Recent advances in Complexity CIS 6930/CIS 4930	September 17, 2002

Lecture 7

Lecturer: Dr. Meera Sitharam — Scribe: Zia Uddin

Today we will finish the proof of the stronger version of Hastad's Lemma started in Lecture 6. We will include the Lecture 6 notes to have the complete proof in one set of notes. Recall that $min(C)$ denotes the maximum possible length of a minterm of the function computed by the circuit C. And given a Boolean function F and a random distribution ρ, we let $F|_\rho$ denote the restriction of F to those variables that are assigned the value 1 by ρ.

Lemma 1 (Stronger Hastad Lemma) *Let* $G = \bigwedge_{i=1}^{w} G_i$ *be a Boolean circuit of* n *variables with an AND gate at the top, where the* G_i*s are circuits with OR gates on top and of fan-in* $\leq t$ *to these OR gates. Let* $F(x_1, \ldots, x_n)$ *be a Boolean function on the same* n *variables, and let* ρ *be a random distribution in* $\mathcal{R}_p$, $p > 0$. *Then for every* $s \geq 1$, *we have* $\text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1] \leq \alpha^s$, *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: I mentioned the following a couple of times when we used Hastad's lemma to prove Theorem 2, Theorem 3 in Lecture 5,6, i.e, the desired exponential (constant depth) circuit size lower bound for parity: Any function and in particular the function computed by the restricted circuit G_ρ above can be written as an OR of ANDS, where the ANDS are the function's minterms. Therefore, for an appropriate choice of s and p, Hastad's switching lemma actually says that *there exists* a restriction (of not too many variables)which allows us to convert an AND of ORS which have small fan-in to an OR of ANDs of small fan in. And this is what we use for Theorem 2 and Theorem 3.

The "there exists" above follows from non-zero probability of the minterm size event being estimated above; minterm size is exactly the bottom level AND fan-in of the resulting OR of ANDs circuit.

Proof: We proceed by induction on w. If $w = 0$, then $G \equiv 1$ and the lemma is clearly true.

Now assume the lemma is true when the number of G_is is $w - 1$ or less. Let G_1 be the rightmost "OR gate." (See Figure 1.) Then we have $\text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1] \leq \max\{\text{I}, \text{II}\}$, where $\text{I} = \text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \equiv 1]$ and $\text{II} = \text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$.

We shall now examine I. Let $F' = F \wedge G_1$. We observe that if $G_1 \equiv 1$,

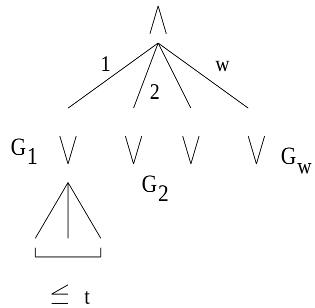

Figure 1: The Circuit G

then $G|_\rho = \bigwedge_{i=1}^{w} G_i|_\rho = \bigwedge_{i=2}^{w} G_i|_\rho$. We have I = Prob$[min(G|_\rho) \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \equiv 1]$ = Prob$[min(G|_\rho) \geq s \mid (F \wedge G_1)|_\rho \equiv 1]$. Thus I is the probability that $\bigwedge_{i=2}^{w} G_i|_\rho$ has a minterm of size at least s given $F'|_\rho \equiv 1$. By the induction hypothesis, we have I $\leq \alpha^s$.

Now we examine II = Prob$[min(G|_\rho) \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$. Suppose that the variables "going into" G_1 belong to a set $T \subseteq \{x_1, \ldots, x_n\}$, where $|T| \leq t$. Write $\rho = \rho_1 \circ \rho_2$, where $\rho_1 : T \to \{0, 1, *\}$ is the restriction of ρ to the variables in T, and $\rho_2 : \{x_1, \ldots, x_n\} \to \{0, 1, *\}$ is the restriction of ρ to the variables not in T and assigns $*$ to the variables in T. We now have $G_1|_\rho \not\equiv 1$ if and only if $G_1|_{\rho_1} \not\equiv 1$. Since G_1 is an OR circuit, $G_1|_\rho \not\equiv 1$ if and only if ρ_1 assigns all the variables in T the values 0 and $*$ only. Thus we in fact have $\rho_1 : T \to \{0, *\}$. Since G is an AND of ORs circuit, every minterm of $G|_\rho$ must make G_1 true. Hence for every minterm σ of $G|_\rho$, there exists a variable $x_i \in T$ such that x_i is part of σ and such that if $\sigma = 1$, then $x_i = 1$. In other words, every minterm of $G|_\rho$ must nontrivially intersect T. Hence we can partition the minterms of $G|_\rho$ according to those variables in T to which the minterms give the values 0 or 1. Now suppose that for a minterm σ of $G|_\rho$, we have $\sigma \cap T = Y$. Then the fact that σ gives the value 0 or 1 to the variables in Y means that all the variables in Y are left unfixed (i.e., assigned $*$) by ρ_1. We will write this event as $\rho_1(Y) = *$. And we will let "$min^Y(G|_\rho) \geq s$" denote the event that $G|_\rho$ has a minterm of size at least s, whose restriction to the variables in T assigns values (0 or 1) to precisely those variables of T that are in Y.

Recall the fact from elementary probability theory that Prob$[A \wedge B \mid C]$ = Prob$[B|C] \cdot$ Prob$[A \mid B \wedge C]$. (This is true because from a diagram of three intersecting circles A, B, and C, it readily follows that $|A \cap B \cap C|/|C| = |B \cap C|/|C| \cdot |A \cap B \cap C|/|B \cap C|$.) Using this and letting A, B, and C denote the events $min^Y(G|_\rho) \geq s$, $\rho_1(Y) = *$, and $F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1$, respectively, we now have:

$$\begin{aligned}
\text{II} &= \text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \\
&\leq \sum_{Y \subseteq T, Y \neq \emptyset} \text{Prob}[min(G|_\rho)^Y \geq s \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \\
&\leq \sum_{Y \subseteq T, Y \neq \emptyset} \text{Prob}[min(G|_\rho)^Y \geq s \wedge \rho_1(Y) = * \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]
\end{aligned}$$

$$= \sum_{Y \subseteq T, Y \neq \emptyset} \text{Prob}[A \wedge B \mid C]$$
$$= \sum_{Y \subseteq T, Y \neq \emptyset} \text{Prob}[B|C] \cdot \text{Prob}[A \mid B \wedge C]$$
$$= \sum_{Y \subseteq T, Y \neq \emptyset} \text{Prob}[\rho_1(Y) = * \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$$
$$\cdot \text{Prob}[min(G|_\rho)^Y \geq s \mid \rho_1(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1].$$

Let P $=$ $\text{Prob}[\rho_1(Y) = * \mid F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$ and let Q $=$ $\text{Prob}[min(G|_\rho)^Y \geq s \mid \rho_1(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$ for notational convenience.

We will now proceed to obtain an upper bound for P using the following three claims:

Claim 1: Looking at P and ignoring the condition $F|_\rho \equiv 1$, we arrive at $\text{Prob}[\rho_1(Y) = * \mid G_1|_\rho \not\equiv 1] = [2p/(1+p)]^{|Y|}$.
Proof of Claim 1: The condition $G_1|_\rho \not\equiv 1$ is equivalent to saying that all variables "going into" G_1 are assigned 0 or $*$ by ρ_1. The probability of a variable going into G_1 being assigned a 0 or a $*$ is $(1-p)/2 + p = (p+1)/2$. Hence the probability of a variable in Y being assigned a $*$, *given* that all variables going into G_1 are assigned 0 or $*$, is $p/[(p+1)/2] = 2p/(p+1)$. It follows that the probability of *every* variable in Y being assigned a $*$, given that all variables going into G_1 are assigned 0 or $*$, i.e., $\text{Prob}[\rho_1(Y) = * \mid G_1|_\rho \not\equiv 1]$, is $[2p/(1+p)]^{|Y|}$.

Claim 2: $\text{Prob}[A \mid B \wedge C] \leq \text{Prob}[A|C]$ if and only if $\text{Prob}[B \mid A \wedge C] \leq \text{Prob}[B|C]$.
Proof of Claim 2: From a diagram of three intersecting circles A, B, and C, it is evident that we have $\text{Prob}[A \mid B \wedge C] \leq \text{Prob}[A|C]$ if and only if we have $|A \cap B \cap C|/|B \cap C| \leq |A \cap C|/|C|$. But we have $|A \cap B \cap C|/|B \cap C| \leq |A \cap C|/|C|$ if and only if $|A \cap B \cap C|/|A \cap C| \leq |B \cap C|/|C|$ if and only if $\text{Prob}[B \mid A \wedge C] \leq \text{Prob}[B|C]$.

Claim 3: $\text{Prob}[F|_\rho \equiv 1 \mid \rho_1(Y) = * \wedge G_1|_\rho \not\equiv 1] \leq \text{Prob}[F|_\rho \equiv 1 \mid G_1|_\rho \not\equiv 1]$.
Proof of Claim 3: The condition $\rho_1(Y) = *$ does not affect the event $F|_\rho \equiv 1$.

Now let A, B, and C denote the events $\rho_1(Y) = *$, $F|_\rho \equiv 1$, and $G_1|_\rho \not\equiv 1$, respectively. Then P $= \text{Prob}[A \mid B \wedge C]$. By Claim 1, we have $\text{Prob}[A|C] = [2p/(1+p)]^{|Y|}$. Thus P $\leq [2p/(1+p)]^{|Y|}$ if and only if $\text{Prob}[A \mid B \wedge C] \leq \text{Prob}[A|C]$. But $\text{Prob}[A \mid B \wedge C] \leq \text{Prob}[A|C]$ if and only if $\text{Prob}[B \mid A \wedge C] \leq \text{Prob}[B|C]$, which is true by Claim 3. Thus we have established an upper bound for P, i.e., the fact that P $\leq [2p/(1+p)]^{|Y|}$.

We will now proceed to obtain an upper bound for Q. Our method will utilize the induction hypothesis. We first need to explain some notation. Let $\sigma \in \{0,1\}^Y$ be an assignment of the variables in Y to 0 and 1. Let "$min^{Y \leftarrow \sigma}(G|_\rho) \geq s$" denote the event that $G|_\rho$ has a minterm of size at least s, whose restriction to the variables in T assigns σ to precisely those variables of T that are in Y, and fixes no other variables in T.

We have Q $=$ $\mathrm{Prob}[min(G|_\rho)^Y \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \leq \sum_{\sigma \in \{0,1\}^Y, \sigma \neq 0^Y} \mathrm{Prob}[min(G|_\rho)^{Y \leftarrow \sigma} \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$.
This is because if $G|_\rho$ has a minterm of size at least s, whose restriction to the variables in T assigns 0 and 1 to precisely those variables of T that are in Y, then this value assignment is some $\sigma \in \{0,1\}^Y$. Hence the sum of probabilities for *all* such σ (excluding $\sigma = 0^Y$ since a minterm must fix some variable in Y to 1) must be an upper bound.

Now fix σ. We have $\mathrm{Prob}[min(G|_\rho)^{Y \leftarrow \sigma} \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \leq \max_{\rho_1} \mathrm{Prob}[min(G|_\rho)^{Y \leftarrow \sigma} \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$. This is because the maximum is taken over all ρ_1 (not to be confused with the specific ρ_1 that we were concerned with earlier) assigning 0s and $*$s to the variables in T and only $*$s to the variables in Y.

Having already fixed σ, we now fix ρ_1 (again, not necessarily the specific ρ_1 that we were concerned with earlier). Let W be the set of variables in $T \setminus Y$ that are assigned $*$ by this ρ_1. Let $\tau \in \{0,1\}^W$ and let $\overline{G}$ be G without G_1, i.e., $\overline{G} = \bigwedge_{i=2}^{w} G_i|_\rho$. Suppose the variables in Y take the assignments given by our fixed σ. Now the phrase "$min((\overline{G}|\sigma \circ \tau \circ \rho_1)|_{\rho_2}) \geq s$" makes sense since $\sigma \circ \tau \circ \rho_1$ fixes all the variables in T, thereby "getting rid off" G_1 and allowing us to use the induction hypothesis. We have $\mathrm{Prob}[min(G|_\rho)^{Y \leftarrow \sigma} \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \leq \max_{\tau \in \{0,1\}^W} \mathrm{Prob}[min((\overline{G}|\sigma \circ \tau \circ \rho_1)|_{\rho_2}) \geq s \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1] \leq \max_{\tau \in \{0,1\}^W} \mathrm{Prob}[min((\overline{G}|\sigma \circ \tau \circ \rho_1)|_{\rho_2}) \geq s - |Y| \mid \rho(Y) = * \wedge F|_\rho \equiv 1 \wedge G_1|_\rho \not\equiv 1]$. This is because the probability of a minterm having a certain length and certain properties is less than the probability of a minterm having the same properties but shorter length. Furthermore, the events $\rho(Y) = *$ and $G_1|_\rho \not\equiv 1$ do not depend on ρ_2, and hence can be dropped. So if we fix the maximizing $\tau \in \{0,1\}^W$, we obtain

$$Q \leq \sum_{\sigma \in \{0,1\}^Y, \sigma \neq 0^Y} \max_{\rho_1} \mathrm{Prob}[min((\overline{G}|\sigma \circ \tau \circ \rho_1)|_{\rho_2}) \geq s - |Y| \mid (F|_\rho)_{\rho_2} \equiv 1]$$

$$\leq \sum_{\sigma \in \{0,1\}^Y, \sigma \neq 0^Y} \max_{\rho_1} \alpha^{s-|Y|} \quad \text{(induction hypothesis)}$$

$$= \sum_{\sigma \in \{0,1\}^Y, \sigma \neq 0^Y} \alpha^{s-|Y|}$$

$$= (2^{|Y|}-1)\alpha^{s-|Y|}$$ ($2^{|Y|}$ ways of assigning 0s and 1s to variables in Y, including the all 0 assignment).

We now have $\mathrm{II} \leq \sum_{Y \subseteq T, Y \neq \emptyset} \mathrm{PQ} \leq \sum_{Y \subseteq T, Y \neq \emptyset} [2p/(1+p)]^{|Y|}(2^{|Y|}-1)\alpha^{s-|Y|}$

$$= \alpha^s \sum_{Y \subseteq T, Y \neq \emptyset} \left(\frac{2p}{\alpha(1+p)}\right)^{|Y|} (2^{|Y|}-1)$$

$$= \alpha^s \sum_{Y \subseteq T, Y \neq \emptyset} \left(\frac{4p}{\alpha(1+p)}\right)^{|Y|} - \alpha^s \sum_{Y \subseteq T, Y \neq \emptyset} \left(\frac{2p}{\alpha(1+p)}\right)^{|Y|}$$

$$= \alpha^s \sum_{i=1}^{|T|} \binom{|T|}{i} \left(\frac{4p}{\alpha(1+p)}\right)^i - \alpha^s \sum_{i=1}^{|T|} \binom{|T|}{i} \left(\frac{2p}{\alpha(1+p)}\right)^i$$

(number of ways of choosing i-element nonempty subsets of T)

$$
\begin{aligned}
&\leq \alpha^s \sum_{i=1}^{t} \binom{t}{i} \left(\frac{4p}{\alpha(1+p)}\right)^i - \alpha^s \sum_{i=1}^{t} \binom{t}{i} \left(\frac{2p}{\alpha(1+p)}\right)^i \quad (|T| \leq t) \\
&= \alpha^s \left[\left(1 + \frac{4p}{\alpha(1+p)}\right)^t - \left(1 + \frac{2p}{\alpha(1+p)}\right)^t \right] \quad \text{(Binomial Theorem)} \\
&= \alpha^s,
\end{aligned}
$$

since α is the solution to the equation mentioned in the statement of the lemma.

It follows that $\text{Prob}[min(G|_\rho) \geq s \mid F|_\rho \equiv 1] \leq \max\{ \text{I, II}\} \leq \alpha^s$, and the proof is complete.

■

Aside: Two events A and B are *independent* if and only if $\text{Prob}[A \wedge B] = \text{Prob}[A] \cdot \text{Prob}[B]$, i.e, $|A \cap B|/|U| = |A||B|/|U|^2$, where U is the universe. All events are subsets of the universe. This is also equivalent to saying that $\text{Prob}[A|B] = \text{Prob}[A]$, i.e, $|A \cap B|/|B| = |A|/|U|$. Now recall the two facts concerning conditional probabilities that were used in the proof above:
Fact 1: $\text{Prob}[A \wedge B \mid C] = \text{Prob}[B|C] \cdot \text{Prob}\ [A \mid B \wedge C]$.
Fact 2: $\text{Prob}[A \mid B \wedge C] \leq \text{Prob}[A|C]$ if and only if $\text{Prob}[B \mid A \wedge C] \leq \text{Prob}[B|C]$.

Exercise 1 *These two facts concerning conditional probabilities and events A, B, and C hold irrespective of whether the three events are independent or not. In particular, the two facts hold even if the three events A, B, and C are all the identical event, say, A.*

We have seen that constant-depth $\{\wedge, \vee, \neg\}$-circuits must have exponential size in order to compute PARITY, which is a mod 2 computation. In the next lecture, we shall see the Razborov-Smolensky result, which extends this result using the oracle technique to show that for any prime p, constant-depth $\{\wedge, \vee, \neg, \text{mod}p\}$-circuits also must have exponential size in order to compute MAJORITY and to carry out mod q computations for any $q \neq p^k$. This will involve showing that $\wedge$, $\vee$, $\neg$, and mod p can be approximated by low-degree polynomials, while MAJORITY and mod q computations require polynomials of large degree. Meanwhile we have an

Exercise 2 *Show that the PARITY lower bound also applies to MAJORITY. In particular,*
(i) Show exactly where to change the proof for PARITY to prove that constant-depth $\{\wedge, \vee, \neg\}$-circuits must have exponential size in order to compute MAJORITY.
(ii) Characterize the class of functions for which the argument in part (i) holds.