Lecture 9

Decision Trees, Certificate Complexity, and Håstad's Switching Lemma

April 29, 2008
Lecturer: Paul Beame
Notes: Ben Birnbaum

9.1 Decision Tree Complexity

Definition 9.1 A Boolean *decision tree* is a rooted binary tree with each internal node labeled by the name of a variable and each leaf labelled by either 0 or 1.

A decision tree as computes a function $f : \{0,1\}^n \to \{0,1\}$ as follows: On input x, the following process is repeated starting at the root v of the tree: The variable x_i that labels v is queried. If x_i is 1, then the right child of v is queried; otherwise, the left child of v is queried. The value of $f(x)$ is the label at the leaf to which this process leads.

Definition 9.2 For a function $f : \{0,1\}^n \to \{0,1\}$, let the *decision tree complexity* of f, $D(f) = \min\{height(T) \mid T \text{ computes } f\}$.

Clearly, $D(OR_n) = D(Parity_n) = n$. One way that a function may have small decision tree complexity is if it does not depend on all its inputs. In fact, there are natural functions f that depend on all their inputs for which $D(f)$ is $o(n)$. Consider, for example an indexing function in which the first $\log n$ bits of the input represent an integer that gives the offset into the rest of the input. The function returns the bit at that offset. In other words, the input is $(\langle i \rangle, x)$ and the output is x_i. Clearly, there is a decision tree that computes this function of depth $\log_2 n + 1$.

9.2 Certificate Complexity

Recall that a partial Boolean assignment ρ on n variables, which is also known as a *restriction*, may be represented as a function $\rho : \{0,1\}^n \to \{0,1,*\}$. For a restriction ρ, we use the notation $|\rho|$ to denote the number of values that ρ assigns. We say that ρ is *consistent with* an input x iff for all i, either $\rho(i) = x_i$ or $\rho(i) = *$.

Definition 9.3 For $b \in \{0,1\}$, we say that a partial assignment ρ is a *b-certificate* for a function $f : \{0,1\}^n \to \{0,1\}$ if $f|_\rho(x) = b$ for all $x \in \{0,1\}^n$.

Definition 9.4 For $b \in \{0,1\}$ and a function $f : \{0,1\}^n \to \{0,1\}$, let

$$C_b(f) = \max_{x \in f^{-1}(b)} \min_{\{} |\rho| : \rho \text{ is } b\text{-certificate consistent with } x\}$$

be the *b-certificate complexity* of f.

Note that $C_1(OR_n) = 1$, $C_0(OR_n) = n$, and $C_1(Parity_n) = C_0(Parity_n) = n$. Also, note that an equivalent definition of C_1 and C_0 is

$$\begin{aligned} C_1(f) &= \min_{\text{DNF formulas } F \equiv f} \{\text{length of longest term in } F\} \\ C_0(f) &= \min_{\text{CNF formulas } F \equiv f} \{\text{length of longest clause in } F\} \ . \end{aligned}$$

Proposition 9.5 For a function $f : \{0,1\}^n \to \{0,1\}$, a 0-certificate for f, ρ_0, and a 1-certificate for f, ρ_1, there must be at least one $i \in [n]$ such that $\rho_0(i) \neq \rho_1(i)$ and both $\rho_0(i)$ and $\rho_1(i)$ are in $\{0,1\}$. In other words, any 0-certificate and 1-certificate must intersect and be inconsistent.

Theorem 9.6 For any function $f : \{0,1\}^n \to \{0,1\}$,

1. $C_0(f) \leq D(f)$ and $C_1(f) \leq D(f)$, and
2. $D(f) \leq C_0(f) \cdot C_1(f)$.

Proof To prove the first part, let T be a minimum-height decision tree for f. Let x be any input for which $f(x) = b$ for $b \in \{0,1\}$. The partial assignment given by the path determined by x is a b-certificate for f, and it has size less than $D(f)$. Hence $C_b(f) \leq D(f)$.

We now prove the second part. Let $F = T_1 \vee T_2 \vee \cdots \vee T_k$ be a DNF formula for f such that the maximum term size is $C_1(f)$. We will build a decision tree for f term by term. Let $T_1 = (x_{i_1}^{b_1} \wedge \ldots \wedge x_{i_\ell}^{b_\ell})$ where $x_i^b = x_i$ if $b = 0$ and $x_i^b = \neg x_i$ if $b = 1$.

The part of the decision tree that corresponds to T_1 will have ℓ levels, with variable x_{i_j} being queried at each node at level j. The path in this tree that satisfies T_1 will end at a leaf labelled 1. For every other assigned π to the variables $x_{i_1}, \ldots, x_{i_\ell}$ leading to a leaf, we add the subtree that is computed recursively starting at T_2 in $F|_\pi$. (Note that π can be thought of either as a path or a partial assignment.)

Each step in this recursive construction adds at most $C_1(f)$ levels, so if we can prove that there are most $C_0(f)$ recursive steps, then we are done. This follows easily from Observation 9.5: the set of variables in T_1 (or any other term) can be part of a 1-certificate; therefore every 0-certificate must contain at least one of these variables. Hence, for any path π in the part of the tree corresponding to T_1, we have $C_0(f|_\pi) \leq C_0(f) - 1$. Since the 0-certificate complexity decreases by 1 at each level in the recursion, the recursion can have a depth of at most $C_0(f)$. □

In the above proof, the construction of the decision tree from the DNF formula is done in canonical fashion given any fixed ordering of the terms of F and ordering of the literals in each term. We will analyze this canonical construction in the context of constant-depth circuits.

9.3 Constant-depth Circuits and Håstad's Switching Lemma

We know that $Parity_n$ requires DNF or CNF formulas of size at least $n2^{n-1}$. These are depth 2 circuits. We will see that $Parity$ requires exponential-size circuits of any constant depth.

Theorem 9.7 (Furst-Saxe-Sipser, Ajtai) $Parity \notin \mathsf{AC}^0$.

The basic idea of the argument is that under any restriction ρ of the input values $D(Parity_n|_\rho) = |unset(\rho)| = n - |\rho|$ whereas $D(OR_n) = 0$ for most restrictions. However, by repeated application of restrictions we can simplify each layer of unbounded fan-in gates in any small constant-depth circuit and therefore such circuits cannot compute parity.

Originally this was done using a so-called "switching lemma" that allowed one to replace ORs or small ANDs by ANDs of small ORs after applying a restriction. The strongest of these is due to Håstad. We find the following form of switching lemma more convenient.

Let R_n^ℓ be the set of all restrictions on n variables that leave precisely ℓ variables unset.

Lemma 9.8 (Håstad's Switching Lemma) Let $f : \{0,1\}^n \to \{0,1\}$ be a function with $C_1(f) \leq r$ then for ρ chosen uniformly at random from R_n^ℓ,

$$\Pr[D(f|_\rho) \geq s] < \left(\frac{8\ell r}{n-\ell}\right)^s.$$

Proof Let F be a DNF formula of term size at most r that represents f. We use a counting argument that is a variant of one suggested by Razborov. We will show that the set of all $\rho \in R_n^\ell$ such that $D(f|_\rho) \geq s$ is small by giving a 1-1 map from such ρ to the set

$$R_n^{\ell-s} \times stars(r,s) \times \{0,1\}^s$$

where $stars(r,s) \subseteq (\{*,-\}^r - \{-\}^r)^*$ is the set of sequences of length r strings of stars and dashes that have s total stars and at least one star per string. This will let us specify which variables in terms of F are unset and therefore potential contributing variables to $D(f_\rho)$.

We first describe the map. Given F, we use some canonical ordering of the terms of F. The map is based on the canonical conversion of $F|_\rho$ to a decision tree which will clearly compute $f|_\rho$. Let π denote the left-most (partial) path in this tree that has length at least s. This path corresponds to a partial assignment that we also denote by π. By construction, this canonical decision tree is produced by taking the first term of $F|_\rho$ which is the first term of F that is not assigned to 0 by ρ and building a tree that queries all variables in that term that are assigned stars by ρ. Let σ_1 be the partial assignment to those variables that satisfies that first term and π_1 be the portion of π assigning values to those variables. Then by construction, the sub-tree below π_1 is the canonical decision tree for $(F|_\rho)|_{\pi_1} = F|_{\rho\pi_1}$ where the concatenation of two partial assignments on disjoint sets of variables is the partial assignment that assigns values for both of them. This sub-tree is produced beginning with the first term of $F|_{\rho\pi_1}$ that is not set to 0. Let σ_2 be the assignment that saitsfies this second term, π_2 be the corresponding portion of π and repeat until all of π is exhausted. (Note that since we cut π at length s, the last σ_k and corresponding π_k might not assign values to all unset variables.) A picture of the tree and the corresponding part of the construction is shown in Figure 9.1.

The first component of the map of ρ is $\rho' = \rho\sigma_1 \cdots \sigma_k$. In total $\sigma_1 \cdots \sigma_k$ gives values to the same s variables that π does so $\rho' \in R_n^{\ell-s}$. The second component of the map will be a sequence

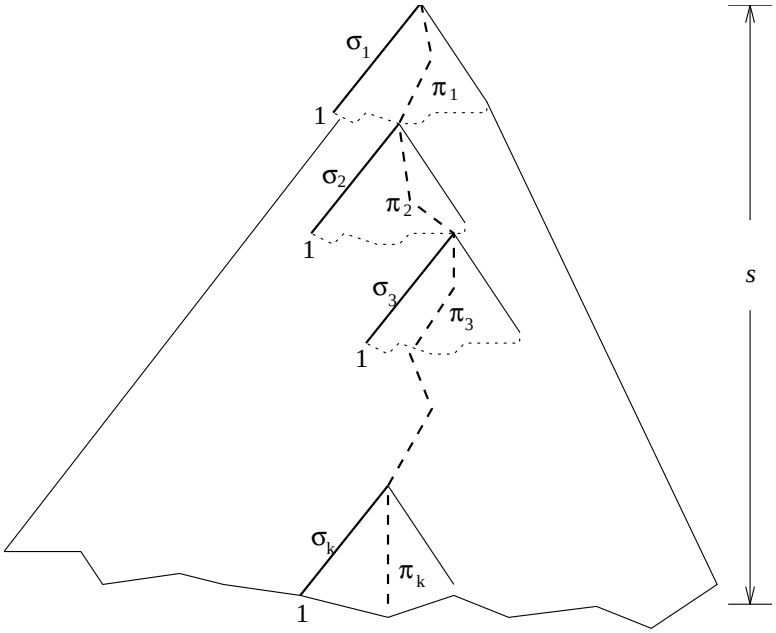

Figure 9.1: Canonical decision tree for $F|_\rho$.

of k strings of length r, each of which has a $*$ in the positions of the unset variables contributing to the σ_i and a $-$ in the other positions. A $*$ will appear a total of s times in this sequence so the result is in $stars(r,s)$. The last component is a binary string of length s that says how $\sigma_1 \cdots \sigma_k$ differs from $\pi = \pi_1 \cdots \pi_k$.

Observe that $|stars(r,s)| \leq 2^{s-1} r^s$. This follows easily by encoding an element of $stars(r,s)$ by listing the position of each $*$ within its string of length r and by an additional bit for each $*$ (but the last) that will be 1 if the next $*$ is in the same term and 0 otherwise. (Sharper calculations can show that $|stars(r,s) \leq (r/\ln 2)^s$ but we will not use them.) Therefore, if we can show that the map is 1-1 then the probability that $D(f_\rho) \geq s$ is at most

$$\begin{aligned}
\frac{|R_n^{\ell-s} \times stars(r,s) \times \{0,1\}^s}{|R_n^\ell|} &\leq \frac{|R_n^{\ell-s}| \cdot 2^{s-1} r^s \cdot 2^s}{|R_n^\ell|} \\
&= \frac{\binom{n}{\ell-s} 2^{n-\ell+s} \cdot 2^{s-1} r^s \cdot 2^s}{\binom{n}{\ell} 2^{n-\ell}} \\
&< \frac{\binom{n}{\ell-s} \cdot (8r)^s}{\binom{n}{\ell}} \\
&= \frac{(n-\ell)!\ell!}{(n-\ell+s)!(\ell-s)!} \cdot (8r)^s \\
&\leq \frac{\ell^s}{(n-\ell)^s} \cdot (8r)^s,
\end{aligned}$$

which is the bound we require.

It remains to show that the map is 1-1 given F. To do so, we show how given access to F, $\rho' \in R_n^\ell$ and the elements of $stars(r,s)$ and $\{0,1\}^s$ we can reconstruct ρ uniquely.

Since $\rho' = \rho\sigma_1 \cdots \sigma_k$, the first term of F not set to 0 by ρ' will in fact be the first term of F not set to 0 by ρ. This is because σ_1 will force the first term not set to 0 by ρ to have value 1 (unless $k = 1$ and π has been truncated to be shorter than the number of unset variables, in which case the term certainly won't set it to 0). Knowledge of F has identified the term; now the first string in the image in $stars(r, s)$ will identify the variables in that term that are unset by ρ and therefore identify which portion of ρ' is σ_1. The initial segment of $|\sigma_1|$ bits in the length s binary string of the third component will allow us to identify π_1. We can now change ρ' to $\rho'' = \rho\pi_1\sigma_2 \cdots \sigma_k$. Since the sub-tree below π_1 is the canonical tree for $F|_{\rho\pi_1}$ we are in the same position to do the analogous determination in the next round. In the i-th round we find the first term of F not set to 0 by $\rho\pi_1 \cdots \pi_{i-1}\sigma_i \cdots \sigma_k$. Then we use the i-th string in $stars(r, s)$ to identify σ_i and the next $|\sigma_i|$ bits of the third component of the map to identify π_i. Eventually we find all σ_i, π_i, and then we can determine ρ itself. Therefore the map is 1-1 as required and the lemma follows. □

To prove the lower bound we will apply this switching lemma to all nodes in the circuit and use the probabilistic method to argue that there exists a restriction so that the output of a small circuit of small depth is smaller than suffices to compute parity. We complete the argument in our next class.