Depth-3 Arithmetic Circuits over Fields of Characteristic Zero*

Amir Shpilka † Avi Wigderson ‡

June 10, 2001

Abstract

In this paper we prove quadratic lower bounds for depth-3 arithmetic circuits over fields of characteristic zero. Such bounds are obtained for the elementary symmetric functions, the (trace of) iterated matrix multiplication, and the determinant. As corollaries we get the first nontrivial lower bounds for computing polynomials of constant degree, and a gap between the power of depth-3 arithmetic circuits and depth-4 arithmetic circuits. We also give new shorter formulae of constant depth for the elementary symmetric functions

The main technical contribution relates the complexity of computing a polynomial in this model to the wealth of partial derivatives it has on every affine subspace of small co-dimension. Lower bounds for related models utilize an algebraic analog of Nečhiporuk lower bound on Boolean formulae.

1 Introduction

Arithmetic circuits are a very natural model for computing polynomials. Like most computational models, almost no lower bounds are known for this one. The best size lower bound known is the classical $\Omega(n \log d)$ (for some natural degree d polynomials over n variables) of [1]. No nontrivial lower bounds are known for depth. For a survey of known results see [14, 3] and the introduction to [10].

Our intuition suggests that arithmetic circuits (being more "structured") are weaker than Boolean circuits, and thus lower bounds for the former should be easier to prove. Our experience with monotone analogs of both models certainly justifies this intuition. However, it is shattered by the simple problem of computing majority in depth-3 circuits. We know that in the Boolean model this requires exponential size. However, Ben-Or proved that the majority polynomial has simple, quadratic-size depth-3 arithmetic formula!

In this paper we deal mainly with depth-3 arithmetic circuits. Such a circuit can be viewed as a sum of products of linear functions of the variables. This is a very restricted model, but it can clearly compute any multivariate polynomial, some of which surprisingly cheaply. Being the "simplest" nontrivial model, it has received significant attention as detailed below.

Despite its innocence, no super-linear lower bounds are known for this model when the field is large, except of the degree lower bound of [1]. This state of affairs is in contrast with what is known for Boolean circuits with $mod\, q$ gates, and arithmetic circuits over finite fields. In the first model [11] and [13] proved exponential lower bounds e.g. for the majority function for any constant

*a preliminary version of the paper appeared in 14th Computational Complexity, pp 87-96, 1999.

†Email: amirs@cs.huji.ac.il

‡This research was supported by grant number 69/96 of the Israel Science Foundation, founded by the Israel Academy for Sciences and Humanities. Email: avi@cs.huji.ac.il

depth. In the second model (for depth-3 only) [5] and [6] recently proved exponential lower bounds for some symmetric functions.

These techniques cannot be extended to give analogous results for large fields, as by the above mentioned result of Ben-Or [2], they are simply false. One path which was taken to handle large fields was to further restrict the model [9, 10]. They consider homogeneous circuits, which for depth-3 circuits amounts to requiring a homogeneous linear function at every gate in the bottom level. In that model [10] were able to prove exponential lower bounds, even for the majority polynomial. Thus, the general model is sometimes exponentially more powerful than its homogeneous variant.

In this paper we study the general model. We prove near quadratic lower bounds for a number of natural functions, such as the elementary symmetric functions, the (trace of) iterated matrix product, and the determinant. We show in particular the following lower bounds (where n denotes the total number of variables):

- $\Omega(n^2)$ lower bound for the elementary symmetric functions of degree $\Omega(n)$ (Thus proving that the construction in [2] is essentially optimal).
- $n^{2-\epsilon}$ lower bound for polynomials of constant degree (No super-linear bound was known before).
- $\tilde{\Omega}(n^2)$ lower bound for the determinant (But we believe that the true complexity is exponential).

The proof combines the idea of partial derivatives of [10] and the idea of approximating "high rank" multiplication gates of [5]. Some of these lower bounds, (together with upper bounds) provide near quadratic separation between depth-3 and depth-4 circuits. Of special interest is a new (depth-6) formula of size $O(nd^3 \log d)$ for the elementary symmetric polynomials of degree d (which beats the Ben-Or construction for small degrees).

We prove two general theorems that express a lower bound for the depth-3 complexity of a polynomial f, in terms of parameters k and D, where D measures the wealth of partial derivatives f has on every affine subspace of co-dimension k. These are stated and proved in section 3. Specific lower bounds follow from proving lower bounds on D for specific polynomials and well chosen values of k. These are stated and proved in section 4.

In section 5 we prove new upper bounds for the elementary symmetric polynomials.

In section 6 we look at different (but related) models of computation. The first model is formulae without any depth restriction, whose inputs are allowed to be not only the variables, but also arbitrary univariate polynomials. For this model we prove a quadratic lower bound for the Discriminant function (which is the determinant of a Vandermonde matrix). The proof uses an algebraic analog of a Nečhiporuk-like argument (see [8]).

Finally, we study the computation of polynomials of the form $t^m f$, where f is our target (homogeneous) polynomial, and t is a new indeterminate. We show general computation of such products of f is no stronger than *homogeneous* computation of f. This result has the same flavor of results on the monotone vs general computation of "slice functions". It yields an exponential gap between the depth-3 homogeneous complexity of a polynomial and the homogeneous complexity of its derivative with respect to a single variable.

2 Definitions and Tools

2.1 Arithmetic Circuits

Definition 2.1 *An* arithmetic circuit *is a labeled directed acyclic graph. The inputs (nodes of in-degree zero) are labeled from the set of variables X. The output nodes are the nodes of out-degree zero. A constant from F (the base field) 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 outputs of the circuit are the polynomials computed at the output nodes. A* formula *is a circuit which all its nodes have out-degree one (namely, a tree). We consider unbounded fan-in circuits. The* size *of a circuit is the number of edges in it, and the size of a formula is the number of its leafs. The* depth *of the circuit is the length of a longest path between the output node and an input node.*

The main model we shall deal with in the paper is the following.

Definition 2.2 *A $\Sigma\Pi\Sigma$ circuit is a leveled depth-3 circuit with a plus gate at the top, multiplication gates at the middle level, and plus gates at the bottom. A homogeneous $\Sigma\Pi\Sigma$ circuit is a $\Sigma\Pi\Sigma$ circuit that is allowed to compute only homogeneous linear functions in the bottom level.*
For a polynomial f, we denote by $s_3(f)$ (respectively $L_3(f)$) the size of the smallest $\Sigma\Pi\Sigma$ circuit (respectively formula) computing f, and by $s_3^H(f)$ (respectively $L_3^H(f)$) the size of the smallest homogeneous $\Sigma\Pi\Sigma$ circuit (respectively formula) computing f.

Comment 1 *Clearly, for all variants, $s \leq L$. All our lower bounds will be on s, and all our upper bounds on L.*

From the definition it is clear that if f is computed by such circuit then f has a representation of the form: $f = \sum_{j=1}^{s} M_j$ where $M_j = \prod_{i=1}^{deg(M_j)} \ell_{i,j}$, each $\ell_{i,j}$ is a linear function in the variables, and $deg(M_j)$ is the fan-in of the jth multiplication gate. We stress that this linear function may involve a constant term (and indeed without this ability the model is homogeneous), and that different multiplication gates may use the same linear function. It will be useful for us to separate out the homogeneous part of such functions.

Definition 2.3 *For a linear function ℓ we let ℓ^h be the homogeneous part of ℓ, and $\ell^0 = \ell - \ell^h$ is the constant term.*
Let $M = \prod_{i=1}^{deg(M)} \ell_i$. We denote $M^h = \{\ell_1^h, ..., \ell_{deg(M)}^h\}$ and let $\dim(M^h)$ be the dimension of the linear span of the set M^h.

All our lower bounds for this model will trade off the number of edges that fan into gates at the middle level (multiplication gates) with the number of gates in that level. We will use the obvious

Proposition 2.1 *Every $\Sigma\Pi\Sigma$ circuit with multiplication gates $\{M_1, ..., M_s\}$ has size $\Omega(\sum_{j=1}^{s} deg(M_j))$.*

2.2 Partial Derivatives

We now recall basic definitions and results from [10] on partial derivatives, which will be key for our lower bounds.

Let F be a field of characteristic zero. We will consider polynomials over a set of variables X.

Definition 2.4 *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 in V over F).*

Remark 1 *Observe that $dim(V)$ is invariant under any full rank linear transformation on the set of variables X.*

Definition 2.5 *[10] Let f be a polynomial and d an integer. We let $\partial_d(f)$ denote the set of partial derivatives of order d of f.*

Example 1

$$\partial_2(x^2y) = \{0, 2x, 2y\} .$$

A fundamental observation of [10] is that this dimension commutes with the arithmetic operations, i.e it is invariant under scalar multiplication and is sub-additive. Here we'll use it only for addition gates.

Proposition 2.2 *[10] For every $f_1, f_2, \cdots, f_r \in F[X]$ and $\alpha \in F$, $\alpha \neq 0$ we have:*

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

For multiplication gates, we'll use stronger bounds than those in [10], which apply only to product of linear functions.

Proposition 2.3 *For a multiplication gate M with $\dim(M^h) = m$, and for every d, $dim(\partial_d(M)) \leq \binom{m+d}{d}$.*

Proof: From the definition it is clear that M is a function of m linearly independent linear functions (w.l.o.g. - the first m), i.e $M = M(\ell_1, ..., \ell_m)$. Therefore $\frac{\partial M}{\partial x} = \sum_{i=1}^{m} \frac{\partial M}{\partial \ell_i} \frac{\partial \ell_i}{\partial x}$. Since ℓ_i is a linear function, $\frac{\partial \ell_i}{\partial x}$ is a scalar, so $\frac{\partial M}{\partial x}$ is a linear combination of the $\frac{\partial M}{\partial \ell_i}$'s. A similar thing happens when we look at derivatives of order d, each derivatives lies in the linear span of order d derivatives of M with respect to the ℓ_i-s. Therefore to bound $dim(\partial_d(M))$ it is sufficient to bound the number of such derivatives. Since M is a polynomial the order in which we take derivatives doesn't matter, i.e $\frac{\partial^2 M}{\partial \ell_1 \partial \ell_2} = \frac{\partial^2 M}{\partial \ell_2 \partial \ell_1}$. Therefore the number of order d partial derivatives is at most the number of ways to write d as a sum of m integers (the value of the i-th integer corresponds to the order we take derivatives of M w.r.t. ℓ_i). □

Proposition 2.4 *For a multiplication gate M with $deg(M) = m$, and for every d, $dim(\partial_d(M)) \leq \binom{m}{d}$.*

Proof: Write $M = \prod_{i=1}^{m} \ell_i$. Denote $[m] = \{1, ..., m\}$. Since each ℓ_i is a linear function we get,

$$\partial_d(M) \subset \ span\{\prod_{i \in T} \ell_i | T \subset [m], |T| = m - d\}.$$

The result follows since $\dim(span\ \{\ \prod_{i \in T} \ell_i \mid T \subset [m], |T| = m - d\ \}) \leq \binom{m}{d}$. □

2.3 Restrictions to affine subspaces

A key ingredient of our lower bound technique will be to study the dimension of a set of partial derivatives not over the whole vector space F^n, but over affine subspaces of it.

Definition 2.6 *Let A be an affine subspace of F^n with the set of coordinates $(x_1, ..., x_n)$. Call $B \subset X$ a base for A if A can be represented by the set of equations:*

$$\{x_b = \ell_b \mid b \in B\}$$

where ℓ_b is a linear function on the set of variables $X - B$. Note that every A has such a base B, so lets fix one such base.
Define $\phi_B : F[X] \mapsto F[X \setminus B]$ to be the homomorphism which assigns to every variable x_b with $b \in B$ the linear function ℓ_b (and leaves the other variables untouched). This map is extended by multiplicativity to monomials and then by additivity to polynomials.

Definition 2.7 *Let A be an affine subspace of F^n, for a polynomial f we denote by $f|_A^{(B)}$ the restriction $\phi_B(f)$ o f f to A. For a set of polynomials V, $V|_A^{(B)} = \{f|_A^{(B)} : f \in V\}$.*

Definition 2.8 *Let C be an arithmetic circuit, and ϕ_B as above. Then $C|_A^{(B)}$ is the circuit obtained by applying ϕ_B to the inputs, and removing sub-circuits whose output becomes identically zero.*

Clearly, restrictions commute with arithmetic operations. Thus

Proposition 2.5 *Assume C computes f. Then for every affine subspace A (chosen together with a base B), $C|_A^{(B)}$ computes $f|_A^{(B)}$.*

The complexity of computing a polynomial f will be related in the next section to the dimension of one of two different sets of polynomials derived from f.

1. $(\partial_d(f))|_A^{(B)}$ is the set of restrictions of all order d partial derivatives on f to A.
2. $\partial_d(f|_A^{(B)})$ is the set defined by first restricting f itself to A, and then taking all order d partial derivatives. To formally define it, we need an "inverse" to ϕ_B below.

Definition 2.9 *Let's assume that $X - B = \{x_{i_1}, ..., x_{i_{n-k}}\}$. Define $\ell_B : F^{n-k} \mapsto F^n$ by*

$$\ell_{B(i)} = \begin{cases} l_b & i = b \in B \\ x_i & i \notin B \end{cases}$$

where $\ell_{B(i)}$ is the i-th coordinate of ℓ_B.

Definition 2.10 *Let $\partial_d(f|_A^{(B)}) = \partial_d(f \circ \ell_B)$.*

We conclude this subsection with a remark on the arbitrariness of the choice of the base B of the affine subspace A. Clearly, this choice affect the sets defined above, as well as the restricted circuit, at least in the obvious sense that the restricted polynomials will be over different sets of indeterminates.

Still, the choice of B has no affect on the proofs, so from this point of view it can certainly be arbitrary. Moreover, it turns out that while the sets of polynomials defined by (1) and (2) above change with different choices of B, the dimension of each is invariant under this choice.

Proposition 2.6 *For every B, B' there's a full rank linear transformation $P : F^{n-k} \mapsto F^{n-k}$ such that $\ell_B = \ell_{B'} \circ P$.*

Claim 2.7 *For every $B \neq B'$ bases for A, $dim(\partial_d(f|_A^{(B)})) = dim(\partial_d(f|_A^{(B')}))$.*

Proof: We show the proof for the case $d = 1$ but for larger d-s it's the same. According to Proposition 2.6 there's a full rank linear transformation P such that $\ell_{B'} = \ell_B \circ P$. For convenience let $P(X_1, ..., X_{n-k}) = (Y_1, ..., Y_{n-k})$, we get:

$$\frac{\partial}{\partial X_i}(f \circ \ell_{B'}) = \frac{\partial}{\partial X_i}(f \circ \ell_B \circ P) = \sum_{j=1}^{n-k}((\frac{\partial}{\partial Y_j}(f \circ \ell_B)) \circ P)\frac{\partial Y_j}{\partial X_i} \ .$$

Since $\frac{\partial Y_j}{\partial X_i}$ is a scalar in F we get that $\frac{\partial}{\partial X_i}(f \circ \ell_{B'}) \in span((\frac{\partial}{\partial Y_j}(f \circ \ell_B)) \circ P)$. Therefore $dim(\partial_1^{B'}(f)) \leq dim(\partial_1^{B}(f))$. From symmetry (since P has full rank) the result follows. □

From now on we shall assume that every affine subspace A comes together with a base B. For short we define:

Definition 2.11 $\partial_d(f|_A) = \partial_d(f|_A^{(B)})$.

3 Main Theorems

The following theorems will be our main tool in deriving the lower bounds. Observe that in both theorems, the lower bound is the minimum of two functions. It will be clear from their proofs that they both actually give a trade-off: either the first function bounds from below the number of edges that fan into the middle level (multiplication gates), or the second lower bounds the number of gates in that level.

Theorem 3.1 *Let f be a polynomial. Assume that for some integers d, k, D, for every affine subspace A of co-dimension k, $dim((\partial_d(f))|_A) > D$. Then*

$$s_3(f) \geq \min(\frac{k^2}{d}, \frac{D}{\binom{k+d}{d}}) \ .$$

Theorem 3.2 *Let f be a polynomial. Assume that for some integers d, k, D, for every affine subspace A of co-dimension k, $dim(\partial_d(f|_A)) > D$. Then for every m*

$$s_3(f) \geq \min(km, \frac{D}{\binom{m}{d}}) \ .$$

The way to apply these theorems is to prove, that for certain polynomials, D can be chosen large, for appropriate values of d, k. We note that the theorems are incomparable – each can give a better lower bound than the other for some polynomials.

To prove these theorems we will first need a technical lemma. It basically shows how to define an affine subspace that nullifies high rank multiplication gates.

Lemma 3.3 *Fix an integer z. Let $\mathcal{C}$ be a $\Sigma\Pi\Sigma$ circuit, and let $S = \{j \mid \dim(M_j^h) \geq k\}$. If $|S| < \frac{k}{z}$ then there is an affine subspace A of co-dimension k such that in $\mathcal{C}|_A$ all the multiplication gates are of dimension less then k. Moreover, in A all the multiplication gates in S have a zero of order $\geq z$.*

Proof: Let's assume w.l.o.g. that $S = \{1, 2, ..., r\}$ for some $0 \leq r < \frac{k}{z}$. Now lets take for each $1 \leq j \leq r$, z linear functions, $\ell_{j,1}, ..., \ell_{j,z}$, from M_j, such that all the $\ell_{j,i}^h$'s are linearly independent. It's possible to make such a choice since $r < \frac{k}{z}$ and $\dim(M_j^h) \geq k$. Because of the independence there's an affine subspace A of co-dimension at most rz such that

$$\forall \vec{x} \in A, \quad \ell_{j,i}^h(\vec{x}) = -\ell_{j,i}^0 \quad \text{(i.e } \ell_{j,i}(\vec{x}) = 0 \text{ for the relevant } (j,i)\text{-s)}.$$

So now in $\mathcal{C}|_A$ all the multiplication gates are of dimension less than k, and $M_1, ..., M_r$ each have a zero of order $\geq z$ on A. □

We can now turn to the proofs of the main theorems.

Proof of Theorem 3.1: Fix any k and d, and set $z = d + 1$. Assume that $\mathcal{C}$ is a depth-3 circuit computing f. If the assumption of Lemma 3.3 does not hold, then by Proposition 2.1 we have a $\frac{k^2}{d}$ lower bound on the number of edges that fan into the multiplication gates. Otherwise the assumption holds, and let A be the subspace guaranteed by the lemma.

Now we know that all the gates in $\mathcal{C}$ with dimension $\geq k$ have a zero of degree $d + 1$ when restricted to A. Thus for each such multiplication gate M, $dim(\partial_d(M)|_A) = 0$. By Proposition 2.3 all the other multiplication gates have $dim(\partial_d(M)|_A) \leq \binom{k+d}{d}$ so by Proposition 2.2 there must be at least $\frac{D}{\binom{k+d}{d}}$ multiplication gates in $\mathcal{F}$. □

Proof of Theorem 3.2: Assume that we are given d, k, m. We now look at an optimal $\Sigma\Pi\Sigma$ circuit for f. If there are more than k multiplication gates of degree greater than m then we are done. Otherwise we can find an affine subspace A of co-dimension k so if we restrict the circuit to A all the remaining gates will be of degree $\leq m$. Using Propositions 2.2, 2.4 we see that there are at least $\frac{D}{\binom{m}{d}}$ multiplication gates in the restricted circuit. □

It is clear from our theorems that the lower bounds for a polynomial f are expressed in terms of the dimensions of the sets $\partial_d(f|_A)$ and $(\partial_d f)|_A$, for some arbitrary affine subspace A. These sets can be quite complicated even for very simple polynomials. To bound the dimensions, we establish two simple lemmas relating these to each other and to the more easily understood set, $\partial_d(f)$.

Lemma 3.4 *For every polynomial f and affine subspace A with a base B we have*

$$dim((\partial_d f)|_A) \geq dim(\partial_d(f) \cap F[X \setminus B]) \ .$$

Proof: The polynomials in $\partial_d(f)$ which do not depend on variable from B are not effected by the restriction to A. □

Definition 3.1 *For a set of variables B, let $\partial_d^B f$ denote all order d derivatives of f, where at least one of the d derivations are with respect to a variable in B.*

Lemma 3.5 *For every polynomial f and affine subspace A with a base B we have*

$$dim(\partial_d(f|_A)) \geq dim((\partial_d f)|_A) - dim(\partial_d^B f) \ .$$

The proof actually shows that

$$(\partial_d f)|_A \subset \text{span}\left(\partial_{\mathrm{d}}(\mathrm{f}|_{\mathrm{A}}) \cup (\partial_{\mathrm{d}}^{\mathrm{B}}(\mathrm{f}))|_{\mathrm{A}}\right) .$$

Proof: We will demonstrate the proof for $|B| = 1$ and $d = 1$. w.l.o.g. assume that $B = \{x_1\}$ and A is given by $x_1 = \ell(x_2, ..., x_n)$. According to the chain rule we get:

$$\partial_1(f|_A) = \{\frac{\partial f(\ell, x_2, ..., x_n)}{\partial x_i} | 1 < i\} = \{\frac{\partial f}{\partial x_1}(\ell, x_2, ..., x_n) \cdot \frac{\partial \ell}{\partial x_i} + \frac{\partial f}{\partial x_i}(\ell, x_2, ..., x_n) \mid 1 < i\} .$$

Since $\partial_1^B f = \frac{\partial f}{\partial x_1}$, the set $(\partial_1 f)|_A$ (which is actually the set $\{\frac{\partial f}{\partial x_i}(\ell, x_2, ..., x_n)|1 \leq i\}$) is spanned by the set $\partial_1(f|_A) \cup \{\frac{\partial f}{\partial x_1}(\ell, x_2, ..., x_n)\}$.

The proof for larger values of $|B|, d$ is achieved using similar arguments. View the restriction to A as a composition of a polynomial with a set of linear functions. Then from the chain rule of partial derivatives, and from the fact that all derivatives of linear functions are constants, it follows that the set $(\partial_d f)|_A$ is spanned by the two sets $\partial_d(f|_A)$ and $\partial_d^B f$. □

4 Lower Bounds

4.1 Elementary symmetric functions

Definition 4.1

$$S_n^d(X) = \sum_{\substack{T \subset [n] \\ |T| = d}} \prod_{i \in T} x_i .$$

X is the set of variables $\{x_1, ..., x_n\}$. This is the d'th elementary symmetric function.

Theorem 4.1 *For every $\log n \leq d \leq 2n/3$ we have $s_3(S_n^d) \geq \max(\Omega(\frac{n^2}{d}), \Omega(nd))$.*

This lower bound is tight (for large values of d) in view of the following theorem (which will be proved in section 5):

Theorem 4.2 (Ben-Or) *For every d, $s_3(S_n^d) \leq O(n^2)$.*

To prove Theorem 4.1 we first need to prove the following theorem:

Theorem 4.3 *$s_3(S_n^{2d}) \geq \Omega(\frac{n^2}{d})$ where $n/10 \geq d \geq \log n$.*

The proof is based on a Lemma showing that S_n^{2d} has lots of partial derivatives even when restricted to affine subspaces of small co-dimension.

Lemma 4.4 *For every n, k, d and for every affine subspace A of co-dimension k,*

$$dim((\partial_d(S_n^{2d}))|_A) \geq \binom{n-k}{d} .$$

Proof of Lemma 4.4: w.l.o.g. assume that A is defined by the following equations:

$$\forall 1 \leq j \leq k \;\; x_j = \ell_j \; .$$

The set $\partial_d(S_n^{2d})$ is actually the set $\{S_{n-d}^d(X - \{x_i | i \in T\})\}$, which for convenience we denote by $\{S_{n-d}^d(X-T)\}$ (recall Definition 4.1) where T ranges over the $\binom{n}{d}$ subsets of $[n]$ of size d. By [4] (as used in [10]), this set is spanned by all degree d multilinear monomials. There are $\binom{n-k}{d}$ linear independent degree d multilinear monomials in the variables of the set $X - B$, using Lemma 3.4 the result follows. □

Proof of Theorems 4.3: If we plug Lemma 4.4 into Theorem 3.1, we get a lower bound of

$$s_3(S_n^{2d}) \geq \min(\frac{k^2}{d}, \frac{\binom{n-k}{d}}{\binom{k+d}{d}}) \;\; \forall k, d \; .$$

Let's estimate $\frac{\binom{n-k}{d}}{\binom{k+d}{d}}$ by $(\frac{n-k}{k+d})^d$ and take $k = n/9, \;\; n/10 \geq d \geq \log n$. We now get $(\frac{n-k}{k+d})^d \geq (\frac{8n/9}{2n/9})^{\log n} = n^2$. Therefore $s_3(S_n^{2d}) \geq \min(\Omega(\frac{n^2}{d}), n^2) = \Omega(\frac{n^2}{d})$ and the lower bound follows. □

This theorem already gives us the desired result (of Theorem 4.1) for $\log n \leq d \leq \sqrt{n}$. The next lemma together with Theorem 3.2 gives the rest of the Theorem, for the right choice of parameters.

Lemma 4.5 *For every n, k, d, and for every subspace A of co-dimension k with $k \leq \min(d/2, n-d)$*

$$dim(\partial_{\frac{d-2k}{2}}(S_n^d|_A)) \geq \binom{n-2k}{\frac{d-2k}{2}} \; .$$

We will derive this lemma from the following one:

Lemma 4.6 *For every affine subspace A of co-dimension $\leq \min(d/2, n-d)$ we have $(S_n^d)|_A \neq 0$.*

To prove this we need some new notations:

Definition 4.2 *For a linear function $\ell = c_1x_1 + c_2x_2 + ... + c_nx_n + c_0, \;\; \ell \neq 0$ denote:*

$$LM(\ell) = \begin{cases} c_j x_j & \text{for the first } 0 < \text{j with } \text{c}_\text{j} \neq 0 \\ c_0 & \text{no such j exists} \end{cases}$$

$$LT(\ell) = \begin{cases} x_j & \text{for the first } 0 < \text{j with } \text{c}_\text{j} \neq 0 \\ 1 & \text{no such j exists} \end{cases}$$

Proof of Lemma 4.6: Let co-dimension $(A) = k$. w.l.o.g. assume that A is defined by the following linear equations:

$$x_i = \begin{cases} 0 & 1 \leq i \leq r \\ \ell_i(x_{d+1}, ..., x_n) & r+1 \leq i \leq k \end{cases}$$

where the ℓ_i are nonzero linear functions. The first r equations reduce $S_n^d(X)$ to the function $S_{n-r}^d(X \setminus [r])$. So assume w.l.o.g. that $r = 0$. Denote $Y = \{LT(\ell_i) \;\; | \;\; LT(\ell_i) \neq 1 \; , \; 1 \leq i \leq k\}$,

$|Y| = t$. We get that $(S_n^d)|_A$ contains all terms of the form $(\prod_{i=1}^{k}(\ell_i))S_{n-k}^{d-k}(X \setminus [k])$. In particular it contains all the monomials of the polynomial

$$(\prod_{i=1}^{k} LM(\ell_i))(\prod_{x_j \in Y} x_j)S_{n-(k+t)}^{d-(k+t)}(X \setminus ([k] \cup Y)).$$

But it is also easy to see that every monomial $\mathcal{M}$ of degree $d-(k-t)$ appearing in $(S_n^d)|_A$, such that $(\prod_{i=1}^{k} LT(\ell_i))(\prod_{x_j \in Y} x_j)$ divides $\mathcal{M}$, comes from $(\prod_{i=1}^{k}(\ell_i))S_{n-k}^{d-k}(X \setminus [k])$. Therefore nothing cancels

$$(\prod_{i=1}^{k} LM(\ell_i))(\prod_{x_j \in Y} x_j)S_{n-(k+t)}^{d-(k+t)}(X \setminus ([k] \cup Y)) \ .$$

and we get $(S_n^d)|_A \neq 0$. □

Proof of Lemma 4.5: Take all derivatives of order $(d-(k+t))/2$ (same t as in the above proof) with respect to the variables from $X \setminus ([k] \cup Y)$. The set of all derivatives contains polynomials of the form

$$\left\{ (\prod_{i=1}^{k} LT(\ell_i))(\prod_{x_j \in Y} x_j)S_{n-(k+d+t)/2}^{(d-(k+t))/2}(X \setminus ([k] \cup Y \cup R)) + Q_R \right\} \ such\ that$$

$$R \subset X \setminus ([k] \cup Y),\ \mid R \mid = (d-(k+t))/2 \ ,$$

where Q_R is a polynomial in which no monomial is divisible by

$$(\prod_{i=1}^{k} LT(\ell_i))(\prod_{x_j \in Y} x_j) \ .$$

These derivatives are linearly independent, so the dimension of their span is $\binom{n-(k+t)}{(d-(k+t))/2}$ and it is minimized when $t = k$. □

Proof of Theorem 4.1: For $\log n \leq d \leq \sqrt{n}$ we can use Theorem 4.3. So let's assume that $\sqrt{n} \leq d \leq 2n/3$. Now plug Lemma 4.5 into Theorem 3.2 with $k = d/4$, $m = n/100$ to get:

$$s_3(S_n^d) \geq \min(dn/100, \frac{\binom{n-d/2}{\frac{d}{4}}}{\binom{n/100}{\frac{d}{4}}}) \geq \min(dn/100, \frac{\binom{2n/3}{\frac{d}{4}}}{\binom{n/100}{\frac{d}{4}}}) \geq \min(dn/100, 2^d) \geq \Omega(nd).$$ □

In [12] Lemma 4.6 was strengthened and a stronger result for depth-3 circuits was proved:

Theorem 4.7 *[12]* $s_3(S_n^d) = \Omega(n^2)$ *for every* $d = \alpha n$, *where* $\alpha < 1$ *is a constant.*

4.2 Product of inner products

Definition 4.3

$$PIP_n^d(X,Y) = \prod_{i=1}^{d}(\sum_{j=1}^{n} x_j^{(i)} y_j^{(i)}) \ .$$

This is the product of d *inner products.* $X = \cup_{i=1}^{d} X^{(i)}$ *and* $Y = \cup_{i=1}^{d} Y^{(i)}$ *with each* $X^{(i)}$ *and* $Y^{(i)}$ *containing* n *variables.*

We have $2nd$ variables, and the lower bound that we show is strongest when $d \leq \sqrt{\log n}$.

Theorem 4.8 *$s_3(PIP_n^d) \geq \Omega(n^{2\frac{d}{d+2}})$ for $d \leq \frac{1}{2}\sqrt{\log n}$.*

$s_3(PIP_n^d) \geq \Omega(n^{2-\frac{4}{\sqrt{\log n}}})$ for $d > \frac{1}{2}\sqrt{\log n}$.

The proof, again, is based on a lemma that gives a lower bound for the wealth of partial derivatives that PIP has on affine subspaces.

Lemma 4.9 *$dim(\partial_d(PIP_n^d|_A)) \geq n^d - 2^{2d+1}kn^{d-1}$ for any affine subspace A, of co-dimension k.*

Proof: Let B be any basis for A, $|B| \leq k$. Consider only order d derivatives of PIP_n^d with respect to x variables, one from each of the sets $X^{(i)}$. The set of the derivatives is the set of all monomials of degree d containing exactly one variable from each of the sets $Y^{(i)}$.

Since every y-variable appears in exactly n^{d-1} of these monomials, we have $dim(\partial_d PIP_n^d) \cap F[X \setminus B]) \geq n^d - kn^{d-1}$.

Similarly, since every monomial has at most 2^{2d} partial derivatives of order d, we also get $dim(\partial_d^B PIP_n^d) \leq 2^{2d}n^{d-1}$.

Combining these with Lemma 3.4 and Lemma 3.5 we get the bound. □

Proof of Theorem 4.8: It suffices to prove the bound for $d \leq \frac{1}{2}\sqrt{\log n}$, since for higher d the bound follows by a simple restriction to $PIP^n_{\frac{1}{2}\sqrt{\log n}}$. Use Lemmas 4.9 and Theorem 3.2 to get

$$s_3(PIP_n^d) \geq \min(k^2, \frac{n^d - 2^{2d+1}kn^{d-1}}{\binom{k}{d}}) .$$

If we take $k = n^{\frac{d}{d+2}}$ we get

$$s_3(PIP_n^d) \geq \min(n^{2\frac{d}{d+2}}, \frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{\binom{n^{\frac{d}{d+2}}}{d}}) .$$

The way to evaluate the second term inside the min is:

$$\frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{\binom{n^{\frac{d}{d+2}}}{d}} \geq \frac{n^d - 2^{2d+1}n^{d-\frac{2}{d+2}}}{n^{\frac{d^2}{d+2}}} \geq \frac{1}{6}(n^{2\frac{d}{d+2}})$$

for $d < (1/2)\sqrt{\log n}$. □

Remark 2 *We showed a lower bound for $\Sigma\Pi\Sigma$ circuits calculating $PIP_n^d(X,Y)$, but there is a trivial $\Pi\Sigma\Pi$ circuit for it of linear size. This is an unusual example where $\Pi\Sigma\Pi$ circuit is more efficient then $\Sigma\Pi\Sigma$ circuit, and provides a (near) quadratic gap between the two.*

Another separation we can deduce is the following separation between depth-4 circuits and depth-3 circuits:

Theorem 4.10

$$\forall d,\ L_4(PIP_n^d(X,Y) + PIP_n^d(W,Z)) = O(nd) ,$$

where $L_4(f)$ is the size of the smallest depth 4 formulae ($\Sigma\Pi\Sigma\Pi$ or $\Pi\Sigma\Pi\Sigma$) for f.

$$for\ \ d \leq \frac{1}{2}\sqrt{\log n},\ \ s_3(PIP_n^d(X,Y) + PIP_n^d(W,Z)) \geq \Omega(n^{2\frac{d}{d+2}}) .$$

4.3 (Trace of) Iterated matrix multiplication

Another function which has a similar lower bound, with a similar proof technique is the trace of product of matrices.

Definition 4.4

$$TR_n^{2d}(X^{(1)},...,X^{(2d)}) = \sum_{i_1,...,i_{2d}} x^{(1)}_{i_1,i_2} \cdot x^{(2)}_{i_2,i_3} \cdots x^{(2d)}_{i_{2d},i_1} .$$

This is the trace of the product of $2d$ $n \times n$ *matrices (each* $X^{(i)}$ *is an* $n \times n$ *matrix).*

Here we have $2dn^2$ variables and the lower bound depends on d and n, as in the previous example.

Theorem 4.11 $s_3(TR_n^{2d}) \geq \Omega(n^{4\frac{d}{d+2}})$ *for* $d \leq \frac{1}{2}\sqrt{\log n}$.
$s_3(TR_n^{2d}) \geq \Omega(n^{4-\frac{8}{\sqrt{\log n}}})$ *for* $d > \frac{1}{2}\sqrt{\log n}$.

The proof will go in the usual manner:

Lemma 4.12 $dim(\partial_d(TR_n^{2d}|_A)) \geq n^{2d} - 2^{2d+1}kn^{2d-2}$ *for any affine subspace* A, *of co-dimension* k.

Proof: The proof is very similar to that of Lemma 4.9. Let B be any basis for A, $|B| \leq k$. Consider only order d derivatives of of TR_n^{2d} with respect to "even" sets $X^{(2i)}$ (one variable from each). These derivatives form the set of all monomials in the "odd" sets $X^{(2i-1)}$ (one variable from each).

Since every variable appears in exactly n^{2d-2} of these monomials, we have

$$dim(\partial_d TR_n^{2d}) \cap F[X \setminus B]) \geq n^{2d} - kn^{2d-2} .$$

Again every monomial has at most 2^{2d} partial derivatives of order d so we also get $dim(\partial_d^B TR_n^{2d}) \leq 2^{2d}n^{2d-2}$.

Combining these with Lemma 3.4 and Lemma 3.5 we get the bound

$$dim(\partial_d(TR_n^{2d}|_A)) \geq (n^{2d} - kn^{2d-2}) - n^{2d-2}2^{2d} \geq n^{2d} - kn^{2d-2}2^{2d+1} .$$

□

Proof of Theorem 4.11: Again, it suffices to prove the bound for $d \leq \frac{1}{2}\sqrt{\log n}$. We use Lemmas 4.12 and Theorem 3.2 to get

$$s_3(TR_n^{2d}) \geq \min(k^2, \frac{n^{2d} - 2^{2d+1}kn^{2d-2}}{\binom{k}{d}}) .$$

Using similar estimates (and replacing n with n^2) in the proof of Theorem 4.8 we get

$$s_3(TR_n^d) \geq \min(n^{4\frac{d}{d+2}}, \Omega(n^{4\frac{d}{d+2}})) ,$$

where k equals $n^{2\frac{d}{d+2}}$. □

Although we can only prove near quadratic lower bounds on depth 3 circuits computing TR_n^{2d} we suspect that any $\Sigma\Pi\Sigma$ circuit for it has size n^{2d} (and there is a trivial $\Sigma\Pi\Sigma$ circuit of size n^{2d} computing TR_n^{2d}).

4.4 Determinant

The last lower bound that we show is for the determinant function. We show an almost n^4 lower bound, where the number of variables is n^2. Note that it is a very weak lower bound compared to what is known over finite fields. Getting a super-polynomial lower bound for this function seems to be the main next challenge[1].

Definition 4.5

$$DET_n(X) = \sum_{\sigma \in S_n} sign(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)} \ ,$$

where S_n is the symmetric group on n variables.

Here X is $n \times n$ matrix.

Theorem 4.13 $s_3(DET_n(X)) \geq \Omega(n^4/\log n)$.

Denote by $s_3^m(f)$ the size of the smallest depth-3 circuit computing f in which the degree of every multiplication gate is at most m. We have a stronger lower bound for the case where $m = o(\frac{n^2}{\log n})$.

Theorem 4.14 $s_3^m(DET_n(X)) \geq \Omega(e^{n^2/me})$.

Proof of Theorem 4.14: This is an immediate corollary of Proposition 2.4 and the fact that $\partial_d(DET_n) = \binom{n}{d}^2$. These facts give:

$$s_3^m(Det_n) \geq \frac{\binom{n}{d}^2}{\binom{m}{d}} \approx (\frac{n^2}{md})^d.$$

Maximizing d we get $d = \frac{n^2}{me}$ and the lower bound follows. □

Using tools like Theorems 3.2, 3.1 yields a $\frac{n^3}{\log n}$ lower bound. But a straight forward induction together with Theorem 4.14 yields a $\frac{n^4}{\log n}$ lower bound. The following simple observation is needed for the induction:

Proposition 4.15 *If $\mathcal{C}$ is a $\Sigma\Pi\Sigma$ circuit that computes $f(X)$, where $f(X)$ is a linear function in $x \in X$, i.e $f(X) = xg(X \setminus \{x\}) + h(X \setminus \{x\})$, and x appears in N multiplication gates, each of degree at most m, then there is a $\Sigma\Pi\Sigma$ circuit that computes $g(X \setminus \{x\})$ with $2N$ multiplication gates each of degree at most m.*

Proof: Since $g(X \setminus \{x\}) = f(x+1, X \setminus \{x\}) - f(x, X \setminus \{x\})$, we can construct a new circuit from $\mathcal{C}$ in the obvious manner, and all the gates not including x will be canceled. □

Proof of Theorem 4.13: The proof is by induction on n. We begin with $x_{1,n}$, if all the multiplication gates including it are of degree $< \frac{n^2}{4e \log n}$ then we make the following assignment:

$$x_{i,n} = 0 \ \ \forall i > 1.$$

[1] In view of [12], getting a super-linear lower bound for the symmetric model (see there) seems the right challenge.

Using Proposition 4.15 we get a new circuit of degree $< \frac{n^2}{4e\log n}$ for DET_{n-1} and according to Theorem 4.14 there are at least $\frac{1}{2}n^4$ multiplication gates in it. Otherwise we can find a multiplication gate of degree $\geq \frac{n^2}{4e\log n}$ containing $x_{1,n}$ and a linear function $x_{1,n} = \ell_n(X \setminus \{x_{1,n}\})$ that nullifies it. We do this for $x_{1,n}...x_{1,2}$. If at any stage there is no multiplication gate of degree $\geq \frac{n^2}{4e\log n}$ involving the variable we look at, then we make the appropriate assignment (at the $k'th$ stage we put $x_{i,n-k+1} = 0 \quad \forall i > 1$). After doing so for $x_{1,2}$ we put $x_{i,1} = 0 \quad \forall i > 1, \quad x_{1,1} = 1$. Thus our restricted circuit computes DET_{n-1}. Therefore we get the recursion:

$$s_3(DET_n) \geq min(s_3(DET_{n-1}) + (n-1)\frac{n^2}{4e\log n}, \frac{n^4}{4})$$

Solving we get $s_3(DET_n) \geq \Omega(\frac{n^4}{\log n})$. □

4.4.1 Restricting the fan-in of the gates on level 1

Since it is hard to prove lower bounds even for such a restricted model, we will restrict the model even further, we will allow the linear functions that are computed at the first level to consist of only one variable, i.e every function will have the form:

$$\ell = \alpha_i x_i + \ell^0$$

The reason to consider this model is that it is strong enough for the construction of [2] for the elementary symmetric polynomials. For this model we can prove an exponential lower bound for DET_n.

Theorem 4.16 *Every restricted depth-3 circuit that computes DET_n must have at least $\Omega(\frac{2^n}{n})$ edges.*

Proof: Let's write each multiplication gate in the form:

$$M = \prod_{i=1}^{n}\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$$

But, in DET_n no two variables from the same column appear in the same monomial, therefore from each of the gates we only need to collect the multilinear monomials in which there are no two variables from the same column. Therefore from each of the terms: $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$, only the part of degree 1 can be used to create a monomial that will appear in the result. So we replace each $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j} + \alpha_{i,j}^0)$ with its degree one monomial: $\sum_{j=1}^{n}\beta_{i,j}x_{i,j}$. Now each multiplication gate looks like

$$M = \prod_{i=1}^{n}(\sum_{j=1}^{n}\beta_{i,j}x_{i,j})$$

The rest of the proof is similar to the proof of Theorem 2.5 from [10]. By Lemma 2.2 for each multiplication gate M, $dim(\partial_{\frac{n}{2}}(M)) \leq 2^n$, and we know that $dim(\partial_{\frac{n}{2}}(DET_n)) \geq \binom{n}{\frac{n}{2}}^2$ therefore Lemma 2.2 gives us a lower bound on the number of multiplication gate of $\frac{\binom{n}{\frac{n}{2}}^2}{2^n} \approx \frac{2^n}{n}$. □

Remark 3 *Notice that replacing each* $\prod_{j=1}^{n}(\alpha_{i,j}x_{i,j}+\alpha_{i,j}^{0})$ *with* $\sum_{j=1}^{n}\beta_{i,j}x_{i,j}$ *may change the number of addition gates in the bottom level, but it doesn't raise the degree of each multiplication gates! Since we count the number of multiplication gates and the sum of their degrees, this replacement doesn't change the result.*

5 Upper bounds for the Elementary symmetric function

Ben-Or showed the following construction for Depth-3 circuits:

Theorem 5.1 (Ben-Or) *For every* d, $L_3(S_n^d) \leq O(n^2)$.

Proof: S_n^d is the coefficient of t^{n-d} in the polynomial $\prod_{i=1}^{n}(t+x_i)$. Using interpolating at n+1 distinct points we can get this coefficient. Notice that evaluation at a point can be computed by a single product gate of degree n, and since we have n+1 points of interpolation the circuit is of size $O(n^2)$. □

In contrast, allowing depth-6 or even depth-4, the construction of Ben-Or can be greatly improved for small d. With our lower bound above for $d = \log n$, it provides a near quadratic separation between depth-3 and depth-4 circuits.

Theorem 5.2 *For every* d, $L_4(S_n^d) \leq n2^{O(\sqrt{d})}$.

Theorem 5.3 *For every* d, $L_6(S_n^d) \leq O(nd^3 \log d)$.

Proof of Theorem 5.3: Let $T_n^d(X) = \sum_{i=1}^{n} x_i^d$. The well known Newton Identities provide a polynomial relation expressing each S_n^d in terms of $\{T_n^k : k \leq d\}$. More precisely, let the degree d truncation of the series for e^y be the polynomial $E^d(y) = \sum_{j=0}^{d} y^j/j!$. Then $S_n^d(X)$ is the coefficient of z^d in the polynomial

$$\Pi_{k=1}^{d} E^{\lfloor d/k \rfloor}((-1)^{k-1}T_n^k z^k)$$

Clearly, this is a $\Pi\Sigma$ circuit of size $d \log d$ in the variables T_n^k, each of which is a $\Pi\Sigma\Pi$ circuit of size at most nd in the original X variables. Finally, interpolation over the variable z (as in Ben-Or's construction) requires $d+1$ values for this variable and gives a $\Sigma\Pi\Sigma\Pi\Sigma\Pi$ circuit of size at most $nd^3 \log d$ for S_n^d. □

Proof of Theorem 5.2: We proceed in the same manner to get:

$$\prod_{k=1}^{d} E^{\lfloor d/k \rfloor}((-1)^{k-1}T_n^k z^k) =$$

$$\prod_{k=1}^{d} \sum_{j=0}^{\lfloor d/k \rfloor} \frac{(-1)^{j(k-1)}}{j!}(T_n^k)^j z^{kj}.$$

To get the coefficient of z^d we sum over all the partitions of d:

$$\sum_{\sum_{k=1}^{d} kj_k = d} \prod_{k=1}^{d} \frac{(-1)^{j_k(k-1)}}{j!}(T_n^k)^{j_k}.$$

Since each $(T_n^k)^{j_k}$ has a $\Pi\Sigma\Pi$ formula of size at most nkj_k, and there are at most $2^{O(\sqrt{d})}$ partitions of d ([15] Theorem 15.7), we get a formula of size at most

$$\sum_{\sum_{k=1}^{d} kj_k = d} \sum_{k=1}^{d} nkj_k = 2^{O(\sqrt{d})} nd.$$

□

6 Strange models

In this section we assume that the field is either $\mathcal{R}$ or $\mathcal{C}$.

6.1 A Nečhiporuk-like lower bound

We will now define a different kind of formula, for which we will prove quadratic lower bounds. Our proof uses similar ideas to those of [8]. In his paper Nečhiporuk proves lower bounds on boolean formula by considering the number of sub functions that a certain function has (a sub function is a the function resulting from restricting some of the variables). In our proof we consider the discriminant polynomial and it's sub functions resulting from restricting all but one variable.

Definition 6.1 *A PolyFormula in the set of variables X over a field F is a binary tree whose internal nodes are labeled by addition or multiplication gate, its inputs are labeled by polynomials in one of the variables of X or by constants from the field. The size of this formula is the number of input nodes that are labeled with polynomials in the variables. We denote the smallest size of a PolyFormula that computes a polynomial f with $\hat{L}(f)$.*

Notice the difference between the definition of PolyFormula and the definition of a Formula where we allow edges to be labeled with constants.

Claim 6.1 *If we change the definition of a Formula to a definition where we do not allow constants on edges then this variant changes only affects the number of inputs labeled with constants.*

Proof: We can push constants down towards the inputs until we are left with constants only on edges leaving inputs. Any constant on such edge can be replaced by an input gate labeled with it which fans out to a multiplication gate that will multiply this constant with the appropriate input. □

Since we count only the number of polynomials in x in the definition of PolyFormula, one may assume that there might be many constants too, but the following lemma bounds the number of constants in terms of $\hat{L}$.

Lemma 6.2 *We can always assume that the number of constants in a PolyFormula is at most 3 times the number of inputs labeled with polynomials in X.*

Proof: We can change the PolyFormula (without changing its size or output) such that in the new PolyFormula each constant only enters a plus or a multiplication gate whose other input is an X-input, or it enters a plus gate that adds it to a multiplication gate of two X-polynomials. Otherwise we could delete this constant and change the other constants (we are just throwing away constant sub-formulas). After making this change we can easily prove the claim by an induction on the formula size. □

Definition 6.2

$$Disc(X) = \prod_{i<j}(x_j - x_i)$$

X is the set of variables $x_1, ..., x_n$. This is the discriminant polynomial and it is the determinant of the Vandermonde matrix ($a_{i,j} = {x_i}^{j-1}$).

Theorem 6.3 $\hat{L}(Disc_n) \geq (1/3)n(n-1)$.

Definition 6.3 *A $d-$scheme is a PolyFormula in one variable, x, with a set of constants $C = \{c_1, ..., c_k\}$, so if we run over all the values for C in F^k we get all deg d monic[2] polynomials in x.*
We denote by $s(d)$ the minimal size of a PolyFormula that is a $d-$scheme.

A trivial upper bound for $s(d)$ is d.

Theorem 6.4 *If F is algebraically closed then $s(3d) \leq 2d$.*

Proof: Let's look at the following PolyFormula:

$$(x^2 + b)(x + a) + (c - ab) = x^3 + ax^2 + bx + c.$$

Thus using only two polynomials in x we can generate every monic deg 3 polynomial. Since F is algebraically closed every monic polynomial of degree $3d$ is the multiplication of d deg 3 monic polynomials whose coefficients are in F. Therefore we can take d distinct copies of this formula, multiply all of them, to get a general deg $3d$ monic polynomial. □

Definition 6.4 *We say that a PolyFormula in one variable x, with a set of constants $C = \{c_1, ..., c_k\}$ is of dimension d if running over all possible assignments to C yields a manifold of polynomials in x of dimension d.*

An easy observation is:

Claim 6.5 *A PolyFormula of dimension d is a polynomial mapping from F^k to a manifold of dimension d, where k is the number of constants in the PolyFormula.*

Since every polynomial mapping is a C^∞ mapping and F^k is a manifold of dimension k, we get by Theorem 6.6 (see [7]) that the dimension of a PolyFormula with k constants is at most k.

Theorem 6.6 *If M,N are manifolds of dimensions m, n respectively, where $m < n$ and $\Phi : M \to N$ is a C^1 mapping, then $\Phi(M)$ has measure 0 (Provided M has only countably many components).*

Together with Lemma 6.2 we get the following corollary.

Corollary 6.1 $\hat{L} \geq k/3$ *for every PolyFormula of dimension k.*

We can now deduce the following lower bound.

Theorem 6.7 $s(d) \geq \frac{d}{3}$.

[2] A monic polynomial is a polynomial whose leading coefficient equals 1.

Proof: Follows from Definition 6.1, Corollary 6.1 and the fact that every d-scheme is a PolyFormula of dimension d. □

We now prove our main theorem.

Proof of Theorem 6.3: Let's view a PolyFormula for $Disc(x_1, ..., x_n)$ as a PolyFormula in x_n over the field $F(x_1, ..., x_{n-1})$ and remove all unnecessary gates. Clearly it is a PolyFormula of dimension $n-1$ (when we run over all substitutions of elements from F to $x_1, ..., x_{n-1}$). According to Corollary 6.1 the number of input gates labeled with polynomials in x_n is at least $\frac{n-1}{3}$. Since we could do it for every x_i we get $\hat{L}(Disc_n) \geq (1/3)n(n-1)$. □

6.2 The derived homogeneous formula

Theorem 6.8 (Observation) *If $f(X)$ is an homogeneous polynomial computed by a $\Sigma\Pi\Sigma$ circuit of size s then there's an m such that $g(X,t) = t^m f(X)$ can be computed by an* **homogeneous** *$\Sigma\Pi\Sigma$ circuit of size $\leq s^2$.*

Proof: Let d be the highest degree of all multiplication gates in the $\Sigma\Pi\Sigma$ circuit computing f. We will now change the circuit in the following way:

- Replace each linear function $\ell = \sum_{i=1}^{n} \alpha_i x_i + \ell^h$ with $\hat{\ell} = \sum_{i=1}^{n} \alpha_i x_i + \ell^h t$.
- Multiply each multiplication gate of degree $d-k$ with t^k.

Each monomial of degree k in the former circuit is now multiplied by t^{d-k}, since f is homogeneous the result will be $t^{d-deg(f)} f$. □

Definition 6.5 *Let f be an homogeneous polynomial, the derived homogeneous circuit of f is a homogeneous circuit of smallest size that computes $t^m f$ for some m. We call it $\mathcal{F}_D(f)$ and we denote by $s_3^D(f)$ the size of a smallest $\mathcal{F}_D(f)$ for f.*

Corollary 6.2 *For S_n^{2d}*
1. $s_3^D(S_n^{2d}) \leq n^2$
2. $s_3^H(S_n^{2d}) \geq \binom{n}{d}^2 / 2^{2d}$

Proof: According to Theorem 5.1 there is a $\Sigma\Pi\Sigma$ circuit of size n^2 that computes S_n^{2d}, therefore according to Theorem 6.8 there is an homogeneous $\Sigma\Pi\Sigma$ circuit of size n^2 (the size remains the same under this transformation) that computes $t^{n-2d} S_n^{2d}$.

In [10] it was proved that each homogeneous $\Sigma\Pi\Sigma$ circuit that computes S_n^{2d} requires size $\geq \binom{n}{d}^2 / 2^{2d}$. □

Corollary 6.3 *There is an exponential gap between computing homogeneously an homogeneous polynomial f, and computing homogeneously $t^m f$ for some m.*

Corollary 6.4 *Taking derivatives of an homogeneous polynomial f with respect to a single variable or assigning a fixed value (say 1) to a variable may exponentiate the size of the depth-3 homogeneous circuit computing it.*

The corollary above is an unconditional result (for a weaker model) of a phenomenon which was observed before. In [14] it was shown that assuming PERM is hard, taking multiple derivatives w.r.t. many variables may exponentiate size of general circuits. One way of seeing it is by looking at the following restriction of $PIP(X,Y)$:

Definition 6.6 $PIP(X,\vec{y}) = \prod_{i=1}^{n} \sum_{j=1}^{n} x_{i,j} y_j$.

Now we get $PERM(X) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} PIP(X,\vec{y})$.

7 Conclusion

To conclude we will represent some related open problems.

7.1 Depth-3 reductions

While quite basic and powerful, depth-3 circuits seem too weak to perform reductions between interesting functions. Most known reductions when depth is not restricted do not seem to carry through when depth is restricted.

The first problem concerns self reductions of elementary symmetric functions. We have proved a lower bound for the $\log n$ elementary symmetric function. The proof doesn't work for S_n^d where $d \geq \log n$, and so we have no lower bound for higher d. In the Boolean setting there is a trivial reduction from smaller to higher values of d (setting variables to 1). Is there an algebraic analog?

Problem 7.1 *Does* $s_3(S_n^d) \leq s_3(S_n^r)$ *for* $d < r$*?*

For polynomially related functions we define below a natural notion of reduction, and list the most frustrating problems of this type.

Definition 7.1 *For two polynomials* f, g *we say that* $f \leq_3 g$ *if* $s_3(f) \leq s_3(g)^{O(1)}$.

A very interesting and nontrivial reduction for circuits without depth restriction is between the Determinant and Iterated Matrix Multiplication (see [3]). Again, we do not know if it extends to the bounded depth case.

Problem 7.2 *Does* $DET \leq_3 IMM$ *?*

Another curiosity is with the Permanent function.

Definition 7.2

$$PERM_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)}$$

where X *is an* $n \times n$ *matrix and* S_n *is the symmetric group on* n *variables.*

PERM is known to be $\#P$ complete, and so we expect it to be harder than DET in any model. Surprisingly we have a more efficient depth-3 circuit for PERM than for DET.
The following theorem is due to it Ryser, see e.g [15].

Theorem 7.1 (Ryser) *There is an homogeneous* $\Sigma\Pi\Sigma$ *circuit for* $PERM(X)$ *of size* $O(n2^n)$.

Proof: $PERM(X) = \sum_{T \subset [n]} (-1)^{|T|} \prod_{i=1}^{n} \sum_{j \in T} x_{i,j}$. □

Problem 7.3 *Does* $DET \leq_3 PERM$*, or maybe* $PERM \leq_3 DET$*?*

Problem 7.4 *Find a simply exponential (*$2^{O(n)}$ *size)* $\Sigma\Pi\Sigma$ *circuit for DET.*

7.2 Strange models

In Theorem 4.16 we showed a lower bound for depth-3 circuits which use linear functions of only one variable. What happens if we allow linear functions of two variables ?

Problem 7.5 *Show an exponential lower bound for $\Sigma\Pi\Sigma$ circuits of the form:*

$$\sum_{j=1}^{m} \prod_{i=1}^{deg(M_j)} (\alpha_{i,j,1} x_{k_1} + \alpha_{i,j,2} x_{k_2} + c_{i,j})$$

.

8 Acknowledgments

We would like to thank Michael Ben-Or, Dima Grigoriev and Alexander A. Razborov for helpful discussions during various stages of this work.

References

[1] W. Baur and V. Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[2] M. Ben-Or. Private communication.

[3] J. v. z. Gathen. Algebraic complexity theory. *Ann. Rev. Computer Science*, pages 317–347, 1988.

[4] D. H. Gottlieb. A certain class of incidence matrices. In *Amer. Math. Soc.*, volume 17, pages 1233–1237, 1966.

[5] D. Grigoriev and M. Karpinski. An approximation algorithm for the number of zeros or arbitrary polynomials over GF(q). In IEEE, editor, *Proceedings: 32nd annual Symposium on Foundations of Computer Science, San Juan, Puerto Rico, October 1–4, 1991*, pages 662–669, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1991. IEEE Computer Society Press.

[6] D. Grigoriev and A. A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. In IEEE, editor, *39th Annual Symposium on Foundations of Computer Science: proceedings: November 8–11, 1998, Palo Alto, California*, pages 269–278, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1998. IEEE Computer Society Press.

[7] J. W. Milnor. *Topology from the differentiable viewpoint.* Princeton Univ. Pr., 1997.

[8] E. I. Nečhiporuk. A boolean function. *Sov. Math. Dokl.*, 7(4):999–1000, 1966.

[9] N. Nisan. Lower bounds for non-commutative computation. In B. Awerbuch, editor, *Proceedings of the 23rd Annual ACM Symposium on the Theory of Computing*, pages 410–418, New Orleans, LS, May 1991. ACM Press.

[10] N. Nisan and A. Wigderson. Lower bound on arithmetic circuits via partial derivatives. *Computational Complexity*, 6:217–234, 1996.

[11] A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Math. Notes*, 41:333–338, 1987.

[12] A. Shpilka. Affine projections of symmetric polynomials. In *16th Annual Conference on Computational Complexity*, 2001. to appear.

[13] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 77–82, New York City, 25–27 May 1987.

[14] V. Strassen:. *Handbook of Th. Comp. Science*, volume A, chapter Algebraic complexity theory, pages 633–672. Elsevier and MIT Press, 1990.

[15] J. H. van Lint and R. M. Wilson. *A course in combinatorics.* Cambridge Univ Pr, 1992.