The shifted partial derivative complexity of Elementary Symmetric Polynomials

Hervé Fournier¹, Nutan Limaye², Meena Mahajan³, and Srikanth Srinivasan⁴

¹ IMJ-PRG, Univ Paris Diderot, France. fournier@math.univ-paris-diderot.fr
² Dept. of Computer Science and Engg., IIT Bombay, India. nutan@cse.iitb.ac.in
³ The Institute of Mathematical Sciences, Chennai, India. meena@imsc.res.in
⁴ Dept. of Mathematics, IIT Bombay, India. srikanth@math.iitb.ac.in

Abstract. We continue the study of the shifted partial derivative measure, introduced by Kayal (ECCC 2012), which has been used to prove many strong depth-4 circuit lower bounds starting from the work of Kayal, and that of Gupta et al. (CCC 2013).

We show a strong lower bound on the dimension of the shifted partial derivative space of the Elementary Symmetric Polynomials of degree $d$ in $N$ variables for $d < \log N / \log \log N$. This extends the work of Nisan and Wigderson (Computational Complexity 1997), who studied the partial derivative space of these polynomials. Prior to our work, there have been no results on the shifted partial derivative measure of these polynomials. Our result implies a strong lower bound for Elementary Symmetric Polynomials in the homogeneous $\Sigma \Pi \Sigma \Pi$ model with bounded bottom fan-in.

This strengthens (under our degree assumptions) a lower bound of Nisan and Wigderson who proved the analogous result for homogeneous $\Sigma \Pi \Sigma$ model (i.e. $\Sigma \Pi \Sigma \Pi$ formulas with bottom fan-in 1).

Our main technical lemma gives a lower bound for the ranks of certain inclusion-like matrices, and may be of independent interest.

1 Introduction

Motivation. In an influential paper of Valiant [23] the two complexity classes $VP$ and $VNP$ were defined, which can be thought of as algebraic analogues of Boolean complexity classes $P$ and $NP$, respectively. Whether $VP$ equals $VNP$ or not is one of the most fundamental problems in the study of algebraic computation. It follows from the work of Valiant [23] that a super-polynomial lower bound for arithmetic circuits computing the Permanent implies $VP \neq VNP$.

The best known lower bound on uniform polynomials for general arithmetic circuits is $\Omega(N \lg N)$ [3] which is unfortunately quite far from the desired super-polynomial lower bound. Over the years, though there has been no stronger lower bound for general arithmetic circuits, many super-polynomial lower bounds have been obtained for special classes for arithmetic circuits [15,17,16].

* This research was supported by IFCPAR/CEFIPRA Project No 4702-1(A) and research grant compA ANR-13-BS02-0001-01.
A very interesting such subclass of arithmetic circuits is the class of bounded-depth arithmetic formulas. The question of proving lower bounds for bounded-depth formulas and in particular depth 3 and 4 formulas has received a lot of attention subsequent to the recent progress in efficient depth reduction of arithmetic circuits [24,1,11,22]. This sequence of results essentially implies that “strong enough” lower bounds for depth-4 homogeneous formulas suffice to separate VP from VNP. More formally, it proves that any sequence \( \{f_N\}_N \) of homogeneous \( N \)-variate degree \( d = N^{O(1)} \) polynomials in VP has depth-4 homogeneous formulas of size \( N^{O(\sqrt{d})} \). Hence, proving an \( N^{\omega(\sqrt{d})} \) lower bound for depth-4 homogeneous formulas suffices to separate VP from VNP.

Even more can be said about the depth-4 formulas obtained in the above results. For any integer parameter \( t \leq d \), they give a \( \Sigma \Pi \Sigma \Pi \) formula for \( f_N \) where the layer 1 product gates (just above the inputs) have fan-in at most \( t \) and the layer 3 gates are again \( \Pi \) gates with fan-in \( O(d/t) \). We will refer to such formulas as \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) formulas. The depth-reduction results mentioned above produce a depth-4 homogeneous \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) formula of size \( N^{O(d/t)} \) and top fan-in \( N^{O(d/t)} \); at \( t = \lceil \sqrt{d} \rceil \), we get the above depth-reduction result.

The tightness of these results follows from recent progress on lower bounds for the model of \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) circuits. A flurry of results followed the groundbreaking work of Kayal [7], who augmented the partial derivative method of Nisan and Wigderson [15] to devise a new complexity measure called the shifted partial derivative measure, using which he proved an exponential lower bound for a special class of depth-4 circuits. Building on this, the first non-trivial lower bound for \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) formulas was proved by Gupta, Kamath, Kayal, and Saptharishi [5] for the determinant and permanent polynomials. This was further improved by Kayal, Saha, and Saptharishi [9] who gave a family of explicit polynomials in VNP the shifted partial derivative complexity of which was (nearly) as large as possible and hence showed a lower bound of \( N^{\Omega(d/t)} \) for the top fan-in of \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) formulas computing these polynomials. Later, a similar result for a polynomial in VP was proved in [4] and this was subsequently strengthened by Kumar and Saraf [12], who gave a polynomial computable by homogeneous \( \Pi \Sigma \Pi \) formulas which have no \( \Sigma \Pi [O(d/t)] \Sigma \Pi [t] \) formulas of top fan-in smaller than \( N^{\Omega(d/t)} \). Finally, using a variant of the shifted partial derivative measure, Kayal et al. [8] and Kumar and Saraf [13] were able to prove similar lower bounds for general depth-4 homogeneous formulas as well.

In this work, we investigate the shifted partial derivative measure of the Elementary Symmetric Polynomials, which is a very natural family of polynomials whose complexity has been the focus of many previous works [15,21,20,6]. Nisan and Wigderson [15] proved tight lower bounds on the depth-3 homogeneous formula complexity of these polynomials. Shpilka and Wigderson [21] and Shpilka [20] studied the general (i.e. possibly inhomogeneous) depth-3 circuit complexity of these polynomials, and showed that for certain degrees, the \( O(N^2) \) upper bound due to Ben-Or (see [21]) is tight.

---

5 For \( O(1) \) depth, the sizes of formulas and circuits are polynomially related.

6 i.e., as large as it can be for a “generic” or “random” polynomial. (See Remark 3.)
Under some degree constraints, we show strong lower bounds on the dimension of the shifted partial derivative space of these polynomials, which implies that the Elementary symmetric polynomial on \( N \) variables of degree \( d \) cannot be computed by a \( \Sigma \Pi^{O(d/t)} \Sigma \Pi^{[d]} \) circuit of top fan-in less than \( N^{\Omega(d/t)} \). This strengthens the result of Nisan and Wigderson \[15\] for these degree parameters. By the upper bound of Ben-Or mentioned above, this also gives the first example of an explicit polynomial with small \( \Sigma \Pi \Sigma \) circuits for which such a strong lower bound is known.

**Results.** We show that, for a suitable range of parameters, the shifted partial derivative measure of the \( N \)-variate elementary symmetric polynomial of degree \( d \) — denoted \( S_N^d \) — is large.

**Theorem 1.** Let \( N, d, k \in \mathbb{N} \) be such that \( d \leq (1/10) \log N / \log \log N \) and \( k = \lfloor d/(\tau + 1) \rfloor \) for some \( \tau \in \mathbb{N} \) satisfying \( \tau \equiv 1 \pmod{4} \). For \( \ell = \lceil N^{1-1/2\tau} \rceil \), \( \dim(\partial_k S_N^d) \leq \ell \geq \frac{(1-o(1)) \binom{N+\ell}{\ell} \binom{N-\ell}{k}}{(3\sqrt{N/2})^{d/2d^2}}. \)

It was observed by \[5\] that for any homogeneous multilinear polynomial \( f \) on \( N \) variables of degree \( d \), we have \( \dim(\partial_k S_N^d) \leq \binom{N+\ell}{\ell} \cdot \binom{N-\ell}{k} \), which is close to the numerator in the above expression. The theorem above should be interpreted as saying that the dimension is not too far from this upper bound.

A corollary of our main result is an \( N^{\Omega(d/t)} \) lower bound on the top fan-in of any \( \Sigma \Pi^{O(d/t)} \Sigma \Pi^{[d]} \) formula computing \( S_N^d \).

**Theorem 2.** Fix \( N, d, D, t \in \mathbb{N} \) and constant \( \epsilon > 0 \) s.t. \( d \leq \frac{\log N}{10 \log \log N} \), \( D \leq N^{1-\epsilon} \). Any \( \Sigma \Pi^{[D]} \Sigma \Pi^{[d]} \) circuit of top fan-in \( s \) computing \( S_N^d \) satisfies \( s = N^{\Omega(d/t)} \).

It is worth noting that in most lower bounds of this flavour, the upper product gates have fanin \( D \) bounded by \( O(d/t) \). Our lower bound works for potentially much larger values of \( D \).

**Corollary 1.** Let parameters \( N, d, t \) be as in Theorem 2. Any \( \Sigma \Pi^{[O(d/t)]} \Sigma \Pi^{[d]} \) computing \( S_N^d \) must have top fan-in at least \( N^{\Omega(d/t)} \). In particular, any homogeneous \( \Sigma \Pi \Sigma \Pi \) circuit \( C \) with bottom fan-in bounded by \( t \) computing \( S_N^d \) must have top fan-in at least \( N^{\Omega(d/t)} \).

By the above depth reduction results, this lower bound is tight up to the constant factor in the exponent. Before our work, \[[15\] proved a lower bound for \( S_N^d \) of \( N^{\Omega(d)} \) for all \( d \), however with respect to \( \Sigma \Pi \Sigma \) circuits (i.e. the case \( t = 1 \)).

**Techniques.** The analysis of the shifted partial derivative measure for any polynomial essentially requires the analysis of the rank of a matrix arising from the shifted partial derivative space. In this work, we analyse the matrix arising from the shifted partial derivative space of the symmetric polynomials. Our analysis is quite different from previous works (such as \[14\]), which are based on either monomial counting (meaning that we find a large identity or upper triangular submatrix inside our matrix) or an analytic inequality of Alon \[2\].

---

7 It is known (see, e.g., \[19\] Proof of Corollary 5.8) that any homogeneous \( \Sigma \Pi \Sigma \Pi^{[d]} \) circuit \( C \) can be converted to a \( \Sigma \Pi^{[O(d/t)]} \Sigma \Pi^{[d]} \) circuit with the same top fan-in.
In our analysis of the shifted partial derivative space, we define a more complicated version of the Inclusion matrix (known to be full rank) and lower bound its rank by using a novel technique, which we describe in the next section.

Disjointness and inclusion matrices arise naturally in other branches of theoretical computer science such as Boolean circuit complexity [15], communication complexity [14, Chapter 2] and also in combinatorics [25, 10]. Therefore, we believe that our analysis of the Inclusion-like matrix arising from the symmetric polynomial may find other applications.

**Organisation of the paper.** In Section 2 we set up basic notation, fix the main parameters, and give a high-level outline of our proof of Theorem 1. In Section 3 we give the actual proof. The formula size lower bounds from Theorem 2 is established in Section 4. Due to space constraints, several proofs are omitted.

## 2 Proving Theorem 1: High-level outline

**Notation:** For a positive integer \( n \), we let \( [n] = \{1, \ldots, n\} \). Let \( X = \{x_1, \ldots, x_N\} \). For \( A \subseteq [N] \) we define \( X_A = \prod_{i \in A} x_i \). The elementary symmetric polynomial of degree \( d \) over the set of variables \( X \) is defined as \( S_N^d(X) = \sum_{A \subseteq [N], |A| = d} X_A \). In the following \( S_N^d(X) \) is abbreviated with \( S_N^d \).

For \( k, \ell \in \mathbb{N} \) and a multivariate polynomial \( f \in \mathbb{F}[x_1, \ldots, x_n] \), we define

\[
\langle \partial_k f \rangle_{\leq \ell} = \text{span} \left\{ x_1^{i_1} \cdots x_n^{i_n} \cdot \frac{\partial^k f}{\partial x_1^{j_1} \cdots \partial x_n^{j_n}} \mid i_1 + \cdots + i_n = k, j_1 + \cdots + j_n \leq \ell \right\}.
\]

Our complexity measure is the dimension of this space, i.e., \( \text{dim}(\langle \partial_k f \rangle_{\leq \ell}) \).

For a monomial \( m = \prod_{i=1}^{N} x_i^{n_i} \), \( \text{deg}(m) = n_1 + n_2 + \cdots + n_N \) is the total degree of \( m \). We denote by \( \text{deg}_{x_i}(m) \) the degree of the variable \( x_i \) in \( m \) (here \( \text{deg}_{x_i}(m) = n_i \)). We define the support of \( m \) as \( \text{supp}(m) = \{ i \in [N] \mid n_i > 0 \} \).

For a monomial \( m \) and \( p > 0 \), let \( \text{supp}_p(m) = \{ i \in [N] \mid \text{deg}_{x_i}(m) = p \} \).

Let \( \mathcal{M}_N^d \) the set of monomials of degree at most \( \ell \) over the variables \( X \). For integers \( n_1, \ldots, n_p \), let

\[
\mathcal{M}_N^d(n_1, \ldots, n_p) = \{ m \in \mathcal{M}_N^d, |\text{supp}_p(m)| = n_i \text{ for } i \in [p] \}.
\]

Given \( p > 0 \), a monomial \( m \in \mathcal{M}_N^d \) can be uniquely written as \( m = \tilde{m} \cdot \prod_{i=1}^{p} (x_{\text{supp}_i(m)})^{i} \). We write \( m \equiv [\tilde{m}, S_1, \ldots, S_p] \) if \( S_i = \text{supp}_i(m) \) for all \( i \in [p] \) and \( m = \tilde{m} \cdot \prod_{i=1}^{p} (X_{S_i})^{i} \).

For a finite set \( S \), let \( \mathcal{U}(S) \) denote the uniform distribution over the set \( S \).

We assume that we are working over a field \( \mathbb{F} \) of characteristic zero. Our results also hold in non-zero characteristic, but the first step of our proof (Lemma 1) becomes a little more cumbersome (this part is omitted in this version).

**Proof Outline** Our lower bound on \( \text{dim}(\langle \partial_k S_N^d \rangle_{\leq \ell}) \) proceeds in 3 steps.

**Step 1:** We choose a suitable subset \( S \) of the partial derivative space. It is convenient to work with a set that is slightly different from the set of partial
derivatives themselves. To understand the advantage of this, consider the simple setting where we are looking at the partial derivatives of the degree-2 polynomial \( S^2_N \) of order 1. It is not difficult to show that the partial derivative with respect to variable \( x_i \) is \( r_i := \sum_{j \neq i} x_j \). Over characteristic zero, this set of polynomials is known to be linearly independent. One way to show this is by showing that each polynomial \( x_i \) can be written as a linear combination of the \( r_j \); explicitly, one can write \( x_i = \frac{1}{n-1} \left( \sum_{j \in [n]} r_j \right) - r_i \). Since the \( x_i \)s are distinct monomials, they are clearly linearly independent and we are done. This illustrates the advantage in moving to a “sparser” basis for the partial derivative space. We do something like this for larger \( d \) and \( k \) (Lemma 1).

**Step 2:** After choosing the set \( S \), we construct the set \( P \) of shifts of \( S \) (actually, we will only consider a subset of \( P \)) and lower bound the rank of the corresponding matrix \( M \). To do this, we also prune the set of rows of the matrix \( M \). In other words, we consider a carefully chosen set of monomials \( M \) and project each polynomial in \( P \) down to these monomials. The objective in doing this is to infuse some structure into the matrix while at the same time preserving its rank (up to small losses). Having chosen \( M \), we show that the corresponding submatrix can be block-diagonalized into matrices each of which is described by a simple inclusion pattern between the (tuples of) sets labelling its rows and columns. This is done in Lemmas 4, 5, 6.

**Step 3:** The main technical step in the proof is to lower bound the rank of the inclusion pattern matrix mentioned above with an algebraic trick. We first find a full-rank matrix that is closely related to our matrix and then show that the columns of our matrix can (with the aid of just a few other columns) generate the columns of the full-rank matrix.

**The main parameters** Let \( N, d \) and \( t \) be fixed. Throughout, we assume that \( d \leq (1/10) \log N / \log \log N \). Let \( \tau = 4t + 1 \), \( \delta = 1/(2\tau) \) and \( \ell = \lfloor N^{1-\delta} \rfloor \). Finally, let \( k \) be such that \( d - k = \tau k \). (In particular, we assume that \( 4t + 2 \leq d \).)

The following are easy to verify for our choice of parameters:

**Fact 1.** \( \tau^2 = o(\ell) \) and \( \tau = o(N^\delta) \).

**Remark 1.** In the above setting of parameters, \( d \) has to be divisible by \( \tau + 1 = 4t + 2 \). For ease of exposition, we present the proof for these parameters. Our proof can be modified so that it works for any \( d \) large enough as compared to \( \tau \) (this part is omitted in this short version).

### 3 Proving Theorem 1: Details

#### 3.1 Choice of basis: Step 1 of the proof

**Lemma 1.** Let \( k \leq d \). The vector space spanned by the set of \( k \)-partial derivatives of \( S^d_N \), that is \( \langle \partial_k S^d_N \rangle_{\leq 0} \), contains \( \{ p_T \mid T \subseteq [N], |T| = k \} \) where

\[
p_T = \sum_{T \subseteq A \subseteq [N], |A| = d-k} X_A \quad \text{(that is, } p_T = X_T \cdot S_{N-k}^{d-2k}(X \setminus T))\].
Let \( \mathcal{P} = \{ m \cdot p_T \mid T \subseteq [N], |T| = k, m \in \mathcal{M}_N, \text{supp}(m) \cap T = \emptyset \} \). From Lemma 1, \( \mathcal{P} \subseteq \langle \partial_k S^d_N \rangle_{\leq \ell} \). Hence, a lower bound on the dimension of \( \text{span} \mathcal{P} \) is also a lower bound on \( \dim(\langle \partial_k S^d_N \rangle_{\leq \ell}) \).

### 3.2 Choice of shifts: Step 2 of the proof

Instead of considering arbitrary shifts \( m \) as in the definition of \( \mathcal{P} \), we will consider shifts by monomials \( m \) with various values of \( |\text{supp}(m)| \) for \( i \in [\tau] \). We first present a technical lemma that is needed to establish the lower bound. It is a concentration bound for support sizes in random monomials.

**Definition 1.** For \( i \in [\tau] \), \( \hat{s}_i \) denotes the average number of variables with degree exactly \( i \); \( \hat{s}_i = \mathbb{E}_{m \sim \mathcal{U}(\mathcal{M}_N)}[|\text{supp}(m)|] \).

**Definition 2 (Good signature).** Given \( m \in \mathcal{M}_N^\ell \), the signature of \( m \), \( s(m) \), is the tuple \( (s_1, \ldots, s_{\tau}) \) such that \( m \in \mathcal{M}_N(s_1, \ldots, s_{\tau}) \). We call the signature \( (s_1, \ldots, s_{\tau}) \) a good signature if for each \( i \in [\tau] \), we have \( \hat{s}_i / 2 \leq s_i \leq 3 \hat{s}_i / 2 \).

**Lemma 2.** For our choice of the main parameters, the following statements hold: (i) \( \hat{s}_i = \mathcal{O}(\frac{\tau}{N_{\tau+1}}) \) \((1 - o(1)) \). In particular, \( \hat{s}_\tau \leq \sqrt{N} \).

(ii) \( \hat{s}_i \geq \hat{s}_\tau \geq N^{-1/10}(1 - o(1)) \geq \frac{N^{1/10}}{2} \), and \( \frac{\hat{s}_i}{\hat{s}_{i+1}} \geq \frac{N^d}{2} \). (iii) Furthermore, \( \mathbb{P}_{m \sim \mathcal{U}(\mathcal{M}_N)}[s(m) \text{ is a good signature}] = 1 - o(1) \). (1)

**Remark 2.** By Lemma 2 for any good signature \( (s_1, \ldots, s_{\tau}) \), we have \( \frac{s_i}{s_{i+1}} = \Omega(N^d) \), \( s_\tau = \Omega(N^{1/10}) \), and also \( \frac{|\bigcup_{s_1, \ldots, s_{\tau} \text{ good}} \mathcal{M}_N(s_1, \ldots, s_{\tau})|}{|\mathcal{M}_N^\ell|} = 1 - o(1) \).

Given a signature \( (s_1, \ldots, s_{\tau}) \), let \( \mathcal{P}(s_1, \ldots, s_{\tau}) \) denote the set of polynomials \( \{ m \cdot p_T \mid T \subseteq [N], |T| = k, m \in \mathcal{M}_N(s_1, \ldots, s_{\tau}), \text{supp}(m) \cap T = \emptyset \} \). Note that all polynomials in \( \mathcal{P}(s_1, \ldots, s_{\tau}) \) are homogeneous of degree at most \( \ell + d - k \).

**Definition 3.** For any signature \( s = (s_1, \ldots, s_{\tau}) \), let \( r_i(s) = s_i \) for \( 1 \leq i \leq \tau - 1 \) and \( r_\tau(s) = s_\tau + k \); also, let \( r(s) = \sum_i r_i(s) = \sum_i s_i + k \). Usually the signature \( s \) will be clear from context, and we use \( r_i \) and \( r \) instead of \( r_i(s) \) and \( r(s) \) respectively. The matrix \( M(s_1, \ldots, s_{\tau}) \) is the matrix whose columns are indexed by polynomials \( m \cdot p_T \in \mathcal{P}(s_1, \ldots, s_{\tau}) \) and rows by the monomials \( w \in \mathcal{M}_N^{\ell+d-k}(r_1, \ldots, r_{\tau}) \). The coefficient in row \( w \) and column \( m \cdot p_T \) is the coefficient of the monomial \( w \) in the polynomial \( m \cdot p_T \).

Note that the columns of \( M(s_1, \ldots, s_{\tau}) \) are simply the polynomials in \( \mathcal{P}(s_1, \ldots, s_{\tau}) \) projected to the monomials that label the rows. In particular, a lower bound on the rank of \( M(s_1, \ldots, s_{\tau}) \) implies a lower bound on the rank of the vector space spanned by \( \mathcal{P}(s_1, \ldots, s_{\tau}) \).

It is not too hard to see that \( M(s_1, \ldots, s_{\tau}) \) has \( |\mathcal{P}(s_1, \ldots, s_{\tau})| \) columns but only \( \frac{|\mathcal{P}(s_1, \ldots, s_{\tau})|}{\binom{\ell + d - k}{\tau}} \) rows. Hence, the rank of the matrix is no more than the number of rows in the matrix. The following lemma, proved in Section 3.3, shows a lower bound that is quite close to this trivial upper bound.
Lemma 3. With parameters as above, for any good signature \( s = (s_1, \ldots, s_\tau) \),
\[
\text{rank}(M(s_1, \ldots, s_\tau)) \geq \frac{|\mathcal{P}(s_1, \ldots, s_\tau)|}{\binom{d-k}{\tau}} (1 - o(1)).
\]

Since \( \mathcal{P}(s_1, \ldots, s_\tau) \subseteq \mathcal{P} \subseteq (\partial_k f)_{\leq \ell} \), the above immediately yields a lower bound on \( \dim((\partial_k f)_{\leq \ell}) \). Our final lower bound, which further improves this, is proved by considering polynomials corresponding to a set of signatures.

Definition 4. Given a set of signatures \( \mathcal{S} \), define \( M_N^\tau(\mathcal{S}) = \bigcup_{s \in \mathcal{S}} M_N^\tau(s) \) and \( \mathcal{P}(\mathcal{S}) = \bigcup_{s \in \mathcal{S}} \mathcal{P}(s) \). Also define the matrix \( M(\mathcal{S}) \) as follows: the columns of \( M(\mathcal{S}) \) are labelled by polynomials \( q \in \mathcal{P}(\mathcal{S}) \) and the rows by monomials \( w \in \bigcup_{s \in \mathcal{S}} M_N^\tau + d - k(r_1(s), \ldots, r_\tau(s)) \). The \((w, q)\)th entry is the coefficient of \( w \) in \( q \).

Note that a lower bound on the rank of \( M(\mathcal{S}) \) immediately lower bounds the dimension of the space spanned by \( \mathcal{P}(\mathcal{S}) \) and hence also \( \dim((\partial_k S_N^\tau))_{\leq \ell} \).

Definition 5. A set of signatures is \( \mathcal{S} \) well-separated if given any distinct signatures \( s = (s_1, \ldots, s_\tau) \) and \( s' = (s'_1, \ldots, s'_{\tau'}) \) from \( \mathcal{S} \), \( \max_{i \in [\tau]} |s_i - s'_i| \geq 2d \).

To analyze the rank of \( M(\mathcal{S}) \), we observe that for a well-separated set of signatures \( \mathcal{S} \), the matrix \( M(\mathcal{S}) \) is block-diagonalizable with \( |\mathcal{S}| \) blocks, where the blocks are the matrices \( M(s) \) for \( s \in \mathcal{S} \). Since we already have a lower bound on the ranks of \( M(s) \) (for good \( s \)), this will allow us to obtain a lower bound on the rank of \( M(\mathcal{S}) \) as well.

Lemma 4. Let \( \mathcal{S} \) be a well-separated set of signatures. Then, the matrix \( M(\mathcal{S}) \) is block-diagonalizable with blocks \( M(s) \) for \( s \in \mathcal{S} \).

3.3 Bounding the rank of \( M \): Step 3 of the proof

We now prove the lower bound on the rank of the matrix \( M(s_1, \ldots, s_\tau) \) as claimed in Lemma 3. We block diagonalize it with matrices that have a simple combinatorial structure (their entries are 0 or 1 depending on intersection patterns of the sets that label the rows and columns). We then lower bound the ranks of these matrices: this is the main technical step in the proof.

Lemma 5. Fix any signature \( (s_1, \ldots, s_\tau) \). The entry of \( M(s_1, \ldots, s_\tau) \) in row \( w = [\hat{w}, R_1, \ldots, R_\tau] \) and column \( m \cdot p_T \) with \( m = [\hat{m}, S_1, \ldots, S_\tau] \) belongs to \( \{0, 1\} \) and is not zero if and only if \( \hat{w} = \hat{m} \) and the following system is satisfied: \( T \subseteq R_1, S_1 \subseteq R_1 \cup R_2, S_2 \subseteq R_2 \cup R_3 \ldots S_{\tau-1} \subseteq R_{\tau-1} \cup R_\tau, \) and \( S_\tau \subseteq R_\tau \).

Moreover, the system above implies that \( T \cup S_1 \cup \ldots \cup S_\tau = R_1 \cup \ldots \cup R_\tau \).

Proof. The entry in row \( w \) and column \( m \cdot p_T \) belongs to \( \{0, 1\} \) and is not zero if and only if there exists \( A \subseteq [N] \) such that \( T \subseteq A \), \( |A| = d - k \) and \( X_A \cdot m = w \). Assume there is such an \( A \).

Say \( w = [\hat{w}, R_1, \ldots, R_\tau] \) and \( m = [\hat{m}, S_1, \ldots, S_\tau] \). Let \( \overline{m} = \prod_{i=1}^\tau (X_{R_i})^i \) and \( \overline{w} = \prod_{i=1}^\tau (X_{R_i})^i \) be the degree at most \( \tau \) parts of \( m \) and \( w \) respectively.

Note that \( \deg(\overline{w}) - \deg(\overline{m}) = \sum_{i=1}^\tau ir_i - \sum_{i=1}^\tau is_i = \tau k = d - k \) by our choice of parameters \( r_\tau \) and \( k \). Putting this together with the fact that \( w = X_A \cdot m \) for
$|A| = d - k$, we see that $X_A$ can only `contribute' to the “degree at most $\tau$” part of $m$: formally, $\bar{w} = X_A : \bar{m}$ and hence, $\bar{w} = \bar{m}$.

Further, since $X_A : \bar{m} = X_A : \bar{T} \prod_{i=1}^\tau (X_{S_i})^i = \prod_{i=1}^\tau (X_{R_i})^i = \bar{w}$, and $T \cap (S_1 \cup \ldots \cup S_\tau) = \emptyset$, we have $T \subseteq R_1$. Since $X_A$ is multilinear, $S_i \subseteq R_i \cup R_{i+1}$ for all $i \in [\tau - 1]$; $S_\tau \subseteq R_\tau$ is obvious.

Conversely, assume that $\bar{w} = \bar{m}$ and the inclusions $T \subseteq R_1, S_i \subseteq R_i \cup R_{i+1}$ for all $i \in [\tau - 1]$ and $S_\tau \subseteq R_\tau$ are satisfied. Then $T \cup S_1 \cup \ldots \cup S_\tau \subseteq R_1 \cup \ldots \cup R_\tau$. Since $|T \cup S_1 \cup \ldots \cup S_\tau| = k + \sum_{i=1}^\tau s_i = \sum_{i=1}^\tau r_i = |R_1 \cup \ldots \cup R_\tau|$, we get $T \cup S_1 \cup \ldots \cup S_\tau = R_1 \cup \ldots \cup R_\tau$. Let $A_i = R_i \setminus S_i$ for $i \in [\tau]$ and $A = A_1 \cup \ldots \cup A_\tau$. The sets $A_i$ are disjoint (because the $R_i$ are disjoint). Moreover, $|A_i| = |R_i \setminus S_i| = r_\tau - s_\tau = k$; and by induction, $|A_i| = |R_i \setminus S_i| = |(R_i \cup \ldots \cup R_\tau) \setminus (S_i \cup \ldots \cup S_\tau)| = \sum_{j=1}^\tau r_j - \sum_{j=1}^\tau s_j = k$. Hence $|A_i| = k$ for all $i \in [\tau]$. Then $|A| = \tau k = d - k$. Moreover, $T = A_1 \subseteq A$. And it holds that $X_A \prod_{i=1}^\tau (X_{S_i})^i = \prod_{i=1}^\tau (X_{R_i})^i$. Since $\bar{w} = \bar{m}$, it follows that $X_A : m = w$. Hence the entry in row $w$ and column $m : p_T$ is 1.

Lemma 6. Let $(s_1, \ldots, s_\tau)$ be any signature. The matrix $M(s_1, \ldots, s_\tau)$ is block diagonalizable with blocks of size $(r_1, \ldots, r_\tau) \times (s_1, s_2, \ldots, s_\tau)$.

We now lower bound the rank of each block in the block diagonalization.

Lemma 7. For a good signature $s = (s_1, \ldots, s_\tau)$, and the corresponding $(r_1, \ldots, r_\tau)$ as in Definition 3, $\sum_{s'_1 \geq s_1, \ldots, s'_{\tau-1} \geq s_{\tau-1}} (s'_1 s_2 \ldots s_{\tau-1} r - \sum_{i=1}^{\tau-1} s'_i) (1 + O(1)) = (s_1 s_2 \ldots s_{\tau-1} r - \sum_{i=1}^{\tau-1} s_i) (1 + O(1))$.

Lemma 8 (Main Technical lemma). Fix any good signature $(s_1, \ldots, s_\tau)$. The rank of any diagonal block of $M(s_1, \ldots, s_\tau)$ is $(s_1 s_2 \ldots s_{\tau+k}) (1 - o(1))$.

Proof Sketch. Let $M'$ be a diagonal block of the matrix $M(s_1, \ldots, s_\tau)$. Recall from Lemma 6 that such a diagonal block is defined by a monomial $\bar{w}$ and a subset $R \subseteq [N]$. Rows of this block are labelled with all monomials $w \equiv [\bar{w}, R_1, \ldots, R_\tau]$ such that $R_1 \cup \ldots \cup R_\tau = R$ and columns of this block are labelled with all polynomials $m : p_T$ where $m \equiv [\bar{w}, S_1, \ldots, S_\tau]$ is such that $T \cup S_1 \cup \ldots \cup S_\tau = R$. First, we set up some notation.

For a partition $B = (B_1, \ldots, B_p)$ of $R$, let $b = (b_1, \ldots, b_p)$ be the tuple of part sizes, $b_i = |B_i|$. We say that $b$ is the signature of $B$.

We say $(a_1, \ldots, a_p) \preceq (b_1, \ldots, b_p)$ if $a_i \leq b_i$ for all $i \in [p]$, and $(a_1, \ldots, a_p) \prec (b_1, \ldots, b_p)$ if $(a_1, \ldots, a_p) \preceq (b_1, \ldots, b_p)$ but $(a_1, \ldots, a_p) \neq (b_1, \ldots, b_p)$.

Define the following collections of partitions of $R$: $X = \{ \bar{R} = (R_1, \ldots, R_\tau) \mid \text{signature}(\bar{R}) = (r_1, \ldots, r_\tau) \}, \ Y = \{ \bar{S} = (S_1, \ldots, S_\tau, T) \mid \text{signature}(\bar{S}) = (s_1, \ldots, s_\tau, k) \}, \ Z' = \{ \bar{Q} = (Q_1, \ldots, Q_\tau) \mid \text{signature}(\bar{Q}) = (q_1, \ldots, q_\tau); (s_1, \ldots, s_\tau-1) \preceq (q_1, \ldots, q_{\tau-1}) \}, \ Z = \{ \bar{Q} = (Q_1, \ldots, Q_\tau) \mid \text{signature}(\bar{Q}) = (q_1, \ldots, q_\tau); (s_1, \ldots, s_\tau-1) \prec (q_1, \ldots, q_{\tau-1}) \}.

Note that $|X| = (s_1 s_2 \ldots s_{\tau+k})$. Also, $Z' \setminus Z$ is precisely $X$. By Lemma 7, $|Z'| = |X| (1 + o(1))$. Hence $|Z| = |X| o(1)$.

The rows and columns of $M'$ are indexed by elements of $X$ and $Y$ respectively (Lemma 5). Let $I$ be the identity matrix with rows/columns indexed by elements
of $X$. We define two auxiliary matrices $M_1$ and $M_2$ as follows. The rows and columns of $M_1$ are indexed by elements of $X$. The entries of $M_1$ are in $\{0, 1\}$ and are defined as follows:

$$M_1(\tilde{R}, \tilde{R}') = \begin{cases} 1 & \text{if } R'_i \subseteq R_i \cup R_{i+1} \text{ for each } i \in [\tau - 1] \\ 0 & \text{otherwise.} \end{cases}$$

The rows and columns of $M_2$ are indexed by elements of $X$ and $Z$ respectively. The entries of $M_2$ are in $\{0, 1\}$ and are defined as follows:

$$M_2(\tilde{R}, \tilde{Q}) = \begin{cases} 1 & \text{if } Q_i \subseteq R_i \cup R_{i+1} \text{ for each } i \in [\tau - 1] \\ 0 & \text{otherwise.} \end{cases}$$

Our proof proceeds as follows:

1. Show that the columns of $M'$ and $M_2$ together span the columns of $M_1$.
2. Show that the columns of $M_1$ and $M_2$ together span the columns of $I$.

It then follows that

$$\text{rank}(M') \geq \text{rank}(M_1) - \text{rank}(M_2) \geq \text{rank}(I) - 2\text{rank}(M_2) \geq |X|(1 - o(1))$$

which is what we had set out to prove.

To prove steps 1 and 2, we describe columns of $M'$, $M_1$, $M_2$, $I$ using functions that express whether two partitions are related in a certain way. In particular, we express the inclusion relations described in Lemma 5, which characterise the non-zeroes in $M'$. The functions we use are multivariate polynomials, whose evaluations at the characteristic vectors of row indices give the entries in the rows. A careful choice of a small basis for these functions yields the result (details are omitted in this short version).

Lemma 3 can now be proved using the block-diagonal decomposition (Lemma 6) and the rank lower bound (Lemma 8).

### 3.4 Putting it together

We now have all the ingredients to establish that the shifted partial derivative measure of $S_N^d$ is large.

**Proof.** (of Theorem 1) By Lemma 4, $\dim(\partial_k S_N^d)_{\leq \ell} \geq \dim(\text{span}(P))$. This in turn is at least as large as $\text{rank}(M(S))$, since $M(S)$ is a submatrix of the matrix that describes a basis for $P$. We now choose a well-separated set of good signatures $S$ and apply Lemmas 4, 5, and 2 to lower bound the rank of $M(S)$. This will allow us our lower bound on $\dim(\partial_k S_N^d)_{\leq \ell}$.

Let us see how to choose $S$. Let $S_0$ denote the set of all good signatures. For integers $d_1, \ldots, d_\tau \in [2d]$, denote by $S(d_1, \ldots, d_\tau)$ the signatures $(s_1, \ldots, s_\tau) \in S_0$ such that $s_i \equiv d_i \pmod{2d}$ for each $i \in [\tau]$. It is easily checked that for any choice of $d_1, \ldots, d_\tau \in [2d]$, the set of signatures $S(d_1, \ldots, d_\tau)$ is well separated. Since there are $(2d)^\tau$ choices for $d_1, \ldots, d_\tau$, there must be one such that

$$|\mathcal{M}_{\mathcal{N}}(S(d_1, \ldots, d_\tau))| \geq \frac{|\mathcal{M}_{\mathcal{N}}(S_0)|}{(2d)^\tau}. \quad (2)$$
We fix \(d_1,\ldots,d_r\) so that the above holds and let \(S = S(d_1,\ldots,d_r)\). This is the set of signatures we will consider.

By Lemma 3, we know that the rank of \(M(S)\) is equal to the sum of the ranks of \(M(s)\) for each \(s \in S\). Hence, by Lemma 3, we have

\[
\text{rank}(M(S)) \geq (1 - o(1)) \cdot \sum_{s \in S} \frac{|\mathcal{P}(s)|}{(s+\ell)}
\]

Recall that \(\mathcal{P}(s) = \{m \cdot p_T \mid m \in M_N^\ell(s), |T| = k, \text{supp}(m) \cap T = \emptyset\}\). Hence, for each choice of \(m \in M_N^\ell(s)\), the number of choices of \(T\) such that \(m \cdot p_T \in \mathcal{P}(s)\) is given by \((N - |\text{supp}(m)|) \geq \binom{N-\ell}{k}\). Adding over all \(m \in M_N^\ell(s)\), we see that \(|\mathcal{P}(s)| \geq |M_N^\ell(s)| \cdot \binom{N-\ell}{k}\).

Plugging this bound into (3), we see that

\[
\text{rank}(M(S)) \geq (1 - o(1)) \cdot \frac{\left(\sum_{m \in M_N^\ell(s)} |M_N^\ell(s)| \cdot \binom{N-\ell}{k}\right)}{(3\sqrt{N}/2)}
\]

where the second inequality follows from the fact that all signatures \(s \in S\) are good, so \(s_\ell \leq 3s_\ell/2\) and hence \(s_\ell^k \leq (3s_\ell/2)^k \leq (3\sqrt{N}/2)^k\) (using Lemma 2); the final inequality is a consequence of Equation (2).

Finally, by Lemma 2 (see also Remark 2), \(|M_N^\ell(S_0)| \geq |M_N^\ell|(1 - o(1)) \geq (1 - o(1)) \cdot \binom{N}{\ell}\), which along with the above computation yields the claimed lower bound on \(\text{rank}(M(S))\). \(\square\)

**Remark 3.** For any multilinear polynomial \(F(X)\) on \(N\) variables, the quantity \(\dim(\partial_k F)_{\leq \ell}\) is at most the number of monomial shifts — which is \(\binom{N}{\ell}\) — times the number of possible partial derivatives of order \(k\), which is at most \(\binom{N}{k}\). Our result says that this trivial upper bound is (in some sense) close to optimal for the polynomial \(S_N^d\) (the \((\sqrt{N})^k\) factor in the denominator can be made \(N^{\varepsilon k}\) for any constant \(\varepsilon > 0\), see the discussion at the end of the proof of Theorem 2).

All previous lower bound results using the shifted partial derivative method also obtain similar statements [11,12,13].

### 4 Lower bound on the size of depth four formulas

In this section, we establish the lower bounds claimed in Theorem 2 As in [5], we say that a \(\Sigma\Pi\Sigma\Pi\) formula \(C\) is a \(\Sigma\Pi^{(D)}\Sigma\Pi^{(\emptyset)}\) formula if the product gates at level 1 (just above the input variables) have fan-in at most \(t\) and the product gates at level 3 have fan-in bounded by \(D\).

The following is implicit in [5] and is stated explicitly in [9].
Lemma 9 ([9], Lemma 4). Let $P$ be a polynomial on $N$ variables computed by a $\Sigma \Pi^{[D]} \Sigma \Pi^{[t]}$ circuit of top fan-in $s$. Then, we have

$$\dim(\langle \partial_k P \rangle \leq \ell) \leq s \cdot \binom{D + \ell}{k} \cdot \left( \frac{N + \ell + (t-1)k}{\ell + (t-1)k} \right).$$

Proof. (of Theorem 2) For notational simplicity, we present the proof only for $\varepsilon = 1/4$. Let $\tau = 4t + 1$. Let $k = d/(\tau + 1)$ and $\delta = 1/(2\tau)$. Assume there exists a $\Sigma \Pi^{[D]} \Sigma \Pi^{[t]}$ circuit computing $S^d_N$. When $N$ is large enough, from Theorem 1 and Lemma 9 it holds that

$$s \geq \frac{N^{\ell-k}}{\binom{D}{k}} \cdot \binom{N + \ell + (t-1)k}{\ell + (t-1)k} \cdot \frac{1 - o(1)}{(3\sqrt{N}/2)^k (2d)^\tau}.$$ 

For large enough $N$, we have $\frac{N^{\ell-k}}{\binom{D}{k}} \geq \left( \frac{N^{\ell-k}}{2D} \right)^k \geq \left( \frac{N}{2D} \right)^k$. Since $kt < d = o(\lg N)$, for large enough $N$ we have $\frac{N + \ell + (t-1)k}{\ell + (t-1)k} \geq \frac{\ell}{N + \tau} \geq \left( \frac{1}{2\tau} \right)^k \geq N^{-\delta t k - o(k)}$. Note that $(2d)^\tau \leq (\lg N)^\tau \leq N^{1/10}$.

Putting it all together, we have obtained that asymptotically,

$$s \geq \left( \frac{N}{2D \cdot N^\delta t + o(1)} \cdot (3\sqrt{N}/2) \right)^k \cdot \frac{1 - o(1)}{N^{1/10}} \geq \frac{1}{N^{1/10}} \cdot \left( \frac{N^{1/2 - \delta t}}{3D \cdot N^{o(1)}} \right)^k.$$ 

(4)

By our choice of parameters, $t \leq \tau/4$ and hence $1/2 - \delta t \geq 1/2 - 1/8 = 3/8$. Since $D \leq N^{1/4}$, we see that (4) yields a lower bound of $N^{\Omega(k)} = N^{\Omega(d/t)}$.

It is not hard to see that the above proof idea (with some changes in parameters) can be made to give lower bounds of $N^{\Omega(d/t)}$ for $D \leq N^{1-\varepsilon}$ for any constant $\varepsilon > 0$. Specifically, choose $\tau = \Theta \left( \frac{1}{\varepsilon} \right)$ (instead of $4t + 1$) and $\delta$ such that $\delta \tau = 1 - \Theta(\varepsilon)$ (instead of $\frac{1}{2}$) in the entire proof. We omit the details. \hfill \Box

References

1. M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In FOCS, pages 67–75, 2008.
2. N. Alon. Perturbed identity matrices have high rank: Proof and applications. Comb. Probab. Comput., 18(1-2):3–15, Mar. 2009.
3. W. Baur and V. Strassen. The complexity of partial derivatives. Theor. Comput. Sci., 22:317–330, 1983.
4. H. Fournier, N. Limaye, G. Malod, and S. Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In Symposium on Theory of Computing, STOC, pages 128–135, 2014.
5. A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Approaching the chasm at depth four. In Conference on Computational Complexity (CCC), 2013.
6. P. Hrubes and A. Yehudayoff. Homogeneous formulas and symmetric polynomials. Computational Complexity, 20(3):559–578, 2011.
7. N. Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:81, 2012.

8. N. Kayal, N. Limaye, C. Saha, and S. Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. In *Foundations of Computer Science (FOCS)*, 2014.

9. N. Kayal, C. Saha, and R. Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *STOC*, pages 146–153, 2014.

10. P. Keevash and B. Sudakov. Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. *SIAM Journal on Discrete Mathematics*, 18(4):713–727, 2005.

11. P. Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.

12. M. Kumar and S. Saraf. The limits of depth reduction for arithmetic formulas: it’s all about the top fan-in. In *STOC*, pages 136–145, 2014.

13. M. Kumar and S. Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *FOCS*, pages 364–373, 2014.

14. E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

15. N. Nisan and A. Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997.

16. R. Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006.

17. R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2), 2009.

18. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

19. R. Saptharishi. *Unified Approaches to Polynomial Identity Testing and Lower Bounds*. PhD thesis, Chennai Mathematical Institute, 2013.

20. A. Shpilka. Affine projections of symmetric polynomials. *J. Comput. Syst. Sci.*, 65(4):639–659, 2002.

21. A. Shpilka and A. Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

22. S. Tavenas. Improved bounds for reduction to depth 4 and 3. In *Mathematical Foundations of Computer Science (MFCS)*, 2013.

23. L. G. Valiant. Completeness Classes in Algebra. In *11th ACM symposium on Theory of Computing (STOC)*, pages 249–261, New York, NY, USA, 1979.

24. L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM J. Comput.*, 12(4):641–644, 1983.

25. R. M. Wilson. A diagonal form for the incidence matrices of $t$-subsets vs. $k$-subsets. *European Journal of Combinatorics*, 11(6):609 – 615, 1990.