THEORY OF COMPUTING, Volume 13 (9), 2017, pp. 1–34
www.theoryofcomputing.org

The Shifted Partial Derivative Complexity of Elementary Symmetric Polynomials

Hervé Fournier* † Nutan Limaye* Meena Mahajan*
Srikanth Srinivasan*

Received September 29, 2015; Revised September 1, 2016; Published September 25, 2017

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 $d < \lg N / \lg\lg N$. 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 the homogeneous $\Sigma\Pi\Sigma$ model (i. e., $\Sigma\Pi\Sigma\Pi$ circuits with bottom fan-in 1).

Our main technical lemma gives a lower bound for the ranks of certain inclusion-like matrices.

ACM Classification: F.1.3, F.2.2

AMS Classification: 68Q17, 68Q15, 68Q25

Key words and phrases: arithmetic circuits, depth-4 circuits, inclusion matrices

---

An extended abstract of this paper appeared in the Proceedings of the 40th International Symposium on Mathematical Foundations of Computer Science, 2015 [5].

*Supported by IFCPAR/CEFIPRA Project No 4702-1(A).

†Supported by ANR project CompA (project number: ANR-13-BS02-0001-01).

 DOI: 10.4086/toc.2017.v013a009

1 Introduction

1.1 Motivation

In an influential paper of Valiant [26] 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 [26] 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 $\Omega(N \lg N)$ [4] 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 [19, 21, 20].

A very interesting such subclass of arithmetic circuits is the class of *bounded-depth* arithmetic circuits. The question of proving lower bounds for bounded-depth circuits and in particular depth-3 and depth-4 circuits has received a lot of attention subsequent to the recent progress in efficient depth reduction of arithmetic circuits [27, 1, 15, 25]. This sequence of results essentially implies that "strong enough" lower bounds for depth-4 homogeneous circuits suffice to separate VP from VNP. More formally, it proves that any sequence $\{f_N\}_N$ of homogeneous N-variate polynomials of degree $d = N^{O(1)}$ in VP has depth-4 homogeneous circuits of size $N^{O(\sqrt{d})}$. Hence, proving an $N^{\omega(\sqrt{d})}$ lower bound for depth-4 homogeneous circuits suffices to separate VP from VNP.

Even more can be said about the depth-4 circuits obtained in the above results. For any integer parameter $t \leq d$, they give a $\Sigma\Pi\Sigma\Pi$ circuit 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 Π gates with fan-in $O(d/t)$. We will refer to such circuits as $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits. The depth-reduction results mentioned above produce a depth-4 homogeneous $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit of size $N^{O((d/t)+t)}$ 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 [11], who augmented the *partial derivative method* of Nisan and Wigderson [19] to devise a new complexity measure called *the shifted partial derivative measure*. He used this measure to prove an exponential lower bound for a special class of depth-4 circuits. Building on this, the first non-trivial lower bound for $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits was proved by Gupta, Kamath, Kayal, and Saptharishi [9] for the determinant and permanent polynomials. This was further improved by Kayal, Saha, and Saptharishi [13] who gave a family of explicit polynomials in VNP for which the shifted partial derivative complexity is (nearly) *as large as possible*[1] and hence showed a lower bound of $N^{\Omega(d/t)}$ for the top fan-in of $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing these polynomials. Later, a similar result for a polynomial in VP was proved in [6] and this was subsequently strengthened by Kumar and Saraf [17], who gave a polynomial computable by homogeneous $\Pi\Sigma\Pi$ circuits such that any $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing it must have top fan-in $N^{\Omega(d/t)}$. Finally, using a variant of the shifted partial derivative measure, Kayal et al. [12] and Kumar and Saraf [16] were able to prove similar lower bounds for general depth-4 homogeneous circuits as well.

[1] i. e., as large as it can be for a "generic" or "random" polynomial (as explained after Theorem 1.1).

In this paper, we investigate the shifted partial derivative measure of the *elementary symmetric polynomials*, a natural family of polynomials whose complexity has been the focus of much previous work [19, 24, 23, 10]. Nisan and Wigderson [19] proved tight lower bounds on the depth-3 homogeneous circuit complexity of these polynomials. Shpilka and Wigderson [24] and Shpilka [23] 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 [24]) is tight.

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 $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit of top fan-in less than $N^{\Omega(d/t)}$. This strengthens the result of Nisan and Wigderson [19] 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.

1.2 Our results

We show that, in 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 1.1. *Let $\alpha \in (0,1/2)$ be a constant. Let $N,d,k,\tau \in \mathbb{N}$ be such that $4k \le d \le \alpha \lg N/\lg\lg N$, $k = \lfloor d/(\tau+1) \rfloor$, and $\tau \ge 3$ is an odd number. Over a field of characteristic zero, for any δ satisfying $\alpha \le 1-\delta(\tau+1) < 1-\delta\tau \le 1-\alpha$, and for $\ell = \lfloor N^{1-\delta} \rfloor$, the following holds.*

$$\dim\langle \partial_k S_N^d \rangle_{\le \ell} \geqslant \frac{(1-o(1)) \cdot \binom{N+\ell}{\ell} \cdot \binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k \cdot (d+1)^\tau} .$$

Here the $o(1)$ term goes to 0 as $N \to \infty$. For any multilinear polynomial $F(X)$ on N variables, the quantity $\dim\langle \partial_k F \rangle_{\le \ell}$ is at most the number of monomial shifts—which is $\binom{N+\ell}{\ell}$—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 $(N^{1-\delta\tau})^k$ factor in the denominator can be made $N^{\varepsilon k}$ for any constant $\varepsilon > 0$. To get an arbitrarily small exponent $1-\delta\tau$, the constant α must be small. For, say, $d \in o(\lg N/\lg\lg N)$, and for sufficiently large N, any $\alpha > 0$ works.) All previous lower bound results using the shifted partial derivative method also obtain similar statements [9, 6, 17, 16].

To illustrate the ideas of the proof, we first prove a special case of Theorem 1.1, namely the case when $2k$ divides d and τ equals $d/k-1$ (all other conditions on the parameters are the same). This result is stated in Theorem 3.1 in Section 3. (Note that in the statement of Theorem 1.1, the values of d,k do not fix τ. E. g., for $(d,k) = (40,2)$, τ can be 13, 15, 17, or 19. The proof of Theorem 3.1, as given, works when τ is fixed to be $d/k-1$.) In Section 4, we describe the modifications needed to carry out the proof when $k = \lfloor d/(\tau+1) \rfloor$ for some odd number $\tau \ge 3$, establishing Theorem 1.1.

The proof of Theorem 1.1 requires that the field over which the polynomial and the circuits are defined should have characteristic 0. In Section 5 we state Theorem 5.1, our most general result, i. e., for general parameters (no divisibility constraint) and over any characteristic, and sketch its proof. Due to the positive characteristic, in Theorem 5.1 we incur a loss of $k+1$ in the denominator (compared to Theorem 1.1).

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

Theorem 1.2. *Let $\varepsilon \in (0,1)$ be a constant. Let $N,d,D,t \in \mathbb{N}$ be such that*

$$\frac{10t}{\varepsilon} \le d \le \frac{\varepsilon \lg N}{5 \lg\lg N}, \qquad D \le N^{1-\varepsilon}.$$

Any $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ 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 fan-in D bounded by $O(d/t)$. Our lower bound works for potentially much larger values of D.

It is known that for every $d \le N$, the elementary symmetric polynomial S_N^d has depth-4 circuits of size $N2^{O(\sqrt{d})}$ [24, Theorem 5.2]. A closer look at the construction there shows that the $N2^{O(\sqrt{d})}$ size circuits are $\Sigma\Pi^{[O(d)]}\Sigma\Pi^{[d]}$ circuits with bottom fan-in up to *and including d*. In contrast, our bound shows that for small d, if the bottom fan-in t is restricted to be a fraction of d, then, even allowing for much larger D, the top fan-in and hence size of an $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ circuit shoots up to $N^{\Omega(d/t)}$.

As a corollary to Theorem 1.2, we obtain a lower bound for homogeneous depth-4 circuits with bounded bottom fan-in.

Corollary 1.3. *Let parameters N,d,t be as in Theorem 1.2. Any $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit 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, Nisan and Wigderson $\binom{N}{d}/2^d$ for all d, however with respect to homogeneous $\Sigma\Pi\Sigma$ circuits (i. e., the case $t = 1$).

1.3 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 paper, we analyse the matrix arising from the shifted partial derivative space of the symmetric polynomials.

Our analysis is quite different from those in previous papers (such as [6, 12, 16]), 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]. Neither of these techniques seems to be applicable in our case. This is already visible from the work of Nisan and Wigderson [19], who analyse the *partial derivative matrix* (without shifts) of the elementary symmetric polynomials. This matrix turns out to be the well-known *Disjointness matrix*, defined as follows: for fixed parameters $N,s,t \in \mathbb{N}$ such that $s+t \le N$, the rows and columns of this matrix are labelled by subsets of $[N]$ of size s and t respectively; the (S,T) entry in the matrix is 1 if $S \cap T = \emptyset$ and 0 otherwise.[2] It is known (see [18, Sec. 2.3] for example) that this matrix is full rank (i. e., has rank equal to the minimum of the number of rows and columns) in characteristic 0 and almost full rank in other characteristics [28]. However, it is not clear how to use either of the two techniques mentioned above to prove this result.

[2]Variants allowing sets of size *at most s* and *t* have also been considered.

In our analysis of the shifted partial derivative space, we block-diagonalize our matrix[3] into matrices each of which is a more complicated version of the *Inclusion matrix* (similar to the Disjointness matrix mentioned above and also known to be full rank), and bound its rank from below by using a technique that, to the best of our knowledge, has not been used in this context before.

Our method of bounding the rank of the matrices, in spirit, resembles the techniques in the papers [7, 8] by Frankl and Wilson. In these papers, the authors consider the problem of bounding the sizes of some interesting families of sets. They reduce this problem to analyzing ranks of certain classes of matrices and use *the linear algebra method* [3] to bound the ranks. Though our method seems similar to these techniques in a general sense, there are some differences. In their work, the results pertaining to ranks of the matrices are extremal in nature. However, our matrix is very concrete. On the other hand, the matrices considered in their work are simpler than the matrix we obtain while analyzing the rank of the shifted partial derivative space of elementary symmetric polynomials. We give a brief overview of our technique in the next section.

Disjointness and inclusion matrices arise naturally in other branches of theoretical computer science such as Boolean circuit complexity [22], communication complexity [18, Chapter 2] and also in combinatorics [28, 14]. Therefore, we believe that our analysis of the inclusion-like matrix arising from the symmetric polynomial may find other applications.

1.4 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 3.1. In Section 3 we give the details of the actual proof. The circuit size lower bound from Theorem 1.2 is established in Section 6. Recall that Theorem 3.1 is stated over fields of characteristic zero when some parameters exhibit some divisibility property (which makes the combinatorics nicer). The way to handle more general parameters is explained in Section 4. In Section 5 we describe how to extend our results to arbitrary fields.

2 Proving Theorem 3.1: high-level outline

2.1 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 X of variables is defined as

$$S_N^d(X) = \sum_{A \subseteq [N], |A| = d} X_A\,,$$

and 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

$$\langle \partial_k f \rangle_{\leq \ell} = \operatorname{span} \left\{ x_1^{j_1} \ldots x_n^{j_n} \cdot \frac{\partial^k f}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}} \;\middle|\; i_1 + \cdots + i_n = k,\ j_1 + \ldots + j_n \leq \ell \right\}.$$

[3]Actually, we only work with a carefully chosen submatrix of the overall matrix.

Our complexity measure is the dimension of this space, i. e., $\dim(\langle \partial_k f \rangle_{\leq \ell})$ [11, 9].

For a monomial $m = \prod_{i=1}^{N} x_i^{n_i}$, $\deg(m) = n_1 + n_2 + \cdots + n_N$ is the total degree of m. We denote by $\deg_{x_i}(m)$ the degree of the variable x_i in m (here $\deg_{x_i}(m) = n_i$). We define the support of m as

$$\mathrm{supp}(m) = \{i \in [N] \mid n_i > 0\}.$$

For a monomial m and $p > 0$, let

$$\mathrm{supp}_p(m) = \{i \in [N], \deg_{x_i}(m) = p\}.$$

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

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

Given $p > 0$, a monomial $m \in \mathcal{M}_N^{\ell}$ can be uniquely written as

$$m = \tilde{m} \cdot \prod_{i=1}^{p} (X_{\mathrm{supp}_i(m)})^i.$$

We write $m \equiv [\tilde{m}, S_1, \ldots, S_p]$ if $S_i = \mathrm{supp}_i(m)$ 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 (see Section 5), but the first step of our proof (Lemma 3.2) becomes a little more cumbersome.

2.2 Proof outline

We want to lower bound $\dim(\langle \partial_k S_N^d \rangle_{\leq \ell})$ for suitable k, ℓ. An alternate way of looking at the vector space $\langle \partial_k S_N^d \rangle_{\leq \ell}$ is as follows. We fix some spanning set $\mathcal{S}$ for the set of all partial derivatives of S_N^d of order k and consider the set $\mathcal{P}$ of all the polynomials obtained by multiplying the polynomials in $\mathcal{S}$ with monomials of degree at most ℓ. We define a matrix M whose columns contain the polynomials in the set $\mathcal{P}$ (seen as vectors of coefficients of the various monomials). Lower bounding $\dim(\langle \partial_k S_N^d \rangle_{\leq \ell})$ is equivalent to lower bounding $\mathrm{rank}(M)$.

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

Step 1: We choose a suitable subset $\mathcal{S}$ 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_N^2 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_js; explicitly, one can write

$$x_i = \frac{1}{n-1}\left(\sum_{j\in[n]} r_j\right) - r_i\,.$$

Since the x_is 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 3.2).

Step 2: After choosing the set $\mathcal{S}$, we construct the set $\mathcal{P}$ of shifts of $\mathcal{S}$ (actually, we will only consider a subset of $\mathcal{P}$) 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 $\mathcal{M}$ and project each polynomial in $\mathcal{P}$ 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 $\mathcal{M}$, 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 Lemma 3.13, Lemma 3.16, and Lemma 3.17.

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 illustrate this technique here with a toy example. Fix parameters $N,s\in\mathbb{N}$ with $s\le N/2$ and define the $\binom{N}{s}\times\binom{N}{s}$ matrix $\mathrm{Disj}_{N,s}$ whose rows and columns are labelled by sets of size s from the universe $[N]$ and the (S,T) entry is 1 if $S\cap T=\emptyset$ and 0 otherwise. We can similarly also define the $\binom{N}{s}\times\binom{N}{s}$ matrix $\mathrm{Inc}_{N,s}$ similarly with the only difference being that the (S,T) entry is 1 if and only if $S\subseteq T$; note that $\mathrm{Inc}_{N,s}$ is simply the identity matrix with the required dimensions and is hence clearly full rank. It is also known that $\mathrm{Disj}_{N,s}$ is full rank over fields of characteristic 0 (see, e. g., [18, Chapter 2]). We prove a weaker statement here in order to illustrate our proof method: we show that when $s=o(N)$, $\mathrm{Disj}_{N,s}$ has rank $\binom{N}{s}(1-o(1))$.

To see this, consider the following alternate way of looking at the above matrices. We identify the labels of the rows—which are elements of $\binom{[N]}{s}$—with their characteristic vectors, which are elements of the N-dimensional hypercube $\{0,1\}^N$ of weight exactly s. Each column is associated with a polynomial p over 0-1 variables $y_1,\ldots,y_N$ such that the entry in the column at the row labelled by $a\in\{0,1\}^N$ is equal to $p(a)$. Specifically, in the matrix $\mathrm{Inc}_{N,s}$, a column labelled $T\subseteq[N]$ is associated with the monomial $m_T=\prod_{i\in T} y_i$; it should be clear that this monomial evaluates to 1 at row a only if a encodes a subset contained in T (which must be T itself). Similarly, in the matrix $\mathrm{Disj}_{N,s}$ the column corresponding to T is associated with the polynomial $q_T=\prod_{i\in T}(1-y_i)$. Now consider the following simple identity.

$$m_T=\prod_{i\in T} y_i=\prod_{i\in T}(1-(1-y_i))=\sum_{T'\subseteq T}(-1)^{|T'|}q_{T'}\,.$$

The above tells us that the columns of $\mathrm{Inc}_{N,s}$ are spanned by the set of all column vectors corresponding to polynomials of the form $Q=\{q_{T'}\mid |T'|\le s\}$. Since $\mathrm{Inc}_{N,s}$ has rank $\binom{N}{s}$, the set of column vectors in Q must have rank at least $\binom{N}{s}$. The subset of these columns corresponding to $|T'|=s$ are exactly the columns of $\mathrm{Disj}_{N,s}$. Note that the remaining columns (corresponding to $|T'|<s$) are only $\sum_{i<s}\binom{N}{i}$ in number and this is only $o(\binom{N}{s})$ since $s=o(N)$. Hence, the columns of $\mathrm{Disj}_{N,s}$ must have rank at least $\binom{N}{s}(1-o(1))$.

The main technical lemma (Lemma 3.21) is a generalization of the above trick to our setting. Given the matrix whose rank we wish to lower bound (like $\mathrm{Disj}_{N,s}$ above), 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.

2.3 The main parameters

For proving Theorem 1.1, recall the parameters: $\alpha \in (0,1/2)$, N,d,k,τ satisfying

$$4k \leq d \leq \alpha \lg N/\lg\lg N, \quad k = \left\lfloor \frac{d}{\tau+1} \right\rfloor, \quad \tau \geq 3 \text{ odd}.$$

Our parameter choices are any δ satisfying $\alpha \leq 1-\delta(\tau+1)$ and $1-\delta\tau \leq 1-\alpha$, and $\ell = \lfloor N^{1-\delta} \rfloor$. The following are easy to verify for our choice of parameters.

Fact 2.1. *$\tau^2 = o(\ell)$, $\tau = o(N^{\delta})$, and $\tau\ell = o(N)$. Also, $(\lg N)^{\tau} = O(N^{\alpha})$, and $N^{\delta(\tau+1)} = O(N^{1-\alpha}) = o(N)$.*

These facts also hold in the settings of Theorem 3.1 and Theorem 5.1.

3 Proving Theorem 1.1: details of a simpler case

In this section we first establish Theorem 3.1 given below. This is a restriction of Theorem 1.1 to the special case when $2k$ divides d. This case illustrates all the ideas and technical constructs used. Small modifications, described in the next section, establish Theorem 1.1.

Theorem 3.1. *Let $\alpha \in (0,1/2)$ be a constant. Let $N,d,k \in \mathbb{N}$ be such that $4k \leq d \leq \alpha \lg N/\lg\lg N$ and $2k \mid d$. Over a field of characteristic zero, for $\tau = d/k-1$, for any δ satisfying $\alpha \leq 1-\delta(\tau+1) < 1-\delta\tau \leq 1-\alpha$, and for $\ell = \lfloor N^{1-\delta} \rfloor$, the following holds.*

$$\dim\langle \partial_k S_N^d \rangle_{\leq \ell} \geqslant \frac{(1-o(1)) \cdot \binom{N+\ell}{\ell} \cdot \binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k \cdot (d+1)^{\tau}}.$$

3.1 Choice of basis: Step 1 of the proof

Lemma 3.2. *Let $k \leq d \leq N$. Over fields of characteristic 0, the vector space spanned by the set of k-partial derivatives of S_N^d, that is, $\langle \partial_k S_N^d \rangle_{\leq 0}$, contains $\{p_T \mid T \subseteq [N], |T| = k\}$ where*

$$p_T = \sum_{T \subseteq A \subseteq [N], |A| = d-k} X_A = X_T \cdot S_{N-k}^{d-2k}(X \setminus T).$$

Proof. If $d < 2k$ then the statement is vacuously true (there is no such p_T). So now assume that $d \geq 2k$.

For any set $S \subseteq [N]$ of size $|S| = k$, let r_S be the polynomial obtained by deriving S_N^d with respect to all the variables in S (once each). Note that r_S contains all degree-$(d-k)$ multilinear monomials that *avoid* variables x_i for $i \in S$. Similarly, for $S \subseteq [N]$ of size $|S| < k$, define the polynomials r_S as sums of degree-$(d-k)$ multilinear monomials *avoiding* x_i for $i \in S$. Correspondingly, define the complementary polynomials: for $T \subseteq [N]$ with $|T| \le k$, p_T consists of all degree-$(d-k)$ multilinear monomials that *include* x_i $(i \in T)$. Formally, for $S, T \subseteq [N]$ with $|S|, |T| \le k$,

$$r_S(x) = \sum_{|A|=d-k,\ A\cap S=\emptyset} X_A; \quad p_T(x) := \sum_{T\subseteq B\subseteq[N],|B|=d-k} X_B\,.$$

The claim is that linear combinations of the partial derivative polynomials, r_S with $|S| = k$, generate the polynomials p_T $(|T| = k)$. We can show this in two steps.

The first step is to show that the $r_{S'}$ $(|S'| = k)$ generate the polynomials r_S $(|S| \le k)$. Fix any S with $|S| = s < k$. Let $\mathcal{F} = \{S' \subseteq [N] \mid |S'| = k, S' \supseteq S\}$. A simple computation shows that any monomial that avoids all the variables in S appears in the same positive number M of all the polynomials $r_{S'}$ $(S' \in \mathcal{F})$ (in fact, it can be checked that $M = \binom{N-(d-k+s)}{k-s}$, though this will not be important for us). Hence, we have $r_S = (1/M)\sum_{S'\in\mathcal{F}} r_{S'}$, which shows that r_S is indeed in the span of $r_{S'}$ $(|S'| = k)$. Note that this step assumes that we are working in characteristic 0.

Finally, once we have r_S for all set sizes in $[k]$, we generate p_T for $|T| = k$ using inclusion-exclusion. The monomials of P_T are the monomials of S_N^d, minus monomials that miss at least one element of T, plus the monomials that miss at least two elements of T (since we overcounted), and so on. The monomials to be added/removed are exactly the monomials of various polynomials r_S. To set this up formally, let $\{0,1\}_t^N$ denote the set of all 0-1 vectors of Hamming weight exactly t. We use the natural correspondence between this set and the set of all subsets of $[N]$ of size exactly t.

$$\begin{aligned}
p_T(x) &= \sum_{T\subseteq B\subseteq[N],|B|=d-k} X_B \\
&= \sum_{y\in\{0,1\}_{d-k}^N} \left(\prod_{i:y_i=1} x_i\right)\left(\prod_{i\in T} y_i\right) = \sum_{y\in\{0,1\}_{d-k}^N} \left(\prod_{i:y_i=1} x_i\right)\left(\prod_{i\in T}(1-(1-y_i))\right) \\
&= \sum_{y\in\{0,1\}_{d-k}^N} \left(\prod_{i:y_i=1} x_i\right)\left(\sum_{S\subseteq T}(-1)^{|S|}\prod_{i\in S}(1-y_i)\right) \\
&= \sum_{S\subseteq T}(-1)^{|S|} \sum_{y\in\{0,1\}_{d-k}^N} \left(\prod_{i:y_i=1} x_i\right)\left(\prod_{i\in S}(1-y_i)\right) = \sum_{S\subseteq T}(-1)^{|S|} r_S(x)\,.
\end{aligned}$$

This shows that each p_T $(|T| = k)$ is a linear combination of the r_S $(|S| \le k)$ and hence also of the derivative polynomials $r_{S'}$ $(|S'| = k)$. □

Let

$$\mathcal{P} = \left\{ m \cdot p_T \,\middle|\, T \subseteq [N], |T| = k, m \in \mathcal{M}_N^{\ell}, \mathrm{supp}(m) \cap T = \emptyset \right\}.$$

From Lemma 3.2, $\mathcal{P} \subseteq \langle \partial_k S_N^d \rangle_{\le \ell}$. Hence, a lower bound on the dimension of span $\mathcal{P}$ is also a lower bound on $\dim(\langle \partial_k S_N^d \rangle_{\le \ell})$.

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 $|\mathrm{supp}_i(m)|$ 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 3.3. For $i \in [\tau]$, $\hat{s}_i$ denotes the average number of variables with degree exactly i. That is,

$$\hat{s}_i = \mathbb{E}_{m \sim \mathcal{U}(\mathcal{M}_N^\ell)} \left[|\operatorname{supp}_i(m)| \right] .$$

Definition 3.4 (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 $m \in \mathcal{M}_N^\ell(s_1, \ldots, s_\tau)$. We call the signature $(s_1, \ldots, s_\tau)$ a *good signature* if for each $i \in [\tau]$, we have $\hat{s}_i/2 \le s_i \le 3\hat{s}_i/2$. Let $\mathcal{S}_0$ denote the set of all good signatures.

The following lemma shows that for our choice of parameters, the average values $\hat{s}_i$ for $i \in [\tau+1]$ are significantly large, and most monomials in $\mathcal{M}_N^\ell$ in fact have good signatures. This lemma imposes the most stringent condition on d, τ: if τ is too large, then we cannot use the bounds from Fact 2.1, we cannot get tight estimates on $\hat{s}_i$, and hence we cannot show that most monomials have good signatures.

Lemma 3.5. *For our choice of the main parameters, the following statements hold.*

1. *For* $i \in [\tau - 1]$, $\dfrac{\hat{s}_i}{\hat{s}_{i+1}} \ge N^\delta$.
2. *For* $i \in [\tau]$, $\hat{s}_i = N^{1-i\delta}(1 - o(1))$.
3. $\Pr[s(m) \in \mathcal{S}_0] = 1 - o(1)$.

Proof. Pick a random monomial of degree at most ℓ uniformly at random; $m \sim \mathcal{U}(\mathcal{M}_N^\ell)$.

Consider the following random variables.

1. For $i \in [\tau+1]$ and $j \in [N]$, $Z_{i,j}$ denotes the 0-1 random variable that is 1 if and only if the variable x_j has degree *at least* i in m.
   Let p_i denote $\mathbb{E}_m[Z_{i,j}]$. (The average is the same for all j.)
2. For $i \in [\tau+1]$, let $Z_i = \sum_{j \in [N]} Z_{i,j}$ be the random variable denoting the number of variables that have degree at least i in m.
   Let μ_i denote $\mathbb{E}_m[Z_i]$; clearly, $\mu_i = Np_i$.
3. For $i \in [\tau]$, the number of variables of degree *exactly* i is given by $Y_i := Z_i - Z_{i+1}$.
   Then $\hat{s}_i = \mathbb{E}_m[Y_i]$; clearly, $\hat{s}_i = \mu_i - \mu_{i+1}$.

Consider p_i, the probability that the variable x_j has degree at least i in a monomial $m \sim \mathcal{U}(\mathcal{M}_N^\ell)$. We see that any such monomial can be written uniquely as $m = m'(x_j)^i$ where m' is a monomial of degree at most $\ell - i$ in the same set of variables X. Since the number of such monomials is exactly $|\mathcal{M}_N^{\ell-i}|$, we have

$$p_i = \frac{|\mathcal{M}_N^{\ell-i}|}{|\mathcal{M}_N^\ell|} = \frac{\binom{N+\ell-i}{\ell-i}}{\binom{N+\ell}{\ell}} .$$

In particular,

$$\frac{\mu_i}{\mu_{i+1}} = \frac{p_i}{p_{i+1}} = \frac{N+\ell-i}{\ell-i} \ge \frac{N+\ell}{\ell} = 1 + N^\delta \ge N^\delta .$$

Hence,

$$\frac{\hat{s}_i}{\hat{s}_{i+1}} = \frac{p_i - p_{i+1}}{p_{i+1} - p_{i+2}} \geq \frac{p_i - p_{i+1}}{p_{i+1}} = \frac{p_i}{p_{i+1}} - 1 \geq N^{\delta},$$

proving the first part of the lemma.

Next, we use a simple fact about binomial coefficients.

Fact 3.6. *For any $N, \ell, i \in \mathbb{N}$ such that $i < \ell$, we have*

$$\left(\frac{\ell - i}{N+\ell}\right)^i \leq \left(\frac{\ell - i}{N+\ell-i}\right)^i \leq \frac{\binom{N+\ell-i}{\ell-i}}{\binom{N+\ell}{\ell}} \leq \left(\frac{\ell}{N+\ell}\right)^i.$$

By Fact 3.6, we have

$$p_i \leq \left(\frac{\ell}{N+\ell}\right)^i \leq \left(\frac{\ell}{N}\right)^i \leq N^{-i\delta}; \quad \mu_i \leq N^{1-i\delta}; \quad \hat{s}_i \leq \mu_i \leq N^{1-i\delta}.$$

Also by Fact 3.6, we have

$$\begin{aligned} p_i &\geq \left[\left(\frac{\ell - i}{\ell}\right)\left(\frac{\ell}{N}\right)\left(\frac{N}{N+\ell}\right)\right]^i = N^{-i\delta}(1 - o(1))\left[\frac{(1 - \frac{i}{\ell})}{(1 + \frac{\ell}{N})}\right]^i \\ &\geq N^{-i\delta}(1 - o(1)) \exp\left(-O\left(\frac{i^2}{\ell} + \frac{i\ell}{N}\right)\right). \end{aligned}$$

By our choice of parameters (Fact 2.1), we have $i^2/\ell \leq (\tau+1)^2/\ell = o(1)$ and $i\ell/N = o(1)$. Hence, we have

$$p_i \geq N^{-i\delta}(1 - o(1)); \quad \mu_i \geq N^{1-i\delta}(1 - o(1)); \quad \hat{s}_i \geq \mu_i \left(1 - \frac{\mu_{i+1}}{\mu_i}\right) \geq N^{1-i\delta}(1 - o(1)).$$

Putting together the upper and lower bounds proves the second part of the lemma.

In order to prove the third part of the lemma, we use the second moment method. To do this, we will need to bound the second moment of Z_i. In order to do this, we will need the following claim.

Claim 3.7. *For any distinct j_1, j_2, we have $\mathbb{E}[Z_{i,j_1} Z_{i,j_2}] \leq \mathbb{E}[Z_{i,j_1}]\, \mathbb{E}[Z_{i,j_2}] = p_i^2$.*[4]

The claim is proved below. First, we use this claim to finish the proof. The following standard analysis will bound the second moment of Z_i.

$$\begin{aligned} \mathbb{E}\left[Z_i^2\right] &= \sum_{j_1, j_2} \mathbb{E}[Z_{i,j_1} Z_{i,j_2}] = \sum_{j_1} \mathbb{E}\left[Z_{i,j_1}^2\right] + \sum_{j_1 \neq j_2} \mathbb{E}[Z_{i,j_1} Z_{i,j_2}] \\ &\leq N p_i + \sum_{j_1 \neq j_2} p_i^2 \leq N p_i + N^2 p_i^2 = \mu_i + \mu_i^2, \end{aligned}$$

[4]This claim posits a weak form of negative dependence on the degrees of distinct variables. It is actually not too hard to prove much stronger forms of negative dependence which yield stronger probability estimates than the ones we give here. However, these estimates suffice for our purposes.

where the first inequality follows from the fact that $Z_{i,j_1}^2 = Z_{i,j_1}$ and Claim 3.7.

Hence, we can bound the variance of Z_i as follows.

$$\mathbb{E}\left[(Z_i-\mu_i)^2\right] = \mathbb{E}\left[Z_i^2\right] - \mu_i^2 \leq \mu_i\,.$$

Thus, by the Chebyshev inequality, we have

$$\Pr[|Z_i-\mu_i| > \mu_i/4] \leq \frac{\mathbb{E}\left[(Z_i-\mu_i)^2\right]}{(\mu_i/4)^2} \leq \frac{16}{\mu_i}\,.$$

Union bounding over all $i \in [\tau+1]$, we have

$$\begin{aligned}\Pr[\exists i, |Z_i-\mu_i| > \mu_i/4] &\leq 16 \sum_{i\in[\tau+1]} \frac{1}{\mu_i} = \frac{16}{N} \sum_{i\in[\tau+1]} \frac{1}{p_i} \\ &\leq \frac{16}{N} \sum_{i\in[\tau+1]} N^{i\delta}\,(1+o(1)) \leq \frac{16N^{(\tau+1)\delta}\,(1+o(1))}{N} \\ &\leq \frac{16\,(1+o(1))}{N^{\alpha}} = o(1)\,.\end{aligned}$$

Thus with probability $(1-o(1))$ we have $|Z_i-\mu_i| \leq \mu_i/4$ for all $i \in [\tau]$. When this event occurs, we see that for all $i \in [\tau]$,

$$|Y_i-\hat{s}_i| = |Z_i - Z_{i+1} - (\mu_i-\mu_{i+1})| \leq \frac{\mu_i}{4} + \frac{\mu_{i+1}}{4} < \frac{\mu_i}{4}(1+N^{-\delta}) < \frac{\mu_i}{4}(1+o(1)) \leq \frac{\hat{s}_i(1+o(1))}{4} < \frac{\hat{s}_i}{2}$$

for sufficiently large N. So $s(m) \in \mathcal{S}_0$, which proves the third part of the lemma.

Finally, it remains to prove the Claim.

Proof of Claim. Without loss of generality, assume that $j_1 = 1$ and $j_2 = 2$. Then, $\mathbb{E}\left[Z_{i,j_1}Z_{i,j_2}\right]$ is the probability that a random $m \sim \mathcal{M}_N^{\ell}$ is divisible by the monomial $x_1^i x_2^i$. Such a monomial m may be uniquely factored as $x_1^i x_2^i m'$ where m' is a monomial of degree at most $\ell - 2i$. Thus, the probability of this event is precisely

$$\frac{|\mathcal{M}_N^{\ell-2i}|}{|\mathcal{M}_N^{\ell}|} = \frac{\binom{N+\ell-2i}{\ell-2i}}{\binom{N+\ell}{\ell}}\,.$$

The statement of the claim says that the above quantity may be bounded by

$$p_i^2 = \left(\frac{\binom{N+\ell-i}{\ell-i}}{\binom{N+\ell}{\ell}}\right)^2\,.$$

Rearranging, we see that this is equivalent to the following inequality

$$\frac{\binom{N+\ell-2i}{\ell-2i}}{\binom{N+\ell-i}{\ell-i}} \leq \frac{\binom{N+\ell-i}{\ell-i}}{\binom{N+\ell}{\ell}}\,.$$

Expanding the binomials above and canceling common terms between the numerators and denominators, we may rewrite the inequality as

$$\left(\frac{\ell-2i+1}{N+\ell-2i+1}\right)\cdots\left(\frac{\ell-i}{N+\ell-i}\right)\leq\left(\frac{\ell-i+1}{N+\ell-i+1}\right)\cdots\left(\frac{\ell}{N+\ell}\right)$$

which is easy to verify since each term on the left hand side is upper bounded by the corresponding term on the right hand side. This proves the claim. $\square$

This completes the proof of Lemma 3.5. $\square$

Remark 3.8. By Lemma 3.5, for any good signature $(s_1,\ldots,s_\tau)$, we have

$$s_i/s_{i+1}=\Omega(N^{\delta}) \qquad \text{and} \qquad s_\tau=\Omega(N^{1-\tau\delta})=\Omega(N^{\alpha})\,.$$

Also

$$\frac{|\bigcup_{(s_1,\ldots,s_\tau)\text{ good}}\mathcal{M}_N^{\ell}(s_1,\ldots,s_\tau)|}{|\mathcal{M}_N^{\ell}|}=1-o(1)\,.$$

Given a signature $(s_1,\ldots,s_\tau)$, let $\mathcal{P}(s_1,\ldots,s_\tau)$ denote the set of polynomials

$$\mathcal{P}(s_1,\ldots,s_\tau)=\left\{m\cdot p_T\;\middle|\;T\subseteq[N],|T|=k,m\in\mathcal{M}_N^{\ell}(s_1,\ldots,s_\tau),\operatorname{supp}(m)\cap T=\emptyset\right\}.$$

Note that all polynomials in $\mathcal{P}(s_1,\ldots,s_\tau)$ are homogeneous of degree at most $\ell+d-k$.

Definition 3.9. For any signature $s=(s_1,\ldots,s_\tau)$, let $r_i(s)=s_i$ for $1\leqslant i\leqslant\tau-1$ 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 $\mathcal{P}(s_1,\ldots,s_\tau)$.

The number of columns of the matrix $M(s_1,\ldots,s_\tau)$ is equal to

$$\begin{aligned}|\mathcal{P}(s_1,\ldots,s_\tau)| &= \binom{N}{s_1\;\ldots\;s_\tau\;N-(s_1+\cdots+s_\tau)}\cdot\binom{N-(s_1+\cdots+s_\tau)}{k}\\ &= \binom{N}{s_1\;\ldots\;s_\tau\;k\;N-(s_1+\cdots+s_\tau+k)}\end{aligned}$$

while the number of its rows is equal to

$$\begin{aligned}\binom{N}{r_1\;\ldots\;r_\tau} &= \binom{N}{s_1\;\ldots\;s_{\tau-1}\;s_\tau+k\;N-(s_1+\cdots+s_\tau+k)}\\ &= \frac{1}{\binom{s_\tau+k}{k}}\cdot\binom{N}{s_1\;\ldots\;s_\tau\;k\;N-(s_1+\cdots+s_\tau+k)}.\end{aligned}$$

Hence $M(s_1,\ldots,s_\tau)$ has $|\mathcal{P}(s_1,\ldots,s_\tau)|$ columns but only $|\mathcal{P}(s_1,\ldots,s_\tau)|/\binom{s_\tau+k}{k}$ rows, and 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 3.10. *With parameters as above, for any good signature $s=(s_1,\ldots,s_\tau)$,*

$$\operatorname{rank}(M(s_1,\ldots,s_\tau)) \geqslant \frac{|\mathcal{P}(s_1,\ldots,s_\tau)|}{\binom{s_\tau+k}{k}}(1-o(1)).$$

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

Definition 3.11. Given a set of signatures $\mathcal{S}$, define

$$\mathcal{M}_N^\ell(\mathcal{S})=\bigcup_{s\in\mathcal{S}}\mathcal{M}_N^\ell(s) \qquad \text{and} \qquad \mathcal{P}(\mathcal{S})=\bigcup_{s\in\mathcal{S}}\mathcal{P}(s).$$

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

$$w\in\bigcup_{s\in\mathcal{S}}\mathcal{M}_N^{\ell+d-k}(r_1(s),\ldots,r_\tau(s)).$$

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 $\dim\langle\partial_k S_N^d\rangle_{\leq\ell}$.

Definition 3.12. A set of signatures $\mathcal{S}$ is *well-separated* if given any distinct signatures $s=(s_1,\ldots,s_\tau)$ and $s'=(s'_1,\ldots,s'_\tau)$ from $\mathcal{S}$, $\max_{i\in[\tau]}|s_i-s'_i|\geqslant d+1$.

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

Proof. The proof is straightforward. Since the rows and columns of $M(\mathcal{S})$ are labelled by elements of

$$\bigcup_{s\in\mathcal{S}}\mathcal{M}_N^{\ell+d-k}(r_1(s),\ldots,r_\tau(s)) \qquad \text{and} \qquad \bigcup_{s\in\mathcal{S}}\mathcal{P}(s)$$

respectively, we can group them in blocks in a natural way: corresponding to each $s\in\mathcal{S}$, we consider the rows corresponding to $\mathcal{M}_N^{\ell+d-k}(r_1(s),\ldots,r_\tau(s))$ and columns corresponding to $\mathcal{P}(s)$. This is possible since for $s\neq s'$, the set of monomials $\mathcal{M}_N^{\ell+d-k}(r_1(s),\ldots,r_\tau(s))$ and $\mathcal{M}_N^{\ell+d-k}(r_1(s'),\ldots,r_\tau(s'))$ are disjoint (because the mapping $s \mapsto (r_1(s), \dots, r_\tau(s))$ is one-to-one). Clearly the diagonal block corresponding to $s \in \mathcal{S}$ is exactly the matrix $M(s)$.

To argue that the matrix is block-diagonal, consider the entry in row w and column $m \cdot p_T$, where for some signatures $s, s' \in \mathcal{S}$, the monomials m, w are in the sets $m \in \mathcal{M}_N^\ell(s)$ and $w \in \mathcal{M}_N^\ell(r(s'))$. (Recall from the end of Section 3.1 that we only consider columns where the support of m is disjoint from T.) Assume that this entry is 1. Hence for some $A \subseteq [N]$ of size $d-k$ containing T, the monomial w has the form $w = m \cdot X_A$. Thus, any monomial w appearing in $m \cdot p_T$ has the property that

$$\Big| |\operatorname{supp}_i(w)| - |\operatorname{supp}_i(m)| \Big| \le d-k\,.$$

Thus, for $i < \tau$, we have $|s'_i - s_i| \le (d-k) < d$, and for $i = \tau$, we have $s_\tau - (d-k) \le s'_\tau + k \le s_\tau + (d-k)$. Hence $|s'_\tau - s_\tau| \le d$.

Since $\mathcal{S}$ is well-separated, both s and s' are in $\mathcal{S}$, and $|s'_i - s_i| \le d$ for all $i \in [\tau]$, it must be the case that $s' = s$. □

This allows us to give a good bound on the matrix $M(\mathcal{S})$ if $\mathcal{S}$ is well-separated.

Lemma 3.14. *For a well-separated set $\mathcal{S}$ of good signatures,*

$$\operatorname{rank}(M(\mathcal{S})) \ge \frac{(1-o(1))\binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k} \cdot |\mathcal{M}_N^\ell(\mathcal{S})|\,.$$

Proof.

$$\begin{aligned}
\operatorname{rank}(M(\mathcal{S})) &= \sum_{s\in\mathcal{S}} \operatorname{rank}(M(s)) && \text{(by Lemma 3.13, block-diagonalizability)}\\
&\ge \sum_{s\in\mathcal{S}} \frac{|\mathcal{P}(s)|}{\binom{s_\tau+k}{k}}(1-o(1)) && \text{(by Lemma 3.10).}
\end{aligned}$$

Consider the numerator, the number of columns in $M(s)$. For each monomial m, the set T generating column $m \cdot p_T$ can be chosen in $\binom{N-|\operatorname{supp}(m)|}{k}$ ways. So

$$|\mathcal{P}(s)| \ge |\mathcal{M}_N^\ell(s)| \binom{N-\ell}{k}\,.$$

Next consider the denominator. Using the fact $k = o(\lg N)$ (Fact 2.1), while $s_\tau \ge \hat{s}_\tau/2 = \Omega(N^{1-\delta\tau})$ and $\hat{s}_\tau \le N^{1-\delta\tau}$ from Lemma 3.5, we have

$$\binom{s_\tau+k}{k} \le \frac{s_\tau^k}{1-o(1)} \le \left(\frac{3\hat{s}_\tau}{2}\right)^k \frac{1}{1-o(1)} \le \left(\frac{3N^{1-\delta\tau}}{2}\right)^k \frac{1}{1-o(1)}\,.$$

Putting these back into our expression for $\operatorname{rank}(M(\mathcal{S}))$, we get

$$\operatorname{rank}(M(\mathcal{S})) \ge \sum_{s\in\mathcal{S}} \frac{|\mathcal{M}_N^\ell(s)|\binom{N-\ell}{k}}{\left(3N^{1-\delta\tau}/2\right)^k}(1-o(1)) = \frac{(1-o(1))\binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k}|\mathcal{M}_N^\ell(\mathcal{S})|\,. \qquad \square$$

Finally, we observe that there is a well-separated set $\mathcal{S}$ of good signatures such that the matrix $M(\mathcal{S})$ is quite large. Recall from Definition 3.4 that $\mathcal{S}_0$ is the set of all good signatures.

Proposition 3.15. *There is a well-separated set of good signatures, $\mathcal{S} \subseteq \mathcal{S}_0$, satisfying*

$$|\mathcal{M}_N^\ell(\mathcal{S})| \geq \frac{|\mathcal{M}_N^\ell(\mathcal{S}_0)|}{(d+1)^\tau}.$$

Proof. Let D be the set $D = \{0,1,\dots,d\}$. Define the mapping $f : \mathcal{M}_N^\ell(\mathcal{S}_0) \longrightarrow D^\tau$ as follows: for $m \in \mathcal{M}_N^\ell(\mathcal{S}_0)$ with signature $s = (s_1,\dots,s_\tau)$, set $f(m) = (d_1,\dots,d_\tau) = d$ where $d_i \equiv s_i \bmod (d+1)$. Then there must be a $\hat{d} \in D^\tau$ such that

$$|f^{-1}(\hat{d})| \geq \frac{|\mathcal{M}_N^\ell(\mathcal{S}_0)|}{(d+1)^\tau}.$$

Define $\mathcal{S}$ to be this set $f^{-1}(\hat{d})$; it is easy to see that $\mathcal{S}$ is well-separated. □

3.3 Bounding the rank of M: Step 3 of the proof

We now prove the lower bound on the rank of the matrix $M(s_1,\dots,s_\tau)$ as claimed in Lemma 3.10. We first 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 3.16. *Fix any signature $(s_1,\dots,s_\tau)$. The entry of $M(s_1,\dots,s_\tau)$ in row $w \equiv [\tilde{w}, R_1,\dots,R_\tau]$ and column $m \cdot p_T$ with $m \equiv [\tilde{m}, S_1,\dots,S_\tau]$ belongs to $\{0,1\}$ and is not zero if and only if $\tilde{w} = \tilde{m}$ and the following system is satisfied.*

$$\begin{cases} T \subseteq R_1 \\ S_1 \subseteq R_1 \cup R_2 \\ S_2 \subseteq R_2 \cup R_3 \\ \vdots \\ S_{\tau-1} \subseteq R_{\tau-1} \cup R_\tau \\ S_\tau \subseteq R_\tau \end{cases}$$

Moreover, the system above implies that $T \cup S_1 \cup \cdots \cup S_\tau = R_1 \cup \cdots \cup R_\tau$.

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 $A \subseteq [N]$ with $|A| = d-k$ such that $T \subseteq A$ and $X_A \cdot m = w$. Assume there is such an A.

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

Note that

$$\deg(\overline{w}) - \deg(\overline{m}) = \sum_{i=1}^{\tau} i r_i - \sum_{i=1}^{\tau} i s_i = \tau k = d-k$$

by our choice of parameters r_τ 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 τ" part of m: formally, $\overline{w} = X_A \cdot \overline{m}$ and hence, $\tilde{w} = \tilde{m}$.

Further, since

$$X_A \cdot \overline{m} = X_{A\setminus T} X_T \prod_{i=1}^{\tau} (X_{S_i})^i = \prod_{i=1}^{\tau} (X_{R_i})^i = \overline{w},$$

and $T \cap (S_1 \cup \cdots \cup S_\tau) = \emptyset$, we have $T \subseteq R_1$. Since X_A is multilinear, $S_i \subseteq R_i \cup R_{i+1}$ for all $i \in [\tau - 1]$, and $S_\tau \subseteq R_\tau$ is obvious.

Conversely, assume that $\tilde{w} = \tilde{m}$ 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 $T \cup S_1 \cup \cdots \cup S_\tau \subseteq R_1 \cup \cdots \cup R_\tau$. Since

$$|T \cup S_1 \cup \cdots \cup S_\tau| = k + \sum_{i=1}^{\tau} s_i = \sum_{i=1}^{\tau} r_i = |R_1 \cup \cdots \cup R_\tau|,$$

we get $T \cup S_1 \cup \cdots \cup S_\tau = R_1 \cup \cdots \cup R_\tau$. Let $A_i = R_i \setminus S_i$ for $i \in [\tau]$ and $A = A_1 \cup \cdots \cup A_\tau$. The sets A_i are disjoint (because the R_i are disjoint). Moreover, $|A_\tau| = |R_\tau \setminus S_\tau| = r_\tau - s_\tau = k$; and by induction,

$$|A_i| = |R_i \setminus S_i| = |(R_i \cup \cdots \cup R_\tau) \setminus (S_i \cup \cdots \cup S_\tau)| = \sum_{j=i}^{\tau} r_j - \sum_{j=i}^{\tau} s_j = k.$$

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 $\tilde{w} = \tilde{m}$, it follows that $X_A \cdot m = w$. Hence the entry in row w and column $m \cdot p_T$ is 1. □

Lemma 3.17. *Let $(s_1, \ldots, s_\tau)$ be any signature. The matrix $M(s_1, \ldots, s_\tau)$ is block-diagonalizable with blocks of size*

$$\binom{r}{r_1\ r_2\ \ldots\ r_\tau} \times \binom{r}{s_1\ s_2\ \ldots\ s_\tau\ k}.$$

Proof. Recall that $r = s_1 + \cdots + s_\tau + k$ (Definition 3.9).

Given a monomial $\tilde{w}$ and $R \subseteq [N]$, a block of rows of $M(s_1, \ldots, s_\tau)$ is defined by the set of monomials w such that $w \equiv [\tilde{w}, R_1, \ldots, R_\tau]$ for some $R_1, \ldots, R_\tau$ satisfying $R_1 \cup \cdots \cup R_\tau = R$. In the same way, given $\tilde{m}$ and $S \subseteq [N]$, a block of columns is defined by the set of polynomials $m \cdot p_T$ such that $m \equiv [\tilde{m}, S_1, \ldots, S_\tau]$ for some $S_1, \ldots, S_\tau$ such that $T \cup S_1 \cup \cdots \cup S_\tau = S$.

By Lemma 3.16, all blocks such that $\tilde{w} \neq \tilde{m}$ or $R \neq S$ are zero. Hence the matrix $M(s_1, \ldots, s_\tau)$ is diagonal by blocks of size

$$\binom{r}{r_1\ r_2\ \ldots\ r_\tau} \times \binom{r}{s_1\ s_2\ \ldots\ s_\tau\ k}.$$ □

To obtain a lower bound on the rank of each block in the block diagonalization, we first establish a technical lemma involving a multinomial inequality.

Lemma 3.18. *For a good signature $s = (s_1, \ldots, s_\tau)$, and the corresponding r as in Definition 3.9,*

$$\sum_{s'_1 \geqslant s_1, \ldots, s'_{\tau-1} \geqslant s_{\tau-1}} \binom{r}{s'_1\ s'_2\ \ldots\ s'_{\tau-1}\ r - \sum_{i=1}^{\tau-1} s'_i} = \binom{r}{s_1\ s_2\ \ldots\ s_{\tau-1}\ r - \sum_{i=1}^{\tau-1} s_i} (1 + o(1)).$$

Proof. Note that $\tau = o(N^\delta)$ (Fact 2.1). Also, from the definition of a good signature and from Remark 3.8, $s_i = \Omega(s_{i+1}N^\delta)$ for $i \in [\tau]$. Also, by our choice of parameters k, τ, we have $r - (s_1 + \cdots + s_{\tau-1}) = s_\tau + k \in O(s_\tau)$, so $s_{\tau-1} = \Omega((s_\tau + k)N^\delta)$.

Thus for sufficiently large N we can find an absolute constant $K > 0$ such that

$$\max\left\{\frac{r-(s_1+\cdots+s_{\tau-1})}{s_{\tau-1}}, \frac{s_{\tau-1}}{s_{\tau-2}}, \cdots, \frac{s_2}{s_1}\right\} \le \frac{K}{N^\delta} \le \frac{1}{20\tau}.$$

Define

$$S_p(b, a_1, \ldots, a_p) = \sum_{\delta_1, \ldots \delta_p \in \mathbb{N}} \binom{b}{a_1+\delta_1 \ a_2+\delta_2 \ \ldots \ a_p+\delta_p \ b - \sum_{i=1}^p (a_i+\delta_i)}.$$

The claimed result is a bound on $S_{\tau-1}(r, s_1, s_2, \ldots, s_{\tau-1})$, and is a special case of the following, with $A = K/N^\delta$. (For our choice of parameters, $K\tau/N^\delta = o(1)$.)

Claim 3.19. *Let A be a non-negative real number and $p, b, a_1, \ldots, a_p$ be positive integers such that $b \ge a_1 + \cdots + a_p$ and*

$$\max\left\{\frac{b-(a_1+\cdots+a_p)}{a_p}, \frac{a_p}{a_{p-1}}, \frac{a_{p-1}}{a_{p-2}}, \cdots, \frac{a_2}{a_1}\right\} \le A \le \frac{1}{20p}.$$

Then the following holds.

$$S_p(b, a_1, \ldots, a_p) \leqslant \binom{b}{a_1 \ a_2 \ \ldots \ a_p \ b - \sum_{i=1}^p a_i} (1 + 5Ap).$$

Proof. We prove this by induction on p.

For the base case $p = 1$, we prove the following slightly stronger statement that will be used in the inductive step as well.

Claim 3.20. *Let R be a non-negative real number, and a, b be integers with $1 \le a \le b$, satisfying $(b-a)/a \le R \le 1/20$. Then*

$$S_1(b,a) \triangleq \sum_{\delta' \in \mathbb{N}} \binom{b}{a+\delta'} \leqslant \binom{b}{a}(1+2R).$$

Proof. By hypothesis, $b - a \le aR$. Notice that for $\delta' \geqslant 0$,

$$\frac{\binom{b}{a+\delta'+1}}{\binom{b}{a+\delta'}} = \frac{b-a-\delta'}{a+\delta'+1} \leqslant \frac{b-a}{a} \leqslant R.$$

Hence,

$$S_1(b,a) = \sum_{\delta' \in \mathbb{N}} \binom{b}{a+\delta'} \leqslant \binom{b}{a}\left(1 + \sum_{\delta'=1}^{\infty} (R)^{\delta'}\right) \leqslant \binom{b}{a}(1+2R). \qquad \square$$

Note that Claim 3.20 implies the base case of induction for Claim 3.19.

Now let $p \geq 2$, and assume that the claim holds for all $p' < p$. We have

$$S_p(b, a_1, \ldots, a_p) = \sum_{\delta_1 \in \mathbb{N}} \binom{b}{a_1 + \delta_1} S_{p-1}(b - (a_1 + \delta_1), a_2, \ldots, a_p) \,.$$

Note that for all non-negative δ_1,

$$(b - (a_1 + \delta_1)) - (a_2 + \cdots + a_p) \leq (b - a_1) - (a_2 + \cdots + a_p) = b - (a_1 + a_2 + \cdots + a_p) \leq a_p A \,.$$

Hence the induction hypothesis is applicable to all the S_{p-1} terms, giving

$$\begin{aligned}
S_p(b, a_1, \ldots, a_p) &\leqslant \sum_{\delta_1 \in \mathbb{N}} \binom{b}{a_1 + \delta_1} \binom{b - (a_1 + \delta_1)}{a_2\ a_3\ \ldots\ a_p\ b - (\sum_{i=1}^p a_i) - \delta_1} (1 + 5A(p-1)) \\
&\leqslant \sum_{\delta_1 \in \mathbb{N}} \binom{b}{a_1 + \delta_1} \binom{b - a_1}{a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5A(p-1)) \\
&\quad \text{(because } \binom{b - (a_1 + \delta_1)}{a_2\ a_3\ \ldots\ a_p\ b - (\sum_{i=1}^p a_i) - \delta_1} \text{ is a decreasing function of } \delta_1\text{)} \\
&= \binom{b - a_1}{a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5A(p-1)) \sum_{\delta_1 \in \mathbb{N}} \binom{b}{a_1 + \delta_1} \\
&= \ast \binom{b - a_1}{a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5A(p-1)) S_1(b, a_1) \,.
\end{aligned}$$

We need to show that Claim 3.20 is applicable to $S_1(b, a_1)$. Note that

$$\begin{aligned}
\frac{b - a_1}{a_1} &= \frac{b - \sum_{i=1}^p a_i + \sum_{i=2}^p a_i}{a_1} = \frac{b - \sum_{i=1}^p a_i}{a_1} + \frac{a_p}{a_1} + \cdots + \frac{a_2}{a_1} \\
&= \left(\frac{b - \sum_{i=1}^p a_i}{a_p} \right) \left(\frac{a_p}{a_{p-1}} \right) \cdots \left(\frac{a_2}{a_1} \right) + \left(\frac{a_p}{a_{p-1}} \right) \cdots \left(\frac{a_2}{a_1} \right) + \cdots + \frac{a_2}{a_1} \\
&\leqslant \sum_{i=1}^p A^i \leqslant 2A \ \text{ because } \ A \leq 1/20 \,.
\end{aligned}$$

Also,

$$2A \leq \frac{1}{10p} \leq \frac{1}{20} \,.$$

So we can use Claim 3.20 with $R = 2A$, $a = a_1$, and b. Continuing our derivation, we get

$$\begin{aligned}
S_p(b, a_1, \ldots, a_p) &\leqslant \binom{b - a_1}{a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5A(p-1)) \binom{b}{a_1} (1 + 4A) \\
&= \binom{b}{a_1\ a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5A(p-1)) (1 + 4A) \\
&\leq \binom{b}{a_1\ a_2\ a_3\ \ldots\ a_p\ b - \sum_{i=1}^p a_i} (1 + 5Ap)
\end{aligned}$$

(using the assumption $p \leq 1/(20A)$). □

This completes the proof of Lemma 3.18. □

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

Lemma 3.21 (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*

$$\binom{r}{s_1\ s_2\ \ldots\ s_\tau+k}(1-o(1))\,.$$

Proof. Let M' be a diagonal block of the matrix $M(s_1,\ldots,s_\tau)$. Recall from Lemma 3.17 that such a diagonal block is defined by a monomial $\tilde{w}$ and a subset $R\subseteq[N]$. Rows of this block are labelled with all monomials $w\equiv[\tilde{w},R_1,\ldots,R_\tau]$ such that $R_1\cup\cdots\cup R_\tau=R$ and columns of this block are labelled with all polynomials $m\cdot p_T$ where $m\equiv[\tilde{w},S_1,\ldots,S_\tau]$ is such that $T\cup S_1\cup\cdots\cup S_\tau=R$. First, we set up some notation.

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

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

$$(a_1,\ldots,a_p)\preceq(b_1,\ldots,b_p)\qquad\text{but}\qquad(a_1,\ldots,a_p)\neq(b_1,\ldots,b_p)\,.$$

Define the following collections of partitions of R.

$$\begin{aligned}
X &= \{\tilde{R}=(R_1,\ldots,R_\tau)\mid \text{signature}(\tilde{R})=(r_1,\ldots,r_\tau)\}\,,\\
Y &= \{\tilde{S}=(S_1,\ldots,S_\tau,T)\mid \text{signature}(\tilde{S})=(s_1,\ldots,s_\tau,k)\}\,,\\
Z' &= \{\tilde{Q}=(Q_1,\ldots,Q_\tau)\mid \text{signature}(\tilde{Q})=(q_1,\ldots,q_\tau);(s_1,\ldots,s_{\tau-1})\preceq(q_1,\ldots,q_{\tau-1})\}\,,\\
Z &= \{\tilde{Q}=(Q_1,\ldots,Q_\tau)\mid \text{signature}(\tilde{Q})=(q_1,\ldots,q_\tau);(s_1,\ldots,s_{\tau-1})\prec(q_1,\ldots,q_{\tau-1})\}\,.
\end{aligned}$$

Note that

$$|X|=\binom{r}{s_1\ s_2\ \ldots\ s_\tau+k}\,.$$

Also, $Z'\setminus Z$ is precisely X. By Lemma 3.18, $|Z'|=|X|(1+o(1))$. Hence $|Z|=|X|\cdot o(1)$.

For any $\tilde{S}\in Y$, define the partition $\tilde{S_X}=(S_1,\ldots,S_{\tau-1},S_\tau\cup T)\in X$. We say that $\tilde{S}$ "extends" $\tilde{S_X}$.

The rows and columns of M' are indexed by elements of X and Y respectively (Lemma 3.17).

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.

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

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.

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

Let I be the identity matrix with rows and columns indexed by elements of X.

Our proof proceeds by establishing the following two claims.

Claim 3.22. *The columns of M' and M_2 together span the columns of M_1; hence*

$$\operatorname{rank}(M_1) \le \operatorname{rank}(M') + \operatorname{rank}(M_2)\,.$$

Claim 3.23. *The columns of M_1 and M_2 together span the columns of I; hence*

$$\operatorname{rank}(I) \le \operatorname{rank}(M_1) + \operatorname{rank}(M_2)\,.$$

It then follows that

$$\operatorname{rank}(M') \ge \operatorname{rank}(M_1) - \operatorname{rank}(M_2) \ge \operatorname{rank}(I) - 2\operatorname{rank}(M_2) \ge |X| - 2|Z| = |X|(1-o(1))$$

which is what we had set out to prove.

To prove the claims, we set up some notation. For $A \subseteq [\tau]$, define the function $\varphi_A : X \times 2^R \longrightarrow \{0,1\}$ as follows.

$$\varphi_A(\tilde{R}, S) = \begin{cases} 1 & \text{if } S \subseteq \bigcup_{i\in A} R_i, \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $S = \emptyset$, then $\varphi_A(\tilde{R}, S) = 1$ for every A and $\tilde{R}$.

With some abuse of notation, for sets of size 1 or 2 we drop the set notation. eg $\varphi_{i_1,i_2}(\tilde{R}, j)$ is the same as $\varphi_{\{i_1,i_2\}}(\tilde{R}, \{j\})$.

Since τ is odd and at least 3, we can express the functions $\varphi_A(\cdot,\cdot)$ for singleton sets A in terms of the functions $\varphi_B(\cdot,\cdot)$ where $|B| = 2$. In particular, for $A = \{1\}$ and for $A = \{\tau\}$, we write

$$\begin{array}{rrcl} & \varphi_\tau(\tilde{R}, j) & = & 1 - \varphi_{1,2}(\tilde{R}, j) - \varphi_{3,4}(\tilde{R}, j) - \cdots - \varphi_{\tau-2,\tau-1}(\tilde{R}, j) \\ \text{and} & \varphi_1(\tilde{R}, j) & = & 1 - \varphi_{2,3}(\tilde{R}, j) - \varphi_{4,5}(\tilde{R}, j) - \cdots - \varphi_{\tau-1,\tau}(\tilde{R}, j)\,. \end{array}$$

For $A = \{1\}$ or $A = \{\tau\}$ and $S \subseteq [N]$ we write

$$\varphi_A(\tilde{R}, S) \quad = \quad \prod_{j\in S} \varphi_A(\tilde{R}, j)\,.$$

We use these functions to compactly describe the columns of M', M_1, M_2, I. By definition,

$$\begin{aligned} M_1[\tilde{R}, \tilde{R}'] &= \prod_{i=1}^{\tau-1} \varphi_{i,i+1}(\tilde{R}, R'_i) \;\; ; \\ M_2[\tilde{R}, \tilde{Q}] &= \prod_{i=1}^{\tau-1} \varphi_{i,i+1}(\tilde{R}, Q_i) \;\; ; \\ I[\tilde{R}, \tilde{R}'] &= \left(\prod_{i=1}^{\tau} \varphi_i(\tilde{R}, R'_i) \right) = \left(\prod_{i=1}^{\tau-1} \varphi_{i,i+1}(\tilde{R}, R'_i) \right) \varphi_\tau(\tilde{R}, R'_\tau) = M_1[\tilde{R}, \tilde{R}']\varphi_\tau(\tilde{R}, R'_\tau) \;\; ; \end{aligned}$$

where the second equality follows from the fact that $\tilde{R}, \tilde{R}'$ have the same signature. (RHS = 1 $\Rightarrow$ $\varphi_\tau(\tilde{R}, R'_\tau) = 1 \Rightarrow R'_\tau \subseteq R_\tau \Rightarrow R'_\tau = R_\tau$ because the sets are equi-sized. Then RHS = 1 further $\Rightarrow$ $\varphi_{\tau-1,\tau}(\tilde{R}, R'_{\tau-1}) = 1 \Rightarrow R'_{\tau-1} \subseteq R_{\tau-1} \cup R_\tau$. But $R'_{\tau-1}$ is disjoint from $R'_\tau = R_\tau$. So $R'_{\tau-1} \subseteq R_{\tau-1}$, and since they are equi-sized, they must be the same. Continuing this way, we conclude $\tilde{R} = \tilde{R}'$.)

Proof of Claim 3.22. Starting with Lemma 3.16,

$$\begin{aligned}
M'[\tilde{R},\tilde{S}] &= \varphi_{1,2}(\tilde{R},S_1)\varphi_{2,3}(\tilde{R},S_2)\ldots\varphi_{\tau-1,\tau}(\tilde{R},S_{\tau-1})\varphi_{\tau}(\tilde{R},S_\tau)\varphi_1(\tilde{R},T) \\
&= \left(\prod_{i=1}^{\tau-1}\varphi_{i,i+1}(\tilde{R},S_i)\right)\left(\prod_{j\in S_\tau}\varphi_\tau(\tilde{R},j)\right)\left(\prod_{j\in T}\varphi_1(\tilde{R},T)\right) \\
&= \left(\prod_{i=1}^{\tau-1}\varphi_{i,i+1}(\tilde{R},S_i)\right) \\
&\quad\times\left(\prod_{j\in S_\tau}\left[1-\varphi_{1,2}(\tilde{R},j)-\varphi_{3,4}(\tilde{R},j)-\cdots-\varphi_{\tau-2,\tau-1}(\tilde{R},j)\right]\right) \\
&\quad\times\left(\prod_{j\in T}\left[1-\varphi_{2,3}(\tilde{R},j)-\varphi_{4,5}(\tilde{R},j)-\cdots-\varphi_{\tau-1,\tau}(\tilde{R},j)\right]\right) \\
&= \sum_{\substack{\text{partition } \tilde{Q}=(Q_1,\ldots,Q_{\tau-1},Q_\tau)\,: \\ \forall i\in[\tau-1], S_i\subseteq Q_i}} \prod_{i=1}^{\tau-1}(-1)^{|Q_i\setminus S_i|}\varphi_{i,i+1}(\tilde{R},Q_i) \\
&= \sum_{\tilde{Q}\in Z'}\alpha_{\tilde{S},\tilde{Q}}\prod_{i=1}^{\tau-1}\varphi_{i,i+1}(\tilde{R},Q_i)\,.
\end{aligned}$$

The coefficients $\alpha_{\tilde{S},\tilde{Q}}$ are all in $\{-1,0,1\}$. Observe that:

- The coefficient $\alpha_{\tilde{S},\tilde{S_X}}$ is 1, and the corresponding term is precisely $M_1[\tilde{R},\tilde{S_X}]$.
- The coefficient $\alpha_{\tilde{S},\tilde{Q}}$ is 0 for all $\tilde{Q}\in X\setminus\{\tilde{S_X}\}$. (One of the requirements $S_i\subseteq Q_i$ must be violated.)
- All other $\tilde{Q}$ are in Z, and the corresponding term is precisely $M_2[\tilde{R},\tilde{Q}]$.

Hence

$$M'[\tilde{R},\tilde{S}] = M_1[\tilde{R},\tilde{S_X}] + \sum_{\tilde{Q}\in Z}\alpha_{\tilde{S},\tilde{Q}}M_2[\tilde{R},\tilde{Q}]\,.$$

Since the coefficients in this combination do not depend on the row $\tilde{R}$, we obtain

$$M'[*,\tilde{S}] = M_1[*,\tilde{S_X}] + \sum_{\tilde{Q}\in Z}\alpha_{\tilde{S},\tilde{Q}}M_2[*,\tilde{Q}]\,.$$

For every $\tilde{R}'\in X$, arbitrarily pick any $\tilde{S}\in Y$ extending it. Then

$$M_1[*,\tilde{R}'] = M'[*,\tilde{S}] - \sum_{\tilde{Q}\in Z}\alpha_{\tilde{S},\tilde{Q}}M_2[*,\tilde{Q}]\,. \qquad \square$$

Proof of Claim 3.23. Starting with I and proceeding in exactly the same way, we obtain

$$\begin{aligned} I[\tilde{R},\tilde{R}'] &= M_1[\tilde{R},\tilde{R}']\varphi_\tau(\tilde{R},R'_\tau) \\ &= M_1[\tilde{R},\tilde{R}']\left(\prod_{j\in R'_\tau}\left[1-\varphi_{1,2}(\tilde{R},j)-\varphi_{3,4}(\tilde{R},j)-\cdots-\varphi_{\tau-2,\tau-1}(\tilde{R},j)\right]\right) \\ &= M_1[\tilde{R},\tilde{R}']+\sum_{\tilde{Q}\in Z}\beta_{\tilde{R}',\tilde{Q}}M_2[\tilde{R},\tilde{Q}] \end{aligned}$$

for some coefficients $\beta_{\tilde{R}',\tilde{Q}}$ independent of $\tilde{R}$. Hence

$$I[*,\tilde{R}'] = M_1[*,\tilde{R}'] + \sum_{\tilde{Q}\in Z}\beta_{\tilde{R}',\tilde{Q}}M_2[*,\tilde{Q}]\,. \qquad \square$$

With both claims established, the proof of Lemma 3.21 is now complete. $\square$

Lemma 3.10 can now be proved using the block-diagonal decomposition (Lemma 3.17) and the rank lower bound (Lemma 3.21).

Proof of Lemma 3.10. By Lemma 3.17, we know that $M(s_1,\ldots,s_\tau)$ can be block-diagonalized into blocks of size

$$\binom{r}{r_1\ r_2\ \cdots\ r_\tau}\times\binom{r}{s_1\ s_2\ \cdots\ s_\tau\ k}.$$

Let B denote the number of blocks in this block diagonalization.

By Lemma 3.21, we know that each block has rank

$$\begin{aligned} (1-o(1))\binom{r}{r_1\ r_2\ \cdots\ r_\tau} &= (1-o(1))\binom{r}{s_1\ s_2\ \cdots\ s_\tau+k} \\ &= (1-o(1))\cdot\frac{1}{\binom{s_\tau+k}{k}}\binom{r}{s_1\ s_2\ \cdots\ s_\tau\ k} \\ &= \frac{1-o(1)}{\binom{s_\tau+k}{k}}\cdot(\#\text{ of columns in each block}) \end{aligned}$$

where the first equality is a result of our choice of parameters and the second follows from the combinatorial identity:

$$\binom{r}{s_1\ s_2\ \cdots\ s_\tau+k} = \frac{1}{\binom{s_\tau+k}{k}}\binom{r}{s_1\ s_2\ \cdots\ s_\tau\ k}.$$

Thus, the rank of the matrix $M(s_1,\ldots,s_\tau)$, which is the sum of the ranks of the blocks, is

$$\begin{aligned} &\frac{1-o(1)}{\binom{s_\tau+k}{k}}\cdot\ (\#\text{ of columns in each block})\cdot B \\ &= \frac{1-o(1)}{\binom{s_\tau+k}{k}}\cdot(\#\text{ of columns in } M(s_1,\ldots,s_\tau)) \\ &= \frac{|\mathcal{P}(s_1,\ldots,s_\tau)|}{\binom{s_\tau+k}{k}}(1-o(1))\,, \end{aligned}$$

since $|\mathcal{P}(s_1,\ldots,s_\tau)|$ is the number of columns in $M(s_1,\ldots,s_\tau)$. $\square$

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.

Theorem 3.1 (Restated)**.** Let $\alpha \in (0,1/2)$ be a constant. Let $N,d,k \in \mathbb{N}$ be such that $4k \le d \le \alpha \lg N/\lg\lg N$ and $2k \mid d$. Over a field of characteristic zero, for $\tau = d/k-1$, for any δ satisfying $\alpha \le 1-\delta(\tau+1) < 1-\delta\tau \le 1-\alpha$, and for $\ell = \lfloor N^{1-\delta} \rfloor$, the following holds.

$$\dim\langle \partial_k S_N^d \rangle_{\le \ell} \geqslant \frac{(1-o(1)) \cdot \binom{N+\ell}{\ell} \cdot \binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k \cdot (d+1)^\tau}\,.$$

Proof of Theorem 3.1. By Lemma 3.2, $\dim\langle \partial_k S_N^d \rangle_{\le \ell} \ge \dim(\mathrm{span}(\mathcal{P}))$. This in turn is at least as large as $\mathrm{rank}(M(\mathcal{S}))$ for any set $\mathcal{S}$ of signatures, since $M(\mathcal{S})$ is a submatrix of the matrix that describes a basis for $\mathcal{P}$. By Proposition 3.15, there is a well-separated set of good signatures $\mathcal{S}$ with large $|\mathcal{M}_N^\ell(\mathcal{S})|$. Choose such a set. Then

$$\begin{aligned}
\dim\langle \partial_k S_N^d \rangle_{\le \ell} &\ge \dim(\mathrm{span}(\mathcal{P})) && \text{(by Lemma 3.2)}\\
&\ge \mathrm{rank}(M(\mathcal{S}))\\
&\ge \frac{(1-o(1))\binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k} \cdot |\mathcal{M}_N^\ell(\mathcal{S})| && \text{(by Lemma 3.14)}\\
&\ge \frac{(1-o(1))\binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k} \cdot \frac{|\mathcal{M}_N^\ell(\mathcal{S}_0)|}{(d+1)^\tau} && \text{(by Proposition 3.15)}\\
&\ge \frac{(1-o(1))\binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k} \cdot \frac{|\mathcal{M}_N^\ell|(1-o(1))}{(d+1)^\tau} && \text{(by Lemma 3.5 and Remark 3.8)}\\
&= \frac{(1-o(1))\binom{N-\ell}{k}\binom{N+\ell}{\ell}}{(3N^{1-\delta\tau}/2)^k(d+1)^\tau}\,.
\end{aligned}$$

□

4 Modification for the general case of parameters

The proof of Theorem 3.1 handles the case when d is divisible by $2k$, and thus, by our choice of τ, it is also divisible by $\tau+1$. In this section, we state the modifications which we make to the proof so that it works in a more general setting, and thus prove Theorem 1.1.

Theorem 1.1 (Restated)**.** Let $\alpha \in (0,1/2)$ be a constant. Let $N,d,k \in \mathbb{N}$ be such that $4k \le d \le \alpha \lg N/\lg\lg N$ and $k = \lfloor d/(\tau+1) \rfloor$ for some odd number $\tau \ge 3$. Over a field of characteristic zero, for any δ satisfying $\alpha \le 1-\delta(\tau+1) < 1-\delta\tau \le 1-\alpha$, and for $\ell = \lfloor N^{1-\delta} \rfloor$, the following holds.

$$\dim\langle \partial_k S_N^d \rangle_{\le \ell} \geqslant \frac{(1-o(1)) \cdot \binom{N+\ell}{\ell} \cdot \binom{N-\ell}{k}}{(3N^{1-\delta\tau}/2)^k \cdot (d+1)^\tau}\,.$$

Proof sketch. We follow the proof outline for Theorem 3.1.

1. The bounds concerning the parameters (Fact 2.1) and good signatures (Lemma 3.5) continue to hold.

2. In Section 2 in Definition 3.9, we had set the parameters $(r_1,\ldots,r_\tau)$ corresponding to a signature $(s_1,\ldots,s_\tau)$. Here we modify this slightly. Let g be the loss due to the floor function; $g \triangleq d-k(\tau+1)$. (Note: earlier, we had $g=0$.) We have $0\le g\le\tau$, and $\tau k+g=d-k$. Now, we let $r_i=s_i$ for $i\in[\tau-2]$, let $r_{\tau-1}=s_{\tau-1}-g$, and let $r_\tau=s_\tau+k+g$. It can be verified that with this choice,
$$\sum_{i=1}^{\tau} r_i=\left(\sum_{i=1}^{\tau} s_i\right)+k \qquad\text{and}\qquad \sum_{i=1}^{\tau} ir_i=\left(\sum_{i=1}^{\tau} is_i\right)+d-k\,.$$

3. We define matrix $M(s_1,s_2,\ldots,s_\tau)$ as in Definition 3.9 but with respect to the new parameter setting. It is easy to see that Lemma 3.16 and Lemma 3.17 hold in this new setting as well, because the proof only uses the fact that
$$\sum_{i=1}^{\tau} ir_i=\sum_{i=1}^{\tau} is_i+d-k \qquad\text{and}\qquad \sum_{i=1}^{\tau} r_i=\sum_{i=1}^{\tau} s_i+k\,.$$

4. Finally, we note that Lemma 3.21 holds with a few modifications to the proof. Recall the overall strategy: for partitions X,Y,Z,Z' and matrices M',M_1,M_2,I, we had shown
$$\operatorname{rank}(M')\ge\operatorname{rank}(M_1)-\operatorname{rank}(M_2)\ge\operatorname{rank}(I)-2\operatorname{rank}(M_2)\ge|X|-2|Z|=|X|(1-o(1))\,.$$
We now define an additional set $\tilde{Y}$ of partitions as follows.
$$\tilde{Y} \quad=\quad \{\tilde{S}=(S_1,\ldots,S_\tau\cup T)\mid \operatorname{signature}(\tilde{S})=(s_1,\ldots,s_\tau+k)\}\,.$$
Notice that $\tilde{Y}=Z\setminus Z'$. Let D be the matrix whose rows are labelled by X and columns by $\tilde{Y}$, and defined in the following way: $D[(R_1,\ldots,R_\tau),(S_1,\ldots,S_{\tau-1},S'_\tau)]=1$ if $S_i\subseteq R_i\cup R_{i+1}$ for all $i\in[\tau-1]$ and $S'_\tau\subseteq R_\tau$; otherwise, it is 0.[5] Now our strategy is to show
$$\begin{aligned}\operatorname{rank}(M')&\ge\operatorname{rank}(M_1)-\operatorname{rank}(M_2)\ge\operatorname{rank}(D)-2\operatorname{rank}(M_2)\\&\ge|\tilde{Y}|(1-o(1))-2|Z|=|\tilde{Y}|(1-o(1))\,.\end{aligned}$$
The first step (columns of M' and M_2 span those of M_1) works exactly as before. So does the second step (columns of M_1 and M_2 span those of D): this is because for $\tilde{R}=(R_1,\ldots,R_\tau)$ and $\tilde{S}=(S_1,\ldots,S_{\tau-1},S'_\tau)$, we have
$$\begin{aligned}D[\tilde{R},\tilde{S}] &= M_1[\tilde{R},\tilde{S}]\varphi_\tau(\tilde{R},S'_\tau)\\ &= M_1[\tilde{R},\tilde{S}]\left(\prod_{j\in S'_\tau}\left[1-\varphi_{1,2}(\tilde{R},j)-\varphi_{3,4}(\tilde{R},j)-\cdots-\varphi_{\tau-2,\tau-1}(\tilde{R},j)\right]\right).\end{aligned}$$
The step showing $\operatorname{rank}(M_2)$ is small follows from Lemma 3.18 which holds as is, since Z,Z' depend only on the signature s and not on how we set r.

[5]Note that, in the original parameter setting when $g=0$, we have $\tilde{Y}=Z\setminus Z'=X$ and D is the identity matrix I. When $g\ne 0$, due to the new setting of parameters, $\tilde{Y}\ne X$ and D is different from I.

We now need one additional step showing that D has rank $|\tilde{Y}|(1-o(1))$, and then we note that

$$|\tilde{Y}| = \binom{s}{s_1\ s_2\ \ldots\ s_\tau + k}.$$

First notice that the matrix D has a block diagonal structure. For a given pair of tuples

$$(R_1, R_2, \ldots, R_{\tau-2}) \qquad \text{and} \qquad (S_1, S_2, \ldots, S_{\tau-2})$$

the block corresponding to this pair is

$$\left\{ (\tilde{R}, \tilde{S}) \;\middle|\; \begin{array}{l} \exists R_{\tau-1}, R_\tau, S_{\tau-1}, S'_\tau \text{ such that } \tilde{R} = (R_1, R_2, \ldots, R_{\tau-2}, R_{\tau-1}, R_\tau) \\ \text{and } \tilde{S} = (S_1, S_2, \ldots, S_{\tau-2}, S_{\tau-1}, S'_\tau) \end{array} \right\}.$$

If $(R_1, R_2, \ldots, R_{\tau-2}) \neq (S_1, S_2, \ldots, S_{\tau-2})$ then any entry of the matrix in the block corresponding to the pair is zero. If $(S_1, S_2, \ldots, S_{\tau-2}) = (R_1, R_2, \ldots, R_{\tau-2})$, then in the block defined by this pair, consider sub-blocks where rows are grouped by $R_{\tau-1} \cup R_\tau$ and columns by $S_{\tau-1} \cup S'_\tau$. Again, entries outside the diagonal sub-blocks are all zeroes.

So now consider a sub-block, with $U = R_{\tau-1} \cup R_\tau = S_{\tau-1} \cup S'_\tau$. Each row can be thought of as labelled by R_τ (this determines $R_{\tau-1}$ as $U \setminus R_\tau$) and each column by S'_τ (again, this determines $S_{\tau-1}$), where $|R_\tau| = r_\tau = s_\tau + g + k$, and $|S'_\tau| = s_\tau + k$. And the entry in the cell labelled by row R_τ and column S'_τ is 1 exactly when $S'_\tau \subseteq R_\tau$. Thus, this sub-block is an inclusion matrix. We use a theorem of Wilson [28] to analyze the rank of each sub-block in the matrix. This theorem is stated below.

Let u, v, w be non-negative integer parameters. We denote by W^u_{vw} the $\binom{u}{v} \times \binom{u}{w}$ inclusion matrix in which each row is labelled by a set of size v from a universe of size u and each column is labelled by a set of size w from the same universe. The (A,B)th entry of the matrix is 1 if $A \subseteq B$ and 0 otherwise.

Theorem 4.1 (Wilson [28])**.** *Assume $u, v, w \in \mathbb{N}$. For $v \leq \min\{w, u-w\}$, the rank of the matrix W^u_{vw} in characteristic $p > 0$ is equal to*

$$\sum \binom{u}{i} - \binom{u}{i-1}$$

where the sum is over indices i such that p does not divide $\binom{w-i}{v-i}$.

When $i = v$, $\binom{w-i}{v-i} = 1$, i. e., this binomial coefficient is not divisible by any p. Therefore, for any p, the term corresponding to $i = v$ appears in the summation. Since each term in the summation is non-negative, the rank of W^u_{vw} is at least $\binom{u}{v} - \binom{u}{v-1}$. (Note also that this holds over any characteristic, a fact that will be useful in Section 5).

Our sub-blocks are transposes of the matrices W^u_{vw}, with $u = s_{\tau-1} + s_\tau + k$, $v = s_\tau + k$, $w = s_\tau + k + g$. Hence each sub-block has rank at least

$$\binom{s_{\tau-1} + s_\tau + k}{s_\tau + k} - \binom{s_{\tau-1} + s_\tau + k}{s_\tau + k - 1}.$$

From Lemma 3.5, $k, g, s_\tau = o(s_{\tau-1})$. Hence the rank of each sub-block is at least

$$\binom{s_{\tau-1}+s_\tau+k}{s_\tau+k}(1-o(1))$$

which is $(1-o(1))$(number of columns in sub-block). Due to the diagonal structure, we can add up over all the sub-blocks and blocks to get

$$\mathrm{rank}(D) \geq (1-o(1))(\text{number of columns in } D) = (1-o(1))|\tilde{Y}|\,.$$

5. Putting together the steps as done in Section 3.4 establishes Theorem 1.1. □

5 Modified proof for non-zero characteristic

In this section we desribe how to adapt our proofs of Theorem 1.1 and Theorem 3.1 to work over fields with positive characteristic. The bound we obtain is slightly, but not significantly, weaker.

We first observe that the only place in our proofs of Theorem 1.1 and Theorem 3.1 where we use characteristic zero is Step 1 in the proof of Lemma 3.2. It is easy to see that all other steps, including the adaptation described in Section 4, are independent of the characteristic.

While working in positive characteristic, we replace Lemma 3.2 by a different statement. Recall the use of Lemma 3.2; it allowed us to lower-bound $\dim\left(\langle \partial_k S_N^d \rangle_{\leq \ell}\right)$ by the dimension of

$$\mathcal{P} = \left\{ m \cdot p_T \,\middle|\, T \subseteq [N], |T| = k, m \in \mathcal{M}_N^\ell, \mathrm{supp}(m) \cap T = \emptyset \right\}.$$

Here, we consider a somewhat different set $\mathcal{P}'$, and also end up with a $1/(k+1)$ factor while lower-bounding $\dim\left(\langle \partial_k S_N^d \rangle_{\leq \ell}\right)$.

Let $N' := N-k$. We work with N' variables, with well-chosen 0-1 settings to the last k variables. Recall that for any set $S \subseteq [N]$ such that $|S| \leq k$, the polynomial $r_S(x)$ is defined as follows.

$$r_S(x) := \sum_{A \subseteq [N], |A|=d-k, A\cap S=\emptyset} X_A\,,$$

where $X_A := \prod_{i\in A} x_i$. Let $\mathcal{D}$ denote the set $\{r_S \mid S \subseteq [N], |S| = k\}$.

Similarly, for any set $S \subseteq [N']$ such that $|S| \leq k$, define the polynomial $r'_S(x)$ as follows.

$$r'_S(x) := \sum_{A \subseteq [N'], |A|=d-k, A\cap S=\emptyset} X_A\,.$$

In Step 1 of the proof of Lemma 3.2 we showed that for any $S' \subseteq [N]$ with $|S'| \leq k$, the polynomial $r_{S'}$ is in the span of $\mathcal{D}$. In the case of non-zero characteristic, we create $k+1$ sets $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k$, each of dimension at most that of $\mathcal{D}$, and show that for each $S' \subseteq [N']$ with $|S'| \leq k$, the union of these sets contains $r'_{S'}$.

For every $0 \leq i \leq k$, let $\pi_i : \{x_1, \ldots, x_N\} \to \{x_1, \ldots, x_{N'}\} \cup \{0, 1\}$ be defined as follows.

$$\pi_i(x_j) = \begin{cases} x_j & \text{if } 1 \leq j \leq N', \\ 1 & \text{if } (N'+1) \leq j \leq (N'+i), \\ 0 & \text{otherwise.} \end{cases}$$

The map π_i naturally extends to a ring homomorphism from $\mathbb{F}[x_1,\ldots,x_N]$ to $\mathbb{F}[x_1,\ldots,x_{N'}]$. For each $0 \le i \le k$, let $\mathcal{D}_i := \pi_i(\mathcal{D}) = \{\pi_i(r_S) \mid r_S \in \mathcal{D}\}$.

Consider any $S' \subseteq [N']$ of size at most k. We augment S' to a set $S'' \subseteq [N]$ of size exactly k, using the last k reserved indices, by defining $S'' = S' \cup \{N'+1,\ldots,N'+(k-|S|)\}$. (If $|S'| = k$, then $S'' = S'$.) Now the projection $\pi_{k-|S'|}$ applied to the augmented-set-polynomial $r_{S''}$ gives back the polynomial $r'_{S'}$; that is, $r'_{S'} = \pi_{k-|S'|}(r_{S''}) \in \mathcal{D}_{k-|S'|}$. Therefore we get

$$\{r'_{S'} \mid S' \subseteq [N'], |S'| \le k\} \subseteq \bigcup_{i=1}^{k} \mathcal{D}_i\,.$$

Let

$$p'_T(x) := \sum_{T \subseteq B \subseteq [N'], |B| = d-k} X_B\,.$$

Let P' denote $\{p'_T \mid T \subseteq [N'], |T| = k\}$. The inclusion-exclusion argument in Step 2 of the proof of Lemma 3.2 works exactly as before, over the set $[N']$, and tells us that P' is contained in

$$\mathrm{span}\{r'_{S'} \mid S' \subseteq [N'], |S'| \le k\}$$

and therefore, in $\mathrm{span}(\cup_i \mathcal{D}_i)$. We define $\mathcal{P}'$ to be set of shifts of P' of degree (at most) ℓ, similar to $\mathcal{P}$ defined in Section 3.1, but restricting even the shifts to variables from $[x_1,\ldots,x_{N'}]$.

$$\mathcal{P}' = \left\{ m \cdot p'_T \,\middle|\, T \subseteq [N'], |T| = k, m \in \mathcal{M}^{\ell}_{N'}, \mathrm{supp}(m) \cap T = \emptyset \right\} \subseteq \{m \cdot q \mid m \in \mathcal{M}^{\ell}_{N'}, q \in P'\}\,.$$

Since $P' \subseteq \mathrm{span} \cup_i \mathcal{D}_i$, we get

$$\begin{aligned} \mathcal{P}' &\subseteq \mathrm{span} \bigcup_{i=0}^{k} \left\{ m \cdot \pi_i(r_S) \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'} \right\} \\ &= \mathrm{span} \bigcup_{i=0}^{k} \left\{ \pi_i(m \cdot r_S) \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'} \right\}. \end{aligned}$$

The equality holds because for monomials $m \in \mathcal{M}^{\ell}_{N'}$, for every $0 \le i \le k$, $m = \pi_i(m)$.

Now, note that for any finite set $X \subseteq \mathbb{F}[x_1 \ldots, x_N]$ and a linear map π between vector spaces $\mathbb{F}[x_1,\ldots,x_N]$ and $\mathbb{F}[x_1,\ldots,x_{N-k}]$, $\dim(\mathrm{span}\,\pi(X)) \le \dim(\mathrm{span}\,X)$. Therefore, we get that

$$\dim\left(\mathrm{span}\left\{\pi_i(m \cdot r_S) \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'}\right\}\right) \le \dim\left(\mathrm{span}\left\{m \cdot r_S \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'}\right\}\right).$$

Thus,

$$\begin{aligned} \dim\left(\mathrm{span}\left\{\mathcal{P}'\right\}\right) &\le \dim\left(\mathrm{span} \bigcup_{i=0}^{k} \left\{\pi_i(m \cdot r_S) \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'}\right\}\right) \\ &\le \sum_{i=0}^{k} \dim\left(\mathrm{span}\left\{\pi_i(m \cdot r_S) \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'}\right\}\right) \\ &\le (k+1)\dim\left(\mathrm{span}\left\{m \cdot r_S \mid r_S \in \mathcal{D}, m \in \mathcal{M}^{\ell}_{N'}\right\}\right) \\ &\le (k+1)\dim\left(\langle \partial_k S^d_N \rangle_{\le \ell}\right). \end{aligned}$$

We can now proceed with the proof of Theorem 1.1 or Theorem 3.1 exactly as before, using $\dim(\text{span}\,\mathcal{P}')$ as our lower bound for $\dim\left(\langle\partial_k S_N^d\rangle_{\le\ell}\right)$. For our choice of parameters, recall that $k,d \in o(\lg N)$, so $N' = N-k = \Omega(N)$ and $d = o(\lg N')$ as well. Hence the earlier proof goes through. We have established the following result.

Theorem 5.1. *Let $\alpha \in (0,1/2)$ be a constant. Let $N,d,k \in \mathbb{N}$ be such that $4k \le d \le \alpha \lg N/\lg\lg N$ and $k = \lfloor d/(\tau+1)\rfloor$ for some odd number $\tau \ge 3$. For any δ satisfying $\alpha \le 1-\delta(\tau+1) < 1-\delta\tau \le 1-\alpha$, and for $\ell = \lfloor N^{1-\delta}\rfloor$, the following holds.*

$$\dim\langle\partial_k S_N^d\rangle_{\le\ell} \geqslant \frac{\dim(\text{span}\{\mathcal{P}'\})}{k+1} \geqslant \frac{(1-o(1))}{k+1}\cdot\frac{\binom{N-k+\ell}{\ell}\cdot\binom{N-k-\ell}{k}}{(3N^{1-\delta\tau}/2)^k\cdot(d+1)^\tau}.$$

6 Lower bound on the size of depth-four circuits

In this section, we establish the lower bounds claimed in Theorem 1.2 and Corollary 1.3. As in [9], we say that a $\Sigma\Pi\Sigma\Pi$ circuit C is a $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ circuit 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 [9] and is stated explicitly in [13].

Lemma 6.1 ([13], 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*

$$\dim(\langle\partial_k P\rangle_{\le\ell}) \le s\cdot\binom{D}{k}\cdot\binom{N+\ell+(t-1)k}{\ell+(t-1)k}.$$

We are now ready to prove

Theorem 1.2 (Restated)**.** Let $\varepsilon \in (0,1)$ be a constant. Let $N,d,D,t \in \mathbb{N}$ be such that

$$\frac{10t}{\varepsilon} \le d \le \frac{\varepsilon\lg N}{5\lg\lg N} \qquad \text{and} \qquad D \le N^{1-\varepsilon}.$$

Any $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ circuit of top fan-in s computing S_N^d satisfies $s = N^{\Omega(d/t)}$.

Proof. Assume there exists a $\Sigma^s\Pi^{[D]}\Sigma\Pi^{[t]}$ circuit computing S_N^d.

We first illustrate the proof for one setting: the field has characteristic zero, $\varepsilon = 3/4$, and $4t+2$ divides d. Choosing $\alpha = 3/20$, $\tau = 4t+1$, $k = d/(4t+2)$, $\delta = 1/(2\tau)$, all the conditions for invoking Theorem 3.1 are met. From Theorem 3.1 and Lemma 6.1, when N is large enough, it holds that

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

For large enough N, since $k,\ell = o(N)$, we have

$$\frac{\binom{N-\ell}{k}}{\binom{D}{k}} \ge \left(\frac{N-\ell-k}{D}\right)^k \ge \left(\frac{N}{2D}\right)^k.$$

Since $kt < d = o(\lg N)$, for large enough N we have

$$\frac{\binom{N+\ell}{\ell}}{\binom{N+\ell+(t-1)k}{\ell+(t-1)k}} \geq \left(\frac{\ell}{N+\ell}\right)^{kt} \geq \left(\frac{1}{2N^{\delta}}\right)^{kt} \geq N^{-\delta tk - o(k)}.$$

By Fact 2.1, $(d+1)^{\tau} \leq (\lg N)^{\tau} \leq N^{\alpha}$.
We are given that $D \leq N^{1-\varepsilon}$.
Putting it all together, we have obtained that asymptotically,

$$s \geq \left(\frac{N}{2D \cdot N^{\delta t + o(1)} \cdot (3N^{1-\delta\tau}/2)}\right)^{k} \cdot \frac{1-o(1)}{N^{\alpha}} \geq \frac{1}{N^{\alpha}} \cdot \left(\frac{N}{N^{1-\varepsilon} \cdot N^{1-\delta\tau} \cdot N^{\delta t} \cdot N^{o(1)}}\right)^{k}. \tag{6.1}$$

By our choice of parameters, $1-\delta\tau = 1/2$. Also, $t \leq \tau/4$, so $\delta t \leq \delta\tau/4 = 1/8$. And $1-\varepsilon = 1/4$. Thus we see that Equation (6.1) yields a lower bound of $N^{\Omega(k)} = N^{\Omega(d/t)}$.

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 $\varepsilon > 0$. Firstly, to handle the absence of a divisibility constraint (in the above setting, we had assumed that $4t+2$ divides d), we should use Theorem 1.1 instead of Theorem 3.1. Then, we must choose α, τ, δ appropriately. It can be verified that if we choose $\alpha = \varepsilon/5$, $\delta = \varepsilon/(10t)$, and let τ be the smallest odd integer such that $1-\delta\tau \leq \varepsilon/2$, everything works out.

Finally, to obtain the same result over fields of positive characteristic, we can follow the same outline, replacing the use of Theorem 3.1 or Theorem 1.1 by Theorem 5.1. This Theorem gives a slightly weaker bound for $\dim(\langle\partial_k P\rangle_{\leq\ell})$. However, in the asymptotic bound stated in Theorem 1.2, the degradation of this bound is irrelevant. □

Corollary 1.3 (Restated)**.** Let parameters N, d, t be as in Theorem 1.2. Any $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit 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 $N^{\Omega(d/t)}$.

Proof. The first statement is an immediate corollary of Theorem 1.2 since $d/t \leq d = N^{o(1)}$. The second follows from the first by a standard trick [9]: given any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit, we can ensure that the fan-in of the layer-3 product gates is at most $O(d/t)$ by repeatedly multiplying out pairs of polynomials of degree at most $t/2$ that feed into it. This does not change the top fan-in of the circuit and ensures that the bottom fan-in remains bounded by t. At the end of this procedure, each Π gate on layer 3 has at most 1 polynomial of degree $< t/2$ feeding into it; homogeneity now entails that the fan-in of the Π-gate must be at most $2d/t+1$. □

Remark 6.2. The above corollary also yields a lower bound for the model of *regular formulas*, which were introduced and studied in the work of Kayal et al. [13]. [13, Theorem 9] shows that if we have a lower bound of $N^{\Omega(d/t)}$ for $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing a polynomial $F[x_1,\ldots,x_N]$ for all $t < d/100$ then it follows that any regular formula computing F must have size $N^{\Omega(\log t)}$. Using this theorem in conjunction with Corollary 1.3 above we immediately get a lower bound of $N^{\Omega(\log d)}$ on the size of any regular formula computing $S_N^d(x_1,\ldots,x_N)$ for d as in Corollary 1.3.

References

[1] MANINDRA AGRAWAL AND V. VINAY: Arithmetic circuits: A chasm at depth four. In *Proc. 49th FOCS*, pp. 67–75. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.32] 2

[2] NOGA ALON: Perturbed identity matrices have high rank: Proof and applications. *Combin. Probab. Comput.*, 18(1-2):3–15, 2009. [doi:10.1017/S0963548307008917] 4

[3] LÁSZLÓ BABAI AND PETER FRANKL: *Linear Algebra Methods in Combinatorics*. 1992. Book manuscript, University of Chicago. 5

[4] WALTER BAUR AND VOLKER STRASSEN: The complexity of partial derivatives. *Theoret. Comput. Sci.*, 22(3):317–330, 1983. [doi:10.1016/0304-3975(83)90110-X] 2

[5] HERVÉ FOURNIER, NUTAN LIMAYE, MEENA MAHAJAN, AND SRIKANTH SRINIVASAN: The shifted partial derivative complexity of elementary symmetric polynomials. In *Proc. 40th Math. Found. Comp. Sci. (MFCS'15)*, volume 9235 of *LNCS*, pp. 324–335. Springer, 2015. Preliminary version in ECCC. [doi:10.1007/978-3-662-48054-0_27] 1

[6] HERVÉ FOURNIER, NUTAN LIMAYE, GUILLAUME MALOD, AND SRIKANTH SRINIVASAN: Lower bounds for depth-4 formulas computing iterated matrix multiplication. *SIAM J. Comput.*, 44(5):1173–1201, 2015. Preliminary version in STOC'14. [doi:10.1137/140990280] 2, 3, 4

[7] PETER FRANKL: Intersection theorems and $\bmod p$ rank of inclusion matrices. *J. Comb. Theory, Ser. A*, 54(1):85–94, 1990. [doi:10.1016/0097-3165(90)90007-J] 5

[8] PETER FRANKL AND RICHARD M. WILSON: Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981. [doi:10.1007/BF02579457] 5

[9] ANKIT GUPTA, PRITISH KAMATH, NEERAJ KAYAL, AND RAMPRASAD SAPTHARISHI: Approaching the chasm at depth four. *J. ACM*, 61(6):33:1–33:16, 2014. Preliminary version in CCC'13. [doi:10.1145/2629541] 2, 3, 6, 29, 30

[10] PAVEL HRUBEŠ AND AMIR YEHUDAYOFF: Homogeneous formulas and symmetric polynomials. *Comput. Complexity*, 20(3):559–578, 2011. [doi:10.1007/s00037-011-0007-3, arXiv:0907.2621] 3

[11] NEERAJ KAYAL: An exponential lower bound for the sum of powers of bounded degree polynomials. *Elect. Colloq. Comput. Complexity (ECCC)*, 19(81), 2012. ECCC. 2, 6

[12] NEERAJ KAYAL, NUTAN LIMAYE, CHANDAN SAHA, AND SRIKANTH SRINIVASAN: An exponential lower bound for homogeneous depth four arithmetic formulas. *SIAM J. Comput.*, 46(1):307–335, 2017. Preliminary version in FOCS'14. [doi:10.1137/151002423] 2, 4

[13] NEERAJ KAYAL, CHANDAN SAHA, AND RAMPRASAD SAPTHARISHI: A super-polynomial lower bound for regular arithmetic formulas. In *Proc. 46th STOC*, pp. 146–153. ACM Press, 2014. Preliminary version in ECCC. [doi:10.1145/2591796.2591847] 2, 29, 30

[14] PETER KEEVASH AND BENNY SUDAKOV: Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. *SIAM J. Discrete Math.*, 18(4):713–727, 2005. [doi:10.1137/S0895480103434634] 5

[15] PASCAL KOIRAN: Arithmetic circuits: The chasm at depth four gets wider. *Theoret. Comput. Sci.*, 448:56–65, 2012. [doi:10.1016/j.tcs.2012.03.041, arXiv:1006.4700] 2

[16] MRINAL KUMAR AND SHUBHANGI SARAF: On the power of homogeneous depth 4 arithmetic circuits. In *Proc. 55th FOCS*, pp. 364–373. IEEE Comp. Soc. Press, 2014. Preliminary version in ECCC. [doi:10.1109/FOCS.2014.46, arXiv:1404.1950] 2, 3, 4

[17] MRINAL KUMAR AND SHUBHANGI SARAF: The limits of depth reduction for arithmetic formulas: It's all about the top fan-in. *SIAM J. Comput.*, 44(6):1601–1625, 2015. Preliminary version in STOC'14. [doi:10.1137/140999220] 2, 3

[18] EYAL KUSHILEVITZ AND NOAM NISAN: *Communication Complexity*. Cambridge Univ. Press, 1997. 4, 5, 7

[19] NOAM NISAN AND AVI WIGDERSON: Lower bounds on arithmetic circuits via partial derivatives. *Comput. Complexity*, 6(3):217–234, 1997. Preliminary version in FOCS'95. [doi:10.1007/BF01294256] 2, 3, 4

[20] RAN RAZ: Separation of multilinear circuit and formula size. *Theory of Computing*, 2(6):121–135, 2006. [doi:10.4086/toc.2006.v002a006] 2

[21] RAN RAZ: Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2):8:1–8:17, 2009. Preliminary versions in STOC'04 and ECCC. [doi:10.1145/1502793.1502797] 2

[22] ALEXANDER A. RAZBOROV: Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Math. Notes of the Acad. Sci. USSR*, 41(4):333–338, 1987. [doi:10.1007/BF01137685] 5

[23] AMIR SHPILKA: Affine projections of symmetric polynomials. *J. Comput. System Sci.*, 65(4):639–659, 2002. Preliminary version in CCC'01. [doi:10.1016/S0022-0000(02)00021-1] 3

[24] AMIR SHPILKA AND AVI WIGDERSON: Depth-3 arithmetic circuits over fields of characteristic zero. *Comput. Complexity*, 10(1):1–27, 2001. Preliminary version in CCC'99. [doi:10.1007/PL00001609] 3, 4

[25] SÉBASTIEN TAVENAS: Improved bounds for reduction to depth 4 and depth 3. *Inform. and Comput.*, 240:2–11, 2015. Preliminary version in MFCS'13. [doi:10.1016/j.ic.2014.09.004, arXiv:1304.5777] 2

[26] LESLIE G. VALIANT: Completeness classes in algebra. In *Proc. 11th STOC*, pp. 249–261. ACM Press, 1979. [doi:10.1145/800135.804419] 2

[27] LESLIE G. VALIANT, SVEN SKYUM, STUART J. BERKOWITZ, AND CHARLES RACKOFF: Fast parallel computation of polynomials using few processors. *SIAM J. Comput.*, 12(4):641–644, 1983. Preliminary version in MFCS'81. [doi:10.1137/0212043] 2

[28] RICHARD M. WILSON: A diagonal form for the incidence matrices of t-subsets vs. k-subsets. *European J. Combin.*, 11(6):609–615, 1990. [doi:10.1016/S0195-6698(13)80046-7] 4, 5, 26

AUTHORS

Hervé Fournier
Institut de Mathématiques de Jussieu-Paris Rive Gauche
Université Paris Diderot, France
fournier@math.univ-paris-diderot.fr
`https://webusers.imj-prg.fr/~herve.fournier`

Nutan Limaye
Department of Computer Science and Engineering
IIT Bombay, Mumbai, India
nutan@cse.iitb.ac.in
`https://www.cse.iitb.ac.in/~nutan`

Meena Mahajan
The Institute of Mathematical Sciences, Chennai, India
meena@imsc.res.in
`http://www.imsc.res.in/~meena`

Srikanth Srinivasan
Department of Mathematics
IIT Bombay, Mumbai, India
srikanth@math.iitb.ac.in

ABOUT THE AUTHORS

HERVÉ FOURNIER obtained his Ph. D. at ENS Lyon under the supervision of Pascal Koiran. His interests include algebraic complexity and its connections to computational geometry.

NUTAN LIMAYE graduated from The Institute of Mathematical Sciences, Chennai, India, in 2009; her advisor was Meena Mahajan. Her thesis focused on the interconnection between language classes and complexity classes. She is interested in Boolean and arithmetic circuit complexity and graph algorithms. She likes all things Japanese and has been trying to learn Japanese for the past year.

MEENA MAHAJAN is a professor in the theoretical computer science group at the Institute of Mathematical Sciences, Chennai, India, reaching there by way of education at the IITs at Mumbai and Chennai. At work, she researches questions concerning computational complexity. As recreation, she loves solving jigsaw puzzles (creating order out of disorder), and also solving combinatorial puzzles, which crop up in real-life situations far oftener than one may think.

SRIKANTH SRINIVASAN got his undergraduate degree from the Indian Institute of Technology Madras, where his interest in the theory side of CS was piqued under the tutelage of N. S. Narayanswamy. Subsequently, he obtained his Ph. D. from The Institute of Mathematical Sciences in 2011; his advisor was V. Arvind. His research interests span all of TCS (in theory), but in practice are limited to circuit complexity, derandomization, and related areas of mathematics. He enjoys running and pretending to play badminton.