Basic Research in Computer Science

BRICS RS-95-43 Nisan & Wigderson: Lower Bounds on Arithmetic Circuits via Partial Derivatives

Lower Bounds on Arithmetic Circuits via Partial Derivatives

(Preliminary Version)

Noam Nisan
Avi Wigderson

BRICS Report Series **RS-95-43**

ISSN 0909-0878 **August 1995**

**Copyright © 1995, BRICS, Department of Computer Science
University of Aarhus. All rights reserved.**

**Reproduction of all or part of this work
is permitted for educational or research use
on condition that this copyright notice is
included in any copy.**

See back inner page for a list of recent publications in the BRICS Report Series. Copies may be obtained by contacting:

**BRICS
Department of Computer Science
University of Aarhus
Ny Munkegade, building 540
DK - 8000 Aarhus C
Denmark**

**Telephone: +45 8942 3360
Telefax: +45 8942 3255
Internet: BRICS@daimi.aau.dk**

BRICS publications are in general accessible through WWW and anonymous FTP:

`http://www.brics.aau.dk/BRICS/`
`ftp ftp.brics.aau.dk (cd pub/BRICS)`

Lower Bounds on Arithmetic Circuits via Partial Derivatives (Preliminary Version)

Noam Nisan * Avi Wigderson †

Abstract

In this paper we describe a new technique for obtaining lower bounds on restriced 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 and iterated matrix products.

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. [9]). To date the best lower bounds known for arithmetic circuit size are only slightly super-linear ($\Omega(n \log n)$,

*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

[4]), 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 [6] (see also [7]), and is believed to hold for the 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 differnt kind of circuit lower bounds, by Smolensky [8]. 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 sections 5 and 6 contains proofs of all lower bounds on iterated matrix product.

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 (indegree zero nodes) 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 plus or multiplication gates, computing the sum and product, resp, of the polynomials on the heads 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 (total) 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 [5] (see [10]) that circuits for homogeneous functions can be made homogeneous with only a polynomials cost in size. However, it turns out that in this construction the depth grows by $O(\log d)$.

Lemma 1 *(Implicit in [5]) 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 $O(h \log d)$.*

We conjecture that the depth increase by $O(\log d)$ is necessary. Natural candidates for exhibiting a gap are the symmetric functions. Ben-Or [1] and others have observed that general circuits of unbounded fanin and depth 3 can compute (via polynomial interpolation) all symmetric polynomials in polynomial size (note the contrast to the exponential lower bounds for Majority in the Boolean model and finite fields).
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 [2] 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 [2] do not apply).
Theorem 0: Any homogeneous depth 3 circuit computing the dth symmetric polynomial on n variables requires $\Omega((n/2d)^{d/2})$ 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.

Definition 2 *Fix a partition $X^1, \cdots, X^d$ of the set of variables. For $T \subseteq [d]$ we call a polynomial T-multilinear if each of its monomials contains exactly one variable from each set X^i for $i \in T$, and in degree 1. A function is multilinear if it is T-multilinear for some subset T. A circuit is multilinear if all of its nodes compute multilinear functions.*

Besides the fact that such circuits are natural, observe that monotone circuits that compute a multilinear function are necessarily 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 $O(h + d \log h)$.*

Proof: Each gate of the original circuit is split into 2^d nodes consisting of the different multilinear parts (one for each subset) computed by the gate. Each gate of the original circuit is then simulated on each of the parts.

An addition gate is replaced by adding each of the different parts. This does not increase depth, as the parts are independent. This is the source of h in the above upper bound.

The source of the $d \log h$ in the bound is the multiplication gates. A multiplication gate $C = A * B$ is simulated by multiplying every multiliner part of A by every multilinear part of B and adding up those which give the same multilinear part of C. To obtain a part of degree d' one needs to add, for each $0 \leq i \leq d'$, $\binom{d'}{i}$ results of multiplications of a degree i term by a degree $d' - i$ term. The depth of this addition tree is $\binom{d'}{i}$.

Consider any path in the original circuit, which has $t \leq h$ multiplication gates. If the child of the smaller degree of the jth multiplication gate has degree d_j, then we have $\sum_j d_j = d$. The total depth of the addition trees

above for this path is at most $\sum_j \log \binom{d}{d_j}$, which is maximized under the above condition when all $d_j = d/t$ to $d \log t \leq d \log h$. ♣

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 indeed 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 $exp(\Omega((d^{1/h})/h))$. Consequently, any polynomial size circuit for this problem requires depth $\Omega(\log d / \log \log d)$.

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 can associate with each node a subset $S \subseteq \{1..d\}$ from which its monomial takes variables.

Definition 3 *A multilinear circuit is called pure if for every two sets S, T associated with nodes in the circuit either $S \subseteq T$ or $S \cap T = \emptyset$.*

Note that natural circuits for iterated matrix product are pure. While pure circuits can certainly compute all multilinear functions, we do not know of any general simulation of general circuits by pure circuits. The techniques of [2] can be extended to give an $\Omega(n)$ lower bound for computing the determinant by pure circuits, but these techniques cannot give any lower bounds for iterated matrix multiplication. 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 $[n] = \{1, 2, \cdots, n\}$.

Let f be a polynomial. We let $\partial(f)$ denote the set of **all** partial derivatives (of all orders) of f. 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 $\alpha \in F$ we have:*

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

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 that $\partial(g)$ is in the linear span of $\{1, g\}$ 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 multiplication fanin exceeding d is useless in homogeneous circuits computing a degree d polynomial, and a lower bound on the dimension of the partials of symmetric functions below. ♣

Lemma 4 *Let SYM_d^n denote the dth elementary symmetric polynomial over a set of n variables. Then*

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

Proof: Extend the notation above to SYM_d^X, denoting the dth elementary symmetric polynomial over a set of variables X. In the sequel let S and T range over the set I all d-subsets of $[n]$. We consider the two vectors of polynomials U, V, indexed by I, defined as follows. $U_S = \Pi_{i \in S} x_i$ are the monomials. $V_T = SYM_d^{[n]-T}$ is the partial derivative of $SYM_{2d}^{[n]}$ with respect to the variables 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 every field, and all monomials in U are linearly independent, we have

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

♣

4 Complexity Measures

4.1 Polynomials

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. The variables in X^i will be denoted x_j^i. Now our polynomials are in the ring $F[X_1, \cdots, X_d]$.

For $T \subseteq [d]$ we use P_T to denote the set of all T-multilinear polynomials (see definition 2). Observe that P_T is a vector space over F, and that $dim(P_T) = n^{2|T|}$.

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. each member of $\partial_S(f)$ is the partial derivative of f with respect to the set of variables of some monomial in P_S.) We clearly have $\partial_S(f) \subseteq P_{T-S}$. A trivial fact of key importance is:

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

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

Proposition 3 *For every $T, R \subseteq [d]$ 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 [7]). 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, $f \in P_T$, $\rho(f) \leq 1$.*

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. 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: Simply take all partial derivatives with respect to the even-numbered blocks of the partition. ♣

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 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 and Theorem 1 in our notation.
Theorem 0: $S_3^H(SYM_d^n) = \Omega(n/2d)^{d/2})$

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

For the proof of Theorem 1 we observe that depth 3 homogeneous circuits can be turned to 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 of the parts in the variables' partition. 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. ♣

Proof:(of Theorem 1): 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.

We will derive our circuit size lower bound via a formula size lower bound. A formula is simply a circuit with fanout 1 at every gate. Let $L_h^M(f)$ denote the size of the smallest multilinear formula of depth h for f. Then we trivially have

Lemma 8 *For every f, h, $L_h^M(f) \leq (S_h^M(f))^h$*

Let $f \in P_{[d]}$ be a function computed by a formula C, and let $R \subseteq [d]$. If we assign constant values to all variables in all $X^i, i \notin R$, we obtain a reduced formula denoted $C|_R$ computing the polynomial denoted $f|_R$ in P_R (the values of the constants are immaterial for the lemmas below, so they do not appear in the notation). 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 9 *Let $\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_d$ be independent, unbiased $0,1$ random variables. For $T \subseteq [d]$ let $\mathbf{z}(T) = (\sum_{i \in T} \mathbf{z}_i) mod 2$. Then for every family of pairwise*

disjoint subsets $T_1, T_2, \cdots, T_r$ *of* $[d]$,
$Pr[\sum_j \mathbf{z}(T_j) < r/3] \leq exp(-r/10)$

Proof: Follows directly from the Chernoff bound. ♣

Lemma 10 *Let* C *be an optimal multilinear formula 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 formula with the following properties.*

- $s \leq L_h^M(f)$
- *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. So by the subadditivity Lemma 5 and the fact that C is a formula we have the third property. ♣

Lemma 11 *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* $L_h^M(f) \leq (1/2)exp(r/10)$. *Then* $Pr[\rho(f|\mathbf{z}) \geq 1/n] \leq 1/2$.

Proof: Let C be an optimal depth h multilinear formula for f, and $g_1, \cdots, g_s$ the gates guaranteed by the previous lemma. Recall that $deg(g_j) \geq r$ and $s \leq (1/2)exp(r/10)$. ♣

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.

Theorem 2 *For all* d, h *we have*

- $L_h^M(IMM_d^2) = exp(\Omega(d^{1/h}))$
- $S_h^M(IMM_d^2) = exp(\Omega((d^{1/h})/h))$

Proof: The circuit lower bound follows from the formula size ove via Lemma 8 To prove the lower bound on formula size, 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. Like the random restrictions of the parity functions left it a parity function on fewer bits, 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 11. ♣

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 [7].

Lemma 12 *Let d be an odd integer, and f be a multilinear polynomial (of degree d) over $X^1, \cdots, X^d$. Then $D^O(f) = \log \rho(f)) + \Omega(\log n \log d)$*

Proof: Assume we are given a multilinear circuit C for f. Assume wlog that the fan-in of every gate is 3 (since here we deal with bounded fanin). 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 $\log \rho(f) + 2 \log_3 n \log_3 d$ plus gates along the path, giving the required bound. ♣

We illustrate the lower bound on two multilinear 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 12 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 12 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 4 *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-transversal to every circuit.*

Lemma 13 *If every path in a some 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) The construction is inductive as follows. Let X be a set of variables, and partition it to three (roughly) equal parts X_1, X_2, X_3. For $i \in \{1,2,3\}$ let g_i be a polynomial with the required property on X_i and h_i a polynomial with the required property on $X \setminus X_i$. Now define $f = g_1h_1 + g_2h_2 + g_3h_3$. It is easy to see that f enjoys this property as well. ♣

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 13 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 inductivly 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. ♣

Acknowledgements

We are greatful to Jiri Sgall for careful reading and many comments on an earlier version of this paper.

References

[1] M. Ben-Or, Private communication.

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

[3] 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.

[4] V. Strassen, "Die Berechnungskomplexitat vo elementarsymmetrischen Funktionen und von Interpolationskoefizienten", *Numer. Math.* 20, pp. 238–251, 1973.

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

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

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

[8] 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.

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

[10] L. Valiant, *Negation can be exponentially powerful*, 1979.

[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.

Recent Publications in the BRICS Report Series

RS-95-43 **Noam Nisan and Avi Wigderson.** ***Lower Bounds on Arithmetic Circuits via Partial Derivatives (Preliminary Version).*** **August 1995. 17 pp. To appear in** ***36th Annual Conference on Foundations of Computer Science*****, FOCS '95, IEEE, 1995.**

RS-95-42 **Mayer Goldberg.** ***An Adequate Left-Associated Binary Numeral System in the λ-Calculus.*** **August 1995. 16 pp.**

RS-95-41 **Olivier Danvy, Karoline Malmkjær, and Jens Palsberg.** ***Eta-Expansion Does The Trick.*** **August 1995. 23 pp.**

RS-95-40 **Anna Ingólfsdóttir and Andrea Schalk.** ***A Fully Abstract Denotational Model for Observational Congruence.*** **August 1995. 29 pp.**

RS-95-39 **Allan Cheng.** ***Petri Nets, Traces, and Local Model Checking.*** **July 1995. 32 pp. Full version of paper appearing in Proceedings of AMAST '95, LNCS 936, 1995.**

RS-95-38 **Mayer Goldberg.** ***Gödelisation in the λ-Calculus.*** **July 1995. 7 pp.**

RS-95-37 **Sten Agerholm and Mike Gordon.** ***Experiments with ZF Set Theory in HOL and Isabelle.*** **July 1995. 14 pp. To appear in** ***Proceedings of the 8th International Workshop on Higher Order Logic Theorem Proving and its Applications*****, LNCS, 1995.**

RS-95-36 **Sten Agerholm.** ***Non-primitive Recursive Function Definitions.*** **July 1995. 15 pp. To appear in** ***Proceedings of the 8th International Workshop on Higher Order Logic Theorem Proving and its Applications*****, LNCS, 1995.**

RS-95-35 **Mayer Goldberg.** ***Constructing Fixed-Point Combinators Using Application Survival.*** **June 1995. 14 pp.**

RS-95-34 **Jens Palsberg.** ***Type Inference with Selftype.*** **June 1995. 22 pp.**