Lower Bounds on Arithmetic Circuits via Partial Derivatives[*]

Noam Nisan [†] Avi Wigderson [‡]

Dedicated to the memory of Roman Smolensky

Abstract

In this paper we describe a new technique for obtaining lower bounds on restricted classes of nonmonotone arithmetic circuits. The heart of this technique is a complexity measure for multivariate polynomials, based on the linear span of their partial derivatives. We use the technique to obtain new lower bounds for computing symmetric polynomials (which hold over fields of characteristic zero) and iterated matrix products (which hold for every field).

1 Introduction

Despite much effort there are still essentially no lower bounds known for general models of computation such as boolean circuits. This sad state of affairs is essentially also true for arithmetic circuits, a natural model for computing arithmetic functions (see e.g. [3]). To date the best lower bounds known for arithmetic circuit size are only slightly super-linear ($\Omega(n \log n)$), and no non-trivial lower bounds are known for arithmetic circuit depth. This is even more humiliating than our lack of knowledge regarding boolean circuits: the arithmetic model is weaker and is not general; even more effort has been put into the arithmetic case; and more mathematical tools are available.

In this paper we are mostly interested in two fundamental problems. The first is computing the symmetric functions. Note that unlike the Boolean case, arithmetic (unbounded fanin) circuits can compute these functions in polynomial size and constant depth (in fact, depth 3 suffices!). The second is computing the product of d $n \times n$ real matrices, the arithmetic analog of graph reachability. Both problems trivially have polynomial size arithmetic circuits of (bounded fanin) depth $O(\log n \log d)$. This trivial depth upper bound is known to be tight for *monotone* arithmetic circuits [8] (see also [9]), and is believed to hold for the

[*] A preliminary version of this paper appeared in STOC 95

[†] Institute of Computer Science, Hebrew University of Jerusalem, Israel. This work was supported by USA-Israel BSF grant 92-00043 and by a Wolfeson research award administered by the Israeli Academy of Sciences. Part of this work was done while visiting BRICS at the University of Aarhus, Denmark

[‡] Institute of Computer Science, Hebrew University of Jerusalem, Israel. This work was supported by USA-Israel BSF grant 92-00106 and by a Wolfeson research award administered by the Israeli Academy of Sciences. Part of this work was done while visiting BRICS at the University of Aarhus, Denmark

general case. In this paper we describe a technique that gives depth lower bounds (and size-depth trade-offs) for computing these functions by various restricted classes of non-monotone circuits.

Our technique is based on measuring the dimension of the vector space spanned by all partial derivatives of the functions computed at the nodes of the circuit. This measure was used, for a different kind of circuit lower bounds, by Smolensky [6]. It is easily seen, and it is crucial for obtaining non-monotone lower bounds, that this measure is not monotone in the set of monomials of a polynomial (namely adding monomials may decrease the dimension). In some sense this measure captures the number of "useful" monomials generated so far by the circuit. The lower bound follows by showing that this measure does not grow too fast in small, shallow circuits, while for symmetric polynomials and iterated matrix product this measure is high. For some of our lower bounds the above argument does not suffice, and we need to use certain restriction arguments, as is common in boolean circuit lower bounds. Here a nice property of iterated matrix product plays a role: like the parity function, fixing some of the input matrices to be the identity matrix leaves us with a smaller iterated matrix product problem.

We obtain a range of lower bounds according to the severeness of the restriction we put on the circuits. In the following section we describe the kinds of restricted classes of circuits we consider; point out the best lower bounds known (to us) for each of these classes; state some simulations between them; and state the lower bounds we can prove using our technique. We also identify the (hopefully) easiest open problem regarding lower bounds for these models and make some conjectures regarding their relative power. In section 3 we illustrate the technique on a simple case: depth 3 homogeneous circuits computing symmetric functions. Section 4 formally defines all complexity measures we use, and section 5 contains proofs of all lower bounds on iterated matrix product. It is interesting to note that despite the use of partial derivatives, all the lower bounds in this paper save Theorem 3 (in Section 3) apply for arithmetic circuits over every field, not just those of characteristic zero!

2 Types of Circuits

2.1 General Arithmetic Circuits

In this paper we consider computing multivariate polynomials in $F[X]$ over a set of variables X and a base field F. Such a function is a sum of monomials, and we like to concentrate on the set of monomials of the function in our study of its complexity.

We use arithmetic circuits to compute these functions. Our main interest is the depth of the circuit needed to compute a function, where the other obvious parameter is the size. As usual, these circuits are direted acyclic graphs. The inputs (nodes of indegree zero) are labeled from the set of variables X. A constant from F can label an edge, which means the polynomial computed at its tail is multiplied by this constant. The internal nodes are labeled by addition or multiplication gates, computing the sum and product, resp, of the polynomials on the tails of incoming edges. (Subtraction is obtained using the constant -1.) The output is the polynomial computed by the circuit. We consider both bounded-fanin circuits, and unbounded fanin ones. Unless stated explicitely, depth refers to the bounded fanin model.

For a function f of degree d on N variables it is easy to prove an $\Omega(\log N + \log d)$ lower bound for depth for bounded fanin circuits. Moreover it is known [11] that if f can be computed by polynomial size circuits then an upper bound for depth is $O(\log n \log d)$. We ask whether a simliar lower bound can be proven.

Essentially no depth lower bounds are known for general circuits. The following open problem, pointed out by Ben-Or, shows the limits of our knowledge:
Open Problem: Find an explicit function which cannot be computed by polynomial size depth 3 circuits with unbounded fanin.

2.2 Homogeneous Functions and Circuits

Definition 1 *We say that a multivariate polynomial is homogeneous if all of its monomials are of the same degree. We say that a circuit is homogeneous if all its nodes compute homogeneous functions.*

It is implicit in [7] (see [10]) that circuits for homogeneous functions can be made homogeneous with only a polynomial cost in size. However, it turns out that in this construction the depth grows by $O(\log d)$.

Lemma 1 *(Implicit in [7]) If a homogeneous function f of degree d can be computed by a circuit of size s and depth h then it can be computed by a homogeneous circuit of size $O(sd^2)$ and depth $O(h \log d)$. For the unbounded fanin model no loss in depth is incurred.*

We conjecture that the depth increase by $O(\log d)$ is necessary. Natural candidates are the symmetric functions. Ben-Or [2] has shown that general circuits of unbounded fanin and depth 3 can compute all symmetric polynomials in polynomial size (note the contrast to the exponential lower bounds for Majority in the Boolean model and finite fileds).
Conjecture: Homogeneous circuits require $\Omega(\log n \log d)$ depth to compute the d'th elementary symmetric polynomial on n variables.

Only very weak lower bounds are known even for homogeneous circuits. In [5] exponential lower bounds are proven for computing the determinant by homogeneous circuits of unbounded fanin and depth 3. We can obtain, using our techniques, exponential lower bounds for the easier problems of computing the elementary symmetric functions and multiplying d $n \times n$ matrices (for which the techniques of [5] do not apply).
Theorem 0: Any homogeneous depth 3 circuit computing the $2d$th symmetric polynomial on n variables over a field of characteristic zero requires $\Omega((n/4d)^d)$ size.
Theorem 1: Any homogeneous depth 3 circuit for multiplying d $n \times n$ matrices requires $\Omega(n^{d-1}/d!)$ size.
Open Problem: Find an explicit homogeneous function which cannot be computed by polynomial size constant depth (or even depth 5) homogeneous circuits.

2.3 Multilinear Functions and Circuits

In many cases the input variables to the function are naturally partitioned into sets $X^1, \cdots, X^d$. An example is the function we are interested in – the product of d $n \times n$ matrices. In this case X^i is all the n^2 variables of the i'th input matrix. Other examples are determinant and permanent – in which X^i is all the variables in the i'th row of the input matrix. Below we define our notion of multilinearity, and caution that it is slightly more restrictive than the standard one.

Definition 2 *Fix a partition $X^1, \cdots, X^d$ of the input variables. For a subset $T \subseteq [d]$ we say that that a T-multilinear monomial is a product of variables (of degree 1 each), exaclty one from each part $\{X^i : i \in T\}$. A function is T-multilinear if it is a linear combination of T-multilinear monomials. A function is multilinear if it is T-multilinear for some subset T. A circuit is multilinear if each of its nodes computes a multilinear function in some subset of the X_i's.*

Besides the fact that such circuits are natural, observe that monotone circuits that compute a multilinear function are necessarily multilinear.

Definition 3 *An arithmetic circuit over the reals (or rationals) is called monotone if all constants labeling its edges are positive.*

Fact 1 *Every monotone arithmetic circuit for a multilinear function is multilinear.*

General circuits can be made multilinear for the following price.

Lemma 2 *If a multilinear function f of degree d can be computed by a circuit of size s and depth h then it can be computed by a multilinear circuit of size $s2^{O(d)}$ and depth $O(h+d \log d)$.*

Proof: Note that every function is a sum of T-multilinear functions, one for every subset T. Each node of the original circuit is split into 2^d nodes consisting of the different multilinear parts (one for each T) computed by the node. Each gate of the original circuit is then simulated on each of the parts. An addition gate is simulated as follows: for every subset T, the T-part of the output is the sum of the two T-parts of the input. This requires depth 1.

A multiplication gate is simulated as follows. For every T, the T part of the output is the sum, over all partitions of T into disjoint subsets R, S, of the product of the R-part of one input and the S-part of the other input.

A simulation depth of d per gate, resulting in total depth $O(hd)$ is trivial. A slighty better result, as stated, requires careful addition of the different terms in the simulation of a multiplication gate.

To obtain a part of degree d' (corresponding to T) one needs to add, for each $0 \leq i \leq d'$ $\binom{d'}{i}$ results of multiplications of a degree i polynomial (corresponding to each R of size i), by a degree $d'-i$ polynomial (corresponding to S of size $d'-i$). For each i, these products are added in a balanced binary tree of depth $O(\log \binom{d'}{i})$. The sums for all $0 < i < d'$ are then added in a balanced binary tree of depth $O(\log d')$.

Denote by $f(h', d')$ the worst case depth that this simulation gives for a gate in the original circuit which is of depth h' and degree d'. The recurrence relation we get for an addition gate is $f(h', d') \leq 1 + f(h'-1, d')$ and for a multiplication gate is $f(h', d') \leq \max_{0 \leq i \leq d'}[f(h'-1, i) + O(\log \binom{d'}{i})]$, which is solved by $f(h', d') = O(h' + d' \log d')$. ♣

While it does not seem that the restriction of being multilinear hampers circuits in computing matrix products, we do think that it is a severe restriction and that the previous lemma is close to optimal. A possible candidate for exhibiting a gap is the determinant function, which can be computed by homogeneous circuits of $O(\log^2 n)$ depth.

Conjecture: Multilinear circuits require $\Omega(n)$ depth to compute the determinant of an n by n matrix.

The best lower bounds we can prove for multilinear circuits are:

Theorem 2: Any depth h unbounded fanin multilinear circuit computing the product of d 2×2 matrices requires size $exp(d^{1/h})$. Consequently, any polynomial size circuit for this problem requires depth $\Omega(\log n / \log\log n)$.

Open Problem: Find an explicit multilinear function which cannot be computed by logarithmic depth (bounded fanin) multilinear circuits.

We are able to prove the "correct" lower bound for the (odd) special case where all multiplication gates have odd fanin. Note that such circuits can compute any odd degree polynomial.

Theorem 3: Any multilinear circuit for multiplying d $n \times n$ matrices, where all multiplication gates have odd fanin requires depth $\Omega(\log n \log d)$.

2.4 Pure circuits

In a multilinear circuit we associate with each node a subset $S \subseteq \{1..d\}$ from which its monomial takes variables. The natural circuits we can think of for iterated matrix multiplication build up these sets in "consistent" way, all across the circuit. This is formalized in

Definition 4 *A Multilinear circuit is called pure if for every two sets S, T associated with nodes in the circuit $S \cap T \in \{\emptyset, S, T\}$.*

While pure circuits can certainly compute all multilinear functions, we do not know of any general simulation of general circuits by pure circuits. Our techniques can give:

Theorem 4: Any pure circuit for computing the product of d $n \times n$ matrices requires depth $\Omega(\log n \log d)$.

3 Theorem 0 - A Motivating Example

In this section we illustrate the technique, proving the exponential lower bound on the size of depth 3 circuits computing elementary symmetric polynomials. We need some notation first, which is refined in the next section for the other lower bounds.

Let F be a field of characteristic zero. We will consider polynomials in n variables X. For any set of polynomials $V \subseteq F[X]$ we use $dim(V)$ for the dimension of the linear span of V (in other words the maximum number of linearly independent polynomials over F in V).

Let f be a polynomial. We let $\partial(f)$ denote the set of **all** partial derivatives (of all orders) of f. Note for example that a single monomial - product of k variables, say - will have 2^k different partial derivatives. The linearity, sum and product formulae for partial derivatives upper bound (respectively) the ability of the different circuit operations to increase the dimension of this set.

Proposition 1 *For every $f_1, f_2, \cdots, f_r \in F[X]$ and $\alpha \in F$, $\alpha \neq 0$ we have:*

- $dim(\partial(\alpha f_1)) = dim(\partial(f_1))$.
- $dim(\partial(\sum_i f_i)) \leq \sum_i dim(\partial(f_i))$.
- $dim(\partial(\Pi_i f_i)) \leq \Pi_i dim(\partial(f_i))$.

This proposition easily bounds the dimension of the output of depth 3 circuits. We assume (wlog, otherwise the results are trivial) that these circuits are leveled, with plus gates at the top and bottom levels and multiplication gates in the middle level.

Lemma 3 *Let f be computed by a depth 3 circuit with fanin s to the top (plus) gate, and fanin d or less at every multiplication gate. Then $dim(\partial(f)) \leq s2^d$.*

Proof: Observe that every linear function g (in particular, those computed at the first level of the circuit) satisfy $\partial(g) = \{1, g\}$ and thus $dim(\partial(g)) \leq 2$. The rest follows by Proposition 1. ♣

Proof of Theorem 0: The conclusion of theorem 0 follows easily from Lemma 3 above, the fact that homogeneous circuits computing a degree d polynomial cannot have multiplication fanin exceeding d, and a lower bound on the dimension of the partials of symmetric functions below. □

Lemma 4 *Let SYM_n^d denote the dth elementary symmetric polynomial. Then*

$$dim(\partial(SYM_n^{2d})) \geq \binom{n}{d}$$

Proof: For a subset $R \subseteq [n]$ we let SYM_R^d be the d'th symmetric polynomial over the variables in R. Thus $SYM_n^d = SYM_{[n]}^d$. In the sequel let S and T range over the set I of all d-subsets of $[n]$. Consider the following two sets of polynomials U, V, indexed by the set I, which we shall order as vectors by fixing an order on the elements of I. The vector U is simply all monomials of length d, i.e. $U_S = \Pi_{i \in S} x_i$. The vector V contains the partial derivatives of $SYM_{[n]}^{2d}$ with respect to d-monomials, which are easily calculated to be $V_T = SYM_{[n]/T}^d$ (this function can be readily understood as the restriction of $SYM_{[n]}^{2d}$ obtained by assigning zeros to all variables in T.) It clearly suffices to lower bound $dim(V)$ for the lemma. But note that $V = DU$, where D is the $I \times I$ disjointness matrix $D_{T,S} = 1$ if $S \cap T = \emptyset$ and 0 otherwise. Since D has full rank over the rationals [4], and all monomials in U are independent, we have

$$dim(\partial(SYM_n^{2d})) \geq dim(V) = dim(U) = \binom{n}{d}$$

♣

4 Complexity Measures

4.1 Polynomials

From this section on, the field F is arbitrary, not necessarily of characteristic 0.

For an integer d we use $[d] = \{1, 2, \cdots, d\}$. Let $X^1, X^2, \cdots, X^d$ be sets of variables of size n^2 each, with $X = \cup_{i \leq d} X^i$. The variables in X^i will be denoted x^i_j. Now our polynomials are in the ring $F[X] = F[X_1, \cdots, X_d]$.

For $T \subseteq [d]$ we use P_T to denote the set of all polynomials in which every monomial is of the form $\Pi_{i \in T} x^i_{j_i}$ (i.e. every monomial is a product of variables, one from each of the variable sets in T). Such polynomials are called multilinear (see Definition 2). Observe that P_T is a vector space over F, and that $dim(P_T) = n^{2|T|}$.

Before using partial derivatives again, we need to define them, as we are not restricting the characteristic of the field anymore. We do it only for multilinear polynomials, as it simplifies matters.

Definition 5 *Let $p = f + yg \in F[X]$, where $y \in X$ is a variable and $f, g \in F[X \setminus \{y\}]$. Then define the partial derivative of p with respect to y by $\frac{\partial}{\partial y} p = g$. Further let $Y = \{y_1, \cdots, y_k\} \subseteq X$, and $p \in F[X]$ a multilinear polynomial. Then define the partial derivative of p with respect to the variables in Y by*

$$\frac{\partial}{\partial Y} p = \frac{\partial}{\partial y_1} \cdots \frac{\partial}{\partial y_k} p$$

It is easy to see that for multilinear polynomials the standard facts regarding derivatives hold over every field.

Fact 2 *For every subset Y of variables we have*

- *$\frac{\partial}{\partial Y}$ is well defined (independent of the order of elements in Y).*
- *$\frac{\partial}{\partial Y}$ is an F-linear function on the multilinear polynomials in $F[X]$.*
- *If f, g and fg are multilinear, then $\frac{\partial}{\partial Y} fg = (\frac{\partial}{\partial Y} f)g + (\frac{\partial}{\partial Y} g)f$.*

Fix T and let f be a polynomial in P_T. For any $S \subseteq T$ we let $\partial_S(f)$ denote the set of partial derivatives of f with respect to all monomials of P_S. (I.e. for every set Y of variables, that forms a monomial of P_S, e.g. $Y = \{x^i_{j_i} \,|\, i \in S\}$, we put $\frac{\partial}{\partial Y} f$ in $\partial_S(f)$.) We clearly have $\partial_S(f) \subset P_{T-S}$. A trivial fact of key importance is:

Proposition 2 *If $f \in P_T$ then $dim(\partial_S(f)) \leq min\,\{n^{2|S|}, n^{2|T-S|}\} \leq n^{2\lfloor |T|/2 \rfloor}$.*

Let $dim(f) = max_{S \subseteq T}\, dim(\partial_S(f))$. Note that $dim(x^i_j) = 1$ for every variable. The proposition below essentially shows that dim is a formal complexity measure for certain arithmetic formulae.

Proposition 3 *For every $T, R \subset [d]$ with $T \cap R = \emptyset$, $f, g \in P_T$, $h \in P_R$, and $\alpha \in F$ we have:*

- *$dim(\alpha f) = dim(f)$*
- *$dim(f + g) \leq dim(f) + dim(g)$*
- *$dim(fh) \leq dim(f)dim(h)$.*

Next we define another complexity measure, ρ, (inspired by the saturation measure in [9]). It is important to note already here that unlike typical complexity measures, ρ will both increase and decrease along the computation of the circuit. First, let $\langle a \rangle$ denote the largest even integer not exceeding a (i.e. $(\langle a \rangle = 2\lfloor a/2 \rfloor)$. For any $f \in P_T$ denote $\rho(f) = dim(f)/n^{\langle |T| \rangle}$. From Proposition 2 we have

Proposition 4 *For every T and $f \in P_T$, $\rho(f) \leq 1$.*

Also, from Propositions 2,3, ρ enjoys the following subadditivity relations.

Proposition 5
- *For every $f \in P_T$ and $\alpha \in F$, $\rho(\alpha f) = \rho(f)$.*
- *For every $f_1, f_2, \cdots, f_r \in P_T$, $\rho(\sum_i f_i) \leq \sum_i \rho(f_i)$.*
- *Let $T_1, T_2, \cdots, T_r$ be pairwise disjont subsets, and $f_i \in P_{T_i}$. If s of the r subsets are of odd size, then*
$$\rho(\Pi_i f_i) \leq n^{-\langle s \rangle} \Pi_i \rho(f_i) \leq n^{-\langle s \rangle} min_i \rho(f_i)$$
.

The multilinear function we shall be most interested in is the product of d $n \times n$ matrices. To consider a single valued function we will concentrate on the $(1,1)$ entry of the product and denote it by IMM_d^n (note that this function depends only on the first "row" of variables in X^1 and the first "column" of variables in X^d matter). It is very rich in partial derivatives:

Proposition 6 *For every n, d, $dim(IMM_d^n) = n^{d-1}$ and thus for odd d $\rho(IMM_d^n) = 1$ and for even d $\rho(IMM_d^n) = 1/n$.*

Proof: Consider $\partial_S(IMM_d^n)$ for $S = \{2i : i \leq d/2\}$. If d is odd then all these n^{d-1} partial derivatives are distinct. When d is even we again get n^{d-1} distinct partial derivatives when considering only those that take from X^d veriables from the first "column". ♣

4.2 Circuits

In the following we derive consequences of the previous subsection to gates of multilinear circuits.

Every gate in a multilinear circuit computes a multilinear polynomial in P_T (of degree $|T|$) for some $T \subseteq [d]$. We will associate the gate with the polynomial it computes. The children (inputs) of a plus gate are in the same P_T, while the children of a times gate define a partition of T. From Proposition 5 we deduce two useful lemmas.

Lemma 5 *If a (plus or times) gate f in a multilinear circuit has inputs $g_1, \cdots, g_m$, then $\rho(f) \leq \sum_{j=1}^{m} \rho(g_j)$.*

Lemma 6 *In any multiplication gate g with m odd degree inputs, $\rho(g) \leq n^{-<m>}$.*

We conclude with our notation for size and depth of the three circuit models we work with. We denote by $S_h^*(f)$ the size (number of gates) of the smallest unbounded fanin circuit for f, whose top gate is a plus, and whose depth (or height) is h. The symbol $*$ takes values from $\{H, M, P\}$ according to whether the circuit is Homogeneous, Multilinear or Pure. Similarly, $D^*(f)$ is the depth of the shallowest bounded depth circuit of each type.

5 Size Lower Bounds for Constant-depth Unbounded-fanin Circuits

5.1 Depth-3 Homogeneous Circuits

We first state again Theorem 0 in our notation.
Theorem 0: $S_3^H(SYM_n^{2d}) = \Omega((n/4d)^d)$

For Theorem 1 we observe that depth 3 homogeneous circuits can be turned into multilinear ones at a reasonable price. We assume (wlog, otherwise the results are trivial) that they are leveled, with plus gates at the top and bottom levels and multiplication gates in the middle level.

Lemma 7 *For every multilinear function f of degree d, $S_3^M(f) \leq d! S_3^H(f)$*

Proof:

Replace each addition gate in the bottom layer by d gates, one for each part of the variables' partition (i.e. each new gate compute the partial sum of variables from one of the d input parts). Now replace each multiplication gate by $d!$ gates, each one multiplying one part from each of the d partitions. Finally add up all these parts in the top addition gate. ♣

Theorem 1 *For all n and d, $S_3^M(IMM_d^n) \geq n^{d-1}$ and also $S_3^H(IMM_d^n) \geq n^{d-1}/d!$*

Proof: By the above simulation lemma it suffices to show only the first lower bound. Each addition gate g on the bottom level of such a circuit computes a function of degree 1. Thus, by lemma 6 and since 1 is an odd number, each multiplication gate $h = g_1 g_2 ... g_d$ has $\rho(h) \leq n^{-<d>}$. So, by lemma 5 we must add at least $\rho(IMM_d^n)/n^{-<d>}$ such h's in order to compute IMM_d^n. The lemma follows by proposition 6. ♣

5.2 Multilinear Circuits

The lower bound of the previous section works because all multiplication gates at the lowest level have odd degree (in fact, degree 1) inputs. This is not the case for multiplication gates "higher" up in the circuit. We will essentially "force " this situation via random restrictions.

Let $f \in P_{[d]}$ be a function computed by a circuit C, and let $R \subset [d]$. If we assign constant values to all variables in all $X^i, i \notin R$, we obtain a reduced circuit denoted $C|_R$ computing the polynomial denoted $f|_R$ in P_R. For a random subset $\mathbf{R} \in [d]$ we similarly define the random variables $C|_{\mathbf{R}}$ and $f|_{\mathbf{R}}$.

We will need the following technical lemmas.

Lemma 8 *Let $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_d$ be independent, unbiased $0, 1$ random variables. For $T \subseteq [d]$ let $\mathbf{z}(T) = \sum_{i \in T} \mathbf{z}_i (mod\, 2)$.*

- *$Pr[\sum_i \mathbf{z}_i < d/3] \leq 1/10$ (for $d \geq 20$.)*
- *For every family of pairwise disjoint subsets $T_1, T_2, \cdots, T_r$ of $[d]$, $Pr[\sum_j \mathbf{z}(T_j) < r/3] \leq 2^{-r/10}$.*

Proof: Follows directly from the Chernoff bound (eg. see [1]). ♣

Lemma 9 *Let C be an optimal multilinear circuit of (multiplication) depth h computing a polynomial $f \in P_{[d]}$. Then there are multiplication gates $g_1, g_2, \cdots, g_s$ in the circuit with the following properties.*

- *$s \leq S_h^M(f)$*
- *For every $j \in [s]$, the fanin of g_i is at least $d^{1/h}$.*
- *$\sum_{j \in [s]} \rho(g_j) \geq \rho(f)$*

Proof: In every possible path in C from the output to an input, take the first (closest to output) multiplication gate of fanin at least $d^{1/h}$ (clearly there is always such a gate). Let $g_1, \cdots, g_s$ be the set of these gates. It clearly satisfies the first two properties. Also, since we took all possible paths, the output is a function of these gates, and by the subadditivity Lemma 5 we have the third property. ♣

Lemma 10 *Fix integers d, h and let $r = d^{1/h} \geq 20$. Let $\mathbf{z} = (\mathbf{z}_1, \cdots, \mathbf{z}_d)$ be a sequence of independent unbiased random variables, and $\mathbf{Z} \subseteq [d]$ the set of 1 positions in $\mathbf{z}$. Let $f \in P_{[d]}$ satisfy $S_h^M(f) \leq (1/2)2^{r/10}$. Then $Pr[\rho(f|_{\mathbf{Z}}) \geq 1/n] \leq 1/2$.*

Proof: Let C be an optimal depth h multilinear circuit for f, and $g_1, \cdots, g_s$ the gates guaranteed by the previous lemma. Recall that $deg(g_j) \geq r$ and $s \leq (1/2)2^{r/10}$. Combining Lemma 8 we deduce that with probability at least $1/2$, all g_j's have at least $r/3$ odd degree inputs after the restriction $\mathbf{Z}$ and using Lemma 6 they all satisfy $\rho(g_j|_{\mathbf{Z}}) \leq n^{-r/3}$. Again from the bound on s and Lemma 5, with probability at least $1/2$ we have $\rho(f|_{\mathbf{Z}}) \leq 2^{r/10} n^{-r/3} < 1/n$ (where we used $n \geq 2$, and $r \geq 20$). ♣

We are now ready for the main theorem of this section. It is interesting to note that the IMM_d^2, the product of d 2×2 matrices, is computable in arithmetic NC^1. An important observation is that this function behaves like the parity function with respect to restrictions: it remains the same function on fewer variables.

Theorem 2 *For all d, h we have $S_h^M(IMM_d^2) = 2^{\Omega(d^{1/h})}$*

Proof: The upper bound on the circuit size is trivial. To prove the lower bound, let us use a random restriction defined by a random vector $\mathbf{z} \in \{0,1\}^d$ as above, where we set each matrix defined by a block of variables outside $\mathbf{Z}$ to the identity matrix. This restriction leaves iterated matrix multiplication the same function on fewer matrices. Precisely, for every value Z of $\mathbf{Z}$ with $|Z| = t$ we have $IMM_d^2|_Z = IMM_t^2$ whose ρ value for every t is at least $1/n$. The rest follows directly from Lemma 10. ♣

6 Depth Lower Bounds for Bounded Fanin Circuits

6.1 The Odd Case

To demonstrate the use of our techniques for proving depth lower bounds, we consider a restricted version of such circuits, which is odd in several ways. First, the circuits have only multiplication gates of *odd* fan-in. Let $D^O(f)$ denote the depth of the smallest such circuit computing f.

Clearly, odd circuits can compute any multilinear polynomial of odd degree, and it seems like this restriction cannot be too severe for such polynomials. Odd as it sounds, simulating regular circuits by these restricted ones is very costly, as our bounds will show. Observe that in the restricted circuit, every gate computes an odd degree polynomial. We thus get for free what we used random restrictions for in the previous section. We make use of that in the main technical lemma below. Note the similarity to the argument used for monotone lower bounds by Tiwari and Tompa [9].

Lemma 11 *Let d be an odd integer, and f be a multilinear polynomial (of degree d) over $X^1, \cdots, X^d$. Then $D^O(f) = \Omega(\log n \log d + \log \rho(f))$*

Proof: Assume we are given an odd circuit C for f. Assume wlog that the fan-in of every gate is 3 (since here we deal with bounded fanin, this costs only a constant factor in the depth). Choose a path from the output to some input inductively as follows.

- The output node is in the path.
- At a times gate, take the child who computes the polynomial with the highest degree.
- At a plus gate, take the child who computes the polynomial with the highest value of ρ.

Consider the values of ρ of the polynomials along the path. They satisfy (using the properties above and Proposition 5):

- The value at the output is $\rho(f)$.
- The value at (the last) input node is $\rho(x_j^i) = 1$.
- At any times gate, the value increases by a factor of at least n^2.
- At any plus gate, the value decreases by at most a factor of 3.

- There are at least $\log_3 d$ times gates along the path.

An immediate consequence of the above is that there must be at least $2\log_3 n \log_3 d$ plus gates along the path, giving the required bound. ♣

We illustrate the lower bound on two clean polynomials. The first is IMM_d^n, iterated matrix multiplication. For this function we get the tight bound (which we expect to hold without the "odd" restriction).

Theorem 3 *For every n and odd d, $D^O(IMM_d^n) = \Theta(\log n \log d)$*

Proof: The upper bound is the trivial one. The lower bound follows from the Lemma 11 and Proposition 6. ♣

The second is PIP^n, the product of inner products. Here we think of (the $2n+1$) input sets as representing vectors, and $PIP^n(X^1, \cdots, X^{2n+1}) = x_1^1 \Pi_{i=1}^n \sum_{j=1}^{n^2} x_j^{2i} x_j^{2i+1}$. This function displays the gap of power between general multilinear and odd circuits.

Gap Theorem:

- $D^M(PIP^n) = \Theta(\log n)$
- $D^O(PIP^n) = \Theta((\log n)^2)$

Proof: The only nontrivial part is the second lower bound, which again follows from Lemma 11 as $\rho(PIP^n) = 1$ (take partials w.r.t. the even numbered blocks of variables). ♣

6.2 Pure Circuits

In trying to handle general multilinear circuits, a lesson from the previous subsections, which easily follows from Lemma 5, is the following useful lemma. Before stating it, we need a definition.

Definition 6 *Let C be a multilinear circuit of fanin 2. A path-transversal of C is the set of all output-input paths in C after removing from every multiplication gate one input wire. Clearly, there are many path-transversals in every circuit.*

Lemma 12 *If every path in a path transversal of a multilinear circuit C for f contains at least t multiplication gates, each with both inputs in C having odd degree, then C must have depth $\log n \log t + \log \rho(f)$.*

Clearly, without the artificial restriction of odd fanin multiplication gates, (indeed assume from now on that all fanins are 2), it is not clear that t will be in general larger than 1 (coming from the bottom level). While it may seem at first sight, that restrictions may force a larger t, a simple example shows that this is not the case.

Proposition 7 *There is a multilinear function f with a multilinear circuit of depth $O(\log n + \log d)$, such that for every restriction $R \subseteq [d]$ and any path-transversal in $C|_R$, there is a path in which only the bottom multiplication gate has two odd (=1) degree inputs.*

Proof: (Sketch) We illustrate this example in the case $n = 1$. Take $f = c\Pi_1^d x^i$ for some constant c. The circuit (in fact, a formula) for this trivial polynomial will not be so trivial – it is constructed recursively from the output as follows. Each node computes f_S, the product of the elements in some subset $S \subseteq [d]$. Start from $S = [d]$. For any S at hand, Partition S into (nearly) equal subsets A, B, C. The the formula for f_S is

$$f_S = f_A \times f_{B \cup C} + f_B \times f_{A \cup C} + f_C \times f_{A \cup B}$$

It is easy to verify that this formula computes f. The property of resiliancy to restriction follows from the simple fact that given any three integers (the sizes of A, B, C), the sum of at least one pair is even. ♣

This stumbling block can be overcome for pure circuits. A nice way to view these circuits (recall the definition from section 2) is that all their subcircuits obey the same recursive partitioning of the input sets $[d]$. More precisely, to every pure circuit C corresponds a binary tree $T(C)$ with d leaves labeled by the elements of $[d]$, and every internal node is labeled by the set of leaf labels in its subtree. Every multiplication gate in C computes a polynomial $g \in P_S$ only for a label S of some node v in $T(C)$. Moreover it does so by multiplying two polynomials from P_A, P_B, with A, B being the labels at the children of v in $T(C)$. Thus $T(C)$ is a "skeleton" of C, as any node v in it with label S represents the sum of many gates in C which multiply polynomials from P_A and P_B.

Theorem 4 $D^P(IMM_d^n) = \Theta(\log n \log d)$

Proof: (Sketch) Again, the upper bound is trivial, as the natural algorithm is pure. For the lower bound, we again use a restriction argument to force the situation of Lemma 12 with $t = (\log d)/2$. The idea is simple: in a pure circuit C every root-leaf path in $T(C)$ induces a highly regular path-transversal in C, arising naturally from the correspondence above. Moreover, the degree of a polynomial computed at a gate of C is the cardinality of the associated label in $T(C)$, so we can easily check its parity. Finally, a restriction $R \subseteq [d]$ corresponds to replacing the leaf labels from R in $T(C)$ by empty sets.

Now construct a restriction R as inductively down from the root of $T(C)$ as follows. Start at the root r, with the set R initially empty. At a node u with children v, w labeled V, W respectively, with $|V| \geq |W|$, add all but one of the elements of W in R, and move to node v. It is easy to see that on the path followed by this procedure (and the restriction R), every second node is labeled by an odd subset, and moreover this path has length at least $\log d$. Thus the corresponding path-transversal has the desired property. ♣

References

[1] N. Alon, J. Spencer, and P. Erdös. The probabilistic method, Wiley, 1991.

[2] M. Ben-Or, Private Communication.

[3] J. von zur Gathen, *Algebraic complexity theory*, Ann. Rev. Comp. Sci. 3:317-47, 1988.

[4] G. H. Gottlieb, *A certain class of incidence matrices*, Proc. AMS 17, 1233–1237, 1966.

[5] N. Nisan, *Lower Bounds for Non-Commutative Computation*, STOC 1991.

[6] R. Smolensky, *On interpolation by analytic functions with special properties and some weak lower bounds on the size of circuits with symmetric gates*, 31st FOCS, pp. 628–631, 1990.

[7] V. Strassen, *Vermeidung von divisionen*, J. Reine Angew Math., 264:184-202, 1973.

[8] E. Shamir and M. Snir, *On the depth complexity of formulas*, Math. Systems theory, 13:301-322, 1980.

[9] P. Tiwari and M. Tompa, *A Direct Version of Shamir and Snir's Lower Bounds on Monotone Circuit Depth*, Information Processing Letters 49, 1994.

[10] L. Valiant, *Negation can be exponentially powerful*, TCS 12, 303–314, 1980.

[11] L. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff, *Fast parallel computation of polynomials using few processors*, Siam J. Comp. 12:641-44, 1983.