6.S897 Algebra and Computation April 25, 2012

Lecture 20

Lecturer: Madhu Sudan *Scribe: Henry Yuen*

1 Overview

Today we talk about algebraic circuit lower bounds. Specifically we will show that computing the function $f(x_1, \ldots, x_n) = (x_1^r, \ldots, x_n^r)$ requires circuits of size $\Omega(n \log r)$. We will describe two proofs of this method: one is due to Strassen, which uses Bezout's theorem, and the other is by Smolensky, which uses elementary combinatorics. Between the two will be a brief interlude about Bezout's theorem.

2 Strassen's Proof

We will be working over algebraically closed fields $\mathbb{K}$. Let C be an algebraic circuit of size s computing the function $f(x_1, \ldots, x_n) = (x_1^r, \ldots, x_n^r)$. The inputs to this circuit are $x_1, \ldots, x_n$, and associate with each gate variables $y_1, \ldots, y_s$, and assume the variables $y_{s-n+1}, \ldots, y_s$ correspond to the output gates.

For each gate y_i we can associate a polynomial $P_i(\mathbf{x}, \mathbf{y})$ such that if gate i were, say, adding the outputs of gates j_1 and j_2, then we set $P_i(\mathbf{x}, \mathbf{y}) = y_i - (y_{j_1} + y_{j_2})$. Similarly, if gate i were multiplying inputs x_1 and x_2 together, then we set $P_i(\mathbf{x}, \mathbf{y}) = y_i - (x_1 \times x_2)$. On input $x_1, \ldots, x_n$, one can view the operation of the circuit as solving the system of equations $\{P_i(\mathbf{x}, \mathbf{y}) = 0\}$ for the unique setting of the variables $\{y_i\}$.

To obtain a lower bound on the size s of the circuit, we restrict our attention to instances when the circuit C will evaluate to the all 1's. That is, we will only consider inputs $x_1, \ldots, x_n$ such that $(x_1^r, \ldots, x_n^r) = (1, \ldots, 1)$. Denote this restricted set of inputs S. Since $\mathbb{K}$ is algebraically closed, we know that $|S| = r^n$. By our characterization above, this restriction is equivalent to restricting the output gates $y_{s_n+1}, \ldots, y_s$ to 1. Then, the P_i's become polynomials $\tilde{P}_i(x_1, \ldots, x_n, y_1, \ldots, y_{s-n})$.

Observe that for every $(x_1, \ldots, x_n) \in S$, there is still a unique setting of the $y_1, \ldots, y_{s-n}$ that satisfy the equations $\tilde{P}_i(\mathbf{x}, y_1, \ldots, y_{s-n}) = 0$. Furthermore, for $(x_1, \ldots, x_n) \notin S$, there is no solution of y_i's that will satisfy the equations. Thus the system of equations $\tilde{P}_i(\mathbf{x}, y_1, \ldots, y_{s-n}) = 0$ has exactly $|S| = r^n$ solutions.

Theorem 1 (Classical Bezout's Theorem) *Let $\mathbb{K}$ be an algebraically closed field. Let $f_1, \ldots, f_m \in \mathbb{K}[x_1, \ldots, x_n]$ be such that total degree of f_i is at most d_i. Then, the number of solutions to the equations $f_1(\mathbf{x}) = 0, \ldots, f_m(\mathbf{x}) = 0$ is either infinite or at most $\prod d_i$.*

We can put Bezout's theorem to use by noting that the total degree of P_i is at most 2, and thus $\deg(\tilde{P}_i) \leq 2$ as well. Since we have argued that the number of common zeroes to $\{\tilde{P}_i\}$ is finite, Bezout's theorem gives that there must be at most 2^s solutions. We know that there are exactly r^n solutions, so this gives that $s \geq n \log r$.

Note about the algebraic closure of $\mathbb{K}$: This lower bound applies even if the field $\mathbb{K}$ were not algebraically closed, but only if the circuit C computes the *formal polynomials* $(x_1^r, \ldots, x_n^r)$. This is because we can "lift" the circuit C to work in the closure of $\overline{\mathbb{K}}$, and since C computes the formal polynomials x_i^r, it will work correctly in the closure, and thus we can apply the Strassen's proof to show that C must have a superlinear size lower bound.

3 Bezout's Theorem

Bezout's theorem is an important result from algebraic geometry that, while apparently simple to state, does not admit a similarly simple proof. Here we will provide some additional remarks about the theorem.

How might one try to prove Bezout's theorem? A natural approach would be to perform induction on the number of polynomials. This could get devilishly complicated. Instead, we will rephrase the theorem in the more modern, more versatile language of algebraic geometry, and show how the classical version of Bezout's theorem (as stated above) can be proved from that.

This will require a few definitions.

Definition 2 (Variety) *Let k be a field. A set $V \in k^n$ is a variety if and only if there exists polynomials $f_1, \ldots, f_m \in k[x_1, \ldots, x_n]$ such that v is the common zero set of the polynomials.*

Definition 3 (Dimension of a variety) *Let k be a field, and let $V \in k^n$ be a variety. The dimension of V is*

$$\dim(V) = \operatorname*{argmin}_d \exists \text{ affine space } A \subseteq k^n \text{ of codimension } d \text{ such that } |A \cap V| \text{ is finite.}$$

Definition 4 (Degree of a variety) *Let k be a field, and let $V \in k^n$ be a variety. The degree of V is*

$$\deg(V) = \max_{\substack{\text{Affine space } A \\ \operatorname{codim}(A)=\dim(V) \\ |A\cap V|\leq\infty}} |A \cap V|$$

Armed with these definitions, we can state the "strong" version of Bezout's theorem, and show how it implies the "classical" theorem.

Theorem 5 (Strong Bezout's Theorem) *Let k be an algebraically closed field. Let V_1, V_2 be varieties over k. Then $\deg(V_1 \cap V_2) \leq \deg(V_1) \cdot \deg(V_2)$.*

Corollary 6 *Strong Bezout's theorem implies the classical version.*

Proof Let k be an algebraically closed field, and let $f_1, \ldots, f_m$ be polynomials in $k[x_1, \ldots, x_n]$. Let $V_i = V(f_i) = \{x \in k^n \mid f_i(x) = 0\}$. Let $V = V(f_1, \ldots, f_m) = V_1 \cap \cdots \cap V_m$. If $\dim(V) > 0$, then $|V|$ must be infinite, by definition. On the other hand, if $\dim(V) = 0$, then $|V|$ is finite (because the only codimension 0 affine space is k^n itself). Observe that in this case, $\deg(V) = |V|$, and by the strong Bezout's theorem, $\deg(V) \leq \prod \deg(V_i)$.

The dimension of each of the V_i is clearly $n - 1$. In general, $|V_i|$ will be infinite, so $\dim(V_i) < n$. Take an affine space A of codimension $n - 1$, i.e., a line. $A \cap V_i$ is the set of 0's of the restriction of f_i to the line A, which is a univariate polynomial. By the Fundamental Theorem of Algebra, this has at most $\deg(f_i)$ roots, so $|A \cap V_i|$ is finite and thus $\dim(V_i) = n - 1$. This also shows that $\deg(V_i) = \deg(f_i)$.

This concludes the proof; the number of common zeroes of $f_1, \ldots, f_m$ is either infinite or at most $\prod \deg(f_i)$. ■

There is another formulation of Bezout's theorem that says the number of *isolated* zeroes of a system of polynomial equations is bounded by the product of the degrees of the polynomials, even if the total number of solutions is infinite.

4 Smolensky's Proof

Smolensky's proof can be found in his short 4-page paper "Easy lower bound for a strange computational model". The lower bound is indeed easy, but the computational model is not *that* strange: instead of considering general algebraic circuits, we consider consider a restricted kind. The gates available for use are: 1) Addition gates with fan-in 2 and fan-out 1, 2) Multiplication gates with fan-in 2 and fan-out 1, and 3) *Duplicator* gates with fan-in 1 and fan-out 2, which will take $x \mapsto x, x$. Smolensky proves an $\Omega(n \log n)$ lower bound on the number of duplicator gates in circuits of this kind that compute $f(x_1, \ldots, x_n) = (x_1^n, \ldots, x_n^n)$ – and hence a lower bound on the circuit size.

This lower bound will then also apply to *general* circuits, because any general circuit C (where fan-out can be arbitrary) of size s can be converted to a restricted circuit C' with $O(s)$ duplicator gates. Suppose gate g in C connected to gates $h_1, \ldots, h_k$. If $k > 2$, we can introduce $k-2$ duplicator gates to reduce the fan-out of g to 2. Each duplicator gate becomes associated with an h_i for $i > 2$. Thus each gate in C has at most 2 duplicator gates associated with it. Thus, if C' must have at least s' duplicator gates, then C must have size $\Omega(s')$.

To obtain the lower bound, we will need the following lemma.

Lemma 7 *Let C be a circuit with*

- *Inputs $x_1, \ldots, x_n$,*
- *Outputs $y_1(x_1, \ldots, x_n), \ldots, y_m(x_1, \ldots, x_n)$, and*
- *Duplicators $D_1, \ldots, D_s$ that duplicate the polynomials $d_1(\mathbf{x}), \ldots, d_s(\mathbf{x})$, where $d_i(x_1, \ldots, x_n)$ is the unique polynomial that describes the input to duplicator gate D_i on input $x_1, \ldots, x_n$.*

Let $T \in \mathbb{F}[z_1, \ldots, z_m]$ with individual degrees less than k. Define $t(x_1, \ldots, x_n) := T(y_1(\mathbf{x}), \ldots, y_m(\mathbf{x}))$. Then there exists a polynomial τ in $n + s$ variables such that

$$t(x_1, \ldots, x_n) = \tau(x_1, \ldots, x_n, d_1^k(\mathbf{x}), \ldots, d_s^k(\mathbf{x}))$$

such that the individual degrees of the first n variables in τ is less than k and at most 1 in the remaining s variables.

We will apply this lemma to argue about circuits C that compute the outputs $y_1 = x_1^n, \ldots, y_n = x_n^n$, as well as the outputs $y_{n+1} = x_1, \ldots, y_{2n} = x_n$ (which we can compute with the help of n extra duplicators). Consider all polynomials $T \in \mathbb{F}[z_1, \ldots, z_{2n}]$ with individual degree less than n in each variable. Define $t(x_1, \ldots, x_n) := T(x_1^n, \ldots, x_n^n, x_1, \ldots, x_n)$.

Note that for every polynomial p in $x_1, \ldots, x_n$ with individual degrees less than n^2, there is a polynomial T (and a corresponding t) that will instantiate p. Thus the dimension of the space of polynomials T is at least $(n^2)^n$. On the other hand, the above lemma gives an upper bound on the dimension of the space of polynomials t. The dimension of the space of polynomials t is at most the dimension of the space of τ polynomials, which is $k^n 2^s$ (by counting the maximum number of distinct monomials of τ). Solving for s, we get that $s \geq \Omega(n \log n)$. Subtracting off the n extra duplicators that we added will not affect this bound asymptotically.

We now prove the lemma.

Proof Imagine that the circuit C is oriented so that the inputs are at the bottom, and the outputs are at the top. We will divide the circuit into levels so that at each level there is exactly one gate (an adder, multiplier, or duplicator). We will prove this by induction on level, starting at the top. At each stage of the induction, we imagine that there is a horizontal line that cuts the circuit into a top half and a bottom half. We will denote the wires that cross this line to be $z_1, \ldots, z_p$, and the duplicator gates above this line as $D_1, \ldots, D_q$, which duplicate the polynomials $d_1(z_1, \ldots, z_p), \ldots, d_q(z_1, \ldots, z_p)$. Intuitively, the top half is a circuit with inputs $z_1, \ldots, z_p$ and duplicators $D_1, \ldots, D_p$. Furthermore, each of the z_i's are polynomials in the inputs $x_1, \ldots, x_n$.

At the top level, the circuit is simply the empty circuits with inputs $z_1 = y_1, \ldots, z_m = y_m$ and outputs $y_1, \ldots, y_m$, and no duplicators. Thus $\tau(z_1, \ldots, z_m) = \tau(y_1, \ldots, y_m)$ is simply $T(y_1, \ldots, y_m)$. Since each of the y_i's have degree less than k in T, the same is true of τ. Thus we have established the base case.

Assume that we have represented the polynomial $t = \tau(z_1, \ldots, z_p, d_1^k, \ldots, d_q^k)$ at some level. We then move the line down, which will move past one gate. Suppose the gate were an adder or a multiplier. The output of this gate was some variable z_i; now, this variable will be replaced by two new variables (corresponding to the inputs to the gate). Call these two new variables z_{i_1} and z_{i_2}. Then, we can define $\tau'(z_1, \ldots, z_{i-1}, z_{i_1}, z_{i_2}, z_{i+1}, \ldots, z_p, d_1^k, \ldots, d_q^k) := \tau(z_1, \ldots, d_{i-1}, d_{i_1} + d_{i_2}, d_{i+1}, \ldots)$ (if the gate were an addition gate). It is easy to verify that τ' indeed satisfies the requirement that each z variable has degree less than k, and that $t = \tau'$.

Now suppose the gate were a duplicator D instead. Suppose the duplicated variables were $z_1 = z_2$. Then we can write $\tau'(z_1, z_3, \ldots, \delta_1, \ldots, \delta_q) = \tau(z_1, z_1, z_3, \ldots, \delta_1, \ldots, \delta_q)$. However, τ''s degree in z_1 might be higher than k – though it will be less than $2k$. Since we have added one more duplicator above the line, we can add another variable δ_{q+1} to create a polynomial $\tilde{\tau}(z_1, z_3, \ldots, \delta_1, \ldots, \delta_q, \delta_{q+1})$ that is τ' but with all factors of z_1^k replaced by δ_{q+1}. This replacement can happen at most once in each monomial of τ'. Since the duplicator D is duplicating the polynomial $d_{q+1} = z_1$, we have that

$$t = \tilde{\tau}(z_1, z_3, \ldots, d_1^k, \ldots, d_q^k, d_{q+1}^k)$$

and because of the replacement, $\tilde{\tau}$ has individual degree in each variable z_i to be less than k, and the degree of each d_j is at most 1.

By induction, at the the last level, the wires $z_1, \ldots, z_p$ are the inputs themselves $x_1, \ldots, x_n$, and we have shown that $t(x_1, \ldots, x_n) = \tau(x_1, \ldots, x_n, d_1^k, \ldots, d_s^k)$ for a polynomial τ with the right degrees. ■