E0 309: Topics in Complexity Theory **Spring 2015**

Lecture 15: Mar 23, 2015

Lecturer: Saravanan K and Chandan Saha *Scribe: Saravanan K*

15.1 Lower Bounds for Depth-3 homogeneous circuits

The theme of this lecture is to prove that any depth three homogeneous circuit computing the $2d^{th}$ elementary symmetric polynomial in n variables must have size at least $(\frac{n}{4d})^d$ over fields of characteristic zero.

Remark :
1) When $d = n/c$ (for some constant c), we obtain an $O(c^n)$ exponential lower bound.
2) It is clear that symmetric polynomial in n variables must have degree d less than n. However the result does not apply for polynomials of degree d much greater than n.

15.2 Partial Derivative Method

Let f be a polynomial. We define the partial derivative measure by

$$PD(f) = dim[span\{\partial f\}]$$

where ∂f is the set of all partial derivatives of f. In other words we write,

$$PD(f) = dim[span\{\partial_S f : S \subseteq [n]\}]$$

where $\partial_S f$ is the partial derivative $\dfrac{\partial f}{\partial x_1 \cdots x_k}$ such that $S = \{x_1, \cdots, x_k\}$.

15.2.1 Properties of Partial Derivatives

Any two polynomials f and g over field $\mathbb{F}$ holds the following properties.

1. Subadditivity : $PD(f + g) \leq PD(f) + PD(g)$.
2. Summultiplicativity : $PD(f.g) \leq PD(f).PD(g)$.
3. $PD(\alpha.f) = \alpha.PD(f)$, for any $\alpha \in \{\mathbb{F}\backslash 0\}$

15.3 Homogeneous Depth three circuits

Consider a homogeneous circuit C computing a homogeneous polynomial of degree d in n variables $\{x_1, x_2, \cdots, x_n\}$. Let C be

$$C = T_1 + T_2 + \cdots + T_s$$

such that $T_i = \prod_{j=1}^{d} l_{ij}$, where l_{ij} are linear forms in variables $x_1, x_2, \cdots, x_n$.

Lemma 15.1 $PD(C) \leq s \cdot 2^d$
Proof:

$$\begin{aligned}
PD(C) &\leq \sum_{i=1}^{s} PD(T_s) && \textit{(using subadditivity)} \\
&\leq \sum_{i=1}^{s} \prod_{j=1}^{d} PD(l_{ij}) && \textit{(using submultiplicativity)} \\
&\leq \sum_{i=1}^{s} 2^d \\
&\leq s.2^d
\end{aligned}$$

Notation : Let S_n^{2d} denote the $2d^{th}$ degree elementary symmetric polynomial in n variables.

Theorem 15.2 $PD(S_n^{2d}) \geq \binom{n}{d}$ *over fields of characteristic zero.*

Before proving the theorem, let us prove the required lower bound. From *theorem* 15.2 and *lemma* 15.1, we get,

$$\begin{aligned}
& \binom{n}{d} \leq s.2^{2d}, \quad \text{if } C \text{ computes the polynomial } S_n^{2d} \\
\implies & s \geq \frac{\binom{n}{d}}{2^{2d}} \\
& \geq (\frac{n}{4d})^d && \text{(using Stirling formula)}
\end{aligned}$$

Hence we have proved the required lower bound. Now the remainder of the proof is to prove *theorem* 15.2.

15.3.1 Proof of Theorem 15.2

Here we restrict our focus only to partial derivatives of order d. Let $T \subseteq \{x_1, \cdots, x_n\}$ and $|T| = d$. Now,

$$\partial_T S_n^{2d} = \sum_{W \subseteq [n] \ \& \ |W|=d \,\&\, W \cap T = \phi} \ \prod_{i \in W} x_i \tag{15.1}$$

Let us define a column vector $\mathbf{m}_{\binom{n}{d} \times 1}$, whose rows are indexed by subsets of $[n]$ of size d. We define the entries by $\partial_T S_n^{2d}$ for any row identified by T. That is

$$\mathbf{m}_{\binom{n}{d} \times 1} = [\partial_T S_n^{2d}] \tag{15.2}$$

From (15.1) and (15.2), we write

$$\mathbf{m} = D.\mathbf{v}$$

where, $D_{\binom{n}{d}\times\binom{n}{d}}$ is a 0/1 matrix whose rows and columns are identified by subsets of $[n]$ of size d such that $D_{T,W} = 1$ iff $T \cap W = \phi$. Here, $\mathbf{v}_{\binom{n}{d}\times 1}$ is a column vector whose rows are identified similar to columns of D such that $\mathbf{v}_W = \prod_{i \in W} x_i$.

From now on let us call D as the disjoint matrix.

Theorem 15.3 *The disjoint matrix D has maximal rank over any field of characteristic zero. That is, $rank(D) = \binom{n}{d}$.*

Since D has maximal rank, we can say that d^{th} order partial derivatives of S_n^{2d}, that is, the entries of $\mathbf{m}$ are all linearly independent. Therefore, we obtain the result $PD(S_n^{2d}) \geq \binom{n}{d}$. Now let us prove *theorem* 15.3.

15.4 Proof of Theorem 15.3

Let $S = \{1, 2, \cdots, n\}$. Let us call any subset of S of size l and k as l-set and k-set respectively. For any two positive integers l, k $(k \leq l)$ we construct a 0/1 incidence matrix B whose rows and columns are indexed by the set of all possible l-sets and k-sets respectively such that, $B_{ij} = 1$, iff the l-set corresponding to the i^{th} row contains the k-set corresponding to the j^{th} column. Clearly B has $\binom{n}{l}$ rows and $\binom{n}{k}$ columns. It has been proved by *Gottlieb* [2] that the rank of such matrix B is maximal. That is $rank(B) = \min\{\binom{n}{l}, \binom{n}{k}\}$.

The idea here is to show a reduction from disjoint matrix D to matrix B and hence claiming that $rank(D)$ is $\binom{n}{d}$. The reduction is as follows.

Suppose $d \leq n/2$. Here the i^{th} row is identified by the set T_i of size d. Also the j^{th} column is identified by a set with d elements (say W_j). We can also identify the same column j by the set $[n]\backslash W_j$. Clearly we observe that, T_i and W_j are disjoint if and only if T_i is contained in the set $[n]\backslash W_j$. Therefore, we claim that the disjoint matrix D is same as an incidence matrix B. Hence D has maximal rank. That is, $rank(D) = \binom{n}{d}$.

Similarly we can prove the case when $d > n/2$.

15.5 The incidence matrix B has maximal rank[2]

We recall the 0/1 matrix B whose rows and columns are indexed by set of all possible l-sets and k-sets respectively. The (i, j) element of matrix B takes value 1, only if the l-set corresponding to the i^{th} row contains the k-set corresponding to the j^{th} column.

Let us define lexicographical ordering for m-sets and n-sets. We represent any m-set by the vector $(a_1, a_2, \cdots, a_m)$, where $a_i < a_{i+1}$, for $i \in [m-1]$. Also we say $(a_1, a_2, \cdots, a_m) < (b_1, b_2, \cdots, b_m)$, if and only if $a_i < b_i$ for the smallest value of i such that $a_i \neq b_i$. Let us call this ordering the canonical ordering.

Now we define the canonical matrix $A_{l,k}^n$ obtained by ordering the rows and columns of B in their canonical order. Clearly $rank(A_{l,k}^n) = rank(B)$. Therefore, it is sufficient to prove that rank of $A_{l,k}^n$ is maximal.

Notations : We use the notation $R^n_{l,k}$, $C^n_{l,k}$ to represent the row null space and column null space of the matrix $A^n_{l,k}$ respectively.

Lemma 15.4 *It is easy to observe that the matrix $A^n_{l,k}$ has,*

1. $\binom{n}{l}$ *rows*
2. $\binom{n}{k}$ *columns*
3. $\binom{l}{k}$ *1's in each row*
4. $\binom{n-l}{l-k}$ *1's in each column*

Lemma 15.5 *We can also verify the following*

1. $A^1_{1,1} = A^1_{1,0} = A^1_{0,0} = A^0_{0,0} = [1]$
2. $A^n_{l,l} = I^n_l$, *where* I^n_l *is the* $\binom{n}{l} \times \binom{n}{l}$ *identity matrix.*

Lemma 15.6 *By definition of $A^n_{l,k}$, we represent $A^n_{l,k}$ by the following partition formula.*

$$A^n_{l,k} = \begin{bmatrix} A^{n-1}_{l-1,k-1} & A^{n-1}_{l-1,k} \\ O & A^{n-1}_{l,k} \end{bmatrix}_{\binom{n}{l} \times \binom{n}{k}}$$

Lemma 15.7 *For $n \geq l \geq p \geq k \geq 0$,*

$$A^n_{l,p} \cdot A^n_{p,k} = \binom{l-k}{p-k} A^n_{l,k}$$

Proof sketch : The above formula is proved by induction on n. Clearly the base case ($n = 1$) can be verified (using *lemma* 15.5). The induction step uses the partition formula (*lemma* 15.6) thus facilitating the multiplication of $A^n_{l,p}$ and $A^n_{p,k}$, which results in $\binom{l-k}{p-k} A^n_{l,k}$.

Theorem 15.8 $dim(R^{n-1}_{l-1,k}) + dim(R^{n-1}_{l,k-1}) = dim(R^n_{l,k})$.

Proof Sketch : Let us define a matrix $T = \begin{bmatrix} I_{l-1} & 0 \\ \dfrac{-1}{l-k} A^{n-1}_{l,l-1} & I_m \end{bmatrix}_{\binom{n}{l} \times \binom{n}{l}}$.

On premultiplying $A^n_{l,k}$ by T we get,

$$T.A^n_{l,k} = \begin{bmatrix} A^{n-1}_{l-1,k-1} & A^{n-1}_{l-1,k} \\ \dfrac{-(l-k+1)}{l-k} A^{n-1}_{l,k-1} & 0 \end{bmatrix}_{\binom{n}{l} \times \binom{n}{k}}$$

Consider a vector $\mathbf{v} = (\mathbf{x}, \mathbf{y})$ such that

$$\mathbf{v}.T.A^n_{l,k} = (\mathbf{x}.A^{n-1}_{l-1,k-1} - \frac{(l-k+1)}{(l-k)} \mathbf{y}.A^{n-1}_{l,k-1},\ \ \mathbf{x}.A^{n-1}_{l-1,k})$$

On solving $\mathbf{v}.T.A^n_{l,k} = 0$, we get $\mathbf{x}.A^{n-1}_{l-1,k} = 0$ and $\mathbf{y}.A^{n-1}_{l,k-1} = 0$. (uses *lemma* 15.7 and the fact that field $\mathbb{F}$ has characteristic zero).
Also since T is a non-singular matrix, we obtain that $\mathbf{v}$ is a direct sum of $\mathbf{x}$ and $\mathbf{y}$, implying

$$dim(R^{n-1}_{l-1,k}) + dim(R^{n-1}_{l,k-1}) = dim(R^n_{l,k}).$$

Theorem 15.9 $dim(C^{n-1}_{l-1,k}) + dim(C^{n-1}_{l,k-1}) = dim(C^n_{l,k})$.
(We can prove the theorem similar to theorem 15.8*).*

Corollary 15.10 .

1. $dim(R^n_{l,k}) = \begin{cases} 0 & \text{if } l+k > n, \\ \binom{n}{l} - \binom{n}{k} & \text{if } l+k \le n. \end{cases}$

2. $dim(C^n_{l,k}) = \begin{cases} \binom{n}{k} - \binom{n}{l} & \text{if } l+k > n, \\ 0 & \text{if } l+k \le n. \end{cases}$

Proof Sketch : We prove this by induction on n. We can easily verify the base case when $n = 1$. We analyze the induction by two cases. Suppose $l+k > n$. This implies $(l-1)+k > n-1$ and $l+(k-1) > n-1$. By induction hypothesis, we get $dim(R^{n-1}_{l-1,k}) = dim(R^{n-1}_{l,k-1}) = 0$. Using *theorem* 15.8 we get, $dim(R^n_{l,k}) = 0$. Also by rank-nullity theorem we obtain $dim(C^n_{l,k}) = \binom{n}{k} - \binom{n}{l}$. Therefore (by rank-nullity theorem) we get, the rank of the matrix is equal to the number of rows, that is $rank(A^n_{l,k}) = \binom{n}{l}$.

Similarly when $l + k \le n$, we get, $rank(A^n_{l,k}) = \binom{n}{k}$. Thus we conclude by stating that the rank of matrix $A^n_{l,k}$ is maximal.

15.6 Motivation for homogeneous circuits

Consider the polynomial

$$\begin{aligned} f(z, x_1, x_2, \cdots, x_n) &= (z + x_1) \cdot (z + x_2) \cdots (z + x_n) \\ &= c_n z^n + c_{n-1} z^{n-1} + \cdots + c_0 \end{aligned}$$

where the coefficient of z^{n-d}, c_{n-d} is exactly the elementary symmetric polynomial $S^d_n(x_1, x_2, \cdots, x_n)$. Let the evaluations of $f(z, x_1, x_2, \cdots, x_n)$ at $z = \alpha_1, \alpha_2, \cdots, \alpha_{n+1}$ be $g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_{n+1}(\mathbf{x})$, where $\alpha_1, \alpha_2, \cdots, \alpha_{n+1}$ are distinct field elements. That is, for all $i \in [n+1]$,

$$\begin{aligned} f(\alpha_i, x_1, x_2, \cdots, x_n) &= g_i(\mathbf{x}) \\ &= c_n \alpha_i^n + c_{n-1} \alpha_i^{n-1} + \cdots + c_0 \end{aligned}$$

In matrix notation we write,

$$\mathbf{g} = A\mathbf{c}$$

where, $\mathbf{g} = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_{n+1}(\mathbf{x}) \end{bmatrix}_{(n+1)\times 1}$, $A = \begin{bmatrix} \alpha_1^0 & \alpha_1^1 & \alpha_1^2 & \cdots & \alpha_1^n \\ \alpha_2^0 & \alpha_2^1 & \alpha_2^2 & \cdots & \alpha_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{n+1}^0 & \alpha_{n+1}^1 & \alpha_{n+1}^2 & \cdots & \alpha_{n+1}^n \end{bmatrix}_{(n+1)\times(n+1)}$, $\mathbf{c} = \cdot \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}_{(n+1)\times 1}$

Since A is a Vandermonde matrix, we know that inverse of A exists. Therefore the coefficient vector $\mathbf{c}$ can be computed by,

$$\mathbf{c} = A^{-1} \cdot \mathbf{g}$$

That is, for $0 \leq j \leq n$,

$$c_j = \sum_{k=1}^{n-1} \beta_k \cdot g_k(\mathbf{x})$$

where, $\beta_k \in \mathbb{F}$ is a field constant. Clearly c_j can be computed by a depth three circuit (not homogeneous) of polynomial size. That is, there exists a depth three circuit of polynomial size that computes the elementary symmetric polynomial. Thus we state the following corollary.

Corollary 15.11 *We cannot homogenize a depth-three circuit without a superpolynomial loss in size.*

References

[1] Noam Nisan and Avi Wigderson, Lower Bounds on Arithmetic Circuits via Partial Derivatives, *Computational Complexity*, 1997

[2] D.H. Gottlieb, A certain class of Incidence Matrices, *Proceedings of the American Mathematical Society, Volume 17, Issue 6*, Dec., 1966