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expander is positive.

Suppose the graph is not an (α, β)-expander with $\alpha = (\frac{1-\epsilon}{\epsilon})$ and $\beta = \epsilon$. Then there exists an $X \subseteq U$, $|X| \leq \epsilon n$ and $Y \subseteq V$, $|Y| < \frac{1-\epsilon}{\epsilon}|X|$ such that $N(X) \subseteq Y$. In other words no matching connects X to a vertex ouside of Y. For a fixed choice of X and Y, with $|X| = k$, the probability of this occurring is

$$\left(\frac{\binom{\alpha k}{k}}{\binom{n}{k}}\right)^{\Delta} \leq \left(\frac{\alpha k}{n}\right)^{k\Delta}$$

Thus, $Pr[\text{G is not an expander}] \leq \sum_{k=1}^{\beta n} \binom{n}{k}(\frac{\alpha k}{n})^{k\Delta}\binom{n}{\alpha k}$.

To insure that this is strictly less than one it suffices to choose

$$\Delta \approx \frac{2H(\epsilon)}{\epsilon \log(\frac{1}{1-\epsilon})} = O(\frac{1}{\epsilon}\log\frac{1}{\epsilon})$$

where $H(\epsilon)$ is the binary entropy of ϵ. ∎

The above nonconstructive proof gives a tight result, i.e. there are no $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expanders of smaller degree. There is also a constructive method which leads to $(\frac{1-\epsilon}{\epsilon}, \epsilon)$ expanders of degree $O(\frac{1}{\epsilon^2})$.