View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dagstuhl Research Online Publication Server

An Exponential Lower Bound for Homogeneous Depth-5 Circuits over Finite Fields*

Mrinal Kumar[1] and Ramprasad Saptharishi[2]

1 **Rutgers University, New Brunswick, NJ, USA**
mrinalkumar08@gmail.com
2 **Tata Institute of Fundamental Research, Mumbai, India**
ramprasad@tifr.res.in

Abstract

In this paper, we show exponential lower bounds for the class of homogeneous depth-5 circuits over all small finite fields. More formally, we show that there is an explicit family $\{P_d\}$ of polynomials in VNP, where P_d is of degree d in $n = d^{O(1)}$ variables, such that over all finite fields $GF(q)$, any homogeneous depth-5 circuit which computes P_d must have size at least $\exp(\Omega_q(\sqrt{d}))$.

To the best of our knowledge, this is the first super-polynomial lower bound for this class for any non-binary field.

Our proof builds up on the ideas developed on the way to proving lower bounds for homogeneous depth-4 circuits [Gupta et al., Fournier et al., Kayal et al., Kumar-Saraf] and for non-homogeneous depth-3 circuits over finite fields [Grigoriev-Karpinski, Grigoriev-Razborov]. Our key insight is to look at the space of shifted partial derivatives of a polynomial as a space of functions from $GF(q)^n$ to $GF(q)$ as opposed to looking at them as a space of formal polynomials and builds over a tighter analysis of the lower bound of Kumar and Saraf [Kumar-Saraf].

1998 ACM Subject Classification I.1.1 Expressions and Their Representation

Keywords and phrases arithmetic circuits, lower bounds, separations, depth reduction

Digital Object Identifier 10.4230/LIPIcs.CCC.2017.31

1 Introduction

Arithmetic circuits are the most natural model to study computations of multivariate polynomials. These are directed acyclic graphs, with a unique sink node called the root or output gate, internal nodes are labeled by addition and multiplication gates[1], and leaves are labeled by variables or constants from the underlying field. The field of arithmetic circuit complexity aims at understanding the hardness of multivariate polynomials in terms of the size of the smallest arithmetic circuit computing it. One of the most important questions in this field of study is to show that there are families of explicit *low-degree*[2] polynomials that require arithmetic circuits of super-polynomial size (in terms of n, the number of variables). It is widely believed that the symbolic $n \times n$ permanent, often denoted by Perm_n, requires circuits of size $\exp(\Omega(n))$ but, as of now, we do not even have a $\Omega(n^2)$ lower bound for any explicit polynomial.

* The first author's research supported in part by a Simons Graduate Fellowship. The second author's research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

[1] Throughout this paper, we consider circuits as having gates of unbounded fan-in.

[2] Where the degree is bounded by a polynomial function in the number of variables.

© Mrinal Kumar and Ramprasad Saptharishi;
licensed under Creative Commons License CC-BY
32nd Computational Complexity Conference (CCC 2017).
Editor: Ryan O'Donnell; Article No. 31; pp. 31:1–31:30
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Depth Reductions

In the absence of much progress on the question of lower bounds for general arithmetic circuits, a natural question to ask is if we can prove good lower bounds for nontrivial restricted classes of circuits. One particular class of circuits which have been widely studied with this aim are the class of bounded depth[3] arithmetic circuits. It turns out that this is not just an attempt to study a simpler model, but there is a formal connection between lower bounds for bounded depth circuits and lower bounds for general circuits. A sequence of structural results, often referred to as *depth reduction* results, show that strong enough lower bounds for bounded depth circuits implies lower bounds for general arithmetic circuits.

The first depth reduction for arithmetic circuits was by Hyafil [10] who showed that any polynomial computed by a polynomial sized arithmetic circuit can be equivalently computed by a circuit of depth $O(\log d)$ and *quasi-polynomial* size. This was improved by Valiant, Skyum, Berkowitz and Rackoff [26], who showed that any n-variate degree d polynomial that can be computed by a circuit of size $(nd)^{O(1)}$ can be equivalently computed by a circuit of depth $O(\log d)$ and size $(nd)^{O(1)}$. Thus, proving super-polynomial lower bounds for $O(\log d)$ depth circuits is sufficient to prove super-polynomial lower bounds for general arithmetic circuits. Agrawal and Vinay [2] further strengthened this to obtain a depth reduction to depth-4 circuits by showing that any n-variate degree d polynomial that can be computed by a $2^{o(n)}$ sized circuit can be equivalently computed by *homogeneous*[4] depth-4 circuit of size $2^{o(n)}$. Their result was strengthened by Koiran [16] and Tavenas [25] to show that any circuit of size s that computes an n-variate degree d polynomial can be computed by a homogeneous depth-4 circuit of size $s^{O(\sqrt{d})}$, and in fact the resulting depth-4 circuits have all multiplication fan-ins bounded by $O(\sqrt{d})$. These results hold over all fields.

Over any field of characteristic zero, Gupta, Kamath, Kayal and Saptharishi [8] showed that any n-variate degree d polynomial computed by a size s circuit can be equivalently computed by a non-homogeneous depth-3 circuit of size $s^{O(\sqrt{d})}$. Thus, these results formally show that proving good enough lower bounds on circuits of bounded depth is sufficient for proving lower bounds for general circuits.

Lower bounds for depth-3 and depth-4 circuits

Nisan and Wigderson [20] proved an $\exp(\Omega(n))$ lower bound for any *homogeneous* depth-3 circuits computing the symbolic $n \times n$ determinant Det_n by studying dimension of the partial derivatives of Det_n as polynomials. Grigoriev and Karpinski [6] and Grigoriev and Razborov [7] extended this to prove an $\exp(\Omega(n))$ lower bound for non-homogeneous depth-3 circuit computing Det_n over any fixed finite field $\mathbb{F}_q$. Chillara and Mukhopadhyay [4] extended this to give an $\exp(\Omega_q(d \log n))$ lower bound for non-homogeneous depth-3 circuits computing an n-variate degree d polynomial in VP. It is worth noting that there is no generic method known to convert a boolean lower bound for $\mathsf{AC}^0[\mathrm{mod} q]$ to lower bounds for arithmetic circuits over $\mathbb{F}_q$ (discussed in more detail in Section 3.1).

The proofs of [6, 4] also studied the dimension of partial derivatives of polynomial, but unlike the proof in [20], they looked at partial derivatives as functions from $\mathbb{F}_q^n \to \mathbb{F}_q$. The proofs in [6], [7] and [4] strongly rely on the fact that we are working over small finite

[3] A depth k arithmetic circuit consists of k layers of alternating sum and multiplication gates with the output being computed by a sum gate.

[4] Which means that all intermediate computations are homogeneous polynomials. Hence the degree of any intermediate computation is bounded by the degree of the output polynomial.

fields, and completely break down over larger fields or fields of large characteristic. Over fields of characteristic zero and over algebraic closure of finite fields, the question of proving super-polynomial lower bounds for non-homogeneous depth three circuits continues to remain wide open.

Even though we had exponential lower bounds for homogeneous depth-3 circuits, the question of proving super-polynomial lower bounds for homogeneous depth-4 circuits remained open for more than a decade. In 2012, Kayal [12] introduced the notion of *shifted partial derivatives*, which is a generalization of the well-known notion of partial derivatives. Shifted partial derivatives have been very influential in a plethora of lower bounds for depth-4 circuits in the past few years. Gupta et. al. [9] used this measure to prove an $\exp(\Omega(\sqrt{n}))$ lower bound for the size of homogeneous depth-4 circuits with multiplication fan-ins bounded by $O(\sqrt{n})$. Subsequently, lower bounds of $\exp(\Omega(\sqrt{d}\log n))$ were proved for other n-variate degree d polynomials computed by almost the same circuit class [15, 5, 18]. (It is worth noting that getting a lower bound of $\exp(\omega(\sqrt{d}\log n))$ would have implied a super-polynomial lower bound for general circuits!) Using a more delicate variant called *projected shifted partials*, Kayal et. al. [13] and Kumar and Saraf [19] proved lower bounds of $\exp(\Omega(\sqrt{d}\log n))$ for homogeneous depth-4 circuits (without any fan-in restrictions) via two very different analyses. The former was an analytic approach and works only over characteristic zero fields, whereas the latter was purely combinatorial and works over any field. These techniques have also been applied to yield lower bounds for non-homogeneous depth-3 circuits with bounded bottom fan-in [14] and homogeneous depth-5 circuits with bounded bottom fan-in [3]. A continuous updated survey [23] contains expositions of many of the lower bounds and depth reduction results listed above.

The results in [19] in fact show that the reduction from general arithmetic circuits to depth-4 circuits with support $O(\sqrt{d})$ cannot be improved, as they give an example of a polynomial in VP for which any depth-4 circuits of support $O(\sqrt{d})$ must be of size $n^{\Omega(\sqrt{d})}$. Further, with the current upper-bounds for the projected shifted partials on such depth-4 circuits, the best we can hope to prove using this measure is an $n^{\Omega(\sqrt{d})}$ lower bound. Hence, it might be insufficient for general arithmetic circuits lower bounds but it could well be the case that we might be able to prove stronger lower bounds for constant depth arithmetic circuits, or arithmetic formulas by variants of this family of measures.

Hence, as a start, the problem of proving lower bounds for homogeneous depth five circuits, seems like the next natural question to explore. This already seems to introduce new challenges as the proofs of lower bounds for homogeneous depth-4 circuits seem to break down for homogeneous depth-5 circuits. In this paper, we pursue this line of enquiry, and prove exponential lower bounds for homogeneous depth-5 circuits over small finite fields. Before stating our results, we first discuss prior results on this question, and the challenges involved in extending the proofs of lower bounds for homogeneous depth four circuits, in the next section.

Lower bounds for depth-5 circuits

Prior to this work, the only known lower bounds for depth-5 circuits that we are aware of are the results of Raz [21], which show super-linear lower bounds for bounded depth circuits over large enough fields, the results of Kalorkoti [11] which show quadratic lower bounds for arithmetic formulas and the results of Kayal and Saha [14] and of Bera and Chakrabarti [3] which show exponential lower bounds for homogeneous depth-5 circuits if the bottom fan-in is bounded.

Given that we have lower bounds for homogeneous depth-4 circuits, it seems natural to try and apply these techniques to prove lower bounds for homogeneous depth-5 circuits.

Unfortunately, the obvious attempts to generalize the proofs in [13, 19] seem to fail for homogeneous depth-5 circuits. We now elaborate on this.

On extending the depth-4 lower bound proofs to depth-5 circuits: To understand these issues, we first need a birds-eye view of the major steps in the proofs of lower bounds for depth-4 circuits [13, 19]. These proofs have two major components.

- **Reduction to depth-4 circuits with bounded bottom support:** In the first step, the circuit C and the polynomial are hit with a random restriction, in which each variable is kept alive independently with some small probability p. The observation is that a bottom level product gate in C of support (the number of distinct variable inputs) at least s survives with probability at most p^s. Therefore, the probability that some bottom product of support at least s in C survives is at most $\text{Size}(C) \cdot p^s$. Now, if the size of C is small (say $\epsilon \cdot 1/p^s$), then this probability is quite small, so with a high probability C reduces to a homogeneous depth-4 circuit with bounded bottom support.
- **Lower bounds for depth-4 circuits with bounded bottom support:** The goal in the second step is to show that the polynomial obtained after random restrictions still remains hard for homogeneous depth-4 circuits with bottom support at most s.

The key point in step 1 is that if $\text{Size}(C)$ is not too large, then we can assume that with a high probability over the random restrictions, all the high support product gates are set to 0. This is where things are not quite the same for depth-5 circuits. When we express a homogeneous depth-5 circuit as a homogeneous depth-4 circuit by expanding the product of linear forms at level four, we might increase the number of monomials a lot (potentially to all possible monomials). Now, the random restriction step no longer works and we do not have a reduction to homogeneous depth-4 circuits with bounded bottom support. If the bottom fan-in of C is bounded, then this strategy does indeed generalize. Kayal and Saha [14] and Bera and Chakrabarti [3] show exponential lower bounds for such cases.

It is not clear to us how fundamental this obstruction is, but our key insight is a strategy for proving lower bounds for homogeneous depth-4 circuits that avoids the random restriction step. Morally speaking, we *do* proceed by a 'reduction' from a depth-5 circuit to a depth-4 circuit, but the meaning of a 'reduction' here is more subtle and largely remains implicit.

Our Contribution

We give an exponential lower bound for homogeneous depth-5 circuits over any fixed finite field $\mathbb{F}_q$. To the best of our understanding, this is the first such lower bound for depth-5 circuits over any field apart from $\mathbb{F}_2$[5]. Stated precisely, we prove the following theorem.

▶ **Theorem 1.** *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$*, with* $\mathsf{Deg}(P_d) = d$*, in the class* VNP *such that for any finite field* $\mathbb{F}_q$*, any homogeneous depth-*5 *circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$*.*

The polynomial P_d is from the Nisan-Wigderson family of polynomials (introduced by [15], Definition 3) with carefully chosen parameters.

Our proof also extends to non-homogeneous depth-5 circuits where the layer of multiplication gates closer to the output have fan-in bounded by $O(\sqrt{d})$ (with no restriction on the fan-in of the other multiplication layer).

[5] For $\mathbb{F}_2$, exponential lower bounds easily follow from the lower bounds of Razborov [22].

▶ **Theorem 2.** *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$*, with* $\mathsf{Deg}(P_d) = d$*, in the class* VNP *such that for any finite field* $\mathbb{F}_q$*, any* $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi\Sigma$ *circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$*.*

It is worth mentioning that for characteristic zero fields, it suffices to prove a lower bound of $\exp(\omega(d^{1/3}\log d))$ for an explicit polynomial computed by such $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi\Sigma$ circuits to separate VP from VNP (by combining the depth reductions of [2, 16, 25] and [8]). We elaborate on this in Section 3.5. Such a phenomenon also happens for non-homogeneous depth three circuits, where over finite fields, we know quite strong lower bounds while much weaker ones would imply $\mathsf{VNP} \neq \mathsf{VP}$ over fields of characteristic zero.

The key technical ingredient of our proof is to look at the space of shifted partial derivatives and the projected shifted partial derivatives of a polynomial. We study them as a space of functions from $\mathbb{F}_q^n \to \mathbb{F}_q$ as opposed to as a space of formal polynomials, as has been the case for the results obtained so far. This perspective allows us the freedom to confine our attention to the evaluations of the shifted partial derivatives of a polynomial on certain well chosen subsets of $\mathbb{F}_q^n$, and this turns out to be critical to our cause. This leads to a new family of complexity measures which could have applications to other lower bound questions as well. Our proof also involves a tighter analysis of the lower bound of Kumar and Saraf [19] (for homogeneous depth-4 circuits) which may be interesting in its own right.

We now give an overview of our proof.

2 An overview of the proof

The proof would consist of the following main steps:

1. Define a function $\Gamma : \mathbb{F}_q[\mathbf{x}] \to \mathbb{N}$. Intuitively, we think of $\Gamma(P)$ to be a measure of the *complexity* of P.
2. For all homogeneous depth-5 circuits C of size at most $\exp(\delta\sqrt{d})$, prove an upper bound on $\Gamma(C)$.
3. For the target hard polynomial P, show that $\Gamma(P)$ is much larger than the upper bound proved in step 2.

The complexity measure: At a high level, the proofs of lower bounds in [20, 9, 15, 5, 18, 13, 19] associate a linear space polynomials to every polynomial in $\mathbb{F}_q[\mathbf{x}]$ and use the dimension of this space over $\mathbb{F}_q$ as a measure of complexity of the polynomial. The mapping from polynomials to linear space of polynomials undergoes subtle changes as we go from the proof of lower bounds for homogeneous depth-3 circuits [20] to lower bounds for homogeneous depth-4 circuits [13, 19].

In this paper, we follow this outline and associate to every polynomial, the space of its shifted partial derivatives as defined in [9]. However, instead of working with this space of polynomials as it is, we study their evaluation vectors over a subset of $\mathbb{F}_q^n$ (similar to [6, 7], where they worked with partial derivatives of a polynomial). The key gain that we have from this change in outlook is that as evaluation vectors, we can choose to confine our attention to evaluations on certain properly chosen subsets of $\mathbb{F}_q^n$. For formal polynomials, it is not clear what should be the correct analog of this approximation. The necessity and the utility of this will be more clear as we go along.

High rank products of linear forms: Consider a polynomial Q which is a product of τ linearly independent (non-constant) linear forms $L_1, L_2, \ldots, L_\tau$.

$$Q = \prod_{i=1}^{\tau} L_i \,.$$

It is not hard to see that

$$\Pr_{\mathbf{a}\in\mathbb{F}_q^n}[Q(\mathbf{a}) \neq 0] \leq \left(1 - \frac{1}{q}\right)^{\tau} .$$

In other words, products of linear forms of rank τ vanish on all but a $o(1)$ fraction of the entire space if $\tau = \omega(1)$. If the size of a depth-5 circuit is not too large as a function of τ (say, at most $\exp(\delta\tau)$ for a small enough $\delta > 0$), then by a union bound, all the products of rank at least τ at the fourth level vanish everywhere apart from a $o(1)$ fraction of the points in $\mathbb{F}_q^n$.

In summary, we just argued that a depth-5 circuit C over $\mathbb{F}_q$ of size at most $\exp(\delta\tau)$ can be approximated by a sub-circuit C' of C which is obtained from C by dropping all products of linear forms of rank at least τ from the bottom level.

Low rank products of linear forms: A second simple observation (Lemma 8) shows that for every product of linear forms of rank at most τ, there is a polynomial of degree at most $(q-1)\tau$, such that they agree at all points in $\mathbb{F}_q^n$. Thus, the circuit C' is equal, as a function from $\mathbb{F}_q^n \to \mathbb{F}_q$ to a depth-4 circuit C'' of bottom fan-in at most $(q-1)\tau$. Moreover, the formal degree and the top fan-in of C'' are upper bounded by the formal degree and top fan-in of C, respectively.

Putting things together This implies that for every homogeneous depth-5 circuit C computing a polynomial of degree d of size at most $\exp(\delta\tau)$ for some τ, there exists a depth-4 circuit C'' of formal degree at most d and top fan-in at most $\exp(\delta\tau)$ such that

$$\Pr_{\mathbf{a}\in\mathbb{F}_q^n}[C(\mathbf{a}) \neq C''(\mathbf{a})] \leq o(1).$$

Therefore, a polynomial P which can be computed by C can be *approximated* by C'' in the point-wise sense. Since we know lower bounds on the top fan-in of homogeneous (and low formal degree) depth-4 circuits with bounded bottom fan-in [9, 15], it seems that we only have a small way to go. Unfortunately, we do not quite know how to make this idea work. The key technical obstacle here is that it seems to be hard to say much about the partial derivatives of C by looking at partial derivatives of C''. As a pathological case, the polynomial $\prod_{i\in[n]} x_i$ has a partial derivative span of size 2^n but is well approximated by the 0 polynomial over $\mathbb{F}_2$.

If we had started with a depth-3 circuit instead of a depth-5 circuit, then such a strategy is indeed known to work [7]. Observe that in this case it is enough to show that that there is an explicit polynomial which cannot be approximated well by a low degree polynomial over $\mathbb{F}_q$. In [7], the authors show this by an adaptation of a similar result of Smolenksy [24] over $\mathbb{F}_2$.

A strengthening of the strategy: The key additional observation that helps us make things work is the fact that not only do high rank product gates at level four of C vanish almost everywhere on $\mathbb{F}_q^n$, but they vanish with a high multiplicity. As we show in Corollary 11, if

the size of C is not *too large*, all the product gates of rank at least τ vanish with a multiplicity $\Omega(\tau)$ at $1 - o(1)$ fraction of points on $\mathbb{F}_q^n$[6].

Therefore, not only can C agree with C' almost everywhere, all the partial derivatives of C of order at most $k = \Omega(\tau)$ agree with all the partial derivatives of C' almost everywhere. This already hints at the fact that if we are to take advantage of this then we should be looking at the evaluation of these derivatives, since our only guarantee is about their evaluations.

In [6], exponential lower bounds were proved for non-homogeneous depth-3 circuits using a very similar strategy. However, adapting the method for shifted partials and projected shifted partials seems to be a challenge.

In Section 5, we show that the dimension of the span of evaluation vectors of shifted partial derivatives of C, when restricted to a properly chosen subset S of $\mathbb{F}_q^n$, is small if the size of the circuit C we started with was small.

As a final step of the proof, we show that with respect to this complexity measure, our target hard polynomial (from the so-called *Nisan-Wigderson* family, defined below) has a large complexity.

▶ **Definition 3** (Nisan-Wigderson polynomial families)**.** Let d, m, e be arbitrary parameters with m being a power of a prime, and $d, e \leq m$. Since m is a power of a prime, let us identify the set $[m]$ with the field $\mathbb{F}_m$ of m elements. Note that since $d \leq m$, we have that $[d] \subseteq \mathbb{F}_m$. The Nisan-Wigderson polynomial with parameters d, m, e, denoted by $\mathsf{NW}_{d,m,e}$ is defined as

$$\mathsf{NW}_{d,m,e}(\mathbf{x}) \quad = \quad \sum_{\substack{p(t) \in \mathbb{F}_m[t] \\ \mathsf{Deg}(p) < e}} x_{1,p(1)} \dots x_{d,p(d)}$$

That is, for every univariate polynomial $p(t) \in \mathbb{F}_m[t]$ of degree less than e, we add one monomial that encodes the 'graph' of p on the points $[d]$. This is a homogeneous, multilinear polynomial of degree d over dm variables with exactly m^e monomials.

This step of the proof builds on a tighter analysis of the lower bound on the dimension of the span of *projected shifted partial derivatives* of the Nisan-Wigderson polynomials in [19]. We show that if the set S is carefully chosen, then we do not incur much loss in the dimension by going from looking at shifted partial derivatives as formal polynomials to looking at them as functions over a small subset of the finite field. We provide the details in Section 6.

One important technicality is the dependency between various parameters involved. For our proof, the choice of k would be $O_q(\tau)$ and would depend on q. The lower bound of [19] would then choose specific parameters for the $\mathsf{NW}_{d,m,e}$. This would mean that for every q, we get a *different* polynomial for which we show a lower bound. We remedy the order of quantifiers and start by fixing specific parameters for $\mathsf{NW}_{d,m,e}$. Then, depending on q, we choose a restriction of $\mathsf{NW}_{d,m,e}$ that would be identical to $\mathsf{NW}_{d',m,e}$ by setting some variables to 0/1. We then apply the [19] argument for this restriction to obtain our lower bound for $\mathsf{NW}_{d',m,e}$ which also yields a lower bound for $\mathsf{NW}_{d,m,e}$. The details are in Section 7.1.

3 Some discussion and open problems

3.1 Connections between arithmetic circuits over $\mathbb{F}_q$ and $\mathsf{AC}^0[\mathrm{mod}\, q]$

Although constant depth arithmetic circuits over $\mathbb{F}_q$ appear to be similar to the class $\mathsf{AC}^0[\mathrm{mod} q]$, they are surprisingly very different with respect to functions computed by them.

[6] In the rest of this discussion, we will think of τ as $\Theta(\sqrt{d})$.

A striking example, due to Agrawal, Allender and Datta [1], is that arithmetic circuits over $\mathbb{F}_3$ can "compute" both the Mod3 function, as well as the Mod2 function via

$$\mathrm{Mod2}(x_1, \ldots, x_n) \quad = \quad \left(2 + \prod_{i=1}^{n}(1 + x_i)\right)^2 .$$

However, it is true that functions computed by arithmetic circuits over $\mathbb{F}_{p^k}$ have strong connections with $\mathsf{AC}^0[\mathrm{mod} p(p^k - 1)]$ but unless we are working over $\mathbb{F}_2$ it seems difficult to lift a lower bound for $\mathrm{AC}^0[\mathrm{mod} p]$ to arithmetic circuits over $\mathbb{F}_p$. For more on this, see [1].

The only exception we know of is the result of Grigoriev and Razborov [7] where they lift Smolensky's [24] lower bound for $\mathrm{AC}^0[\mathrm{mod} p]$ to depth-3 arithmetic circuits over $\mathbb{F}_p$, and this crucially uses the fact that depth-3 arithmetic circuits can be point-wise approximated by a "sparse polynomial". But in general, constant depth arithmetic circuits over $\mathbb{F}_p$ and Boolean circuits in $\mathrm{AC}^0[\mathrm{mod} p]$ seem to be two very different classes.

3.2 Lower bounds for iterated matrix multiplication

Given the results in this paper, one might wonder if the lower bounds in this paper work for a polynomial in VP. One natural candidate polynomial for which one might hope to show such a lower bound would be the iterated matrix multiplication polynomial (IMM). It was shown in [19] that IMM has a large complexity with respect to the measure of projected shifted partial derivatives. Unfortunately, the bounds in [19] only show that the dimension of the space of projected shifted partial derivatives of the IMM (degree d in $d^{O(1)}$ variables) are a factor $\exp(\delta\sqrt{d}\log d)$ close to the maximum possible value for some constant δ. This slack seems to be insufficient for our proofs in this paper to work as in the proof of Lemma 24, we would have to rely on the fact that for the polynomial NW, the projected shifted partials complexity was at most a quasi-polynomial factor away from the largest possible.

3.3 Finer separations for bounded depth circuits?

In [18], it was shown that homogeneous depth-4 circuits are exponentially more powerful than homogeneous depth-4 circuits with bounded bottom fan-in. A natural question to ask is whether such separations can be shown between homogeneous depth-4 circuit and homogeneous depth-5 circuits. One of the first strategies to attempt for this question would be to try and show that there is a homogeneous depth-5 circuit such that its projected shifted partial derivative complexity is quite large. The results in this paper show that the measure can not to be too close to the largest possible value, in particular it needs to be at least a factor $\exp(\Omega(\sqrt{d}))$ away from the largest possible value. If this is bound is tight, then such a separation between homogeneous depth-5 circuits and homogeneous depth-4 circuits can still be shown using projected shifted partial derivatives. However, it is not clear if this is the case. As mentioned before, even the known lower bounds on the dimension of the projected shifted partials for the IMM seem a factor $\exp(\Omega(\sqrt{d}\log d))$ away from the largest possible value.

In a recent result [17], using a different measure (called dimension of shifted projected partials, first used by [13]) such a separation between homogeneous depth-4 and homogeneous depth-5 circuits was shown in the regime when $d = O(\log^2 n)$. Extending this for other regimes of d continues to remain open.

3.4 The tightness of the results and relevance to VP vs. VNP

For homogeneous depth-4 circuits, we know $\exp(\Omega(\sqrt{d}\log d))$ lower bounds [13, 19] and any asymptotic improvement in the exponent would imply general arithmetic circuit lower bounds. In this sense, the lower bounds for homogeneous depth-4 circuits are tight, over all fields. It is natural to ask, if the bounds in this paper are tight in this sense? The answer to this question is far from obvious to us. In particular, it is not clear if we can use computational advantage of having linear forms at the bottom level of the circuit to get a better depth reduction from VP to homogeneous depth-5 circuits, when compared to depth reduction to homogeneous depth-4 circuits.

3.5 Lower bounds over fields of characteristic zero?

One might wonder if the techniques in this paper could be potentially adapted to work for depth-5 circuits over fields of characteristic zero. As in the work of Grigoriev and Karpinski [6], our proof (Lemma 13 in particular) strongly relies on the fact that we are working over a fixed finite field, so it clearly seems hard to generalize over large fields (even when the characteristic is small). In addition to this obvious technical obstruction to generalizing the proof in this paper, there seems to be another reason why the proof strategy in this paper could be hard to replicate over fields of characteristic zero, namely, an analog of Theorem 2 over fields of characteristic zero would imply that $\mathsf{VP} \neq \mathsf{VNP}$. The reason is that over characteristic zero fields, one can obtain better depth reductions to non-homogeneous depth-5 circuits by combining [2, 16, 25] with [8]. Although this is reasonably well known, we give a formal proof here for completeness.

The following lemma is a simple generalization of the proof of depth reduction to depth-4 circuits by Tavenas [25].

▶ **Lemma 4** (Depth reduction to homogeneous depth six circuits). *Let P be a polynomial of degree d in $\mathrm{poly}(d)$ variables which can be computed by an arithmetic circuit C of size $\mathrm{poly}(d)$. Then, there is a homogeneous depth-6 circuit C' which computes P and satisfies*

- *Size(C) $\leq \exp(O(d^{1/3}\log d))$, and*
- *The fan-in of all the product gates in C' is bounded by $O(d^{1/3})$.*

Now, we start with the circuit C' as guaranteed by the lemma above, and for each of the product gates at the second level, look at its inputs. Each such input is a $\Sigma\Pi^{O(d^{1/3})}\Sigma\Pi^{O(d^{1/3})}$ circuit (depth-4 circuit with all product fan-ins being at most $O(d^{1/3})$) of size at most $\exp(O(d^{1/3}\log d))$. We now apply the depth reduction to depth-3 by Gupta et al. [8] to each one of these depth-4 circuits. As a result, each of these depth-4 circuits get reduced to a depth-3 circuit, with at most a factor of $\exp(O(d^{1/3}))$ blow up in size. Plugging these depth-3 circuits back into C', we obtain a depth-5 circuit C'' such that

- Size(C) $\leq \exp(O(d^{1/3}\log d))$, and
- The fan-in of all the product gates at level two of C'' is bounded by $O(d^{1/3})$.

Recall that the depth reduction in [8] only works over fields of characteristic zero. This yields the following depth reduction to non-homogeneous depth-5 circuits.

▶ **Lemma 5** (Depth reduction to non-homogeneous depth-5 circuits). *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a polynomial of degree d in $\mathrm{poly}(d)$ variables over $\mathbb{F}$ which can be computed by an arithmetic circuit C of size $\mathrm{poly}(d)$. Then, there is a depth-5 circuit C'' which computes P and satisfies*

- *Size(C) $\leq \exp(O(d^{1/3}\log d))$, and*
- *The fan-in of all the product gates at level two of C' is bounded by $O(d^{1/3})$.*

Now, observe that an analogue of Theorem 2 over fields of characteristic zero, would imply an $\exp(\Omega(d^{1/2}))$ lower bound for the depth-5 circuits obtained in Lemma 5, and hence imply $\mathsf{VP} \neq \mathsf{VNP}$.

4 Notation

- Throughout the paper, we shall use bold-face letters such as $\mathbf{x}$ to denote a set $\{x_1, \ldots, x_n\}$. Most of the times, the size of this set would be clear from context. We shall also abuse this notation to use $\mathbf{x}^{\mathbf{e}}$ to refer to the monomial $x_1^{e_1} \cdots x_n^{e_n}$.
- For an integer $m > 0$, we shall use $[m]$ to denote the set $\{1, \ldots, m\}$.
- We shall use the short-hand $\partial_{\mathbf{x}^{\mathbf{e}}}(P)$ to denote
$$\frac{\partial^{e_1}}{\partial x_1^{e_1}}\left(\frac{\partial^{e_2}}{\partial x_2^{e_2}}\left(\cdots(P)\cdots\right)\right).$$
- For a set of polynomials $\mathcal{P}$ shall use $\partial^{=k}\mathcal{P}$ to denote the set of all k-th order partial derivatives of polynomials in $\mathcal{P}$, and $\partial^{\leq k}\mathcal{P}$ similarly.
  Also, $\mathbf{x}^{=\ell}\mathcal{P}$ shall refer to the set of polynomials of the form $\mathbf{x}^{\mathbf{e}} \cdot P$ where $\mathsf{Deg}(\mathbf{x}^{\mathbf{e}}) = \ell$ and $P \in \mathcal{P}$. Similarly $\mathbf{x}^{\leq \ell}\mathcal{P}$.
- For a polynomial $P \in \mathbb{F}_q[\mathbf{x}]$ and for a set $S \subseteq \mathbb{F}_q^n$, we shall denote by $\mathsf{Eval}_S(P)$ the vector of the evaluation of P on points in S (in some natural predefined order like say the lexicographic order). For a set of polynomials $\mathcal{P}$, $\mathsf{Eval}_S(\mathcal{P})$ denotes the set $\{\mathsf{Eval}_S(P) : P \in \mathcal{P}\}$. For a set of vectors V, their span over $\mathbb{F}_q$ will be denoted by $\mathsf{Span}(V)$ and their dimension by $\mathsf{Dim}(V)$.
- We shall use $\mathcal{H}$ to denote the set $\{0,1\}^n \subset \mathbb{F}_q^n$.
- A polynomial of the form $P = \alpha_0 + \sum_{i=1}^n \alpha_i x_i$, where each $\alpha_j \in \mathbb{F}$ is referred to as an affine form. If $\alpha_0 = 0$, then P is referred to as a linear form or a linear polynomial over $\mathbb{F}$.

The complexity measure

We now define the complexity measure that we shall be using to prove the lower bound. The measure will depend on a carefully chosen set $S \subset \mathbb{F}_q^n$.

▶ **Definition 6** (The complexity measure). Let k, ℓ be some parameters and let $S \subset \mathbb{F}_q^n$. For any polynomial P, define $\Gamma_{k,\ell,S}(P)$ as
$$\Gamma_{k,\ell,S}(P) \quad := \quad \mathsf{Dim}\left\{\mathsf{Eval}_S\left(\mathbf{x}^{=\ell}\partial^{=k}(P)\right)\right\}.$$

5 Complexity measure on a depth-5 circuit

A depth-5 circuit computes a polynomial of the form
$$C \quad = \quad \sum_a \prod_b \sum_c \prod_d L_{abcd} \tag{5.1}$$
where each L_{abcd} are linear polynomials.

▶ **Definition 7** (Terms of a circuit, and rank). For a depth-5 circuit such as (5.1), we shall denote by $\mathrm{Terms}(C)$ the set
$$\mathrm{Terms}(C) \quad := \quad \left\{\prod_d L_{abcd}\right\}_{a,b,c}$$
which are all products of linear polynomials computed by the bottommost product gates.

For any term $T = \prod_d L_d$, define $\mathsf{Rank}(T)$ to be $\mathsf{Dim}\,\{L_d\}_d$, which is the maximum number of independent linear polynomials among the factors of T. For a depth-5 circuit C, we shall use $\mathsf{Rank}(C)$ to denote $\max_{T \in \mathrm{Terms}(C)} \mathsf{Rank}(T)$.

For a parameter τ, we shall use $\mathrm{Terms}_{>\tau}(C)$ to refer to terms $T \in \mathrm{Terms}(C)$ with $\mathsf{Rank}(T) > \tau$.

Low rank gates are low-degree polynomials

The following Lemma, present implicitly in [6, 7], is a very useful transformation of gates of low-rank (and possibly large degree) when working over a finite field.

▶ **Lemma 8** ([6, 7]). *Let Q be a product of linear polynomials of rank at most τ. Then, there is a polynomial $\tilde{Q}$ of degree at most $(q-1)\cdot\tau$ such that $\tilde{Q}(\mathbf{a}) = Q(\mathbf{a})$ for all $\mathbf{a} \in \mathbb{F}_q^n$.*

Proof. Without loss of generality, we shall assume that the rank is equal to τ, as the degree upper bound will only be better for a smaller rank and let $L_1, \ldots, L_\tau$ be linearly independent. Let

$$Q = \prod_{i=[\tau]} L_i \cdot \prod_{j \notin [\tau]} L_j.$$

Here, each linear form in the second product term is in the linear span of the linear forms $\{L_i : i \in [\tau]\}$, and so can be expressed as their linear combination. Therefore, Q can be expressed as a polynomial in $\{L_i : i \in [\tau]\}$. Let $Q = f(L_1, L_2, \ldots, L_\tau)$. Since we are working over $\mathbb{F}_q$, it follows that for every choice of L_i and for every $\mathbf{a} \in \mathbb{F}_q^n$, we have $L_i^q(\mathbf{a}) = L_i(\mathbf{a})$. So, for every $\mathbf{a} \in \mathbb{F}_q^n$,

$$f(L_1, L_2, \ldots, L_\tau)(\mathbf{a}) = [f(L_1, L_2, \ldots, L_\tau) \mod \langle(\{L_i^q - L_i \ : \ i = 1, \ldots, \tau\}\rangle)](\mathbf{a}).$$

The lemma immediately follows by setting

$$\tilde{Q} := f(L_1, L_2, \ldots, L_\tau) \mod \langle(\{L_i^q - L_i \ : \ i = 1, \ldots, \tau\}\rangle).$$ ◀

High rank gates are almost always zero

Let us assume that $\mathrm{size}(C) \le 2^{\sqrt{d}/100}$. We shall fix a threshold τ and call all terms T with $\mathsf{Rank}(T) > \tau$ as "high rank terms" and the rest as "low rank terms". Under a random evaluation in $\mathbb{F}_q^n$, every non-zero linear polynomial takes value zero with probability $1/q$. Thus, if we have a term that is a product of *many* independent linear polynomials, then with very high probability *many* of them will be set to zero, i.e. the term will vanish with high multiplicity at most points. This is formalized by the following definition and the lemma after it.

▶ **Definition 9** (Multiplicity at a point). For any polynomial P and a point $\mathbf{a} \in \mathbb{F}_q^n$, we shall say that $\mathbf{a}$ vanishes *with multiplicity* t on P if $Q(\mathbf{a}) = 0$ for all $Q \in \partial^{\le t-1}(P)$. In other words, $\mathbf{a}$ is a root of P and all its derivatives up to order $t-1$.

We shall denote by $\mathsf{Mult}(P, \mathbf{a})$ the maximum t such that $\mathbf{a}$ vanishes on $\partial^{\le t-1}(P)$.

It is easy to see that if P is a product of linear polynomials, then $\mathbf{a}$ vanishes with multiplicity t on P if $\mathbf{a}$ vanishes on at least t factors of P.

▶ **Observation 10.** *Let $T = \prod_{i=1}^d L_i$ be a term of rank at least r. Then, for every $\delta > 0$,*

$$\Pr_{\mathbf{a} \in \mathbb{F}_q^n}\left[\mathsf{Mult}(T, \mathbf{a}) \le (1-\delta)\frac{r}{q}\right] \le \exp\left(-\frac{\delta^2 r}{2q}\right).$$

Proof. Without loss of generality, let $L_1, \ldots, L_r$ be linearly independent. Then, the evaluations of $L_1, \ldots, L_r$ at a point $\mathbf{a} \in \mathbb{F}_q^n$ are also linearly independent and $\Pr_{\mathbf{a}}[L_i(\mathbf{a}) = 0] = (1/q)$ for $i = 1, \ldots, r$.

For $i = 1, \ldots, r$, let Y_i be the indicator random variable that is one if $L_i(\mathbf{a}) = 0$ and zero otherwise. Let $Y = \sum_{i \in [r]} Y_i$. Clearly, by linearity of expectations

$$\mathbb{E}[Y] = \sum_{i \in [r]} \mathbb{E}[Y_i] = \frac{r}{q}.$$

Since the events Y_i are linearly independent, by the Chernoff Bound, we know that for every $\delta > 0$

$$\Pr\left[Y \leq (1-\delta)\frac{r}{q}\right] \quad \leq \quad \exp\left(-\frac{\delta^2 r}{2q}\right).$$ ◀

The following corollary is a simple union bound on all high-rank gates in a small circuit.

▶ **Corollary 11.** *Let C be a depth-5 circuit over $\mathbb{F}_q$ such that* $\mathrm{size}(C) \leq 2^{\sqrt{d}/100}$. *Let $\tau = \frac{q\sqrt{d}}{6}$ so that*

$$\exp\left(\frac{\tau}{8 \cdot q}\right) \quad > \quad 2^{\sqrt{d}/50}.$$

Then,

$$\Pr_{\mathbf{a} \in \mathbb{F}_q^n}\left[\exists T \in \mathrm{Terms}_{>\tau}(C) \ : \ \mathsf{Mult}(T, \mathbf{a}) \ \leq \ \frac{\tau}{2q}\right] \quad \leq \quad 2^{-(\sqrt{d}/100)}$$

We shall set our parameter τ as in the above corollary and our parameter $k = \tau/2q^3$.

5.1 Upper bound on complexity measure

For a circuit C of size at most $2^{\sqrt{d}/100}$, let $\mathcal{E}$ refer to the "bad set" of points $\mathbf{a}$ such that there is some $T \in \mathrm{Terms}_{>\tau}(C)$ for which $\mathsf{Mult}(T, \mathbf{a}) \leq k = \tau/2q^3$. By the above corollary, we know that

$$|\mathcal{E}| = \delta \cdot q^n \quad \text{for some } \delta = \exp(-O(\sqrt{d})).$$

Let S be any subset of $\mathbb{F}_q^n \setminus \mathcal{E}$ that is contained in a "translate of a hypercube", that is there exists some $\mathbf{c} \in \mathbb{F}_q^n$ such that

$$S \subset (\mathbf{c} + \mathcal{H}) \setminus \mathcal{E}.$$

The following lemma allows us to "multilinearize" any polynomial as long as we are only interested in evaluations on a translate of a hypercube.

▶ **Lemma 12** (Multilinearization). *Fix a translate of a hypercube $\mathbf{c} + \mathcal{H}$. Then for every polynomial $Q \in \mathbb{F}_q[\mathbf{x}]$, there is a unique multilinear polynomial Q' such that* $\mathsf{Deg}(Q') \leq \mathsf{Deg}(Q)$ *and $Q'(\mathbf{a}) = Q(\mathbf{a})$ for every $\mathbf{a} \in \mathbf{c} + \mathcal{H}$.*

Proof. If $\mathbf{a} \in \mathbf{c} + \mathcal{H}$, then for each $i \in [n]$ we have a_i to be either c_i or $c_i + 1$. Thus, it suffices to replace each x_i^2 by a linear polynomial in x_i that maps c_i to c_i^2 and $c_i + 1$ to $(c_i + 1)^2$. This is achieved by $x_i^2 \mapsto c_i^2 + (x_i - c_i)(2c_i + 1)$. By repeated applications of this reduction, we obtain a multilinear polynomial Q' of degree at most $\mathsf{Deg}(Q)$ that agrees on all points on $\mathbf{c} + \mathcal{H}$.

Another way to state this is by looking at $Q \bmod \mathcal{I}_{\mathbf{c}}$ where $\mathcal{I}_{\mathbf{c}}$ is the ideal defined by

$$\mathcal{I}_{\mathbf{c}} \quad := \quad \big\langle \big(\{x_i^2 - (c_i^2 + (x_i - c_i)(2c_i + 1)) \ : \ i = 1, \ldots, n\}\big\rangle\big).$$

It is easy to check that $\mathcal{I}_{\mathbf{c}}$ vanishes on $\mathbf{c} + \mathcal{H}$, and any Q can be reduced to a multilinear polynomial modulo $\mathcal{I}_{\mathbf{c}}$.

The uniqueness of Q' follows from the fact that no non-zero multilinear polynomial can vanish on all of $\mathbf{c} + \mathcal{H}$. ◀

We remark that *multilinearization* as defined above is similar to the notion of taking *multilinear projections* in [13]. However, the notion defined above is more amenable to work with when we are looking at evaluations of polynomials. It is not clear to us if the same can be done with the original notion of taking multilinear projections, as in [13].

The main lemma of this section would be the following bound on the complexity measure on a depth-5 circuit.

▶ **Lemma 13** (Upper bound on circuit). *Let C be a depth-5 circuit, of formal degree at most $2d$ and $\mathrm{size}(C) \le 2^{\sqrt{d}/100}$, that computes an n-variate degree d polynomial. Let τ and k be chosen as above, and ℓ be a parameter satisfying $\ell + k\tau q < n/2$. If S is any subset of $\mathbb{F}_q^n \setminus \mathcal{E}$ that is contained in a translate of a hypercube, then*

$$\Gamma_{k,\ell,S}(C) \quad \le \quad 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n).$$

Proof. Suppose $C = R_1 + \cdots + R_s$, where $s \le 2^{\sqrt{d}/100}$ and each R_i is a product of depth-3 circuits with $\mathsf{Deg}(R_i) \le 2d$. Since $\Gamma_{k,\ell,S}$ is clearly sub-additive (i.e. $\Gamma_{k,\ell,S}(f+g) \le \Gamma_{k,\ell,S}(f) + \Gamma_{k,\ell,S}(g)$ for any f, g), it suffices to show that for each R_i we have

$$\Gamma_{k,\ell,S}(R_i) \quad \le \quad \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n).$$

For each such R_i, define the $R_i^{\le\tau}$ as the polynomial obtained by "deleting" all terms $T \in \mathrm{Terms}_{>\tau}(R_i)$. That is,

$$\text{if} \quad R_i \quad = \quad \prod_a \sum_b T_{ab} \quad \text{then} \quad R_i^{\le\tau} \quad = \quad \prod_a \sum_{b : Rank(T_{ab}) \le \tau} T_{ab}.$$

The lemma follows from the following two claims which we now prove.

▶ **Claim 14.** *For every $i \in [r]$*

$$\Gamma_{k,\ell,S}(R_i) \quad = \quad \Gamma_{k,\ell,S}(R_i^{\le\tau}).$$

▶ **Claim 15.** *For every $i \in [r]$*

$$\Gamma_{k,\ell,S}(R_i^{\le\tau}) \quad \le \quad \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n).$$ ◀

Proof of Claim 14. For brevity, we shall drop some indices and work with $R = Q_1 \cdots Q_m$. Let $T \in \mathrm{Terms}_{>\tau}(C)$. We shall show if $R' = (Q_1 - T)Q_2 \cdots Q_m$, then for any k-th order partial derivative $\partial_{\mathbf{x}^\alpha}$,

$$\mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R)) \quad = \mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R')).$$

Indeed, consider the difference $R - R' = T \cdot Q_2 \cdots Q_m$. By the chain rule, every term in $\partial_{\mathbf{x}^\alpha}(R - R')$ is divisible by some k'-th order partial derivative of T with $k' \le k$. By the choice of S, we know that every $\mathbf{a} \in S$ satisfies $\mathsf{Mult}(T, \mathbf{a}) > k$, and hence $\mathbf{a}$ vanishes on $\partial^{\le k}(T)$ for any $T \in \mathrm{Terms}_{>\tau}(C)$. Thus, it follows that $\mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R - R'))$ is just the zero vector.

Repeating this argument, we can prune away all terms in $\mathrm{Terms}_{>\tau}(C)$ to get that $\mathsf{Eval}_S(\partial_\alpha(R)) = \mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R^{\le\tau}))$ where $\mathsf{Deg}(\mathbf{x}^\alpha) = k$. Thus, $\Gamma_{k,\ell,S}(R) = \Gamma_{k,\ell,S}(R^{\le\tau})$. ◀

Proof of Claim 15. Let $R^{\le\tau} = Q_1 \cdots Q_d$, with each Q_i being a $\Sigma\Pi\Sigma$ circuit. Some of these Q_is could have degree more than τ although their rank is bounded by τ. Without loss of generality, let $Q_1, \cdots, Q_m$ be all the Q_is with $\mathsf{Deg}(Q_i) > \tau$, and $Q_{m+1}, \ldots, Q_d$ have $\mathsf{Deg}(Q_i) \le \tau$.

We shall modify the "low-degree" Q_is by multiplying together any two of them of degree less than $\tau/2$. This ensures that at most one of the Q_is may have degree less than $\tau/2$ and for $i > m$, all the Q_is have degree at most τ. The sizes of some of the low-degree Q_is do increase in the process but this would not be critical as the degree of any such term is still bounded by τ. The main point is that now we have an expression of the form

$$R^{\le\tau} = Q_1 \cdots Q_m \cdot Q'_1 \cdots Q'_r$$

where each Q_i is a $\Sigma\Pi\Sigma$ circuit of rank at most $\tau - 1$ and $\mathsf{Deg}(Q_i) \ge \tau$, and all but one of the Q'_i satisfies $\tau \ge \mathsf{Deg}(Q'_i) \ge \tau/2$. As $\mathsf{Deg}(R^{\le\tau}) \le 2d$, it follows that $m + r \le \frac{4d}{\tau} + 1$.

As a notational convenience, for any set H we let $Q_H := \prod_{i\in H} Q_i$ and we use $R^{\le\tau}$ and R interchangeably in the calculations that follow. Let us look at any partial derivative $\partial_{\mathbf{x}^\alpha}$ of order k applied on R. By the chain-rule, any such partial derivative can be written as a natural linear combination of terms. For any two monomials $\mathbf{x}^\alpha = \prod_{i=1}^n x_i^{\alpha_1}$ and $\mathbf{x}^\beta = \prod_{i=1}^n x_i^{\beta_i}$, we say that $\mathbf{x}^\alpha \succeq \mathbf{x}^\beta$, if there exists a monomial $\mathbf{x}^\gamma = \prod_{i=1}^n x_i^{\gamma_i}$ such that $\mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma$.

$$\begin{aligned}
\partial_{\mathbf{x}^\alpha}(R) &\in \mathsf{Span}\left\{ \partial_{\mathbf{x}^\beta}(Q_A) \cdot \partial_{\mathbf{x}^\gamma}(Q'_B) \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \ \begin{array}{c} \mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\} \\
&\in \mathsf{Span}\left\{ \partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\le k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \ \begin{array}{c} \mathbf{x}^\alpha \succeq \mathbf{x}^\beta \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\}. \\
\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R) &\subseteq \mathsf{Span}\left\{ \partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\le \ell + k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \ \begin{array}{c} \mathbf{x}^\alpha \succeq \mathbf{x}^\beta \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\}. \\
\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R)) &\subseteq \mathsf{Span}\left\{ \mathsf{Eval}_S\left(\partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\le \ell + k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}}\right) \ : \ \begin{array}{c} \mathbf{x}^\alpha \succeq \mathbf{x}^\beta \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\}.
\end{aligned}$$

If we focus on the term $\partial_{\mathbf{x}^\beta}(Q_A)$, since Q_A is a product of $\Sigma\Pi\Sigma$ circuits of rank at most τ, we have that $\partial_{\mathbf{x}^\beta}(Q_A)$ is a linear combination of terms $T_1 \cdots T_{|A|}$ where each T_i is a product of linear polynomials and has rank at most τ. Using Lemma 8 on each of these T_is,

$$\mathsf{Eval}_S(\partial_{\mathbf{x}^\beta}(Q_A)) \in \mathsf{Span}\left\{ \mathsf{Eval}_S(\mathbf{x}^{\le (q-1)\tau|A|}) \right\}.$$

Therefore,

$$\begin{aligned}
\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R)) &\subseteq \mathsf{Span}\left\{ \mathsf{Eval}_S\left(\partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\le \ell + k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}}\right) \ : \ \begin{array}{c} \mathbf{x}^\alpha \succeq \mathbf{x}^\beta \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\} \\
&\subseteq \mathsf{Span}\left\{ \mathsf{Eval}_S\left(\mathbf{x}^{\le \ell + k\tau + (q-1)k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}}\right) \ : \ \begin{array}{c} A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array} \right\}.
\end{aligned}$$

Finally, Lemma 12 shows for every polynomial f, there is a multilinear polynomial of degree at most $\mathsf{Deg}(f)$ that agrees with f on all evaluations on a translate of a hypercube. Therefore,

in the above span, we may assume that we are only shifting by multilinear monomials of degree $\ell + qk\tau$ as this doesn't change the evaluations $S \subseteq \mathbf{c} + \{0,1\}^n$. Hence,

$$\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \mathsf{Span}\left\{\mathsf{Eval}_S\left(\mathbf{x}_{\text{mult}}^{\leq \ell+qk\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}}\right) \;:\; \begin{array}{c} A \subseteq [m]\,, \\ B \subseteq [r]\,,\ |A|+|B| = k \end{array}\right\}.$$

Therefore, using the fact that $m + r \leq (4d/\tau) + 1$, we get the bound

$$\Gamma_{k,\ell,S}(R) := \mathsf{Dim}\left\{\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial^{=k}(R))\right\} \quad \leq \quad \binom{\frac{4d}{\tau}+1}{k} \cdot \binom{n}{\ell + qk\tau} \cdot n,$$

where the first term corresponds to the number of choices for the subsets A and B, and the last two terms correspond to the number of multilinear monomials of degree at most $\ell + qk\tau$. Recall that by the hypothesis of the lemma, $\ell + qk\tau < n/2$, hence the number of multilinear monomials of degree at most $\ell + qk\tau$ is at most $n \cdot \binom{n}{\ell+qk\tau}$. ◀

▶ Remark 16. Observe that, even if the circuit C is of the form

$$C \quad = \quad \sum_a \prod_{b\in[m]} \sum_c \prod_d L_{abcd}$$

such that $\text{Size}(C) \leq 2^{\sqrt{d}/100}$ and $m = O(\frac{d}{\tau})$, then the upper bound in Lemma 13 continues to hold.[7] In particular, the formal degree of C could be much larger than d but if the product fan-in at level two of C is small, then

$$\Gamma_{k,\ell,S}(C) \quad \leq \quad 2^{\sqrt{d}/100} \cdot \binom{O(\frac{d}{\tau})}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \text{poly}(n).$$

We later use this observation to complete the proof of Theorem 2 in Section 7.

6 Lower bound for the complexity measure for an explicit polynomial

Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. We will be choosing a specific set S that shall be a subset of a translate of the hypercube and disjoint from $\mathcal{E}$. We will fix the precise definition of S shortly once we motivate the requirements.

The polynomial for which was shall prove our lower bound would be from the Nisan-Wigderson family. We would have to set our parameters carefully but for now we shall just be intentionally vague and refer to the polynomial as just NW and fix parameters at a later point.

Associated with our measure $\Gamma_{k,\ell,S}(\mathsf{NW})$ is a natural matrix that we shall call $\Lambda(\mathsf{NW})$:

> The rows of $\Lambda(\mathsf{NW})$ are indexed by a derivative $\partial_{\mathbf{x}^\alpha} \in \partial^{=k}$ of order k, and a monomial $\mathbf{x}^\beta$ of degree equal to ℓ. The columns are indexed by all points $\mathbf{a} \in S$. The entry in $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}, \mathbf{a})$ is the evaluation of $\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW})$ at the point $\mathbf{a}$.

In other words, the matrix is just the vectors $\mathsf{Eval}_S(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW}))$ listed as different rows for each choice of $\mathbf{x}^\alpha$ and $\mathbf{x}^\beta$. Therefore,

$$\mathsf{Rank}\left(\Lambda(\mathsf{NW})\right) = \Gamma_{k,\ell,S}(\mathsf{NW}). \tag{6.1}$$

[7] Essentially, in the proof of Claim 15, we already have an expression of the form $R^{\leq\tau} = Q_1 \cdots Q_m$ with $m = O\left(\frac{d}{\tau}\right)$ and the rest of the proof proceeds as expected.

Recall from Lemma 12 (multilinearization), as long as we only care about evaluations on a translate of a hypercube, we can assume each row is the evaluations of the multilinearization of $\mathbf{x}^\alpha \cdot \partial_{\mathbf{x}^\beta}(\mathsf{NW})$. This does not change the evaluation on any point $\mathbf{a} \in S \subseteq \mathbf{c} + \mathcal{H}$.

Now any such matrix of evaluations can be naturally factorized as a coefficient matrix and an evaluation matrix.

$C_{k,\ell}(\mathsf{NW})$: Each row is indexed by a derivative $\partial_{\mathbf{x}^\alpha}$ of order k, and a monomial $\mathbf{x}^\beta$ of degree ℓ, and each column is indexed by a multilinear monomial m of degree *at most* $\ell + d - k$ over n variables, and the $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}, m)$ entry is the coefficient of monomial m in the multilinearization of $\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW})$ with respect to $\mathbf{c} + \mathcal{H}$ (Lemma 12).

$V_t(S)$: Rows are indexed by multilinear monomials of degree at most $t = \ell + d - k$ over n variables, columns are indexed by $\mathbf{a} \in S$ and $(m, \mathbf{a})$ entry is the evaluation monomial m at $\mathbf{a}$.

Clearly, $\Lambda(\mathsf{NW}) = C_{k,\ell}(\mathsf{NW}) \cdot V_t(S)$. Thus if we can get a good lower bound on the ranks of the matrices $C_{k,\ell}(\mathsf{NW})$ and $V_t(S)$ for a suitable set S, we would then be able to lower bound the rank of $\Lambda(\mathsf{NW})$. This is formalized by the following simple linear algebraic fact often referred to as the Sylvester's rank inequality.

▶ **Lemma 17** (Rank of products of matrices [27]). *If A, B and C are matrices such that $A = B \cdot C$, then* $\mathsf{Rank}(A) \geq \mathsf{Rank}(B) + \mathsf{Rank}(C) - (\#$ *rows of* $C)$.

6.1 Rank of $C_{k,\ell}(\mathsf{NW})$

Let us focus on the matrix $C_{k,\ell}(\mathsf{NW})$ and restrict ourselves a sub-matrix $C'_{k,\ell}(\mathsf{NW})$ to only those columns whose degree is *exactly* $t = \ell + d - k$, and rows indexed by $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha})$ with $\mathbf{x}^\beta$ being a multilinear monomial of degree exactly ℓ.

Since our polynomial NW is multilinear, if we were to read off any row of $C'_{k,\ell}(\mathsf{NW})$, this is just the list of coefficients of all multilinear monomials of $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW}))$. This is because the multilinearization operation in Lemma 12 maps any non-multilinear monomial to a multilinear polynomial of strictly smaller degree and hence these monomials are not included in the columns of $C'_{k,\ell}$.

The key point here is that the matrix $C'_{k,\ell}(\mathsf{NW})$ is just the matrix of *projected shifted partial derivatives* of NW. The results of Kayal et. al [13] and Kumar and Saraf [19] give a lower bound on the rank of this matrix for a suitable choice of the polynomial, but the lower bound of Kumar and Saraf [19] is more relevant as it is true over any field (unlike [13] that works only over characteristic zero fields).

Using a tight analysis of the argument in [19], that we present in Section A we obtain the following lemma for the Nisan-Wigderson polynomial for very carefully chosen parameters.

▶ **Lemma 18** (Tight analysis of [19]). *For every d and $k = O(\sqrt{d})$ there exists parameters m, e, ϵ such that $m = \Theta(d^2)$, $n = md$ and $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ with*

$$\begin{aligned} m^k &\geq (1+\epsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\epsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

For any $\{d, m, e, k, \epsilon\}$ satisfying the above constraints and for $\ell = \frac{n}{2}(1-\epsilon)$, over any field $\mathbb{F}$, we have

$$\mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \mathsf{Rank}(C'_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)).$$

6.2 Rank of $V_t(S)$

Let $\mathcal{H}_{\leq t}$ refer to elements of $\{0,1\}^n$ of Hamming weight at most t. Our first step would be to choose our set S carefully so that we can maximize the rank of $V_t(S)$.

▶ **Observation 19.** *Let $\mathcal{E}$ be a subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. Then for any $0 \leq t \leq n$, there is a vector $\mathbf{c} \in \mathbb{F}_q^n$ such that*

$$|(\mathbf{c} + \mathcal{H}_{\leq t}) \cap \mathcal{E}| \quad \leq \quad \delta \cdot |\mathcal{H}_{\leq t}| \,.$$

Proof. Let $\mathbb{1}_{\mathcal{E}}(\mathbf{a})$ be the indicator function that is 1 if $\mathbf{a} \in \mathcal{E}$, and 0 otherwise. Then,

$$\delta \quad \geq \quad \mathbb{E}_{\mathbf{a} \in \mathbb{F}_q^n} \left[\mathbb{1}_{\mathcal{E}}(\mathbf{a})\right] \quad = \quad \mathbb{E}_{\mathbf{c} \in \mathbb{F}_q^n} \left[\mathbb{E}_{\mathbf{y} \in \mathcal{H}_{\leq t}} \left[\mathbb{1}_{\mathcal{E}}(\mathbf{c} + \mathbf{y})\right]\right] .$$

Thus, there exists some $\mathbf{c} \in \mathbb{F}_q^n$ that still maintains the inequality. ◀

Our set would be $S = (\mathbf{c} + \mathcal{H}_{\leq t}) \setminus \mathcal{E}$, which has the property that $|S \cap (\mathbf{c} + \mathcal{H}_{\leq t})| \geq (1-\delta) \cdot |\mathcal{H}_{\leq t}|$ by the above observation, and $S \cap \mathcal{E} = \emptyset$.

Let $V_t(S - \mathbf{c})$ be the matrix whose rows are indexed by the polynomials $m(\mathbf{x} - \mathbf{c})$, where m is a multilinear monomials in variables $\mathbf{x}$ of degree at most t. The columns of $V_t(S - \mathbf{c})$ are indexed by S and the entries correspond to the evaluation of the polynomial indexing the row at the point in S given by the column. We have the following observation.

▶ **Observation 20.** $\mathsf{Rank}(V_t(S)) = \mathsf{Rank}(V_t(S - \mathbf{c}))$.

Proof. For any multilinear monomial m of degree at most t, the polynomial $m(\mathbf{x} - \mathbf{c})$ is multilinear and has degree at most t. Thus clearly, the row-space of $V_t(S - \mathbf{c})$ is contained in the row-space of $V_t(S)$. The converse also holds trivially as the translation is invertible. ◀

We now prove our next lemma which shows a lower bound on the rank of $V_t(S - \mathbf{c})$.

▶ **Lemma 21.** *For any set $S \subseteq \{0,1\}^n \subset \mathbb{F}_q^n$ and any $0 \leq t \leq n$,*

$$\mathsf{Rank}(V_t(S - \mathbf{c})) \quad = \quad |S| \,.$$

Proof. Since $S \subseteq \mathbf{c} + \mathcal{H}_{\leq t}$, the set $S' := S - \mathbf{c} \subset \mathcal{H}_{\leq t}$. Thus the matrix $V_t(S - \mathbf{c})$ is simply the matrix $V_t(S')$. For any $\mathbf{a} \in \{0,1\}^n$, let $m_{\mathbf{a}}$ refer to the 'characteristic' monomial $\prod_{i:a_i=1} x_i$, and let $m_{\mathbf{0}} = 1$.

Consider the sub-matrix of $V_t(S')$ by restricting to rows indexed by $\{m_{\mathbf{a}} \,:\, \mathbf{a} \in S'\}$. By rearranging the rows and columns in the increasing order of the weight of $\mathbf{a}$, it is evident that the sub-matrix is upper-triangular with ones on the diagonal. It therefore follows that the rank of $V_t(S')$ (which is just $V_t(S - \mathbf{c})$) is at least $|S'| = |S|$. ◀

Combining Observation 20 and Lemma 21, we have our required bound on the rank of $V_t(S)$.

▶ **Lemma 22.** *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\mathsf{Rank}(V_t(S)) \quad \geq \quad (1-\delta) \cdot |\mathcal{H}_{\leq t}| \quad = \quad (1-\delta) \cdot (\#\ \textit{rows of}\ V_t(S)).$$

Putting them together

▶ **Lemma 23** (Rank bound for $\Lambda(\mathsf{NW}_{d,m,e})$). *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$, with $\delta = \exp(-\omega(\log^2 d))$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\mathsf{Rank}(\Lambda(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)),$$

where the parameters d, m, e, k, ℓ are chosen as in Lemma 18.

Proof. Consider the set S chosen in Lemma 22 (for $t = \ell + d - k$). By Lemma 22,

$$\mathsf{Rank}(V_t(S)) - (\#\text{ rows of } V_t(S)) \quad \geq \quad (-\delta)\,|\mathcal{H}_{\leq t}| \quad \geq \quad (-\delta) \cdot n \cdot \binom{n}{\ell + d - k}.$$

Lemma 18 shows that rank of $C_{k,\ell}(\mathsf{NW}_{d,m,e})$ can be lower bounded by

$$\mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)).$$

Thus, since $\Lambda(\mathsf{NW}_{d,m,e}) = C_{k,\ell}(\mathsf{NW}_{d,m,e}) \cdot V_t(S)$ with $t = \ell + d - k$, Lemma 17 implies that

$$\begin{aligned}
\mathsf{Rank}(\Lambda(\mathsf{NW}_{d,m,e})) \quad &\geq \quad \mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) + \mathsf{Rank}(V_t(S)) - (\#\text{ rows of } V_t(S)) \\
&\geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)) - \delta \cdot n \cdot \binom{n}{\ell + d - k} \\
&\geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)) \qquad \text{as } \delta = \exp(-\omega(\log^2 d)). \qquad ◀
\end{aligned}$$

Combining this with Equation 6.1, we get the following lemma.

▶ **Lemma 24** (Measure of $\mathsf{NW}_{d,m,e}$). *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$, with $\delta = \exp(-\omega(\log^2 d))$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\Gamma_{k,\ell,S}(\mathsf{NW}_{d,m,e}) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)).$$

7 Wrapping up the proof

▶ **Theorem 25.** *Let $\mathbb{F}_q$ be the finite field of cardinality q. Let C be a depth-5 circuit of formal degree at most $2d$ which computes the polynomial $\mathsf{NW}_{d,m,e}$ with parameters as in Lemma 18. Then*

$$\mathit{Size}(C) > 2^{\sqrt{d}/100}.$$

Further, the same lower bound holds even if C was a circuit of the form

$$C \quad = \quad \sum_i \prod_{j \in [m]} \sum_k \prod_\ell L_{ijk\ell},$$

with $m = O(\sqrt{d})$.

Proof. We shall prove the above theorem for homogeneous depth-5 circuits. The lower bound for such non-homogeneous circuits would also follow directly since such circuits also have the same upper-bound on the complexity measure (Remark 16).

Assume on the contrary that there is a circuit C computing $\mathsf{NW}_{d,m,e}$ with $\mathrm{Size}(C) \leq 2^{\sqrt{d}/100}$. Let τ be as defined in Corollary 11 and let $k = \tau/2q^3$. Let $\mathcal{E} = \mathcal{E}(\mathcal{C})$ be the set as defined in Section 5.1. We know that

$$|\mathcal{E}| \leq \delta \cdot q^n,$$

for some $\delta = \exp(-O(\sqrt{d}))$. Let $\ell = \frac{n}{2}(1-\epsilon)$ where $\epsilon = \frac{\log d}{c\sqrt{d}}$ is chosen as in Lemma 18. Since $n = d^3$, clearly we satisfy $\ell + k\tau q < n/2$. Let $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ be the set guaranteed by Lemma 24. From Lemma 24, we know that

$$\Gamma_{k,\ell,S}(\mathsf{NW}) \geq \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)).$$

This may be simplified using Lemma 32 to

$$\Gamma_{k,\ell,S}(\mathsf{NW}) \geq \binom{n}{\ell} \cdot (1+\epsilon)^{2d-2k} \cdot \exp(-O(d\epsilon^2)) \cdot \exp(-O(\log^2 d)).$$

Also, from Lemma 13, we know that

$$\Gamma_{k,\ell,S}(C) \leq 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau}+1}{k} \cdot \binom{n}{\ell + qk\tau} \cdot \mathrm{poly}(n).$$

Notice that from our choice of k and τ, $qk\tau = O(d) = O(\sqrt{n})$. Again, using Lemma 32, we get

$$\Gamma_{k,\ell,S}(C) \leq 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau}+1}{k} \cdot \binom{n}{\ell} \cdot (1+\epsilon)^{2qk\tau} \cdot \exp(O(-qk\tau \cdot \epsilon^2)) \cdot \mathrm{poly}(n).$$

Since C computes $\mathsf{NW}_{d,m,e}$, so it must be the case that

$$2^{\sqrt{d}/100} \cdot \mathrm{poly}(n) \quad \geq \quad (1+\epsilon)^{(d-k)+(d-k-2qk\tau)} \cdot \exp(-O_q(\log^2 d)).$$

Since $k = \tau/2q^3$, so $2qk\tau = \tau^2/q^2$. From our choice of τ in Corollary 11, $\tau = \frac{q\sqrt{d}}{6}$. So

$$2qk\tau = \tau^2/q^2 = d/36.$$

Therefore,

$$2^{\sqrt{d}/100} \cdot \mathrm{poly}(n) \geq (1+\epsilon)^{(d-k)} \cdot \exp(-O_q(\log^2 d)).$$

But this is a contradiction since $(1+\epsilon)^{(d-k)} = \exp(\Omega(\sqrt{d}\log d))$ by our choice of parameters. Therefore, the size of C is at least $2^{\sqrt{d}/100}$. ◀

In fact, the above proof gives more. It shows that if we have a depth-5 circuit computing $\mathsf{NW}_{d,m,e}$ over $\mathbb{F}_q$, then either the number of high-rank terms is at least $2^{\sqrt{d}/50}$ or the top fan-in is $\exp(\Omega(\sqrt{d}\log d))$.

7.1 Getting the right order of quantifiers

In our proof so far, we first fix the field $\mathbb{F}_q$ that we are working over and the parameters of our polynomial are then chosen based on q. Thus, as q varies, the polynomial for which we show the lower bound also seems to vary. The ideal scenario would be to construct a fixed polynomial family so that for every q we get a lower bound of $\exp(\Omega_q(\sqrt{d}))$. We do that now, and this would complete the proof of Theorem 1.

We shall be dealing with a lot of parameters and constraints. The following is essentially the "zone" in which we can prove strong lower bounds (Lemma 18).

▶ **Definition 26** (Goldilocks Zone)**.** We shall say that parameters m, d, e, k, ϵ with $k = \Theta(\sqrt{d})$ and $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ lie in the *Goldilocks Zone* if they satisfy

$$\begin{aligned} m^k &\geq (1+\epsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\epsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

Recall that for Lemma 18, and consequently Theorem 25, the parameters m, d, e, k must lie in the Goldilocks zone. The crucial point is that this is a field dependent condition since k (and hence everything else) explicitly depends on q. In the next lemma, we show that we can start with a fixed polynomial such that for every finite field $\mathbb{F}_q$ of fixed size, there exists a $0,1$ projection which lies in the Goldilocks zone.

▶ **Lemma 27.** *Consider the* $\mathsf{NW}_{d,m,e}$ *polynomial with* $m = \Theta(d^2)$ *and* $e = \left\lceil \frac{d}{\log m} \right\rceil$ *so that*

$$m^e = 2^d \cdot \mathrm{poly}(m).$$

Suppose $k = \Theta(\sqrt{d})$ *and* $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ *satisfy the constraint* $m^k > (1+\epsilon)^{2(d-k)}$*. Then, there exists a* $d' \in [d - O(\sqrt{d}\log d), d]$ *such that* $\mathsf{NW}_{d',m,e}$ *is a* $0/1$ *projection of* $\mathsf{NW}_{d,m,e}$ *and the parameters* $\{d', m, e, k, \epsilon\}$ *fall in the Goldilocks Zone.*

Proof. Since $m^e = 2^d \cdot \mathrm{poly}(m)$, $m^k > (1+\epsilon)^{2(d-k)}$ and $(1+\epsilon)^d = \exp(\Theta(\sqrt{d}\log d))$, we have

$$m^{e-k} = \left(\frac{2}{1+\epsilon}\right)^{d-k} \cdot \exp(-\Theta(\sqrt{d}\log d)).$$

The goal now is to find a $d' < d$ such that

$$m^{e-k} = \left(\frac{2}{1+\epsilon}\right)^{d'-k} \cdot O(1).$$

Indeed, since the LHS and RHS differ by just a factor of $\exp(\Theta(\sqrt{d}\log d))$, and decreasing d by 1 reduces the RHS by a constant factor, there exists some $d' \in [d - O(\sqrt{d}\log d), d]$ such that

$$m^{e-k} = \left(\frac{2}{1+\epsilon}\right)^{d'-k} \cdot O(1).$$

Further, since $m^k > (1+\epsilon)^{2(d-k)}$, it follows that $m^k > (1+\epsilon)^{2(d'-k)}$ as $d' < d$. Hence the parameters $\{d', m, e, k, \epsilon\}$ indeed fall in the Goldilocks Zone (Definition 26).

It suffices to show that $\mathsf{NW}_{d',m,e}$ is a projection of $\mathsf{NW}_{d,m,e}$. This is readily seen as setting the variables $x_{ij} = 1$ for all $i \in [d - d']$ and $j \in [m]$ yields $\mathsf{NW}_{d',m,e}$ up to relabeling variables. ◀

With this, we can finally prove our main theorems.

▶ **Theorem 1** (restated). *Consider the polynomial* $\mathsf{NW}_{d,m,e}$ *with parameters chosen such that* $m = \Theta(d^2)$ *and* $m^e = 2^d \cdot \mathrm{poly}(m)$*. Then, for any fixed finite field* $\mathbb{F}_q$*, any homogeneous depth-5 circuit over* $\mathbb{F}_q$ *computing* $\mathsf{NW}_{d,m,e}$ *must have size at least* $2^{\sqrt{d}/200}$*.*

Proof. Fix a field $\mathbb{F}_q$ and let $k = \sqrt{d}/12q^2$.

Suppose on the contrary that there is indeed a homogeneous depth-5 circuit C computing $\mathsf{NW}_{d,m,e}$. Then, by Lemma 27, this also implies there is a projection C' that computes the $\mathsf{NW}_{d',m,e}$ such that there is an $d - O(\sqrt{d}\log d) \le d' \le d$ and there is an $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ for which $\{d', m, e, k, \epsilon\}$ fall in the Goldilocks Zone (Definition 26). Now C' is a circuit of formal degree $d \le d' + O(\sqrt{d}\log d) \le 2d'$ that computes the polynomial $\mathsf{NW}_{d',m,e}$. By Theorem 25, this implies that

$$\mathrm{size}(C) \ge \mathrm{size}(C') > 2^{\sqrt{d'}/100} > 2^{\sqrt{d}/200}.$$ ◀

The proof of this theorem also follows along the same lines.

▶ **Theorem 2** (restated). *Consider the polynomial* $\mathsf{NW}_{d,m,e}$ *with parameters chosen such that* $m = \Theta(d^2)$ *and* $m^e = 2^d \cdot \mathrm{poly}(m)$*. Then, for any fixed finite field* $\mathbb{F}_q$*, any depth-5 circuit over* $\mathbb{F}_q$ *of the form*

$$C \quad = \quad \sum_i \prod_{j\in[m]} \sum_k \prod_\ell L_{ijk\ell}$$

where each $L_{ijk\ell}$ *is a linear polynomial and* $m = O(\sqrt{d})$ *that computes* $\mathsf{NW}_{d,m,e}$ *must have size at least* $2^{\sqrt{d}/200}$*.*

Acknowledgments. We are very grateful to Mike Saks and Shubhangi Saraf for many discussions and much encouragement. The chat with Mike about powering circuits over finite fields was specially insightful. Part of this work was done while the first author was an intern at MSR New England. We are thankful to Madhu Sudan and the other members of the lab, for stimulating discussions and generous hospitality. Many thanks to Madhu for patiently sitting through a presentation of the proof.

We would also like to thank Eric Allender for answering our questions about connections between boolean circuits and arithmetic circuits over finite fields and pointing us to reference [1].

This work was partly motivated by discussions over the questions of the projected shifted partials complexity of homogeneous depth-5 circuits while the authors were at MSR Bangalore in Summer'14. We are grateful to Neeraj Kayal for hosting us and for many insightful conversations.

References

1 Manindra Agrawal, Eric Allender, and Samir Datta. On TC^0, AC^0, and Arithmetic Circuits. *Journal of Computer and System Sciences*, 60(2):395–421, 2000. doi:10.1006/jcss.1999.1675.

2 Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pages 67–75, 2008. doi:10.1109/FOCS.2008.32.

3 Suman K. Bera and Amit Chakrabarti. A Depth-Five Lower Bound for Iterated Matrix Multiplication. In *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 183–197. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.CCC.2015.183.

4 Suryajith Chillara and Partha Mukhopadhyay. On the limits of depth reduction at depth 3 over small finite fields. In *Mathematical Foundations of Computer Science (MFCS)*, pages 177–188, 2014. arXiv:1401.0189, doi:10.1007/978-3-662-44465-8_16.

5 Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 128–135, 2014. doi:10.1145/2591796.2591824.

6 Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC 1998)*, pages 577–582, 1998. doi:10.1145/276698.276872.

7 Dima Grigoriev and Alexander A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Appl. Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000. Preliminary version in the *39th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1998)*. doi:10.1007/s002009900021.

8 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth Three. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)*, pages 578–587, 2013. doi:10.1109/FOCS.2013.68.

9 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. *Journal of the ACM*, 61(6):33:1–33:16, 2014. Preliminary version in the *28th Annual IEEE Conference on Computational Complexity (CCC 2013)*. doi:10.1145/2629541.

10 Laurent Hyafil. On the parallel evaluation of multivariate polynomials. *SIAM Journal of Computing*, 8(2):120–123, 1979. Preliminary version in the *10th Annual ACM Symposium on Theory of Computing (STOC 1978)*. doi:10.1137/0208010.

11 Kyriakos Kalorkoti. A Lower Bound for the Formula Size of Rational Functions. *SIAM Journal of Computing*, 14(3):678–687, 1985. doi:10.1137/0214050.

12 Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. In *Electronic Colloquium on Computational Complexity (ECCC)TR12-081*, 2012. URL: http://eccc.hpi-web.de/report/2012/081/.

13 Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*, 2014. doi:10.1109/FOCS.2014.15.

14 Neeraj Kayal and Chandan Saha. Lower bounds for depth three arithmetic circuits with small bottom fanin. In *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 158–208. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.CCC.2015.158.

15 Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 146–153, 2014. doi:10.1145/2591796.2591847.

16 Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65, 2012. arXiv:1006.4700, doi:10.1016/j.tcs.2012.03.041.

17 Mrinal Kumar and Ramprasad Saptharishi. Finer Separations Between Shallow Arithmetic Circuits. In *36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2016)*, volume 65 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 38:1–38:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.FSTTCS.2016.38.

18 Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 136–145, 2014. doi:10.1145/2591796.2591827.

19 Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*, 2014. doi:10.1109/FOCS.2014.46.

20 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. Available on citeseer:10.1.1.90.2644. doi:10.1007/BF01294256.

21 Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*, pages 659–666, 2010. doi:10.1145/2535928.

22 Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987. doi:10.1007/BF01137685.

23 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015. URL: https://github.com/dasarpmar/lowerbounds-survey/releases/.

24 Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, pages 77–82, 1987. doi:10.1145/28395.28404.

25 Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Inf. Comput.*, 240:2–11, 2015. Preliminary version in the *38th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 2013)*. doi:10.1016/j.ic.2014.09.004.

26 Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM Journal of Computing*, 12(4):641–644, 1983. Preliminary version in the *6th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 1981)*. doi:10.1137/0212043.

27 Marc van Leeuwen. https://math.stackexchange.com/questions/298836/sylvester-rank-inequality-operatornamerank-a-operatornamerankb-leq-o/298944#298944.

A Tight analysis of the [19] lower bound

We recall the measure of *projected shifted partial derivatives* that was used in [13] and [19].

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad = \quad \mathsf{Dim}\left\{\mathrm{mult}\left(\mathbf{x}^{=\ell}\partial^{=k}(P)\right)\right\}$$

where $\mathrm{mult}(f)$ is just the polynomial f restricted to just its multilinear monomials. As mentioned before, this $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$ is precisely $\mathsf{Rank}(C'_{k,\ell}(P))$ as defined in Section 6.1.

The goal of this section would be to prove Lemma 18 that we restate below.

▶ **Lemma 18** (restated). *For every d and $k = O(\sqrt{d})$ there exists parameters m, e, ϵ such that $m = \Theta(d^2)$, $n = md$ and $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ with*

$$\begin{aligned} m^k &\geq (1+\epsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\epsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

For any $\{d, m, e, k, \epsilon\}$ satisfying the above constraints, the polynomial $\mathsf{NW}_{d,m,e}$, if $\ell = \frac{n}{2}(1-\epsilon)$, then over any field $\mathbb{F}$, we have

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e}) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)).$$

The rest of this section would just be a proof of this lemma.

Before we proceed to the lower bound on $\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e})$, let us first show that we can indeed find parameters that satisfy the above constraints. Fix m to be the smallest power of 2 greater than d^2 to get $m = \Theta(d^2)$. Next, we shall fix the constant c in $\epsilon = \frac{\log d}{c\sqrt{d}}$ so that

$$m^k \quad \geq \quad (1+\epsilon)^{2(d-k)}.$$

This is always possible by choosing c to be large enough as $(1+\epsilon)^{d-k} = \exp(O(\sqrt{d}\log d))$ and that is also the order of m^k.

Once we have done that, we shall fix e so as to ensure that

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\epsilon}\right)^{d-k} \cdot \mathrm{poly}(m),$$

which can be done by setting

$$e = \left\lceil \left(\frac{d-k}{\log m}\right) \cdot \log\left(\frac{2}{1+\epsilon}\right) + k \right\rceil.$$

Since changing e by 1 changes the LHS by a factor of m, the above choice would ensure the LHS and RHS are within a multiplicative factor of m. Note that this definition of e also ensures that $e \leq d - k$.

All lower bounds on the dimension of shifted partial derivatives of a polynomial P was obtained by finding a *large* set of *distinct leading monomials*. In [19], they take this approach but require a very careful analysis. The key difference in this setting is the following:

> If β is the leading monomial of a polynomial P, then for any monomial μ, we also have that $\gamma = \beta \cdot \mu$ is the leading monomial of μP.
> However, the leading monomial of $\mathrm{mult}(\mu P)$ could be $\beta' \cdot \mu$ for some $\beta' \neq \beta$ (as higher monomials could be made non-multilinear during the shift by μ).

The multilinear projection makes the task of counting leading monomials much harder and [19] come up with an indirect way to count them. Throughout this discussion, let $\mathrm{LM}(f)$ refer to the leading monomial of f in some natural ordering, say the lexicographic order.

Leading monomials after multilinear projections

Let P the polynomial of degree d for which we are trying to lower bound $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$. For every monomial multilinear monomial α of degree k, and a monomial $\beta \in \partial_\alpha(P)$, define the set $A(\alpha, \beta)$ as

$$A(\alpha, \beta) \quad = \quad \left\{ \gamma \ : \ \begin{array}{c} \mathsf{Deg}(\gamma) = \ell + d - k \text{ and there is a } mu \text{ of degree } \ell \\ \text{such that } \gamma = \mathrm{LM}(\mathrm{mult}(\mu \cdot \partial_\alpha(P))) = \mu \cdot \beta \end{array} \right\}.$$

In other words, these are the set of leading monomials obtained from a specific monomial β in $\partial_\alpha(P)$. We then have

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \geq \quad \left| \bigcup_{\alpha,\beta} A(\alpha, \beta) \right|.$$

Choice of derivatives: Instead of looking at all derivatives in $\partial^{=k}$, we shall restrict ourselves to just a subset of derivatives. Restricting the above union to a subset $\Delta \subset \mathbf{x}^{=k}$ still continues to remain a lower bound for $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$. Keeping in mind that we are dealing with $P = \mathsf{NW}_{d,m,e}$ and that $m^k > (1+\epsilon)^{2(d-k)}$. We shall choose Δ to be a set of size exactly $(1+\epsilon)^{2(d-k)}$ which consists of monomials of the form $x_{1a_1} \cdots x_{ka_k}$ with each $a_i \leq m$. This shall become relevant later.

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \geq \quad \left| \bigcup_{\substack{\alpha \in \Delta \\ \beta \in \partial_\alpha(P)}} A(\alpha,\beta) \right| . \tag{A.1}$$

We are abusing notation here to use $\beta \in \partial_\alpha(P)$ to mean that β is a non-zero monomial in $\partial_\alpha(P)$.

We shall need the following lemma from [19] that is a strengthening of the standard Inclusion-Exclusion principle.

▶ **Lemma 28** (Stronger Inclusion-Exclusion [19]). *Let $A_1, \ldots, A_r$ be sets and suppose*

$$\lambda := \frac{\sum_{i \neq j} |A_i \cap A_j|}{\sum_i |A_i|} \geq 1.$$

Then,

$$\left| \bigcup_i A_i \right| \quad \geq \quad \left(\frac{1}{4\lambda} \right) \cdot \left(\sum_i |A_i| \right) .$$

▶ **Corollary 29.** *Considers sets $A_1, \ldots, A_r$ and let $S_1 = \sum_i |A_i|$ and $S_2 = \sum_{i \neq j} |A_i \cap A_j|$. Then,*

$$\left| \bigcup A_i \right| \quad \geq \quad \frac{S_1}{4} \cdot \min\left(1, \frac{S_1}{S_2}\right) .$$

Estimating $|\bigcup A(\alpha,\beta)|$ via Inclusion-Exclusion

$$\left| \bigcup_{\alpha,\beta} A(\alpha,\beta) \right| \quad \geq \quad \sum_{\alpha,\beta} |A(\alpha,\beta)| \quad - \quad \sum_{(\alpha,\beta) \neq (\alpha',\beta')} |A(\alpha,\beta) \cap A(\alpha',\beta')| .$$

Let us first address the term $\sum |A(\alpha,\beta)|$. As mentioned earlier, it is not an easy task to get a good handle on the set $A(\alpha,\beta)$ for polynomial such as NW, for any reasonable monomial ordering. However, [19] circumvent this difficult by using an indirect approach to estimate this term.

For any derivative α and $\beta \in \partial_\alpha(P)$, define the set $S(\alpha,\beta)$ as the following set of multilinear monomials of degree ℓ that is disjoint from β.

$$S(\alpha,\beta) \quad = \quad \left\{ \mu \ : \ \begin{array}{c} \mu \text{ is multilinear, has} \\ \text{degree } \ell \text{ and } \gcd(\beta,\mu) = 1 \end{array} \right\} .$$

This on the other hand is independent of any monomial ordering, and is also easy to calculate:

$$\text{For every } \alpha,\beta \qquad |S(\alpha,\beta)| \quad = \quad \binom{n-d+k}{\ell} .$$

▶ **Lemma 30** ([19]). *For any α,*

$$\sum_{\beta} |A(\alpha,\beta)| \quad \geq \quad \left|\bigcup_{\beta} S(\alpha,\beta)\right|.$$

Proof. Consider any $\mu \in \bigcup_\beta S(\alpha,\beta)$. By definition, there is at least one multilinear monomial in $\mu \cdot \partial_\alpha(P)$. Thus, in particular $\mathrm{LM}(\mathrm{mult}(\mu \cdot \partial_\alpha(P))$ is non-zero and equal to some $\mu \cdot \beta$ for some monomial $\beta \in \partial_\alpha(P)$. This also implies that $\gamma = \mu \cdot \beta \in A(\alpha,\beta)$. This yields an injective map ϕ

$$\phi : \bigcup_{\beta} S(\alpha,\beta) \quad \rightarrowtail \quad \{(\beta,\gamma) \ : \ \beta \in \partial_\alpha(P) \ , \ \gamma \in A(\alpha,\beta)\}\,.$$

Since the size of the RHS is precisely $\sum_\beta |A(\alpha,\beta)|$, the lemma follows. ◀

Thus, by another use of Inclusion-Exclusion on the $S(\alpha,\beta)$'s, we get

$$\begin{aligned}
\left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \quad &\geq \quad \sum_{\alpha,\beta} |A(\alpha,\beta)| \quad - \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta)\cap A(\alpha',\beta')| \\
&\geq \quad \sum_{\alpha}\left(\sum_{\beta} |S(\alpha,\beta)|\right) \quad - \sum_{\alpha}\left(\sum_{\beta\neq\beta'} |S(\alpha,\beta)\cap S(\alpha,\beta')|\right) \\
&\qquad - \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta)\cap A(\alpha',\beta')|\,.
\end{aligned}$$

Let us call the three terms in the RHS of the last equation as T_1, T_2 and T_3 respectively. Since we know the size of each $S(\alpha,\beta)$ exactly, the value of T_1 is easily obtained.

▶ **Lemma 31** ([19]).

$$T_1(\alpha) \quad := \quad \sum_{\beta} |S(\alpha,\beta)| \quad = \quad (\#\ \textit{mons in a deriv}) \cdot \binom{n-d+k}{\ell}.$$

We shall be simplifying such binomial coefficients very often.

▶ **Lemma 32.** *Let n and ℓ be parameters such that $\ell = \frac{n}{2}(1-\epsilon)$ for some $\epsilon = o(1)$. For any a, b such that $|a|\,,|b| = O(\sqrt{n})$,*

$$\binom{n-a}{\ell-b} \quad = \quad \binom{n}{\ell} \cdot 2^{-a} \cdot (1+\epsilon)^{a-2b} \cdot \exp(-O(b\cdot\epsilon^2)).$$

Proof. The proof of the above lemma would repeatedly use the fact that $n! = (n-a)! \cdot n^a \cdot \mathrm{poly}(n)$ whenever $a = O(\sqrt{n})$ (see [9, Lemma 3.4]).

$$\begin{aligned}
\frac{\binom{n-a}{\ell-b}}{\binom{n}{\ell}} \quad &= \quad \frac{(n-a)!}{n!} \cdot \frac{\ell!}{(\ell-b)!} \cdot \frac{(n-\ell)!}{(n-\ell-a+b)!} \\
&\overset{\mathrm{poly}}{\approx} \quad \frac{1}{n^a} \cdot \ell^b \cdot \frac{(n-\ell)^a}{(n-\ell)^b} \\
&= \quad \frac{\left(\frac{n}{2}\right)^a (1+\epsilon)^a}{n^a} \cdot \frac{(1-\epsilon)^b}{(1+\epsilon)^b} \\
&= \quad 2^{-a} \cdot (1+\epsilon)^{a-2b} \cdot \exp(-O(b\cdot\epsilon^2)).
\end{aligned}$$

◀

Since our of parameters would be $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$, the bound on T_1 can be simplified as

$$\begin{aligned} T_1(\alpha) &= (\#\text{ mons in a deriv}) \cdot \binom{n}{\ell} \cdot \left(\frac{1+\epsilon}{2}\right)^{d-k} \cdot O(1) \\ &= m^{e-k} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\epsilon}{2}\right)^{d-k} \cdot O(1) \\ &= \binom{n}{\ell} \cdot O(1). \end{aligned}$$

where we used the fact that every non-zero k-th order derivative of $\mathsf{NW}_{d,m,e}$ has exactly m^{e-k} monomials and our setting of parameters.

▶ Remark. In this particular instance, the approximation factor was just $O(1)$ but we would often have to deal with situations where this factor is $\exp(O(\log^2 d))$. To avoid writing this factor of $\exp(O(\log^2 d))$, we shall use $\approx$ of $\gtrsim$ or $\lesssim$ to indicate that a factor $\exp(O(\log^2 d))$ is omitted.

We now move on to the calculation of T_2. This is the first place where the choice of the polynomial and parameters becomes crucial.

▶ **Lemma 33** ([19]). *For the polynomial $P = \mathsf{NW}_{d,m,e}$, if $n = md$ and $\ell = \frac{n}{2}(1-\epsilon)$ for $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$, for any $\alpha \in \Delta$*

$$T_2(\alpha) \quad := \quad \sum_{\beta \neq \beta'} |S(\alpha,\beta) \cap S(\alpha,\beta')| \quad \lesssim \quad m^{2(e-k)} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\epsilon}{2}\right)^{2d-2k}.$$

Proof. Recall that $S(\alpha,\beta) \cap S(\alpha,\beta')$ is just set of all multilinear monomials μ of degree ℓ that are disjoint from both β and β'. Hence, for any pair of multilinear degree $(d-k)$ monomials $\beta \neq \beta' \in \partial_\alpha(P)$ such that $\mathsf{Deg}(\gcd(\beta,\beta')) = t$,

$$|S(\alpha,\beta) \cap S(\alpha,\beta')| \quad = \quad \binom{n-2d+2k+t}{\ell}.$$

Thus, if we can count the number of pairs (β,β') that agree on exactly t places, we can obtain $T_2(\alpha)$. Note that for $\mathsf{NW}_{d,m,e}$, any two $\beta,\beta' \in \partial_\alpha(\mathsf{NW}_{d,m,e})$ can agree on at most $e-k$ places. Further, the number of pairs that agree in exactly $0 \le t \le e-k$ places is at most

$$m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t}$$

as there are m^{e-k} choices for β, and $\binom{d-k}{t}$ choices for places where they may agree, and

$(m-1)^{e-k-t}$ choices for β' that agree with β on those t places. Thus,

$$\begin{aligned}
T_2(\alpha) &\leq \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t} \cdot \binom{n-2d+2k+t}{\ell} \\
&\approx \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t} \cdot \binom{n}{\ell} \frac{1}{2^{2d-2k-t}} \cdot (1+\epsilon)^{2d-2k-t} \\
&\leq m^{2(e-k)} \binom{n}{\ell} \left(\frac{1+\epsilon}{2}\right)^{2d-2k} \cdot \sum_{t=0}^{e-k} \binom{d-k}{t} \left(\frac{2}{(1+\epsilon)m}\right)^{t} \\
&\leq m^{2(e-k)} \binom{n}{\ell} \left(\frac{1+\epsilon}{2}\right)^{2d-2k} \cdot \left(1+\frac{2}{(1+\epsilon)m}\right)^{d-k} \\
&= m^{2(e-k)} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\epsilon}{2}\right)^{2d-2k} \cdot O(1) \qquad \text{as } m = \Omega(d).
\end{aligned}$$

◀

Combining this with Lemma 31 and using Inclusion-Exclusion (Corollary 29), we get that for every $\alpha \in \Delta$,

$$\begin{aligned}
\left|\bigcup_{\beta} S(\alpha,\beta)\right| &\gtrsim T_1(\alpha) \cdot \min\left(1, \frac{T_1(\alpha)}{T_2(\alpha)}\right) \\
&\approx T_1(\alpha) \cdot \min\left(1, \frac{\left(\frac{2}{1+\epsilon}\right)^{d-k}}{m^{e-k}}\right) \\
&\approx T_1(\alpha).
\end{aligned}$$

by our choice of parameters. Note that e needs to tailored very precisely to force the above condition! If e is chosen too large or small, we get nothing from this whole exercise!

Thus by Lemma 30 and Lemma 31, we get

$$\sum_{\substack{\alpha\in\Delta \\ \beta\in\partial_\alpha(P)}} |A(\alpha,\beta)| \geq |\Delta| \cdot \left|\bigcup_{\beta} S(\alpha,\beta)\right| \geq |\Delta| \cdot T_1(\alpha) \approx |\Delta| \cdot \binom{n}{\ell}. \tag{A.2}$$

Upper bounding $\sum |A(\alpha,\beta) \cap A(\alpha',\beta')|$

We are still left with the task of upper bounding

$$T_3 = \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta) \cap A(\alpha',\beta')| .$$

As mentioned earlier, we really do not have a good handle on the set $A(\alpha,\beta)$, and certainly not on the intersection of two such sets. Once again, we shall use a proxy that is easier to estimate to upper bound T_3.

The set $A(\alpha,\beta) \cap A(\alpha',\beta')$ consists of multilinear monomials γ of degree $\ell+d-k$ such that there exists multilinear monomials μ, μ' of degree ℓ satisfying

$$\begin{aligned}
\gamma &= \mu\beta = \mu'\beta', \\
\mu\beta &= \mathrm{LM}(\mathrm{mult}(\mu\partial_\alpha(P))) \\
\text{and} \quad \mu'\beta' &= \mathrm{LM}(\mathrm{mult}(\mu'\partial_{\alpha'}(P))).
\end{aligned}$$

This in particular implies that γ must be divisible by both β and β'.

▶ **Observation 34.** *If* $\mathsf{Deg}(\gcd(\beta,\beta')) = t$, *then*

$$|A(\alpha,\beta) \cap A(\alpha',\beta')| \quad \leq \quad \binom{n-2d+2k+t}{\ell-d+k+t}.$$

Proof. Every monomial $\gamma \in A(\alpha,\beta) \cap A(\alpha',\beta')$ must be divisible by β and β'. Since $|\beta \cup \beta'| = 2d-2k-t$, the number of choices of γ is precisely

$$\binom{n-(2d-2k-t)}{(\ell+d-k)-(2d-2k-t)} \quad = \quad \binom{n-2d+2k+t}{\ell-d+k+t}.$$ ◀

One needs a similar argument as in the case of T_2 to figure out how many pairs $(\alpha,\beta) \neq (\alpha',\beta')$ are there with $\mathsf{Deg}(\gcd(\beta,\beta')) = t$ and sum them up accordingly.

▶ **Lemma 35** ([19]). *For the polynomial* $\mathsf{NW}_{d,m,e}$, *and* $n = md$ *and* $\ell = \frac{n}{2}(1-\epsilon)$ *for* $\epsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$,

$$\sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta) \cap A(\alpha',\beta')| \quad \lesssim \quad |\Delta|^2 \cdot \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell}.$$

Proof. Fix a pair of derivatives α, α'. Let

$$T_3(\alpha,\alpha') \quad := \quad \sum_{\substack{\beta\in\partial_\alpha(P)\\ \beta'\in\partial_{\alpha'}(P)\\ (\alpha,\beta)\neq(\alpha',\beta')}} |A(\alpha,\beta) \cap A(\alpha',\beta')|.$$

As before, we shall first count the number of pairs of monomials $\beta \in \partial_\alpha P$ and $\beta' \in \partial_{\alpha'} P$ such that $\gcd(\beta,\beta') = t$. Note that since α may differ from α', we could potentially have $\gcd(\beta_1,\beta_2) = e$. Once again, this is easily seen to be at most

$$m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t}.$$

Therefore, using Observation 34,

$$\begin{aligned}
T_3(\alpha,\alpha') &\leq \sum_{t=0}^{e} m^{e-k} \cdot (m-1)^{e-k-t} \binom{d-k}{t} \binom{n-2d+2k+t}{\ell-d+k+t} \\
&\approx \sum_{t=0}^{e} m^{e-k} \cdot (m-1)^{e-k-t} \binom{d-k}{t} \cdot \binom{n}{\ell} \left(\frac{1}{2}\right)^{2d-2k-t} (1+\epsilon)^t \\
&\leq \frac{m^{2(e-k)}}{2^{2(d-k)}} \cdot \binom{n}{\ell} \cdot \left(1+\frac{2(1+\epsilon)}{m}\right)^{d-k} \\
&\approx \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell}. \\
\implies T_3 &\lesssim |\Delta|^2 \cdot \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell}.
\end{aligned}$$

◀

Recalling that we have chosen our parameters so that

$$\frac{m^{e-k}}{2^{d-k}} \approx \left(\frac{1}{1+\epsilon}\right)^{d-k} \quad \text{and} \quad |\Delta| = (1+\epsilon)^{2(d-k)},$$

the above equation reduces to

$$T_3 = \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta) \cap A(\alpha',\beta')| \lesssim |\Delta| \cdot \binom{n}{\ell}.$$

Combining with (A.2), we obtain the required bound for $|\bigcup A(\alpha,\beta)|$ via Inclusion-Exclusion (Corollary 29).

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e}) \geq \left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \gtrsim \binom{n}{\ell} \cdot (1+\epsilon)^{2d-2k}.$$

The only thing left to observe is that by Lemma 32,

$$\binom{n}{\ell+d-k} \approx \binom{n}{\ell} \cdot (1+\epsilon)^{2d-2k},$$

and that completes the proof of Lemma 18. ◀