Recent advances in Complexity CIS 6930/CIS 4930 September 10, 2002

Lecture 5

Lecturer: Dr. Meera Sitharam Scribe: Zia Uddin

The Lecture 4 notes contain the proof, using Hastad's Switching Lemma, of the fact that the parity function PARITY cannot be computed by a depth d circuit of "small" size $\leq 2^{(1/10)n^{(1/d)}}$ with unbounded fan-in. That proof is in the Boppana-Sipser paper in the Handbook of TCS. As noted by Andrew, there is a mistake in the authors' proof, which has been duly corrected in Andrew's write-up. Today we shall give a more intuitively appreciable proof of the above fact, again using Hastad's Lemma.

We begin with some definitions given last time. Recall that a *restriction*, ρ, is a mapping from a set $\{x_1, \ldots, x_n\}$ of variables to the set $\{0, 1, *\}$. Here $\rho(x_i) = 0$ (resp. 1) means x_i is set to 0 (resp. 1), while $\rho(x_i) = *$ means x_i remains the variable x_i under the action of ρ. Fix $0 < p < 1$. By a *random restriction ρ from $\mathcal{R}_p$*, we mean a restriction ρ that independently assigns to each variable x_i a value in $\{0, 1, *\}$ with the probabilities $\text{Prob}[\rho(x_i) = 0] = (1-p)/2 = \text{Prob}[\rho(x_i) = 1]$ and $\text{Prob}[\rho(x_i) = *] = p$.

Given a function $f(x_1, \ldots, x_n)$, we now define the *induced function* $f|_\rho$ by an example: Let $f(x_1, x_2, x_3, x_4, x_5) = TH_{3,5}$ and suppose $\rho(x_1) = 1$, $\rho(x_2) = *$, $\rho(x_3) = *$, $\rho(x_4) = 1$, and $\rho(x_5) = *$. Then $f|_\rho(x_2, x_3, x_5) = TH_{1,3}$.

Note that the *induced circuit* $C|_\rho$ computing $f|_\rho$ is obtained from the circuit C computing the original function f by eliminating those gates of C that become determined as a result of ρ's setting any of the variables to 0 or 1.

We now restate

Lemma 1 (Hastad's Switching Lemma) *Let G be an AND of ORs circuit of bottom fan-in (i.e., fan-in of the ORs) $\leq t$. Let ρ be a random restriction from $\mathcal{R}_p$. Then for every $s \geq 0$, $G|_\rho$ can be written as an OR of ANDs circuit of bottom fan-in $\leq s$ with probability $\geq 1 - \alpha^s$, where α is the unique positive root of $\left(1 + \frac{4p}{\alpha(1+p)}\right)^t = \left(1 + \frac{2p}{\alpha(1+p)}\right)^t + 1$.*

Remark 1 By looking at $\neg G$, it is possible to convert an OR of ANDs circuit into an AND of ORs circuit subject to the same restrictions: The proof of Hastad's Lemma works just as well.

Remark 2 It can be shown that if $p = o(1)$, then $\alpha = 2pt/\ln\phi$, where $\phi = (1 + \sqrt{5})/2$, the golden ratio.

We now state the theorem we shall prove today. Note that the statement is slightly different from what is stated as Theorem 6 in the Lecture 4 notes. But this difference is not important.

Theorem 2 *For each $n, k \geq 0$, PARITY cannot be computed by a depth k circuit of size (total number of all gates)* $2^{(1/10)^{(k/(k-1))} n^{1/(k-1)}}$.

Remark 3 This is an almost optimal lower bound since by a previous exercise, there do exist circuits of depth d and size $n2^{n^{1/d}}$ that compute PARITY.

As a corollary to Theorem 2, we can prove

Exercise 1 *Any polynomial size cicuit computing PARITY must have depth* $\geq \frac{\log n}{c + \log\log n}$. *(As an aside, recall that by a previous exercise, the corresponding upper bound result is: PARITY can be computed by $O(n)$ size and $O(\log n)$ depth circuits.)*

Before proceeding to prove Theorem 2, we shall use Hastad's Lemma to prove the following

Theorem 3 *PARITY cannot be computed by a depth k circuit containing* $\leq 2^{(1/10)(n^{1/(k-1)})}$ *subcircuits all of depth* ≥ 2 *and bottom fan-in* $\leq (1/10)n^{1/(k-1)}$.

Proof: Note that when we speak of size in this proof, we do not include the bottom level gates. And as for the bottom level gates, we are only interested in their fan-in. Our proof proceeds by induction on the depth k.

For the basis of induction $k = 2$, recall that in the process of proving Lupanov's result in Lecture 3, we proved that each prime implicant for PARITY must have all n of the variables. Consequently, any depth $k = 2$ circuit computing PARITY must have bottom fan-in $\geq n$. This verifies the induction basis.

Now assume, as the induction step, that Theorem 3 holds for all circuits of depth $\leq k - 1$. Suppose for the sake of contradiction that there does exist a PARITY-computing circuit C of depth k containing $\leq 2^{(1/10)(n^{1/(k-1)})}$ subcircuits all of depth ≥ 2 and bottom fan-in $\leq t = (1/10)n^{1/(k-1)}$. We shall use a random restriction and Hastad's Lemma to contradict the induction step.

Let $p = n^{-1/(k-1)}$. Allow C to be acted upon by a random restriction ρ from $\mathcal{R}_p$. By Hastad's Lemma, for every $s \geq 0$, each depth 2 subcircuit (which may be assumed to be an AND of ORs circuit by Remark 1) of C can be written as an OR of ANDs circuit of bottom fan-in $\leq s$ with probabilty $\geq 1 - \alpha^s$. Let $t = (1/10)n^{1/(k-1)}$. Then $pt = 1/10$ and hence by Remark 2, we have $\alpha = 2pt/\ln[(1+\sqrt{5})/2] = \{5\ln[(1+\sqrt{5})/2]\}^{-1} < 1/2$. Now choose $s = (1/10)n^{1/(k-1)} = t$. It follows that with probability $\geq 1 - \alpha^s > 0$, we can replace the depth 2 AND of ORs subcircuits with OR of ANDs subcircuits and still have the bottom fan-in of $C|_\rho$ bounded by $s = (1/10)n^{1/(k-1)}$. Since this replacement results in two adjacent levels of ORs which can be collapsed to a single level, $C|_\rho$ is a circuit of depth $k - 1$ with bottom fan-in $\leq s = (1/10)n^{1/(k-1)}$. However the size of $C|_\rho$ is the same as the size of C since we do not take into account the bottom gates. Thus $C|_\rho$ is a depth $k - 1$ circuit containing $\leq 2^{(1/10)(n^{1/(k-1)})}$ subcircuits all of depth ≥ 2 and bottom fan-in $\leq (1/10)n^{1/(k-1)}$, which is obtained with a probability ≥ 0.

Note that $C|_\rho$ is not a circuit of n vriables: Some of these variables are set by ρ to be 0 or 1. So we must determine the number m of variables that are assigned the value $*$ by ρ. Then we will prove that $C|_\rho$ is a depth $k - 1$ circuit containing $\leq 2^{(1/10)(m^{1/(k-2)})}$ subcircuits all of depth ≥ 2 and bottom fan-in $\leq (1/10)m^{1/(k-2)}$, thereby contradicting the induction hypothesis.

We will first prove the

Claim 1: The expected value of the number of variables assigned $*$ by ρ is $\sum_{r=1}^{n} r\binom{n}{r}p^r(1-p)^{n-r} = n^{(k-2)/(k-1)}$.

Proof of Claim 1: Observe that if r of the n variables are assigned $*$, then there are 2^{n-r} ways of assigning the reamining $n-r$ variables a 0 or a 1. Since $\text{Prob}[\rho(x_i) = 0] = (1-p)/2 = \text{Prob}[\rho(x_i) = 1]$ and $\text{Prob}[\rho(x_i) = *] = p$, it follows that the probability of assigning any r of the n variables a $*$ is $\binom{n}{r}p^r 2^{n-r}\left(\frac{1-p}{2}\right)^{n-r}$. Now by the definition of expected value, the expected number of variables assigned a $*$ is $\sum_{r=0}^{n} r\binom{n}{r}p^r 2^{n-r}\left(\frac{1-p}{2}\right)^{n-r} = \sum_{r=1}^{n} r\binom{n}{r}p^r(1-p)^{n-r}$.

Now we will show that $\sum_{r=1}^{n} r\binom{n}{r}p^r(1-p)^{n-r} = n^{(k-2)/(k-1)}$. Since $p = n^{-1/(k-1)}$, we have $np = nn^{-1/(k-1)} = n^{1-[1/(k-1)]} = n^{(k-2)/(k-1)}$. Hence we will now prove by induction on n that $\sum_{r=1}^{n} r\binom{n}{r}p^r(1-p)^{n-r} = np$. This equality is certainly true for $n = 1$. Suppose as the induction hypothesis that $\sum_{r=1}^{n-1} r\binom{n-1}{r}p^r(1-p)^{n-r-1} = (n-1)p$. Recall the combinatorial fact $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$. Using this fact, we obtain $\sum_{r=1}^{n} r\binom{n}{r}p^r(1-p)^{n-r}$

$$
\begin{aligned}
&= \sum_{r=1}^{n} r\left[\binom{n-1}{r-1} + \binom{n-1}{r}\right]p^r(1-p)^{n-r} \\
&= \sum_{r=1}^{n} r\binom{n-1}{r-1}p^r(1-p)^{n-r} \;+\; \sum_{r=1}^{n} r\binom{n-1}{r}p^r(1-p)^{n-r} \\
&= \sum_{r=0}^{n-1}(r+1)\binom{n-1}{r}p^{r+1}(1-p)^{n-r-1} \;+\; \sum_{r=0}^{n-1} r\binom{n-1}{r}p^r(1-p)^{n-r} \\
&= \sum_{r=0}^{n-1}\binom{n-1}{r}\left[(r+1)p^{r+1}(1-p)^{n-r-1} \;+\; rp^r(1-p)^{n-r}\right] \\
&= \sum_{r=0}^{n-1}\binom{n-1}{r}p^r(1-p)^{n-r-1}\left[(r+1)p + r(1-p)\right] \\
&= \sum_{r=0}^{n-1}\binom{n-1}{r}(r+p)p^r(1-p)^{n-r-1} \\
&= \sum_{r=0}^{n-1}\binom{n-1}{r}rp^r(1-p)^{n-r-1} \;+\; \sum_{r=0}^{n-1}\binom{n-1}{r}pp^r(1-p)^{n-r-1} \\
&= \sum_{r=1}^{n-1}\binom{n-1}{r}rp^r(1-p)^{n-r-1} \;+\; p\sum_{r=0}^{n-1}\binom{n-1}{r}p^r(1-p)^{n-r-1} \\
&= (n-1)p \;+\; p \text{ (induction hypothesis and binomial theorem)} \\
&= np, \text{ as required.}
\end{aligned}
$$

Thus we have proved that the expected value of the number of variables assigned $*$ by ρ is $n^{(k-2)/(k-1)}$. Let $m = \lfloor n^{(k-2)/(k-1)} \rfloor$.

Now we prove the

Claim 2: The restriction ρ leaves m of the n variables free with a probability $\geq 1/3$.

Proof of Claim 2: It suffices to show that the sum of the probabilities for assigning r of the varibles a $*$, with r ranging from $m = \lfloor n^{(k-2)/(k-1)} \rfloor$ (the expected number of unfixed variables for our chosen p) to n, exceeds $1/3$. Hence we will prove by induction on n the inequality $\sum_{r=m}^{n} \binom{n}{r} p^r 2^{n-r} \left(\frac{1-p}{2}\right)^{n-r} = \sum_{r=m}^{n} \binom{n}{r} p^r (1-p)^{n-r} \geq 1/3$. For $n = 1$, we have $m = 1$, and the inequality holds. Let $u = \lfloor (n-1)^{(k-2)/(k-1)} \rfloor$, and assume the induction hypothesis, i.e., $\sum_{r=u}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r-1} \geq 1/3$. Note that $u + 1 = m$. Once again, using the combinatorial fact introduced in the proof of Claim 1, we obtain the following:

$$\sum_{r=m}^{n} \binom{n}{r} p^r (1-p)^{n-r} = \sum_{r=m}^{n} \left[\binom{n-1}{r-1} + \binom{n-1}{r} \right] p^r (1-p)^{n-r}$$

$$= \sum_{r=m}^{n} \binom{n-1}{r-1} p^r (1-p)^{n-r} + \sum_{r=m}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r}$$

$$= \sum_{r=u}^{n-1} \binom{n-1}{r} p^{r+1} (1-p)^{n-r-1} + \sum_{r=m}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r}$$

$$= p \sum_{r=u}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r-1} + \sum_{r=m}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-r}$$

$$\geq \frac{p}{3} + p^{n-1} \quad \text{(induction hypothesis and last term in binomial expansion)}$$

$$= \frac{1}{3n^{\frac{1}{k-1}}} + \frac{1}{n^{\frac{n-1}{k-1}}} \quad \text{(since } p = n^{-1/(k-1)}\text{)}$$

$$= \frac{n+2}{3n^{\frac{n-1}{k-1}}} \quad \text{(arithmetic)}$$

$$\geq \frac{1}{3}, \text{ since } n \geq 2 \text{ certainly implies } (n+2)/n^{(n-1)/(k-1)} \geq 1.$$

Thus we have shown that ρ leaves m of the n variables free with a probability exceeding $1/3$.

To finish the proof, note that we have $n = m^{(k-1)/(k-2)}$. And recall that $C|_\rho$ is a depth $k-1$ circuit containing $\leq 2^{(1/10)(n^{1/(k-1)})}$ subcircuits all of depth ≥ 2 and bottom fan-in $\leq (1/10)n^{1/(k-1)}$. Substituting m for n, we see that $C|_\rho$ is a depth $k-1$ circuit containing $\leq 2^{(1/10)(m^{1/(k-2)})}$ subcircuits all of depth ≥ 2 and bottom fan-in $\leq (1/10)m^{1/(k-2)}$. However this contradicts the induction hypothesis.

■

We can now prove that Theorem 3 implies Theorem 2:

Proof of Theorem 2: Suppose for the sake of contradiction that there is a depth k circuit C of size $2^{(1/10)^{(k/(k-1))} n^{1/(k-1)}}$ which computes PARITY. We

may think of this circuit as a depth $k+1$ circuit with bottom fan-in $= 1$. Let $p = 1/10$, $s = (1/10)^{(k/k-1)} n^{1/(k-1)}$, and let ρ be a random restriction from $\mathcal{R}_p$. This time, $t = 1$ and so $\alpha = 2pt/\ln[(1+\sqrt{5})/2] = \{5\ln[(1+\sqrt{5})/2]\}^{-1} < 1/2$, as in the proof of Theorem 3. By Hastad's Lemma and the reasoning used in the proof of Theorem 3, we know with nonzero probability that $C|_\rho$ is a depth k circuit (because of the collapsing of the adjacent levels of ORs) of size $2^{(1/10)^{(k/(k-1))} n^{1/(k-1)}}$ (since size only decreases after the interchanging of gates; we do not include the bottom level gates, so the new size is the number of depth 3 or higher gates of previous circuit) and bottom fan-in s. We saw in the proof of Theorem 2 that the expected number of variables assigned $*$ by ρ is $m = np = n/10$. Substituting $n = 10m$ into $2^{(1/10)^{(k/(k-1))} n^{1/(k-1)}}$, we get $2^{(1/10)^{(k/(k-1))} 10^{1/(k-1)} m^{1/(k-1)}} = 2^{10^{[(1/(k-1))-(k/(k-1))]} m^{1/(k-1)}} = 2^{(1/10) m^{1/(k-1)}}$. And subsituting m for n in $s = (1/10)^{(k/k-1)} n^{1/(k-1)}$ gives a bottom fan in of $\leq (1/10) m^{1/(k-1)}$. It follows that $C|_\rho$ is a depth k circuit (of m variables) of size $\leq 2^{(1/10) m^{1/(k-1)}}$ and bottom fan-in $\leq (1/10) m^{1/(k-1)}$, which cannot exist by Theorem 3.

■

Note that in the proof of Theorem 3, the random restriction ρ is not constructed explicitly but proved to exist with probability ≥ 0. This leads to an

Open Problem 1: Can we prove Theorem 3 without using probabilistic arguments?

Possibility 1: Produce an algorithm that explicitly constructs, from the description of C in the proof of Theorem 3, the circuit $C|_\rho$ that contradicts the induction hypothesis.

Possibility 2: Use an analytic (e.g., Fourier methods) or algebraic approach and polynomial approximations. So far, we have used purely combinatorial methods (finite sets, unions, intersections), which works since we dealt only with AND OR gates. In fact, we need algebraic or analytic methods to deal with general arithmetic circuits, as we will begin to see when we do the Razborov Smolensky lower bound next.

The idea behind Possibility 2 is that bounded depth circuits composed of OR and AND gates only are analogous to low-degree polynomials. The open problem involves a proper formalization of this. For example, the AND of n variables, i.e., $x_1 \wedge \ldots \wedge x_n$, with domain $\{0,1\}^n$ can be thought of as the monomial $x_1 x_2 \cdots x_n$ defined on the vertices of the unit n-cube. This is because AND and the monomial agree on the vertices. I.e, this monomial interpolates AND at these vertices. Thus if we informally identify AND with $x_1 \wedge \ldots \wedge x_n$, then AND is a polynomial of degree n. Of course we can also use other polynomials, e.g., $(x_1 \wedge \ldots \wedge x_n)^k$ to represent AND as well, but these cannot have lower degree since AND depends on all n variables.

As for PARITY, we claim that we can similarly interpolate PARITY by a polynomial of degree n on the vertices of the unit n-cube.

Infact, recall that to interpolate a univariate polynomial in n points, n degrees of freedom are necessary and sufficient – the n coefficients of the polynomial. The sufficiency is because any polynomial of degree n is a linear combination of the basis elements $1, x, \ldots, x^n$, since they form an independent basis.

None of them is a linear combination of the others, over a set of n distinct points on a line.

We want to show that similarly the 2^n (multilinear) monomials $x_1 x_2 \cdots x_k$, $1 \leq k \leq n$ form an independent basis for the space of functions over the vertices of the cube $\{0,1\}^n$.

Exercise: Here is one fallacious "proof" of this independence, i.e, that no such monomial can be written as linear combinations of others. What is wrong about this proof? Fix it. "If we imagine ordering these monomials first by length and then lexicographically (i.e., $x_{1_1} \cdots x_{1_k} \leq x_{2_1} \cdots x_{2_k}$ if and only if in the first index j where the monomials differ, we have $1_j \leq 2_j$,) then clearly no monomial is a combination of any of the previous ones." *Hint:* This proof would imply that the 2^n multilinear monomials form an independent basis for a space of functions defined over *any* 2^n points in n dimensional space. How about if all 2^n points were on the same line? Or if all 2^n points lay in some $k \leq n$ dimensional affine subspace?

Let's assume we have a proof of independence of the 2^n multilinear monomials over $\{0,1\}^n$. So there is a basis of 2^n multilinear monomials which can be used to interpolate on the 2^n vertices of the unit n-cube. Parity alternates between 0 and 1 as one travels along the "diagonals" of the cube. Hence the multilinear polynomial representing PARITY requires n "turning points," which means it must have degree n.

It may seem odd that PARITY and AND are both interpolated by polynomials of the same degree even though PARITY is a more "complicated" function than AND, and hence "ought" to be analogous to a polynomial of higher degree. But this oddity is resolved if we note that:

(i) The polynomial analogous to AND has only one term while that analogous to PARITY has many terms. Besides we are looking at a particular monomial basis.

(ii) We can approximate AND almost everywhere, i.e., on all except one vertex of $\{0,1\}^n$ by the constant zero polynomial. This certainly is not the case for PARITY. So if we are interested in approximating a circuit gate almost everwhere instead of literally everywhere, then AND is analogous to a degree 0 polynomial while PARITY still requires a degree n polynomial to represent itself.

It is generally believed that the solution to Open Problem 1 will help towards solving a second open problem whose "folklore statement" we give next. We will see this problem in greater detail when we look at the randomness-hardness trade-off later in the semester.

Open Problem 2: "Polylog-wise independent distributions fool any constant depth polynomial size unbounded fan-in $\{\neg, \vee, \wedge\}$-circuit C."

Here the distributions are probability functions $\delta : \{0,1\}^n \to [0,1]$ such that $\sum_{x \in \{0,1\}^*} \delta(x) = 1$. The phrase "$t$-wise independent" means: the distribution is induced by n random variables taking values in $\{0,1\}$ and "any set of t of the n variables is independent. So a uniform distribution over $\{0,1\}^n$ is a n-wise independent distribution. And "polylog-wise" means $t = \log^k(n)$. "Fools a circuit C" means "looks like the uniform distribution to the circuit C." And "looks like" means "$[prob_{x \in \{0,1\}^n} C(x) = 1] \approx \sum_{x \in \{0,1\}^*} C(x)\delta(x)$," where $C(x)$ is the value of the circuit C on x, and $[prob_{x \in \{0,1\}^n} C(x) = 1] = |\{x : C(x) =$

$1\}|/2^n$.

Since t-wise independent distributions are a lot easier to generate than n-wise independent or uniform distributions, they can mimic random number/string generators used by constant depth, polynomial size, unbounded fan-in circuits.