# Lecture 10

# Lower bounds for constant-depth circuits

May 1, 2008
Lecturer: Paul Beame
Notes:

Let $R_n^\ell$ be the set of all restrictions on $n$ variables that leave precisely $\ell$ variables unset. Since decision tree complexity $D(f) = D(\neg f)$, we can restate the decision tree version of Håstad Switching Lemma as follows:

**Lemma 10.1 (Håstad's Switching Lemma)** Let $b \in \{0,1\}$ and $f : \{0,1\}^n \to \{0,1\}$ be a function with $C_b(f) \leq r$ then for $\rho$ chosen uniformly at random from $R_n^\ell$,

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

We use this to derive lower bounds for $\mathsf{AC}$ circuits computing Parity.

**Theorem 10.2** Any $\mathsf{AC}$ circuit computing $Parity_n$ using size $S$ and depth $d$ satisfies $S \geq 2^{n^{1/(d-1)}/17}$.

**Proof** Let $C$ be any $\mathsf{AC}$ circuit of size $S$ and depth $d$. For each node $v$ in $C$ let $f_v$ be the function of the inputs computed at node $v$. We count the height of a node $v$ to be the maximum number of AND or OR gates on any path from $v$ to an input node.

Define $n_1 = n/17$ and, more generally, let

$$n_{i+1} = \frac{n}{17(17\log_2 S)^i}$$

for $0 \leq i < d$. We will show that for each $1 \leq i \leq d$ there is a restriction $\rho_i \in R_n^{n_i}$ such that for every node $v$ at height at most $i$ above the leaves of $C$, $D(f_v|_{\rho_i}) \leq \log_2 S$.

For any node $v$ of height 1, either $C_0(f_v) = 1$ or $C_1(f_v) = 1$. Therefore we can apply the switching lemma with $r = 1$, $s = \log_2 S$ and $\ell = n_1 = n/17$ to say that for $\rho \in R_n^{n_1}$, the probability

$$\Pr[D(f_v|_\rho) \geq \log_2 S] < (\frac{8n/17}{n - n/17})^s = 2^{-s} = 1/S.$$

Therefore by a union bound over the at most $S$ nodes of the probability that there exists a node $v$ of height 1 with $D(f_v|_\rho) \geq \log_2 S$ is strictly less than 1. By the probabilistic method there must exist a restriction $\rho$ satisfying this property for all gates of height 1. Call this restriction $\rho_1$.

For the inductive step, consider nodes $u$ of height $i+1$ for $i > 0$. Observe that if $u$ is an OR node then by the inductive hypothesis, $D(f_v|_{\rho_i}) \leq \log_2 S$ for some $\rho_i \in R_n^{n_i}$ and all nodes $v$ that are inputs to $u$. It follows that $C_1(f_v|_{\rho_i}) \leq \log_2 S$ for all such nodes and therefore $C_1(f_u|_{\rho_i}) \leq \log_2 S$. We apply the switching lemma with $r = \log_2 S$, $s = \log_2 S$ and $\ell = n_{i+1}$ to say that for $\rho$ chosen randomly from $R_{n_i}^{n_{i+1}}$,

$$\Pr[D(f_v|_{\rho_i\rho}) \geq \log_2 S] = \Pr[D((f_v|_{\rho_i})|_\rho) \geq \log_2 S] < (\frac{8n_{i+1}\log_2 S}{n_i - n_{i+1}})^s = (\frac{8n_i/17}{n_i(1-1/(17\log_2 S)})^s < 2^{-s} = 1/S.$$

Similarly, if $v$ is an AND gate we use $C_0$ and if $v$ is a NOT gate we simply complement the decision tree of its child. Since $(f_v|_{\rho_i}|_\rho = f_v|_{\rho_i\rho}$ the probability that a random $\rho$ fails at some node of height $i+1$, given that lower heights have all been successful, is then $< 1$. Again using the probabilitic method we obtain an $\rho_{i+1} = \rho_i\rho$ in $R_n^{n_{i+1}}$ as required.

Now $D((Parity_n)|_{\rho_d}) = n_d$ since $(Parity_n)|_{\rho_d}$ is another parity function or its negation and therefore must have decision tree height $n_d$. Therefore $\log_2 S \geq n_d = \frac{n}{17^d(\log_2 S)^{d-1}}$. Rewriting we obtain that $(17\log_2 S)^d \geq n$ and therefore $S \geq 2^{n^{1/d}/17}$.

We can do a little better by only applying $\rho_{d-1}$ and using the fact that $C_0(Parity_{n_{d-1}}) = C_1(Parity_{n_{d-1}}) = n_{d-1}$ . If the output gate is an AND gate, say, then each node $v$ at height $d-1$ has $C_0((f_v)|_{\rho_{d-1}} \leq \log_2 S$ and therefore the output has $C_0$ value at most $\log_2 S$. If the output is an OR gate then we use that $C_1((f_v)|_{\rho_{d-1}} \leq \log_2 S$ and so the output has 1-certificate complexity $C_1$ at most $\log_2 S$. Therefore $\log_2 S \geq n_{d-1} = \frac{n}{17^{d-1}(\log_2 S)^{d-2}}$. It follows that $(17\log_2 S)^{d-1} \geq n$. Therefore $S \geq 2^{n^{1/(d-1)}}/17$ as required. □

In particular, this proves that Parity is not in $\mathsf{AC}^0$. In fact, polynomial-size $\mathsf{AC}$ circuits for parity must have much more than constant depth.

**Corollary 10.3** The depth complexity of polynomial-size $\mathsf{AC}$ circuits for *Parity* is $\Theta(\log n/\log\log n)$

**Proof** If we set the circuit size $S$ to be polynomial in $n$ then we must have that $n^{1/(d-1)}/17$ is at most $\log_2 S$ which is $O(\log n)$. Therefore $d$ is $\Omega(\log_{\log_n} n) = \Omega(\log n/\log\log n)$. Previously, we showed that there are $\mathsf{AC}$ circuits of polynomial size and $O(\log n/\log\log n)$ depth for any $\mathsf{NC}^1$ function. □

Here's one important corollary that I did not get to in class. It was part of the original motivation for Furst, Saxe, and Sipser.

**Lemma 10.4** There is an oracle $A$ such that $\mathsf{PH}^A \neq \mathsf{PSPACE}^A$.

**Proof** Define the language

$$Parity(A) = \{1^n \ : \ |A \cap \{0,1\}^n| \text{ is odd}\}.$$

Clearly $Parity(A) \in \mathsf{PSPACE}^A$ since a Turing machine with $O(n)$ space can make all $2^n$ calls to $A$ on elements of $\{0,1\}$ and count the number of accepted strings.

Now, as we have seen, we can view each $\Sigma_k^p$ or $\Pi_k^p$ algorithm as an unbounded fan-in circuit with $\vee$'s of fan-in $2^{q(n)}$ for each $\exists$ quantifier and $\wedge$'s of the same fan-in for each $\forall$ quantifier for some polynomial $q$. Moreover, when we add the ability to make oracles calls we can extend the last $\exists$ quantifier to guess the values of all oracle calls so that polynomial-time predicate depends only on the conjunction of the answers to its oracle calls. We can view each oracle answer $A(y)$ for $y \in \{0,1\}^n$ as an input variable to our circuits. Therefore since the input a $\Sigma_k^p$ or $\Pi_k^p$ algorithm with oracle for $A$ computing $Parity(A)$ yields an unbounded fan-in circuit of depth $k+2$ and size $2^{O(kq(n))}$ that computes $Parity_{2^n}$. Letting $N = 2^n$, these have size $2^{\log^{O(1)} N}$ which is impossible for any constant $k$ so $Parity(A) \in \mathsf{PSPACE}^A - \mathsf{PH}^A$. $\square$

## 10.1 Unbounded fan-in circuits with modular counting gates

In the above we have seen that Parity is hard for unbounded fan-in circuits. What happens if we add unbounded fan-in parity gates $\oplus$ to the circuits? These gates compute the sum of the inputs modulo 2. We can generally think about the analogous computation modulo $p$ but since we need Boolean values for the other gates we consider the $MOD_p$ gates given by

$$MOD_p(x_1, \ldots, x_n) = \begin{cases} 0 & \text{if } \sum_{i=1}^n x_i \equiv 0 \pmod p \\ 1 & \text{otherwise.} \end{cases}$$

**Definition 10.5** Let $\mathsf{AC}^0[p]$ denote the set of functions $f : \{0,1\}^* \to \{0,1\}^*$ omputable by constant-depth unbounded fan-in circuits of $\neg$, $\vee$, and $MOD_p$ gates. (For convenience we don't include unbounded fan-in $\wedge$ gates since they are not necessary.) A common alternative notation this is $\mathsf{ACC}_p^0$ where $\mathsf{ACC}$ stands for alternative circuits with counters. Also define $\mathsf{ACC} = \bigcup_p \mathsf{AC}[p] = \bigcup_p \mathsf{ACC}_p^0$.

**Theorem 10.6 (Razborov,Smolensky)** $MOD_p \notin \mathsf{AC}^0[q]$ for all primes $p \neq q$.

Since we easily have $MOD_p \leq_{AC^0} Majority$ we easily have:

**Corollary 10.7** $Majority \notin \mathsf{AC}^0[p]$ for all primes $p$.

Both of the above statements can be extended to prime powers involving distinct primes. We will not prove this in its full generality. For simplicity we will just show that $\oplus \notin \mathsf{AC}^0[q]$ for any odd prime $q$. We will obtain a lower bound nearly as strong as for $\mathsf{AC}^0$.

**Theorem 10.8** Any $\mathsf{AC}[p]$ circuit computing $\oplus$ on $n$ bits in size $S$ and depth $d$ must have $S \geq \frac{1}{50} q^{n^{1/(2d)}/(q-1)}$.

**Proof** The proof of this theorem introduces the Method of Approximation. The general idea of this method is to consider a class of approximating functions and to define an approximator $\tilde{g}$ for each gate $g$ in the given circuit $C$. If gate $g$ has inputs $y_1, \ldots, y_\ell$ where the $y_i$ themselves depend on the input $x$ then we require that $\tilde{g}(y) = g(y)$ for all but at most an $\epsilon$ fraction of $x$. If we denote the output of the circuit $C$ by $C(x)$ then the above will show that $\tilde{C}(x) = C(x)$ except for at most an $S\epsilon$ fraction of inputs $x$. If one can show that any approximator in the class must disagree from the function to be computed in at least a $\delta$ fraction of inputs, then $S\epsilon \geq \delta$ which yields a lower bound of $S \geq \delta/\epsilon$.

The class of approximators we will consider will be polynomials over $\mathbb{F}_q$ of somewhat small total degree.

Observe that by Fermat's Little Theorem since $q$ is prime

$$MOD_q(y_1,\ldots,y_\ell) = (y_1 + \cdots + y_\ell)^{q-1}.$$

In this case there is loss at all and if we have polynomials of degree $d$ for each of the $y_i$ then the degree of $MOD_q(y_1,\ldots,y_\ell)$ is at most $(q-1)d$.

Similarly $\neg y = (1-y)$ which is also exact.

The place we will approximate is in computing $\vee(y_1,\ldots,y_\ell)$. If we wanted to do this exactly we would use the polynomial $1 - \prod_{i=1}^{\ell}(1-y_i)$ which has degree equal to the sum of the degrees of the $y_i$ which might be very large.

Instead, we use the following trick using the probabilistic method due to Razborov, which is a simpler form of the construction of Valiatn-Vazirani.

Choose $\vec{r}$ uniformly at random from $\mathbb{F}_q^\ell$ and consider $\sum_i r_i y_i$ in $\mathbb{F}_q$. Now if $\vee_{i=1}^{\ell} y_i = 0$ then $\sum_{i=1}^{\ell} r_i y_i = 0$. On the other hand if $\vee_{i=1}^{\ell} y_i = 1$ then $\Pr[\sum_{i=1}^{\ell} r_i y_i = 0] = 1/q$. Therefore $\Pr[\sum_{i=1}^{\ell} r_i y_i)^{q-1} \neq \vee_{i=1}^{\ell} y_i] \leq 1/q$.

To improve the approximation for this $\vee$ gate well we will do this $k$ times independently and take the $\vee$ of the result. Therefore,

$$\Pr[1 - \prod_{j=1}^{k}(1 - \sum_{i=1}^{\ell} r_{ij} y_i)^{q-1}) \neq \vee_{i=1}^{\ell} y_i] \leq q^{-k}.$$

Now the $y_i$ depend on the input vector $x$ so for any fixed input over a random choice of the $k$ vectors $\vec{r}$, the expected fraction of errors is at most $q^{-k}$. Therefore averaging over all inputs and random vectors we get an error fraction at most $q^{-k}$. It follows that there is some choice of the random vectors that makes an error on at most a $q^{-k}$ fraction of inputs. Fix that random choce and define the approximator for that gate to be $1 - \prod_{j=1}^{k}(1 - \sum_{i=1}^{\ell} r_{ij} y_i)^{q-1})$. This increases the degree by at most a $k(q-1)$ factor. Putting this all together we have proved the following lemma:

**Lemma 10.9 (Approximation Lemma)** For any integer $k$ and any $\mathsf{AC}[q]$ circuit $C$ of size $S$ and depth $d$ there is a polynomial over $\mathbb{F}_q$ of degree at most $[(q-1)k]^d$ that agrees with $C$ on all but at most an $S/q^k$ fraction of input vectors.

For a circuit $C$ computing $Parity_n$ we choose $k = n^{1/(2d)}/(q-1)$ which implies that there is a degree $\sqrt{n}$ polynomial that agrees with $C$ on all but an $S/q^{n^{1/(2d)}/(q-1)}$ fraction of inputs. We obtain a lower bound on $S$ by the following lemma.

**Lemma 10.10** No polynomial of degree $\sqrt{n}$ over $\mathbb{F}_q$ agrees with $Parity_n$ on more than $\sum_{i \leq n/2+\sqrt{n}} \binom{n}{i} \leq \frac{49}{50} 2^n$ inputs in $\{0,1\}^n$.

**Proof** Let $P$ be a polynomial of degree $\sqrt{n}$. Let $G \subseteq \{0,1\}^n$ be the set of inputs $x$ on which $P(x) = Parity_n(x)$. We find it convenient to use $\{1,-1\}$ rather than $\{0,1\}$ in the representation of the inputs and outputs of our functions where the mapping $\phi$ from $\{0,1\}$ to $\{1,-1\}$ takes bit $b$ to $(-1)^b$. (This representation is also convenient for Fourier analysis of Boolean functions.) Note that

$\phi(Parity_n(x_1,\ldots,x_n)) = (-1)^{\sum_{i=1}^n x_i} = \prod_{i=1}^n \phi(x_i)$. In particular, setting $y_i = \phi(x_i) = (-1)^{x_i}$, note that

$$Parity'_n(y_1,\ldots,y_n) = \phi(Parity_n(\phi^{-1}(y_1),\ldots,\phi^{-1}(y_n))) = y_1 y_2 \cdots y_n,$$

computes a canonical monomial.

Now, despite the fact that it looks like exponentiation, the function $\phi$ is a degree 1 map over $\mathbb{F}_q$; for $x \in \{0,1\}$, we have $\phi(x) = 1-2x$. The same is true for $\phi^{-1}$, since $\phi^{-1}(y) = 2^{-1}(1-y)$ and 2 is invertible in $\mathbb{F}_q$. Let $G' \subseteq \{1,-1\}^n = \{(\phi(x_1),\ldots,\phi(x_n)) \mid (x_1,\ldots,x_n) \in G\}$. Since $\phi$ is 1-1, $|G'| = |G|$. Since $\phi$ and $\phi^{-1}$ are degree 1 maps we can compose them with the polynomial $P$ to produce a polynomial $P'$ in the $y_i$ that has degree $\sqrt{n}$ and equals $Parity'_n(y_1,\ldots,y_n) = y_1 y_2 \cdots y_n$ on all inputs in $G'$.

We use this strange ability to approximate a generic high degree function by the low degree polynomial $P'$ to derive the bound. Let $F_{G'} = \{f : G' \to \mathbb{F}_q\}$ so $|F_{G'}| = q^{|G|}$. Now by simple interpolation, any function defined on $G' \subseteq \{1,-1\}^n$ can be written as a polynomial in the $y_1,\ldots,y_n$ with coefficients in $\mathbb{F}_q$. Moreover, this polynomial's monomials can be assumed to be multilinear in the $y_i$ since $y_i^2 = 1$ for $y_i \in \{1,-1\}$. We use $P'$ to reduce the degree of any such polynomial. We can use the correctness of P' on $G'$ to express any monomial $\prod_{i\in T} y_i$ with $|T| > n/2$ as $y_1 y_2 \cdot y_n \cdot \prod_{i\notin T} y_i = P' \cdot \prod_{i \notin T} y_i$ on $G'$. This shows that any function in $F_{G'}$ can be expressed as a multilinear polynomial of degree at most $n/2 + \sqrt{n}$. The number of such polynomials is $q^{\sum_{i\le n/2+\sqrt{n}} \binom{n}{i}}$. Therefore $|G| \le \sum_{i \le n/2+\sqrt{n}} \binom{n}{i} \le (1-\gamma)2^n$ where $\gamma \ge 1/50$ is a fixed constant by standard properties of the binomial distribution. □

We now complete the proof of the theorem using the two lemmas and choice of $k = n^{1/(2d)}/(q-1)$. Combining the lemmas we have that that $(1-S/q^k)2^n \le \frac{49}{50}2^n$ and thus $S/q^k \ge 1/50$. Therefore $S \ge \frac{1}{50} q^{n^{1/(2d)}/(q-1)}$ as required. □

The above proof completely breaks down when considering moduli. In fact the following question is still open.

**Open Problem 10.11** Is $\mathsf{NP} \subseteq \mathsf{AC}^0[6]$?

The only lower bound we have for all of $\mathsf{ACC}^0$ applies in the uniform case and uses a clever downward translation and diagonalization.

**Theorem 10.12 (Allender-Gore)** $PERM \notin \mathsf{uniform}-\mathsf{ACC}^0$.

## 10.2 Threshold Circuits

Since $Majority$ is not in any $\mathsf{AC}^0[p]$ it is natural to ask what happens if one allows unbounded fan-in $Majority$ gates. One can make this more general still by allowing arbitrary *threshold gate* of the form

$$g(x_1,\ldots,x_\ell) = \begin{cases} 1 & \text{if } \sum_i w_i x_i \ge \theta \\ 0 & \text{otherwise.} \end{cases}$$

Such circuits with smoothed threshold behavior are sometimes called *neural nets.*

**Definition 10.13** Let $\mathsf{TC}^0$ be the set of a functions computed by constant=depth polynomial-size threshold circuits.

It is not hard to show that Iterated-Addition is in $\mathsf{TC}^0$. One can extend this to all the basic arithmetic functions, though the case for Division is fairly complicated. Clearly we also have that $\mathsf{ACC}^0 \subseteq \mathsf{TC}^0$.

It is also known that one can convert any polynomial-size threshold circiut into one that only uses majority gates (all constants $w_i$ are 1) by adding 1 to the depth. Current lower bounds bounds say little beyond depth 2 circuits.

The following theorem was first proved by Allender for $\mathsf{AC}^0$ using little more than the construction from the Razborov-Smolensky proof and then it was extended to $\mathsf{ACC}^0$ by Yao and Beigel-Tarui.

**Theorem 10.14** Any function in $\mathsf{ACC}^0$ can be expressed as a depth 3 $\mathsf{TC}$ circuit of size $2^{\log^{O(1)} n}$ and bottom fan-in $\log^{O(1)} n$. In particular it can be expressed as a symmetric function of $2^{\log^{O(1)} n}$ ANDs of fan-in $\log^{O(1)} n$ of variables and their negations.