“Resistant” Polynomials and Stronger Lower Bounds for Depth-Three Arithmetical Formulas

Maurice J. Jansen
University at Buffalo

Kenneth W.Regan*
University at Buffalo

Abstract

We derive quadratic lower bounds on the $*$-complexity of sum-of-products-of-sums ($\Sigma\Pi\Sigma$) formulas for classes of polynomials f that have too few partial derivatives for the (near-)quadratic lower bound techniques of Shpilka and Wigderson [SW99, Shp01] (after [NW96]). Our techniques introduce a notion of "resistance" which connotes full-degree behavior of f under any projection to an affine space of sufficiently high dimension. They also show stronger lower bounds over the reals than the complex numbers or over arbitrary fields. Separately, by applying a special form of the Baur-Strassen Derivative Lemma tailored to $\Sigma\Pi\Sigma$ formulas, we obtain sharper bounds on $+,*$-complexity than those shown for $*$-complexity in [SW99], most notably for the lowest-degree cases of the polynomials they consider.

1 Introduction

In contrast to the exponential size lower bounds on constant-depth Boolean circuits for majority and related functions [FSS81, Yao85, Hås86], Shpilka and Wigderson [SW99] observed that in *arithmetic complexity*, over fields of characteristic zero, super-quadratic lower bounds are not even known for constant-depth *formulas*. Indeed they are unknown for unbounded fan-in, depth 3 formulas that are sums of products of affine linear functions, which they call $\Sigma\Pi\Sigma$ formulas. These formulas have notable *upper-bound power* because they can carry out forms of Lagrange interpolation. As they ascribed to M. Ben-Or, $\Sigma\Pi\Sigma$ formulas can compute the elementary symmetric polynomials S_n^k (defined as the sum of all degree-k monomials in n variables, and analogous to majority and threshold-k Boolean functions) in size $O(n^2)$ independent of k. Thus $\Sigma\Pi\Sigma$ formulas present a substantial challenge for lower bounds, as well as being a nice small-scale model to study.

Shpilka and Wigderson defined the *multiplicative size* of an arithmetical (circuit or) formula ϕ to be the total fan-in to multiplication gates. We denote this by $\ell^*(\phi)$, and write $\ell(\phi)$ for the total fan-in to all gates, i.e. $+$ gates as well. The best known lower bound for general arithmetical circuits has remained for thirty years the $\Omega(n \log n)$ lower bound on ℓ^* by the "Degree Method" of Strassen [Str73] (see also [BS82, BCS97]). However, this comes nowhere near the exponential lower bounds conjectured by [Val79] for the permanent and expected by many for other NP-hard arithmetical functions. For polynomials f of total degree $n^{O(1)}$, the method is not even capable of $\Omega(n^{1+\epsilon})$ circuit lower bounds, not for any $\epsilon > 0$. Hence it is notable that [SW99] achieved lower bounds on $\ell_3^*(f)$, where the subscript-3 refers to $\Sigma\Pi\Sigma$ formulas, of $\Omega(n^2)$ for $f = S_n^k$ when $k = \Theta(n)$, $n^{2-\epsilon_k}$ for S_n^k with small values of k, and $\Omega(N^2/\operatorname{polylog}(N))$ for the determinant, with $N = n^2$. However,

*Part of this work by both authors was supported by NSF Grant CCR-9821040. Corresponding author: Dept. of CSE at UB, 201 Bell Hall, Buffalo, NY 14260-2000; (716) 645-3180 x114, fax 645-3464; regan@cse.buffalo.edu.

their methods have a similar limitation of $\Omega(n^2)$ for $\Sigma\Pi\Sigma$ formulas. Shpilka [Shp01] got past this only in some further-restricted cases, and also considers a depth-2 model consisting of an arbitrary symmetric function of sums. This barrier provides another reason to study the $\Sigma\Pi\Sigma$ model, in order to understand the obstacles and what might be needed to surpass them.

The techniques in [NW96, SW99, Shp01] all depend on the set of dth-order partial derivatives of f being large. This condition fails for functions such as $f(x_1, \ldots, x_n) = x_1^n + \ldots + x_n^n$, which has only n dth-order partials for any d. We refine the analysis to show the sufficiency of f behaving like a degree-r polynomial on any affine subspace A of sufficiently high dimension (for this f, $r = n$ and any affine line suffices). Our technical condition is that for every polynomial g of total degree at most $r-1$ and every such A, *there exists* a d-th order partial of $f - g$ that is non-constant on A. This enables us to prove an absolutely sharp n^2 bound on $\ell_3^*(f)$ for this f computed over the real or rational numbers, and a lower bound of $n^2/2$ over any field of characteristic zero. Note the absence of "O, Ω" notation. We prove similar tight bounds for sums of powered monomial blocks, powers of inner-products, and functions depending on ℓ_p-norm distance from the origin, and also replicate the bounds of [SW99, Shp01] for symmetric polynomials. Even in the last case, we give an example where our simple existential condition may work deeper than the main question highlighted in [Shp01] on the maximum dimension of subspaces A on which S_n^k vanishes.

In the second half, we prove lower bounds on $+,*$ complexity $\ell_3(f)$ that are significantly higher (but still sub-quadratic) than those given for $\ell_3^*(f)$ in [SW99] when the degree r of f is small. This is done intuitively by exploiting a closed-form application of the Baur-Strassen "Derivative Lemma" [BS82] to $\Sigma\Pi\Sigma$ formulas, showing that f and all of its n first partial derivatives can be computed with only a constant-factor increase in ℓ and ℓ^* over $\Sigma\Pi\Sigma$ formulas for f.

2 Preliminaries

A $\Sigma\Pi\Sigma$-formula is an arithmetic formula consisting of four consecutive layers: a layer of inputs, next a layer of addition gates, then a layer of multiplication gates, and finally the output sum gate. The gates have unbounded fan-in from the previous layer (only), and individual wires may carry arbitrary constants from the underlying field. Given a $\Sigma\Pi\Sigma$-formula for a polynomial p, we can write

$$\begin{aligned} p &= \sum_{i=1}^{s} M_i, \qquad \text{where} \qquad M_i = \Pi_{j=1}^{d_i} l_{i,j} \qquad \text{and} \\ l_{i,j} &= c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}. \end{aligned}$$

Here d_i is the in-degree of the ith multiplication gate (fix any order on the multiplication gates), and $c_{i,j,k}$ is nonzero iff there is a wire from x_k to the addition gate computing $l_{i,j}$. Note that $l_{i,j}$ is homogeneous of degree 1, i.e. *strictly linear*, if $c_{i,j,0} = 0$, and is *affine linear* otherwise.

2.1 Affine Linear Subspaces and Derivatives

An *affine linear* subspace A of F^n is a set of the form $A = V + w = \{ v + w : v \in V \}$, where V is a linear subspace of F^n, and w is a vector in F^n. The dimension of A is defined to be the vector space dimension of V.

Let $X = (x_1, \ldots, x_n)$ be an n-tuple of variables. For any affine subspace A, we can always find a set of variables $B \subset X$, and affine linear forms l_b in the variables $X \setminus B$, for each $b \in B$, such that A is the set of solutions of $\{x_b = l_b : b \in B\}$. This representation is not unique. The set B is called a *base* of A. The size $|B|$ always equals the co-dimension of A.

In the following, whenever we consider an affine linear subspace A, we assume we have fixed some base B of A. Any of our numerical "progress measures" used to prove lower bounds will not depend on the choice of a base. The following notion *does* depend on the choice of a base:

Definition 2.1 ([SW99]). Let A be an affine linear subspace of F^n, and let $f \in F[x_1, \ldots, x_n]$. Then the *restriction of f to A* is the polynomial obtained from f by substituting l_b for the variable x_b for each $b \in B$, is denoted by $f_{|A}$. If W is a set of polynomials, define $W_{|A} = \{f_{|A} \mid f \in W\}$.

For linear polynomials $l = c_1x_1 + \ldots + c_nx_n + c_0$, we denote $l^h = c_1x_1 + \ldots + c_nx_n$. For a set S of linear polynomials, $S^h = \{l^h : l \in S\}$. Observe that if S^h is an independent set, then the set of common zeroes of S is affine linear of dimension $n - |S|$.

3 Resistance of polynomials

We state our new definition in the weakest and simplest form that suffices for the lower bounds, although the functions in our applications all meet the stronger condition of Lemma 3.3 below.

Definition 3.1. A polynomial f in variables $x_1, x_2, \ldots, x_n$ is (d, r, k)-*resistant* if for any polynomial $g(x_1, x_2, \ldots, x_n)$ of degree at most $r - 1$, for any affine linear subspace A of co-dimension k, there exists a dth order partial derivative of $f - g$ that is non-constant on A.

For a multiset X of size d with elements taken from $\{x_1, x_2, \ldots, x_n\}$, we will use the notation $\frac{\partial^d f}{\partial X}$ to indicate the dth-order derivative with respect to the variables in X. As our applications all have $r = \deg(f)$, we call f simply (d, k)-*resistant* in this case. Then the case $d = 0$ says that f itself has full degree on any affine A of co-dimension k, and in most cases corresponds to the nonvanishing condition in [SW99]. We separate our notion from [SW99] in applications and notably in the important case of the elementary symmetric polynomials in Section 3.2 below.

The conclusion of Definition 3.1 is not equivalent to saying that some $(d+1)$st-order partial of $f - g$ is non-vanishing on A, because the restriction of this partial on A need not be the same as a first-partial of the restriction of the dth-order partial to A. Moreover, (d, k)-resistance need not imply $(d-1, k)$-resistance, even for $d, k = 1$: consider $f(x, y) = xy$ and A defined by $x = 0$. We have the following theorem:

Theorem 3.1 *Suppose $f(x_1, x_2, \ldots, x_n)$ is (d, r, k)-resistant, then*

$$\ell_3^*(f) \geq r\frac{k+1}{d+1}.$$

Proof. Consider a $\Sigma\Pi\Sigma$-formula that computes f. Remove all multiplication gates that have degree at most $r - 1$. Doing so we obtain a $\Sigma\Pi\Sigma$ formula $\mathcal{F}$ computing $f - g$, where g is some polynomial of degree at most $r - 1$. Say $\mathcal{F}$ has s multiplication gates. Write:

$$\begin{aligned} f - g &= \sum_{i=1}^{s} M_i, \qquad \text{where} \qquad M_i = \Pi_{j=1}^{d_i} l_{i,j} \qquad \text{and} \\ l_{i,j} &= c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}. \end{aligned}$$

The degree of each multiplication gate in $\mathcal{F}$ is at least r, i.e. $d_i \geq r$, for each $1 \leq i \leq s$. Now select a set S of input linear forms using the following algorithm:

$S = \emptyset$
for $i = 1$ to s **do**
  repeat $d+1$ times:
    if($\exists j \in \{1, 2, \ldots, d_i\}$) such that $S^h \cup \{l_{i,j}^h\}$ is a set of independent vectors **then**
      $S = S \cup \{l_{i,j}\}$

Let A be the set of common zeroes of the linear forms in S. Since S^h is an independent set, A is affine linear of co-dimension $|S| \leq (d+1)s$.

Claim 3.2 *If at a multiplication gate M_i we picked strictly fewer than $d+1$ linear forms, then any linear form that was not picked is constant on A.*

Proof. Each linear form l that was not picked had l^h already in the span of S^h, for the set S built up so far. Hence we can write $l = c + l^h = c + \sum_{g \in S} c_g g^h$, for certain scalars c_g. Since each g^h is constant on A, we conclude l is constant on A. □

We conclude that for each multiplication gate at least one of the following holds:

1. $(d+1)$ input linear forms vanish on A, or
2. fewer than $(d+1)$ linear form vanishes on A, and all others are constant on A.

For each multiset X of size d with elements from $\{ x_1, x_2, \ldots, x_n \}$, the dth order partial derivative

$$\frac{\partial^d (f-g)}{\partial X} \tag{1}$$

is in the linear span of the set

$$\{ \prod_{\substack{j=1 \\ j \notin J}}^{d_i} l_{ij} : \ 1 \leq i \leq s, \ J \subseteq \{1, 2, \ldots, d_i\}, \ |J| = d \ \}$$

This follows from the sum and product rules for derivatives and the fact that a first order derivative of an individual linear form l_{ij} is a constant. Consider $1 \leq i \leq s$ and $J \subseteq \{1, 2, \ldots, d_i\}$ with $|J| = d$. If item 1. holds for the multiplication gate M_i, then

$$\prod_{\substack{j=1 \\ j \notin J}}^{d_i} l_{ij} \tag{2}$$

vanishes on A, since there must be one l_{ij} that vanishes on A that was not selected, given that $|J| = d$. If item 2 holds for M_i, then (2) is constant on A.

Hence, we conclude that (1) is constant on A. Since f is (d, r, k)-resistant, we must have that the co-dimension of A is at least $k+1$. Hence $(d+1)s \geq k+1$. Since each gate in $\mathcal{F}$ is of degree at least r, we obtain

$$\ell_3^*(\mathcal{F}) \geq r \frac{k+1}{d+1}.$$

Since $\mathcal{F}$ was obtained by removing zero or more multiplication gates from a $\Sigma\Pi\Sigma$-formula computing f, we have proven the statement of the theorem. □

To prove lower bounds on resistance, we supply the following lemma that uses the syntactic notion of affine restriction. In certain cases this will be convenient.

Lemma 3.3 *Over fields of characteristic zero, for any $d \leq r$, $k > 0$, and any polynomial $f(x_1, x_2, \ldots, x_n)$, if for every affine linear subspace A of co-dimension k, there exists some dth order partial derivative of f such that*

$$\deg(\left(\frac{\partial^d f}{\partial X}\right)_{|A}) \geq r - d + 1$$

then f is $(d, r+1, k)$-resistant.

Proof. Assume for every affine linear subspace A of co-dimension k, there exists some dth order partial derivative derivative of f such that

$$\deg(\left(\frac{\partial^d f}{\partial X}\right)_{|A}) \geq r - d + 1.$$

Let g be an arbitrary polynomial of degree r. Then

$$\left(\frac{\partial^d f - g}{\partial X}\right)_{|A} = \left(\frac{\partial^d f}{\partial X} - \frac{\partial^d g}{\partial X}\right)_{|A} = \left(\frac{\partial^d f}{\partial X}\right)_{|A} - \left(\frac{\partial^d g}{\partial X}\right)_{|A}.$$

The term $\left(\frac{\partial^d f}{\partial X}\right)_{|A}$ has degree at least $r - d + 1$, whereas the term $\left(\frac{\partial^d g}{\partial X}\right)_{|A}$ can have degree at most $r - d$. Hence $\deg(\left(\frac{\partial^d f - g}{\partial X}\right)_{|A}) \geq r - d + 1 \geq 1$. Since over fields of characteristic zero, syntactically different polynomials define different mappings, we conclude $\frac{\partial^d f - g}{\partial X}$ must be non-constant on A. □

The main difference between Lemma 3.3 and the original Definition 3.1 appears to be the order of quantifying the polynomial "g" of degree $r - 1$ out front in the former, whereas analogous considerations in the lemma universally quantify it later (making a stronger condition). We have not found a neat way to exploit this difference in any prominent application, however.

3.1 Applications

We will now prove lower bounds on the $\Sigma\Pi\Sigma$-formula size of several explicit polynomials.

3.1.1 Sum of Nth Powers Polynomial

Consider $f = \sum_{i=1}^n x_i^n$. For this polynomial we have $\Sigma\Pi$-circuits of size $O(n \log n)$. This can be shown to be optimal using Strassen's degree method. By that method we know any circuit for f has size $\Omega(n \log n)$. The obvious $\Sigma\Pi\Sigma$-formula has additive size n^2 wires in the top linear layer, and has n multiplication gates of degree n. We prove that this is essentially optimal.

Theorem 3.4 *Over fields of characteristic zero, any $\Sigma\Pi\Sigma$-formula for $f = \sum_{i=1}^n x_i^n$ has multiplicative size at least $n^2/2$.*

Proof. We will show that f is $(1, n-1)$-resistant. The result then follows from Theorem 3.1. Let g be an arbitrary polynomial of degree $deg(f) - 1 = n - 1$. Letting $g_1, \ldots, g_n$ denote the first order partial derivatives of g, we get that the ith partial derivative of $f - g$ equals

$$n x_i^{n-1} - g_i(x_1, \ldots, x_n).$$

Note that the g_i's are of total degree at most $n-2$.

We claim there is no affine linear subspace of dimension greater than zero on which all $\partial f/\partial x_i$ are constant. Consider an arbitrary affine line, parameterized by a variable t:

$$x_i = c_i + d_i t,$$

where c_i and d_i are constants for all $i \in [n]$, and with at least one d_i nonzero. Then $\frac{\partial(f-g)}{\partial x_i}$ restricted to the line is given by

$$n(c_i + d_i t)^{n-1} - h_i(t),$$

for some univariate polynomials $h_i(t)$ of degree $\leq n-2$. Since there must exist *some* i such that d_i is nonzero, we know some partial derivative restricted to the affine line is parameterized by a univariate polynomial of degree $n-1$, and thus, given that the field is of characteristic zero, is not constant for all t.

□

In case the underlying field is the real numbers $\mathbf{R}$ and n is even, we can improve the above result to prove an absolutely tight n^2 lower bound. We start with the following lemma:

Lemma 3.5 *Let $f = \sum_{i=1}^n x_i^n$. Over the real numbers, if n is even, we have that for any affine linear subspace A of dimension $k \geq 1$, $deg(f_{|A}) = n$.*

Proof. Since f is symmetric we can assume without loss of generality that the following is a base representation of A:

$$\begin{aligned} x_{k+1} &= l_1(x_1, \ldots, x_k) \\ x_{k+2} &= l_2(x_1, \ldots, x_k) \\ &\vdots \\ x_n &= l_{n-k}(x_1, \ldots, x_k). \end{aligned}$$

Then

$$f_{|A} = x_1^n + \ldots x_k^n + l_1^n + \ldots + l_{n-k}^n.$$

We conclude that $f_{|A}$ must include the term x_1^n, since each l_j^n has a non-negative coefficient for the term x_1^n, since n is even. □

Theorem 3.6 *Over the real numbers, for even n, any $\Sigma\Pi\Sigma$-formula for $f = \sum_{i=1}^n x_i^n$ has multiplicative size at least n^2.*

Proof. Using Lemma's 3.3 and 3.5 we conclude that over the real numbers f is $(0, n-1)$-resistant. Hence, by Theorem 3.1 we get that $\ell_3^*(f) \geq \deg(f)\frac{n}{1} = n^2$. □

Let us note that $f = \sum_{i=1}^n x_i^n$ is an example of a polynomial that, even for large d, has relatively few, namely only n, partial derivatives. This makes application of the partial derivatives technique of [SW99], which we will describe and extend in the next section, problematic.

3.1.2 Blocks of Powers

Let the underlying field have characteristic zero, and suppose $n = m^2$ for some m. Consider the "m blocks of m powers" polynomial

$$f = \sum_{i=1}^{m} \prod_{j=(i-1)m+1}^{im} x_j^m.$$

The straightforward $\Sigma\Pi\Sigma$-formula for f, that computes each term/block using a multiplication gate of degree n, is of multiplicative size $n^{3/2}$. We will show this is tight.

Proposition 3.7 *The blocks of powers polynomial f defined above is $(0, m-1)$-resistant.*

Proof. Consider an affine linear space of co-dimension $m-1$. For any base B of A, restriction to A consists of substitution of the $m-1$ variables in B by linear forms in the remaining variables X/B. This means there is at least one term/block $B_i := \prod_{j=(i-1)m+1}^{im} x_j^m$ of f whose variables are disjoint from B. This block B_i remains the same under restriction to A. Also, for every other term/block there is at least one variable that is not assigned to. As a consequence, B_i cannot be canceled against terms resulting from restriction to A of other blocks. Hence $\deg(f_{|_A}) = \deg(f)$. Hence by Lemma 3.3 we have that f is $(0, m-1)$-resistant. $\square$

Corollary 3.8 *For the blocks of powers polynomial f defined above, $\ell_3^*(f) \geq nm = n^{3/2}$.*

Proof. Follows immediately from Theorem 3.1 and Proposition 3.7. $\square$

Alternatively, one can observe that by substitution of a variable y_i for each variable appearing in the ith block one obtains from a $\Sigma\Pi\Sigma$-formula $\mathcal{F}$ for f a formula for $f' = \sum_{i=1}^{m} y_i^n$ of the same size as $\mathcal{F}$. Theorem 3.4 generalizes to show that $\ell_3^*(f') \geq \frac{1}{2}n^{3/2}$, which implies $\ell_3^*(f) \geq \frac{1}{2}n^{3/2}$.

3.1.3 Polynomials depending on distance to the origin

Over the real numbers, $d_2(x) = x_1^2 + x_2^2 + \cdots + x_n^2$ is the square of the Euclidean distance of the point x to the origin. Polynomials f of the form $q(d_2(x))$ where q is a single-variable polynomial can be readily seen to have high resistance. Only the leading term of q matters.

For example, consider $f = (x_1^2 + x_2^2 + \cdots + x_n^2)^m$. On any affine line L in $\mathbf{R}^n$, $\deg(f_{|_L}) = 2m$. Therefore, by Lemma 3.3, over the reals, f is $(0, n-1)$-resistant. Hence by Theorem 3.1 we get that

Proposition 3.9 *Over the real numbers, $\ell_3^*((x_1^2 + x_2^2 + \cdots + x_n^2)^m) \geq 2mn$.*

Observe that by reduction this means that the "mth-power of an inner product polynomial", defined by $g = (x_1y_1 + x_2y_2 + \cdots + x_ny_n)^m$, must also have $\Sigma\Pi\Sigma$-size at least $2mn$ over the reals numbers. Results for l_p norms, $p \neq 2$, are similar.

3.2 Symmetric Polynomials

The special case of $(0, k)$-resistance implicitly appears in [Shp01], at least insofar as the sufficient condition of Lemma 3.3 is used for the special case $d = 0$ in which no derivatives are taken. For the elementary symmetric polynomial S_n^r of degree $r \geq 2$ in n variables, Theorem 4.3 of [Shp01] implies (via Lemma 3.3) that S_n^r is $(0, n - \frac{n+r}{2})$-resistant. Shpilka proves for $r \geq 2$, $\ell_3(S_n^r) = \Omega(r(n-r))$, which can be verified using Theorem 3.1: $\ell_3(S_n^r) \geq (r+1)(n - \frac{n+r}{2}) = \Omega(r(n-r))$. For $r = \Omega(n)$ this yields a tight $\Omega(n^2)$ bound as observed in [Shp01].

The symmetric polynomials S_n^k collectively have the "telescoping" property that every dth-order partial is (zero or) the symmetric polynomial S_{n-d}^{k-d} on an $(n-d)$-subset of the variables. Shpilka [Shp01] devolves the analysis into the question, "What is the maximum dimension of a linear subspace of $\mathbf{C}^n$ on which S_n^r vanishes?" In Shpilka's answer, divisibility properties of r come into play as is witnessed by Theorem 5.9 of [Shp01]. To give an example case of this theorem, one can check that S_9^2 vanishes on the 3-dimensional linear space given by

$$\{(x_1, \omega x_1, \omega^2 x_1, x_2, \omega x_2, \omega^2 x_2, x_3, \omega x_3, \omega^2 x_3) : x_1, x_2, x_3 \in \mathbf{C}\},$$

where ω can be selected to be any primitive 3rd root of unity. Let

$$\rho_0(f) = \max\{k : \text{ for any linear space } A \text{ of codimension } k,\ f_{|A} \neq 0\}$$

Shpilka proved for $r > n/2$, that $\rho_0(S_n^r) = n - r$, and for $r \geq 2$, that $\frac{n-r}{2} < \rho_0(S_n^r) \leq n - r$. For S_9^2 we see via divisibility properties of d that the value for ρ_0 can get less than the optimum value, although the $\frac{n-r}{2}$ lower bound suffices for obtaining the above mentioned $\ell_3(S_n^r) = \Omega(r(n-r))$ lower bound. We have some indication from computer runs using the polynomial algebra package *Singular* [GPS05] that the "unruly" behaviour seen for ρ_0 because of divisibility properties for $r \leq n/2$ can be made to go away by considering the following notion:

$$\rho_1(f) = \max\{k : \text{ for any linear space } A \text{ of codimension } k, \text{ there exists } i,\ \left(\frac{\partial f}{\partial x_i}\right)_{|A} \neq 0\}$$

One can still see from the fact that S_n^r is homogeneous and using Lemma 3.3 and Theorem 3.1 that $\ell_3^*(S_n^r) \geq \frac{r \cdot (\rho_1(S_n^r)+1)}{2}$. Establishing the exact value of $\rho_1(S_n^r)$, which we conjecture to be $n+1-r$ at least over the rationals, seems at least to simplify obtaining the $\ell_3(S_n^r) = \Omega(d(n-d))$ lower bound. In any case we can prove the following relation between the two notions:

Lemma 3.10 *For $r \geq 2$, $\rho_1(S_{n+1}^{r+1}) \geq \rho_0(S_n^{r-1})$.*

Proof. Suppose A is a linear space of codimension $k = \rho_1(S_{n+1}^{r+1}) + 1$ over $\{x_1, x_2, \ldots, x_{n+1}\}$ such that all partials of S_{n+1}^{r+1} vanish on A, i.e. for every i,

$$S_n^r(x_1, \ldots, \underline{x_i}, \ldots, x_{n+1})_{|A} = 0. \tag{3}$$

Here we denote $\underline{x_i}$ to indicate that the variable x_i is excluded. If $k = n+1$, then $\rho_1(S_{n+1}^{r+1}) = n$, so the statement of the lemma is trivial. Otherwise, some variable is not being substituted for by the restriction "$|_A$". By symmetry we can assume wlog. that this is x_{n+1}. Hence we can write Equation (3) as:

$$x_{n+1} S_{n-1}^{r-1}(x_1, \ldots, \underline{x_i}, \ldots, x_n)_{|A} + S_{n-1}^r(x_1, \ldots, \underline{x_i}, \ldots, x_n)_{|A} = 0,$$

for all i, and

$$S_n^r(x_1, \ldots, x_n)_{|A} = 0.$$

Adding the first n equations and subtracting the last $n-r$ times one obtains

$$\begin{aligned} 0 &= x_{n+1} \sum_{i=1}^{n} S_{n-1}^{r-1}(x_1, \ldots, \underline{x_i}, \ldots, x_n)_{|A} \\ &= (n-r) x_{n+1} S_n^{r-1}(x_1, \ldots, x_n)_{|A} \end{aligned}$$

In other words $S_n^{r-1}(x_1, \ldots, x_n)_{|A} = 0$. Simplify the substitution corresponding to A by taking $x_{n+1} = 0$ in the defining k equations. We obtain a set of k substitutions on k variables from $\{x_1, x_2, \ldots, x_n\}$. This defines a linear space of codimension k on which S_n^{r-1} vanishes. So $\rho_0(S_n^{r-1}) < k = \rho_1(S_{n+1}^{r+1}) + 1$. $\square$

For another example, S_6^3 is made to vanish at dimension 3 not by any subspace that zeroes out 3 co-ordinates but rather by $A = \{\, (u, -u, w, -w, y, -y) : u, w, y \in \mathbf{C} \,\}$. Now add a new variable t in defining $f = S_7^4$. The notable fact is that f 1-resists the dimension-3 subspace A' obtained by adjoining $t = 0$ to the equations for A, upon existentially choosing to derive by a variable other than t, such as u. All terms of $\partial f/\partial u$ that include t vanish, leaving 10 terms in the variables v, w, x, y, z. Of these, 4 pairs cancel under the equations $x = -w$, $z = -y$, but the leftover $vwx + vyz$ part equates to $uw^2 + uy^2$, which not only doesn't cancel but also dominates any contribution from the lower-degree g. Gröbner basis runs using *Singular* imply that S_7^4 is $(1,4)$-resistant over $\mathbf{C}$ as well as the rationals and reals, though we have not yet made this a consequence of a general resistance theorem for all S_n^r.

Hence our $(1,k)$-resistance analysis for S_7^4 is not impacted by the achieved upper bound of 3 represented by A. Admittedly the symmetric polynomials f have $O(n^2)$ upper bounds on $\ell_3(f)$, so our distinction in this case does not directly help surmount the quadratic barrier. But it does show promise of making progress in our algebraic understanding of polynomials in general.

4 Bounds for +,*-Complexity

The partial derivatives technique of [SW99] ignores the wires of the formula present in the first layer. In the following we show how to account for them. As a result we get a sharpening of several lower bounds, though not on ℓ_3^* but on total formula size. We refine the [SW99] result for *-complexity:

Theorem 4.1 ([SW99]) *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\dim(\partial_d(f)_{|A}) > D$. Then*

$$\ell_3^*(f) \geq \min(\frac{\kappa^2}{d}, \frac{D}{\binom{\kappa+d}{d}});$$

—to our result for +,*-complexity:

Theorem 4.2 (new) *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}] > D$. Then*

$$\ell_3(f) \geq \min(\frac{\kappa^2}{d+2}, \frac{D}{\binom{\kappa+d}{d}}).$$

Comparing the two theorems, we see that the result by Shpilka and Wigderson provides a lower bound on multiplicative complexity, while our result gives a lower bound on the *total* number of wires. We do get an extra "factor n" of additions with the $\sum_{i=1}^n \dim[\partial_d(\frac{\partial f}{\partial x_i})_{|A}] > D$ condition compared to just $\dim(\partial_d(f)_{|A}) > D$. Potentially this can lead to improved lower bounds on the *total* size of the formula, better than one would be able to infer from the lower bound on *multiplicative* complexity of Theorem 4.1 alone. We shall see that we can indeed get such kinds of improvements in the applications section below.

We employ the following suite of concepts and lemmas from [SW99] directly.

Definition 4.1 ([SW99]). For $f \in F[x_1, \ldots, x_n]$, let $\partial_d(f)$ be the set of all dth order *formal* partial derivatives of f w.r.t. variables from $\{x_1, \ldots, x_n\}$.

For a set of polynomials $A = \{f_1, \ldots, f_t\}$, let $\mathrm{span}(A) = \{\sum_{i=1}^{t} c_i f_i \mid c_i \in F\}$, i.e., $\mathrm{span}(A)$ is the linear span of A. We write $\dim[A]$ as shorthand for $\dim[\mathrm{span}(A)]$. We have the following elementary sub-additivity property for the measure $\dim[\partial_d(f)]$.

Proposition 4.3 ([SW99]) *For $f_1, f_2 \in F[x_1, \ldots, x_n]$ and constants $c_1, c_2 \in F$,*

$$\dim[\partial_d(c_1 f_1 + c_2 f_2)] \leq \dim[\partial_d(f_1)] + \dim[\partial_d(f_2)].$$

One also needs to bound the growth of $\dim[\partial_d(f)]$ in case of multiplication. For multiplication of affine linear forms, we have the following two bounds.

Proposition 4.4 ([SW99]) *Let $M = \Pi_{i=1}^{m} l_i$, where each l_i is affine linear. Then*

$$\dim[\partial_d(M)] \leq \binom{m}{d}.$$

Proposition 4.5 ([SW99]) *Let M be a product gate with $\dim[M^h] = m$, then for any d,*

$$\dim[\partial_d(M)] \leq \binom{m+d}{d}.$$

Note that for polynomials $f_1, \ldots, f_s$, $\mathrm{span}(f_1, \ldots, f_s)_{|A} = \mathrm{span}(f_{1|A}, \ldots, f_{s|A})$, and that $\dim[W_{|A}] \leq \dim[W]$. Now we modify Proposition 4.3 a little to get a result implicitly used by Shpilka and Wigderson in their arguments.

Proposition 4.6 (cf. [SW99]) *For $f_1, f_2 \in F[x_1, \ldots, x_n]$ and constants $c_1, c_2 \in F$, and affine linear subspace A, we have that* $\dim[\partial_d(c_1 f_1 + c_2 f_2)_{|A}] \leq \dim[\partial_d(f_1)_{|A}] + \dim[\partial_d(f_2)_{|A}]$.

Finally, we require:

Lemma 4.7 ([SW99]) *For every n, κ, d, and every affine subspace A of co-dimension κ, we have that*

$$\dim[\partial_d(S_n^{2d})_{|A}] \geq \binom{n-\kappa}{d}.$$

Now we can prove our sideways improvement of Shpilka and Wigderson's main Theorem 3.1 [SW99].

Proof of Theorem 4.2. Consider a minimum-size $\Sigma\Pi\Sigma$-formula for f with multiplication gates $M_1, \ldots, M_s$. We have that

$$\begin{aligned} f &= \sum_{i=1}^{s} M_i, \qquad \text{where for } 1 \leq i \leq s, \qquad M_i = \Pi_{j=1}^{d_i} l_{i,j} \qquad \text{and} \\ l_{i,j} &= c_{i,j,1} x_1 + c_{i,j,2} x_2 + \ldots + c_{i,j,n} x_n + c_{i,j,0}, \end{aligned}$$

for certain constants $c_{i,j,k} \in F$. Computing the partial derivative of f w.r.t. variable x_k we get

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{s} \sum_{j=1}^{d_i} c_{i,j,k} \frac{M_i}{l_{i,j}}. \tag{4}$$

Let

$$S = \{i : \dim[M_i^h] \geq \kappa\}.$$

If $|S| \geq \frac{\kappa}{d+2}$, then $\ell_3(f) \geq \frac{\kappa^2}{d+2}$. Suppose $|S| < \frac{\kappa}{d+2}$. If $S = \emptyset$, then let A be an arbitrary affine subspace of co-dimension κ. Otherwise, construct an affine space A as follows. Since $|S|(d+2) < \kappa$, and since for each $j \in S$, $\dim[M_i^h] \geq \kappa$, it is possible to pick $d+2$ input linear forms $l_{j,1}, \ldots, l_{j,d+2}$ of each multiplication gate M_j with $j \in S$, such that $\{l_{j,1}^h, \ldots, l_{j,d+2}^h | j \in S\}$ is a set of $|S|(d+2) < \kappa$ independent homogeneous linear forms. Define

$$A = \{x : \ l_{i,j}(x) = 0, \text{ for any } i \in S, \ j \in [d+2]\}.$$

We have that the co-dimension of A is at most κ. W.l.o.g. assume the co-dimension of A equals κ. For each $i \in S$, d+2 linear forms of M_i vanish on A. This implies that

$$\dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}] = 0.$$

for any $i \in S$. For any $i \notin S$, by Proposition 4.5,

$$\dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}] < \binom{\kappa+d}{d}.$$

Let $D_k = \dim[\partial_d(\frac{\partial f}{\partial x_k})_{|A}]$. By Proposition 4.6 and equation (4),

$$D_k \leq \sum_{i \notin S} \sum_{\substack{j \\ c_{i,j,k} \neq 0}} \dim[\partial_d(\frac{M_i}{l_{i,j}})_{|A}].$$

Hence there must be at least $\frac{D_k}{\binom{\kappa+d}{d}}$ terms on the r.h.s., i.e. there are at least that many wires from x_k to gates in the first layer. Hence in total the number of wires to the first layer is at least $\sum_{i=1}^{n} \frac{D_i}{\binom{\kappa+d}{d}} > \frac{D}{\binom{\kappa+d}{d}}$. □

We can apply a similar idea to adapt the other main theorem from [SW99]:

Theorem 4.8 ([SW99]) *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ it holds that for every affine subspace A of co-dimension κ, $\dim(\partial_d(f_{|A})) > D$. Then for every $m \geq 2$,*

$$\ell_3^*(f) \geq \min(\kappa m, \frac{D}{\binom{m}{d}}).$$

Theorem 4.9 (new) *Let $f \in F[x_1, \ldots, x_n]$. Suppose for integers d, D, κ with $d \geq 1$, it holds that for every affine subspace A of co-dimension κ, $\sum_{i=1}^{n} \dim[\partial_d(\frac{\partial f}{\partial x_i}{}_{|A})] > D$. Then for every $m \geq 2$,*

$$\ell_3(f) \geq \min(\frac{1}{2}\kappa m, \frac{D}{\binom{m-1}{d}}).$$

Proof. Consider a minimum size $\Sigma\Pi\Sigma$-formula for f with multiplication gates $M_1, \ldots, M_s$. We have that

$$\begin{aligned} f &= \sum_{i=1}^{s} M_i, \qquad \text{where for } 1 \leq i \leq s, \qquad M_i = \Pi_{j=1}^{d_i} l_{i,j}, \qquad \text{with} \\ l_{i,j} &= c_{i,j,1}x_1 + c_{i,j,2}x_2 + \ldots + c_{i,j,n}x_n + c_{i,j,0}. \end{aligned}$$

If there are $\frac{\kappa}{2}$ multiplication gates M_i of degree greater than m then already $\ell_3(f) > \frac{1}{2}\kappa m$. So suppose the number t of multiplication gates of degree greater than m is less than $\frac{\kappa}{2}$. Wlog. assume these gates are given by

$$M_1, M_2, \ldots, M_t.$$

For $i = 1, 2, \ldots, t$, pick two input linear forms $l_{i,1}, l_{i,2}$ of M_i, such that for the total collection $l_{1,1}, l_{1,2}, \ldots, l_{i,1}, l_{i,2}$ we have that the strictly linear parts $l^h_{1,1}, l^h_{1,2}, \ldots, l^h_{i,1}, l^h_{i,2}$ are independent. It might be that at some $i \leq t$, we cannot find any $l_{i,1}$ or $l_{i,2}$ with $l^h_{i,1}$ or $l^h_{i,2}$ independent from the previously collected linear forms. In this case, we just pick $l_{i,1}$ if that one is still independent, and skip to the next index i. If we can't even find $l_{i,1}$ for which $l_{i,1}$ is independent, we pick no linear form and proceed to the next i.

Let A be the zero set of all the collected input linear forms. Then A has co-dimension at most κ. Wlog. we may assume that the co-dimension of A equals κ. Observe that

$$\frac{\partial f}{\partial x_k}_{|A} = \sum_{i=1}^{s}\sum_{j=1}^{d_i} c_{i,j,k}\left(\frac{M_i}{l_{i,j}}\right)_{|A}. \tag{5}$$

Now for a multiplication gate M_i of degree $\geq m$, there are three cases: either we picked two input linear forms of M_i, or we picked just one, or none at all. In the first case,

$$\left(\frac{M_i}{l_{i,j}}\right)_{|A} = 0$$

in the r.h.s. of (5), for all i, j. In the second and third case, we know that for every input l of M_i that was not picked, l^h is a linear combination of l^h_i's for l_i's that were picked. Hence

$$l^h_{|A} = \sum_{i=1}^{r} c_i(l^h_{i\,|A}) = \text{constant}.$$

As a consequence, $(\frac{M_i}{l_{i,j}})_{|A} =$ constant in the r.h.s. of (5), for all i, j. Since $d \geq 1$, in either three cases, we obtain that $\partial_d(\frac{M_i}{l_{i,j}}{}_{|A}) = 0$. For multiplication gates M_i of degree at most m, Proposition 4.4 gives us that $\dim[\partial_d((\frac{M_i}{l_{i,j}})_{|A})] \leq \binom{m-1}{d}$. Let $D_k = \dim[\partial_d(\frac{\partial f}{\partial x_k}{}_{|A})]$. By Proposition 4.3, we see there are at least $D_k/\binom{m-1}{d}$ terms in (5). This implies that there are at least that many wires fanning out of x_k. Adding up for all variables, we conclude that $\ell_3(f) \geq D/\binom{m-1}{d}$. □

4.1 Some Applications

In [SW99] it was proved that for $d \leq \log n$, $\ell^*_3(S^{2d}_n) = \Omega(\frac{n^{\frac{2d}{d+2}}}{d})$. Note for $d = 2$, this lower bound is only $\Omega(n)$. We can apply Theorem 4.2 to prove the following stronger lower bound on the total formula size of S^{2d}_n. In particular for $d = 2$, we get an $\Omega(n^{\frac{4}{3}})$ bound.

Theorem 4.10 *For* $1 \leq d \leq \log n$, $\ell_3(S_n^{2d}) = \Omega(\frac{n^{\frac{2d}{d+1}}}{d})$.

Proof. For any affine subspace A of co-dimension κ and $d \geq 2$ we have that

$$\sum_{i=1}^{n} \dim[\partial_{d-1}(\frac{\partial S_n^{2d}}{\partial x_i})_{|A}] \geq \dim[\partial_d(S_n^{2d})_{|A}] \geq \binom{n-\kappa}{d}.$$

The latter inequality follows from Lemma 4.7. Applying Theorem 4.2 we get that

$$\ell_3(S_n^{2d}) \geq \min(\frac{\kappa^2}{d+1}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d-1}{d-1}}) = \min(\frac{\kappa^2}{d+1}, \frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{\kappa+d}{d}). \tag{6}$$

Set $\kappa = \frac{1}{9}n^{\frac{d}{d+1}}$. Then we have that

$$\frac{\binom{n-\kappa}{d}}{\binom{\kappa+d}{d}}\frac{\kappa+d}{d} \geq (\frac{n-\kappa}{\kappa+d})^d\frac{\kappa+d}{d} \geq (\frac{8/9n}{2/9n^{\frac{d}{d+1}}})^d\frac{\kappa+d}{d} = 4^d n^{\frac{d}{d+1}}\frac{\kappa+d}{d} \geq \frac{4^d}{9d}n^{\frac{2d}{d+1}} \geq n^{\frac{2d}{d+1}}.$$

Hence (2) is at least $\min(\frac{n^{\frac{2d}{d+1}}}{81(d+1)}, n^{\frac{2d}{d+1}}) = \Omega(\frac{n^{\frac{2d}{d+1}}}{d})$. □

Corollary 4.11 $\ell_3(S_n^4) = \Omega(n^{4/3})$.

Shpilka and Wigderson defined the "product-of-inner-products" polynomial over $2d$ variable sets of size n (superscript indicate different variables, each variable has degree one) by

$$PIP_n^d = \prod_{i=1}^{d}\sum_{j=1}^{n} x_j^i y_j^i.$$

Theorem 4.12 *For any constant* $d > 0$, $\ell_3(PIP_n^d) = \Omega(n^{\frac{2d}{d+1}})$.

Proof. Let $f = PIP_n^d$. Essentially we have that

$$\frac{\partial f}{\partial x_j^i} = y_j^i PIP_n^{d-1},$$

where the PIP_n^{d-1} must be chosen on the appropriate variable set. Let A be an arbitrary affine linear subspace of co-dimension κ. Then

$$\begin{aligned}\sum_{i=1}^{d}\sum_{j=1}^{n}\dim[\partial_{d-1}(\frac{\partial f}{\partial x_j^i}|_A)] &= \sum_{i=1}^{d}\sum_{j=1}^{n}\dim[\partial_{d-1}(y_j^i PIP_n^{d-1}|_A)] \\ &\geq (dn-\kappa)\dim[\partial_{d-1}(PIP_n^{d-1}|_A)]\end{aligned}$$

The last inequality follows because at least $dn - \kappa$ of the y-variables are not assigned to with the restriction to A. From Lemma 4.9 in [SW99] one gets

$$\dim[\partial_{d-1}(PIP_n^{d-1}|_A) \geq n^{d-1} - 2^{2d-1}\kappa n^{d-2}.$$

Using Theorem 4.9 we get

$$\ell_3(f) \geq \min(\frac{\kappa^2}{2}, \frac{(dn-\kappa)(n^{d-1} - 2^{2d-1}\kappa n^{d-2})}{\binom{\kappa-1}{d-1}}).$$

Taking $\kappa = n^{\frac{d}{d+1}}$, one gets for constant d that

$$\ell_3(PIP_n^d) = \Omega(n^{\frac{2d}{d+1}}).$$

□

For comparison, in [SW99] one gets $\ell_3^*(PIP_n^d) = \Omega(n^{\frac{2d}{d+2}})$.

5 Conclusion—Possible Further Tools

We have taken some further steps after [SW99], obtaining absolutely tight (rather than asymptotically so) multiplicative size lower bounds for some natural functions, and obtaining somewhat improved bounds on $+, *$-size for low-degree symmetric and product-of-inner-product polynomials. However, these may if anything enhance the feeling from [SW99, Shp01] that the concepts being employed may go no further than quadratic for lower bounds. One cannot after all say that a function $f(x_1, \ldots, x_n)$ is non-vanishing on an affine-linear space of co-dimension more than n. The quest then is for a mathematical invariant that scales beyond linear with the number of degree-d-or-higher multiplication gates in the formula.

Notably absent in current lower bound techniques for $\Sigma\Pi\Sigma$-formulas are random restriction type arguments, whereas many of the results for *Boolean* constant depth circuits of [Ajt83, FSS81, Yao85, Hås89] proceed using random restrictions. Note that Raz managed to use random restrictions in conjunction with a partial derivatives based technique in his work on *multilinear* arithmetical formulas [Raz04a, Raz04b]. We speculate that the weaker existential requirement in our resistance notion may help it adapt to random-restriction scenarios, although plenitude of kth-partials may still be needed to ensure the function does not collapse too far. In any event, the search for stronger mathematical techniques to prove exponential lower bounds, even in the self-contained $\Sigma\Pi\Sigma$ formula case, continues.

Acknowledgments We thank Avi Wigderson for comments on a very early version of this work, and referees of a later version for very helpful criticism.

References

[Ajt83] M. Ajtai. Σ_1^1 formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[BCS97] P. Bürgisser, M. Clausen, and M.A. Shokrollahi. *Algebraic Complexity Theory.* Springer Verlag, 1997.

[BS82] W. Baur and V. Strassen. The complexity of partial derivatives. *Theor. Comp. Sci.*, 22:317–330, 1982.

[FSS81] M. Furst, J. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *Proc. 22nd Annual IEEE Symposium on Foundations of Computer Science*, pages 260–270, 1981.

[GPS05] G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR 3.0. A Computer Algebra System for Polynomial Computations, Centre for Computer Algebra, University of Kaiserslautern, 2005. `http://www.singular.uni-kl.de`.

[Hås86] J. Håstad. Almost optimal lower bounds for small-depth circuits. In *Proc. 18th Annual ACM Symposium on the Theory of Computing*, pages 6–20, 1986.

[Hås89] J. Håstad. Almost optimal lower bounds for small-depth circuits. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 143–170. JAI Press, Greenwich, CT, USA, 1989.

[NW96] N. Nisan and A. Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6:217–234, 1996.

[Raz04a] R. Raz. Multilinear formulas for permanent and determinant are of super-polynomial size. In *Proc. 36th Annual ACM Symposium on the Theory of Computing*, 2004. to appear; also ECCC TR03-067.

[Raz04b] R. Raz. Multilinear $\mathrm{NC}^1 \neq$ multilinear NC^2. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 344–351, 2004.

[Shp01] A. Shpilka. Affine projections of symmetric polynomials. In *Proc. 16th Annual IEEE Conference on Computational Complexity*, pages 160–171, 2001.

[Str73] V. Strassen. Berechnung und Programm II. *Acta Informatica*, 2:64–79, 1973.

[SW99] A. Shpilka and A. Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. Technical Report 23, ECCC, 1999.

[Val79] L. Valiant. The complexity of computing the permanent. *Theor. Comp. Sci.*, 8:189–201, 1979.

[Yao85] A. Yao. Separating the polynomial-time hierarchy by oracles. In *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.