The Computational Complexity Column

by

V. Arvind

Institute of Mathematical Sciences, CIT Campus, Taramani
Chennai 600113, India
arvind@imsc.res.in
http://www.imsc.res.in/~arvind

It is widely believed that the Permanent polynomial requires superpolynomial size arithmetic circuits. This is the central problem in the area of arithmetic circuit complexity. In the last couple of years there has been exciting progress on this problem with several new arithmetic circuit lower bound results. Ramprasad Saptharishi's nicely written survey article explains these new developments. He shares with the reader intuitions and proof insights.

Recent Progress on Arithmetic Circuit Lower Bounds

Ramprasad Saptharishi
Microsoft Research India
ramprasad@cmi.ac.in

Abstract

The field of arithmetic circuit complexity has lately seen a flurry of results. Several of these results are centered homogeneous depth four circuits, and come tantalizingly close to separating the algebraic analogue of P from the algebraic analogue of NP. In this article, we survey some of the more recent results and present the key intuitions. We also look at some results on depth reduction and some approaches aimed towards proving super-polynomial lower bounds for homogeneous formulas.

1 Introduction

"What is the best way to compute a given polynomial $f(x_1, \ldots, x_n)$ from basic operations such as $+$ and $\times$?" This is the main motivating problem in the field of arithmetic circuit complexity. The notion of *complexity* of a polynomial is measured via the size of the smallest arithmetic circuit computing it. Arithmetic circuits provide a robust model of computation for polynomials. Formally, these are directed acyclic graphs with a unique sink vertex, where internal nodes are labelled by $+$ and $\times$ and each source node labelled with either a variable or a field constant. Each $+$ gate computes the sum of the polynomials computed at its children, and $\times$ gates the product. The unique sink vertex is called the root or the output gate, and the polynomial computed by that gate is the polynomial computed by the circuit.

There are several interesting questions that can be asked about arithmetic circuits, and polynomials that they compute. One category of problems are of the form, "Is there an explicit polynomial $f(x_1, \ldots, x_n)$ that require (perhaps restricted) arithmetic circuits of size $2^{\Omega(n)}$ to compute them?", or questions about proving lower bounds. Another category of problems are of the form,

"Is the given circuit computing the 0 polynomial?", which is also called 'Polynomial Identity Testing (PIT)'. Yet another class of questions are of the form "Given oracle access to a circuit, can you write down the polynomial computed by this circuit?", which are also called 'polynomial reconstruction'. Several of these problems have very strong connections between each other despite being of very different flavours. Formal connections between PIT and lower bounds have been shown by [18, 1]. Further, strong lower bounds for restricted models have often been succeeded by reconstruction algorithms (at least on average). In this article we shall mainly be looking at lower bounds. For more on reconstruction and PIT questions, the author is invited to read other excellent surveys such as [34, 6].

1.1 Arithmetic complexity classes

In the seminal paper of [36], Valiant defined two classes of polynomials which we now call VP and VNP.

Definition 1. *The class* VP *is defined as the set of all polynomial* $f(x_1, \ldots, x_n)$ *with* $\deg(f) = n^{O(1)}$ *that can be computed by an arithmetic circuit of size* $s = n^{O(1)}$.
The class VNP *is defined as the set of all polynomial* $f(x_1, \ldots, x_n)$ *such that there exists a* $g(x_1, \ldots, x_n, y_1, \ldots, y_m)$ *with* $m = n^{O(1)}$ *such that*

$$f(x_1, \ldots, x_n) \quad = \quad \sum_{y_1=0}^{1} \cdots \sum_{y_m=0}^{1} g(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

The class VP is synonymous to what we understand as *efficiently computable* polynomials. The class VNP, whose definition is similar to the boolean class NP, is in some sense a notion of what deem as *explicit*.

Fact 1. *Let* $f(x_1, \ldots, x_n)$ *be a polynomial such that* $\deg(f) = n^{O(1)}$ *and given any exponent vector* $e_1, \ldots, e_n$*, the coefficient of the monomial* $x_1^{e_1} \ldots x_n^{e_n}$ *in* f *can be computed in polynomial time. Then,* $f \in$ VNP.

For example, consider the permanent of a symbolic $n \times n$ matrix. In fact, [36] showed that the symbolic $n \times n$ permanent is in some sense complete for the class VNP. Further, he also showed that the determinant of a symbolic $n \times n$ matrix is (almost) complete for the class VP. Separating the determinant and the permanent is the Holy Grail in the field of arithmetic circuit complexity.

Remark. Note that the above fact merely gives a sufficient condition for a polynomial to be in VNP. There are examples of polynomials f where computing the coefficient of a given monomial is believed to be very hard but $f \in \mathsf{VNP}$.[1] In this article however, all the polynomials we shall be dealing with would have this property that the coefficient of a given monomial can be efficiently computed. For more about completeness classes in arithmetic complexity, [4] is a wonderful text.

1.2 Prior lower bounds

Proving lower bounds is generally considered challenging, in most models of computation. For general circuits, the best lower bound we have for an explicit polynomial is by [5] who prove an $\Omega(n \log n)$ lower bound. For the subclass of arithmetic formulas, [17] has shown a $\Omega(n^{3/2})$ lower bound. On the other hand, we know by standard counting methods that most n-variate degree d polynomials require circuits of size $\Omega\left(\sqrt{\binom{n+d}{d}}\right)$.

To gain better understanding of computation by arithmetic circuits, researchers focused on proving lower bounds for restricted models of computation. One very natural restriction is the depth of the circuit. Proving lower bounds for depth two circuits are trivial. For general depth three circuits, the best lower bound we have is by [33] who present an $\Omega(n^2)$ lower bound. Exponential lower bounds are known with additional restrictions like *homogeneity* [27], *multilinearity* [28, 30], over finite fields [13, 10], *monotonicity* [16] etc.

For multilinear models, more is known for even larger depth. [28] showed an $n^{\Omega(\log n)}$ lower bound for the class of multilinear formulas. [30] extended those techniques to show an $2^{n^{\Omega(1/\Delta)}}$ lower bound for multilinear formulas of depth Δ.

1.3 Relevance of shallow circuits for "VP vs VNP"

The study of lower bounds for shallow circuits is not just an attempt to simplify the problem and gain insight on the larger goal. The class of shallow arithmetic circuits are surprisingly powerful, unlike the boolean case. Shallow circuits in the arithmetic world almost capture the entire computational power of unrestricted circuits!

[1] For example, consider the n^2 variate multilinear polynomial f such that the coefficient $x_{11}^{e_{11}} \dots x_{nn}^{e_{nn}}$ is the permanent of the $n \times n$ matrix $((e_{ij}))_{i,j}$. Turns out $f \in \mathsf{VNP}$. In fact, a necessary and sufficient condition is that the coefficient of a given monomial can be computed in $\#\mathsf{P}/\mathsf{poly}$.

There has been a long series of results that simulate a general arithmetic circuit C by a *shallow* circuit of size comparable to the size of C. This task simulating a circuit but another not-too-large circuit of small depth is called *depth reduction*. The first result in this regard is by [37] who proved the following.

Theorem 2 ([37]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit C of size s. Then, f can be equivalently computed by a homogeneous circuit C' of depth $O(\log d)$ with unbounded fan-in $+$ and $\times$ gates and size $s' = (nds)^{O(1)}$.*

In fact, the resulting circuit C' has the following useful structure.

- The circuit is made up of alternating layers consisting of $+$ and $\times$ gates.
- All multiplication gates have fan-in at most 5.
- If g is the polynomial computed at a multiplication gate, and g' is the polynomial computed at one of its children, then $\deg(g') \leq \deg(g)/2$.

The above theorem allows us to focus on just homogeneous circuits of $O(\log d)$ depth and attempt lower bounds for this model. Any super-polynomial lower bound for the class of $O(\log d)$ depth circuits automatically yields a super-polynomial lower bound for general circuits.

However, if we really hope to prove much stronger lower bounds for Perm_n like say $2^{\Omega(n)}$, maybe we can afford to incur a slightly larger blow-up in size to obtain an even shallower circuit. This line was first pursued by [3], and subsequently strengthened by [20] and [35] to yield the following result.

Theorem 3 ([3, 20, 35]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s. Then f can be computed by a homogeneous $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of size $s' \leq s^{O(\sqrt{d})}$*
More generally, for any $0 \leq r \leq d$, there is a homogeneous $\Sigma\Pi^{[O(d/r)]}\Sigma\Pi^{[r]}$ circuit of top fan-in at most $s^{O(d/t)}$ computing f.

Recall that a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ia} \quad , \quad \text{where } a = O(\sqrt{d}) \text{ and } \deg Q_{ij} \leq \sqrt{d}$$

In other words, if we can prove a lower bound of $n^{\omega(\sqrt{d})}$ for the class of $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits, we would have a super-polynomial lower bound for

the class of general arithmetic circuits! In fact, the model of depth 4 circuits seem so central in that almost all known lower bounds for other restricted models proceed by proving a suitable lower bound for a depth 4 analogue. Several examples of this may be seen in [22].

The first breakthrough was obtained by [11] who showed an $2^{\Omega(\sqrt{d})}$ lower bound for such circuits computing the symbolic $d \times d$ determinant or permanent. Subsequently, there was a flurry of activity[2] towards achieving the goal of proving $n^{\omega(\sqrt{d})}$ lower bounds [25, 9, 23], and this is where we currently stand.

Theorem 4. *There is an explicit homogeneous n-variate degree d polynomial f that can be computed by a homogeneous depth 4 circuit of size $n^{O(1)}$ but any $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ computing it requires top fanin $s = n^{\Omega(\sqrt{d})}$.*

If we could change the $n^{\Omega(\sqrt{d})}$ to $n^{\omega(\sqrt{d})}$ in the above theorem (of course, the polynomial f cannot then have a small arithmetic circuit computing it), we would have proved a super-polynomial lower bound for general arithmetic circuits! The following is the simplest formulation of a lower bound of shallow circuit that would imply lower bounds for general circuits.

Open Problem 1. *Find an explicit n-variate degree d polynomial f such that any expression of the form*

$$f \quad = \quad (Q_1)^{\sqrt{d}} + \cdots + (Q_s)^{\sqrt{d}} \quad , \quad \deg(Q_i) \leq \sqrt{d} \text{ for all } i$$

must have $s = n^{\omega(\sqrt{d})}$.

Subsequent to this line of work, several researchers addressed the task of proving lower bounds for homogeneous depth 4 circuits without any restriction on the fan-ins. It is worth noting that a lower bound for homogeneous depth 4 circuits must be on the total size and not the top fan-in, as otherwise one could just compute the polynomial f in a single gate of the bottom two layers.

[2] so much this is the second survey on arithmetic circuit lower bounds that the author is involved in within a year!

Why another survey?

> So why are super polynomial lower bounds still not proved? Maybe it's because not enough people are working on it.
>
> – Ran Raz (in [29])

We strongly believe that the above statement really hits the nail on the head. Fortunately, over the last few years we have seen such a phenomenal activity in arithmetic circuit lower bounds and an increased optimism that we can indeed soon separate VP and VNP. The open problem stated above is simple enough (to state!) that any one can start thinking about it. Further, we already have an $n^{\Omega(\sqrt{d})}$ lower bound, and we only need to make that $n^{\omega(\sqrt{d})}$. We believe that separating VP and VNP would be solved in the not-so-distant future and the hope is that the recent surveys would assist people familiarize with the known lower bounds and develop the necessary tools. As a student, the surveys of [34, 6] were immensely helpful and this is an attempt to give back to the community.

Recently, with Neeraj Kayal [22], we presented a comprehensive exposition of almost all known lower bounds known until then with nearly complete proofs. We tried to present all of them from a single perspective of constructing *complexity measures* for appropriate depth 4 analogues. Subsequently, there has been fresh lower bounds which, although are modifications of the earlier measures, are much more delicate to analyze and employ several new ideas to assist in the calculations. The goal of this survey is to complement [22] and present the key intuitions in the newer lower bounds for restricted arithmetic circuits. This article would not have complete proofs of the newer lower bounds but would hopefully present the main subtleties involved to help the interested reader to work through the full proofs themselves.

Organization

We first begin with some preliminaries and notation that would be required in Section 2. We then move on to present the depth reduction to depth 4 circuits to put the lower bounds in context. We then outline the general road map followed by almost all lower bound proofs in Section 4 and then proceed to the lower bounds of [19] and [24] in Section 5. In Section 6, we focus on non-homogeneous depth 3 circuits and present the depth reduction of [12] and the recent lower bound of [21] for depth three circuits with small bottom fan-in. We look at some speculative approaches towards proving superpolynomial lower bounds for homogeneous formulas in Section 7 before concluding in Section 8.

2 Notation and preliminaries

2.1 Subclasses of circuits

We shall be considering various subclass of constant depth circuits in this article and it would be useful to fix some notation for the parameters involved.

- A $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_i Q_{i1} \dots Q_{ia} \quad \text{where} \quad \deg Q_{ij} \leq b$$

- A $\wedge$ refers to a layer of exponentiation gates. For example, a $\Sigma \wedge \Sigma$ circuit computes a polynomial of the form

$$f \quad = \quad \ell_1^{d_1} + \cdots + \ell_s^{d_s}$$

  where each ℓ_i is a linear polynomial.

- In general, we shall add super script such as $\Sigma^{[a]}$ or $\Pi^{[a]}$ to denote a bound on the fan-in of gates in that layer. For example, $\Sigma\Pi\Sigma^{[a]}$ would refer to depth three circuits where every linear polynomial depends on at most a variables.

Throughout the article, we would be dealing mainly with multilinear polynomials with zero-one coefficients. Thus, it would be useful to identify such any monomial of such a polynomial $P(x_1, \dots, x_n)$ by the set of variables that divide it. This shall allow us to say "$m_1 \cap m_2$" instead of $\gcd(m_1, m_2)$. Further, we shall abuse notation and say "$m \in P(\mathbf{x})$" to mean that the monomial m has a non-zero coefficient in $P(\mathbf{x})$.

2.2 Some useful estimates

We shall be seeing a lot of binomial coefficients and the following lemma would be useful to have a handle on how large they are.

Lemma 5. *Let n and ℓ be parameters such that $\ell = \frac{n}{2}(1-\varepsilon)$ for some $\varepsilon = o(1)$. For any a, b such that $a, b = O(\sqrt{n})$,*

$$\binom{n-a}{\ell-b} \quad = \quad \binom{n}{\ell} \cdot 2^{-a} \cdot (1+\varepsilon)^{a-2b} \cdot \mathsf{poly}(n)$$

Proof. The proof of the above lemma would repeated use [11, Lemma 6] that

$$(n+a)! \quad = \quad n! \cdot n^a \cdot \mathsf{poly}(n)$$

for any $a = O(\sqrt{n})$.
Hence,

$$\begin{aligned}
\frac{\binom{n-a}{\ell-b}}{\binom{n}{\ell}} &= \frac{(n-a)!}{n!} \cdot \frac{\ell!}{(\ell-b)!} \cdot \frac{(n-\ell)!}{(n-\ell-a+b)!} \\
&\overset{\mathsf{poly}}{\approx} \frac{1}{n^a} \cdot \ell^b \cdot \frac{(n-\ell)^a}{(n-\ell)^b} \\
&= \frac{\left(\frac{n}{2}\right)^a (1+\varepsilon)^a}{n^a} \cdot \frac{(1-\varepsilon)^b}{(1+\varepsilon)^b} \\
&\overset{\mathsf{poly}}{\approx} 2^{-a} \cdot (1+\varepsilon)^{a-2b}
\end{aligned}$$

□

2.3 Polynomials of interest

There are a few polynomials that are the usual suspects while proving lower bounds. The polynomials that we would be dealing with in this article are defined below.

The determinant and the permanent families

The determinant of an $n \times n$ symbolic matrix shall be denoted by Det_n and is defined as

$$\mathsf{Det}_n \quad = \quad \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) x_{1,\sigma(1)} \ldots x_{n,\sigma(n)}$$

The permanent of an $n \times n$ symbolic matrix shall be denoted by Perm_n and is defined as

$$\mathsf{Perm}_n \quad = \quad \sum_{\sigma \in S_n} x_{1,\sigma(1)} \ldots x_{n,\sigma(n)}$$

Both of these polynomials are of degree n and over n^2 variables. We know that Det_n can be computed by a polynomial sized arithmetic circuit and it is widely believed that the permanent requires circuits of size $2^{\Omega(n)}$.

The Nisan-Wigderson polynomial families

Let n, m, d be arbitrary parameters with m being a power of a prime, and $n, d \leq m$. 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 $n \leq m$, we have that $[n] \subseteq \mathbb{F}_m$. The Nisan-Wigderson polynomial with parameters n, m, d, denoted by $\mathrm{NW}_{n,m,d}$ is defined as

$$\mathrm{NW}_{n,m,d}(\mathbf{x}) \quad = \quad \sum_{\substack{p(t) \in \mathbb{F}_m[t] \\ \deg(p) \leq d}} x_{1,p(1)} \dots x_{n,p(n)}$$

That is, for every univariate polynomial $p(t) \in \mathbb{F}_m[t]$ of degree at most d, we add one monomials that encodes the 'graph' of p on the points $[n]$. This is a polynomial of degree n over mn variables.
This monomials of this polynomial satisfy a very useful low-pairwise-intersection property.

Lemma 6. *Let m_1 and m_2 be any two distinct monomials in $\mathrm{NW}_{n,m,d}(\mathbf{x})$. Then, there are at most d variables that divide both m_1 and m_2.*

Proof. Let m_1 and m_2 correspond to univariates $p_1(t), p_2(t) \in \mathbb{F}_m[t]$ of degree at most d. Then if x_{ij} divides m_1, then $p_1(i) = j$, similarly for m_2. But since p_1 and p_2 are two distinct polynomials of degree at most d, they can agree in at most d evaluations. Thus, there can be at most d variables that divide both m_1 and m_2. □

For most generic choices of the parameters, the polynomial $\mathrm{NW}_{n,m,d}$ is believed to require circuits of exponential size to compute them.

The Iterated-Matrix-Multiplication polynomial

For parameters n and d, the Iterated-Matrix-Multiplication polynomial, denoted by $\mathrm{IMM}_{n,d}$, is defined as follows

$$\mathrm{IMM}_{n,d} \quad = \quad \sum_{1 \leq i_1, \dots, i_d \leq n} x^{(1)}_{1,i_1} x^{(2)}_{i_1,i_2} \dots x^{(d-1)}_{i_{d-2},i_{d-1}} x^{(d)}_{i_{d-1},1}.$$

An equivalent way of defining the polynomial as the $(1,1)$-th entry of the product of d generic $n \times n$ matrices:

$$\mathrm{IMM}_{n,d} \quad = \quad \left(\begin{bmatrix} x^{(1)}_{11} & \dots & x^{(1)}_{1n} \\ \vdots & \ddots & \vdots \\ x^{(1)}_{n1} & \dots & x^{(1)}_{nn} \end{bmatrix} \cdots \begin{bmatrix} x^{(d)}_{11} & \dots & x^{(d)}_{1n} \\ \vdots & \ddots & \vdots \\ x^{(d)}_{n1} & \dots & x^{(d)}_{nn} \end{bmatrix} \right)_{(1,1)}.$$

It is often useful to think of this as the polynomial computed by a *generic algebraic branching program* of width n and depth n (where the edge connecting vertex i of layer ℓ to vertex j of layer $\ell+1$ has weight $x_{ij}^{(\ell)}$).
This is a polynomial of degree d and over $n^2(d-2)+2n$ variables. Further, since the polynomial corresponds to a generic algebraic branching program, $\mathrm{IMM}_{n,d}$ can be computed by an arithmetic circuit of size $\mathsf{poly}(n,d)$.

3 A primer on depth reduction to depth 4 circuits

In this section, we shall look at depth reduction for arithmetic circuits. As mentioned earlier, the starting point of all known depth reductions is the result of [37].

Theorem 2 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit C of size s. Then, f can be equivalently computed by a homogeneous circuit C' of depth $O(\log d)$ with unbounded fan-in $+$ and $\times$ gates and size $s' = (nds)^{O(1)}$.*
Further, the circuit C' has the following structure:

- *The circuit is made up of alternating layers consisting of $+$ and $\times$ gates.*
- *All multiplication gates have fan-in at most 5.*
- *If g is the polynomial computed at a multiplication gate, and g' is the polynomial computed at one of its children, then $\deg(g') \leq \deg(g)/2$.*

We shall not be proving this theorem here but with this as the starting point, we shall given an alternate proof of the depth reduction by [3, 20, 35]. This alternate proof was obtained jointly with V Vinay.

Theorem 3 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s. Then f can be computed by a homogeneous $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of size $s' \leq s^{O(\sqrt{d})}$*

Proof. Start with the circuit C' obtained from Theorem 2 computing f of size $s' = s^{O(1)}$. Let g be the polynomial computed at any arbitrary gate in the circuit. From the structure of C', we have that

$$g = \sum_{i=1}^{s'} g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5} \tag{1}$$

with $\deg(g_{i1}) + \cdots + \deg(g_{i5}) \leq \deg(g)$ and $\deg(g_{ij}) \leq \deg(g)/2$ for all i, j.
Key Observation. For each i, there must be at least two j's such that $\deg(g_{ij}) \geq \deg(g)/8$.

Since the above decomposition is true for any gate in the circuit, f, polynomial computed at the root, can be written as

$$f = \sum_{i=1}^{s'} f_{i1} \cdot f_{i2} \cdot f_{i3} \cdot f_{i4} \cdot f_{i5}. \tag{2}$$

The RHS is a $\Sigma\Pi^{[5]}\Sigma\Pi^{[d/2]}$ circuit of top fan-in s'. The goal would be to progressively reduce the bottom fan-in at the cost of increasing the top fan-in slightly. Eventually, we want an expression where all f_{ij}'s involved have degree at most $\sqrt{d}$.
We shall follow an extremely natural strategy:

> Start with (2) for f.
>
> For each summand $f_{i1} \dots f_{ir}$ in the RHS, if the largest degree f_{ij} has degree more than $\sqrt{d}$, expand that f_{ij} with the its corresponding representation using (1).
>
> Repeat this process until all f_{ij}'s on the RHS have degree at most $\sqrt{d}$.

In every round of the above routine, the initial equation (2) for f slowly evolves. At the end of each round, the top fan-in increases by a factor of s' but there is some drop in the degree of terms involved. We now need to show that by $O(\sqrt{d})$ rounds, all of the f_{ij}'s involved would have degree bounded by $\sqrt{d}$. This would imply that the top fan-in of that equation is bounded by $s^{O(\sqrt{d})}$ as claimed.
If we take any term $f_{i1} \dots f_{ir}$ with $\deg(f_{i1}) \geq \sqrt{d}$ and expand f_{ij} via (1), each term in the expansion of f_{i1} must have at least two factors of degree more than $\sqrt{d}/8$ (by the key observation). Thus, in each term obtained by expanding f_{i1} in $f_{i1} \dots f_{ir}$ must have the number of factors of degree more than $\sqrt{d}/8$ increased by at least one. Since we know that no term can have more than $8\sqrt{d}$ such factors, this must imply that the number of rounds is bounded by $8\sqrt{d}$.
Thus we eventually have an equation of the form

$$f = \sum_{i=1}^{s'^{8\sqrt{d}}} f_{i1} \dots f_{ir} \quad \text{where for each } i, j, \quad \deg(f_{ij}) \leq \sqrt{d}$$

To ensure that $r \leq O(\sqrt{d})$, the standard trick is to take any ensure that $\deg(f_{ij}) \geq \sqrt{d}/2$ by multiplying out factors of degree smaller than $\sqrt{d}/2$. Thus, we have a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of top fan-in $s^{O(\sqrt{d})}$ computing f. $\square$

The original proof of Tavenas was not much involved either, but the above proof would be able to offer more insights towards proving homogeneous formula lower bounds. We shall however defer this discussion to Section 7.

Surprisingly, it was shown by [12] that over characteristic zero fields, any n-variate degree d polynomial f can be computed by a depth 3 circuit of size $n^{O(\sqrt{d})}$. We shall present its proof in Section 6 where it would better placed alongside the recent lower bound for depth 3 circuits by [21].

4 'Natural' proof strategies

Most lower bounds follow the plan outlined below. There are a few notable exceptions but by and large this is the general strategy followed by almost all known lower bound proofs.

> **Step 1 (normal forms)** For every circuit in the circuit class $\mathcal{C}$ of interest, express the polynomial computed as a *small sum of simple building blocks*.

For example, every $\Sigma\Pi\Sigma$ circuit is a *small* sum of *products of linear polynomials* which are the building blocks here. In this case, the circuit model naturally admits such a representation. There are cases when obtaining this representation might itself be non-trivial.

> **Step 2 (complexity measure)** Construct a map $\Gamma : \mathbb{F}[x_1, \ldots, x_n] \to \mathbb{Z}_{\geq 0}$ with that is *sub-additive* i.e. $\Gamma(f_1 + f_2) \leq \Gamma(f_1) + \Gamma(f_2)$

This is really the most crucial part of the lower bounds. In most cases, $\Gamma(f)$ is the rank of a large matrix whose entries are linear functions in the coefficients of f. This would be the case in the lower bounds considered in this article as well. In such cases, we immediately get that Γ is sub-additive. The strength of the choice of Γ is determined by the next step.

> **Step 3 (potential usefulness)** Show that if B is a *simple building block*, then $\Gamma(B)$ is *small*. Further, check if $\Gamma(f)$ for a *random polynomial* f is large (potentially).

This would suggest that if any f with large $\Gamma(f)$ is to be written as a sum of $B_1 + \cdots + B_s$, then sub-additivity and the fact that $\Gamma(B_i)$ is small for each i and $\Gamma(f)$ is large immediately imply that s must be large. This implies that the complexity measure Γ does indeed have a potential to prove a lower bound for the class. The next step is just to replace the *random polynomial* by an explicit polynomial.

> **Step 4 (explicit lower bound)** Find an explicit polynomial f for which $\Gamma(f)$ is large.

The bulk of all lower bound proofs goes into this step. In several cases, there are natural candidate polynomials for which one can show $\Gamma(f)$ is large. In some cases, it might be easier to *engineer* a polynomial for which it is easier to show that $\Gamma(f)$ is large.

With this general strategy in mind, we can go ahead to see the lower bounds of [19, 24].

5 Lower bounds for homogeneous depth four circuits

The model for which we shall be interested in proving lower bounds are homogenous depth four circuits. These circuits compute polynomials of the form

$$f \quad = \quad \sum_i Q_{i1} \dots Q_{ia_i}$$

where each Q_{ij} is a homogeneous polynomial. This immediately forces that $\sum_{j=1}^{a_i} \deg(Q_{ij}) = \deg(f)$ for all i.

Goal. Find an explicit polynomial f (of degree d, and over n variables) such that any homogeneous depth four circuit requires size $n^{\Omega(\sqrt{d})}$. That is, if

$$f \quad = \quad \sum_i Q_{i1} \dots Q_{ia_i}$$

for homogeneous polynomials Q_{ij}'s, then the total number of monomials present among the Q_{ij}'s must be $n^{\Omega(\sqrt{d})}$.

Intuition towards the measure - (1)

Consider an expression of the form

$$C \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ia_i}$$

We shall call a summand $Q_{i1} \dots Q_{ia_i}$ *good* if the degree of each $Q_{ij} \leq \sqrt{d}$. Let us split the above sum into *good* terms and the rest.

$$C_1 \quad = \quad \sum_{i=1}^{s_1} Q_{i1} \dots Q_{ia_i} \quad \text{where} \quad \deg(Q_{ij}) \leq \sqrt{d} \quad \text{for all } i, j \tag{3}$$

$$C_2 \quad = \quad \sum_{i=s_1+1}^{s} Q_{i1} \dots Q_{ia_i} \quad \text{where} \quad \deg(Q_{i1}) > \sqrt{d} \quad \text{for all } i > s_1 \tag{4}$$

If one were to just prove a lower bound for (3), then using the dimension of shifted partial derivatives we can obtain a lower bound of $n^{\Omega(\sqrt{d})}$. Hence let us focus on an expression of the form (4) and see if we can come up with a measure that gives a $n^{\Omega(\sqrt{d})}$ lower bound there as well.
Starting with (2), let us expand each Q_{i1} as a sum of monomials to obtain an expression of the form

$$C_2 \quad = \quad \sum_{i=1}^{s'} m_i \cdot Q'_i$$

where each m_i is a monomial of degree greater than $\sqrt{d}$, and Q'_i some polynomial of degree $d - \deg(m_i)$. The number of summands s' would be at most the size of the circuit we started out with.

Key Idea: Suppose the polynomial C_2 was multilinear, i.e. the degree in each variable is bounded by 1. Further, say $s' \leq n^{\sqrt{d}/10}$. Apply a random restriction ρ on the variables by setting each variable independently to zero with probability $p < \frac{1}{n^{1/20}}$.
If m was any monomial that was divisible by $\sqrt{d}$ disjoint variables, then $\rho(m) \neq 0$ with probability at most $\frac{1}{n^{\sqrt{d}/20}}$. Hence, the probability that $\rho(m_i) \neq 0$ for some $i \leq s'$ that is divisible by $\sqrt{d}$ variables is at most $\frac{1}{n^{\sqrt{d}/10}}$. Hence, the only terms that would survive on the RHS are terms of the form $\rho(m_i \cdot Q'_i)$ where m_i is divisible by at most $\sqrt{d}$ distinct variables. But recall that $\deg(m_i) > \sqrt{d}$ and this implies that m_i is non-multilinear. If that is the case, then every monomial on the RHS is non-multilinear! Thus as long as

$\rho(C_2) \neq 0$, there would be at least one multilinear monomial that survives. This would contradict our original assumption that $s' \leq n^{\sqrt{d}/10}$, giving us the lower bound we were after.

Thus, the measure for the sum of *good* terms is the dimension of shifted partial derivatives. The measure for the sum of non-*good* terms was *the number of non-zero multilinear monomials after a random restriction.* Hopefully some combination of these measures would give us a measure for their sum.[3]

Intuition towards the measure - (2)

The idea of using random restrictions as defined above essentially kills all monomials that are divisible by 'too many' variables. Let us consider an extreme case where every monomials in each Q_{ij} is just a power of a single variable. We shall first try to prove a lower bound for expression of the form

$$C \quad = \quad \sum_i Q_{i1} \cdots Q_{ia_i}$$

where every monomial in any Q_{ij} is a power of a single variable, i.e. each Q_{ij} is a sum of univariate polynomials.
Define the operator MultiQuad that acts on any polynomial Q such that MultiQuad(Q) is just the sum of monomials of Q of degree at most 2 in every variable. Then,

$$\begin{aligned} C \quad &= \quad \sum_i \text{MultiQuad}(Q_{i1}) \cdots \text{MultiQuad}(Q_{ia_i}) \quad + \quad \text{other terms} \\ &= \qquad C_1 \quad + \quad C_2 \end{aligned}$$

Notice that C_1 corresponds to a $\Sigma\Pi^{[d/2]}\Sigma\Pi^{[2]}$ circuit since we assume that each Q_{ij} is a sum of univariates. The dimension of shifted partial derivatives would yield a lower bound for such $\Sigma\Pi^{[d/2]}\Sigma\Pi^{[2]}$ circuits. But what really happens to C_2 as we take some partial derivative?

Key Observation. For any multilinear monomial m, the partial derivative $\partial_m(C_2)$ only consists of non-multilinear monomials.

Thus, this points towards the following modification of the traditional dimension of shifted partial derivatives:

[3]There are some instances when this strategy can fail spectacularly. See [23]

For any polynomial P, look at the set of polynomials of obtained as $m_1 \cdot \partial_{m_2}(P)$ where m_1 and m_2 are *multilinear monomials* of a certain degree, and compute the dimension of the *multilinear component* of these polynomials i.e. erase all monomials that are non-multilinear and then compute the dimension of the residual polynomials.

This basically allows us to completely ignore the contribution of C_2 as we have that multilinear component of $m_1\partial_{m_2}(C_2))$ is zero for every m_1 and m_2 that are multilinear.

Both these point us to a modification of the shifted partials, which [19, 24] refer to as *projected shifted partial derivatives*.

Definition 7 (Projected Shifted Partial Derivatives)**.** *Fix parameters $k, \ell > 0$. For any polynomial P, the set of projected shifted partials of P, denoted by $\mathrm{PSD}_{k,\ell}(P)$ is defined as follows*

$$\mathrm{PSD}_{k,\ell}(P) \quad = \quad \left\{ \mathrm{mult}(m_1 \cdot \partial_{m_2}(P)) \ : \ \begin{array}{c} \deg(m_1) = \ell \ , \ \deg(m_2) = k, \\ m_1 \text{ and } m_2 \text{ are multilinear} \end{array} \right\}$$

where $\mathrm{mult}(f)$ refers to the polynomial f projected to only the multilinear monomials of f.
The measure $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$ is defined as the dimension of the above set of polynomials, i.e.

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad = \quad \dim\left(\mathrm{span}(\mathrm{PSD}_{k,\ell})\right)$$

The works of [19, 24] use this measure to prove a lower bound for "*low-support* depth 4 circuits". As sketched earlier, the task of proving lower bounds for general homogeneous depth 4 circuits can be reduced to the *low-support* depth 4 circuits via random restrictions.

5.1 Reducing to 'low-support' depth 4 circuits

We have already seen a sketch of how this can be done via a random restriction but let us formalize this as a lemma.

Lemma 8. *Let P be an n-variate degree d polynomial computed by a homogeneous depth 4 circuit C of size $s \leq n^{c\sqrt{d}}$, for some $c > 0$. Let ρ be a random restriction that sets each variable to zero independently with probability $1 - 1/n^{2c}$. Then with probability at least $(1 - 1/s)$, the polynomial $\rho(P)$ is computed by a homogeneous depth 4 circuit C' with bottom support at most $\sqrt{d}$ and size at most s.*

Proof. Let $\{m_1, \ldots, m_r\}$ be the set of all monomials computed at the lowest layer of the depth 4 circuit C that are divisible by more than $\sqrt{d}$ distinct variables. Since the size of C is at most s, we also have that $r \leq s$. Then,

$$\begin{aligned} \forall i \in [r] \qquad \Pr[\rho(m_i) \neq 0] &\leq \frac{1}{n^{2c\sqrt{d}}} \\ \implies \qquad \Pr[\exists i \ : \ \rho(m_i) \neq 0] &\leq \frac{r}{n^{2c\sqrt{d}}} \leq \frac{1}{n^{c\sqrt{d}}} \leq \frac{1}{s} \end{aligned}$$

Thus, with probability at least $(1 - 1/s)$, all the large support monomials are killed and C reduces to a homogeneous depth 4 circuit of bottom support at most $\sqrt{d}$. □

5.1.1 Handling random restrictions

The previous section outlined how in essence, it would suffice to try and find an explicit polynomial for which we can prove a good enough lower bound for bounded bottom-support depth 4 circuits. Let us say that we have found an explicit polynomial g that requires depth 4 circuits of size at least $n^{\sqrt{d}/100}$. Are we done? Let us write things down formally to see exactly what we need. Say the polynomial we wish to show requires large homogeneous depth 4 circuits is f. Let us assume on the contrary that f can be computed by homogeneous depth 4 circuits of size $s < n^{\sqrt{d}/10000}$. Then, by Lemma 8, $\rho(f)$ can be computed by a homogeneous depth 4 circuits of bottom support bounded by $\sqrt{d}/1000$ of size s. We want to be able to say that this is a contradiction. We might be able to say that if $\rho(f)$ has g as *a projection*, that is, but setting more variables to zero in $\rho(f)$ we obtain g.
Both the results of [19] and [24] proceed by showing that the polynomial g, for which they show a lower bound for bounded bottom support circuits, is robust enough to yield the lower bound even after random restriction. The calculations become trickier because the calculations of $\Gamma^{[\mathrm{PSD}]}_{k,\ell}(\rho(f))$. However, in this survey we shall use an easier approach to generically lift any g to a different polynomial f such that $\rho(f)$ has g as a projection. This trick came up during discussions with Mrinal Kumar.

Lemma 9. *Let ρ be a random restriction that sets each variable to zero independently with probability $1 - p$. For any polynomial $f(y_1, \ldots y_n)$, define $f \circ \mathrm{Lin}_p$ as*

$$f \circ \mathrm{Lin}_p \quad = \quad f\left(\sum_{i=1}^{t} y_{1i}, \cdots, \sum_{i=1}^{t} y_{nt}\right) \quad \text{where } t = \left(\frac{1}{p}\right) n \log n$$

Then, $\rho(f \circ \mathrm{Lin}_p)$ has f as a projection with probability $1 - 1/2^n$.

Proof. For any $i = 1, \ldots, n$

$$\begin{aligned} \Pr[\rho(y_{i1}) = \ldots \rho(y_{it}) = 0] &= (1-p)^t \\ &= \frac{1}{n \cdot 2^n} \\ \implies \Pr[\exists i \ : \ \rho(y_{i1}) = \ldots \rho(y_{it}) = 0] &\leq \frac{1}{2^n} \end{aligned}$$

Hence, with probability at least $1 - 1/2^n$, for each i there is some j such that $\rho(y_{ij}) \neq 0$. Therefore, with probability at least $1 - 1/2^n$, the polynomial f is a projection of $\rho(f \circ \mathrm{Lin}_p)$. □

In all the applications, as in Lemma 8, we would have $p = 1/n^{O(1)}$. Thus, we would only incur a polynomial blow-up in the number of variables from f to $f \circ \mathrm{Lin}_p$. Hence, we can focus on proving a lower bound a homogeneous depth 4 circuit of bottom support at most r (which would eventually be something like $\sqrt{d}/100$).

Lemma 10 ([19])**.** *Let P be an n-variate degree d polynomial computed by a homogeneous depth 4 circuit of size s and bottom-support at most r. Then for any k, ℓ such that $\ell + rk \leq n/2$,*

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \leq \quad s \cdot \binom{\frac{2d}{r} + k}{k} \cdot \binom{n}{\ell + rk}.$$

The proof of this lemma is exactly along the description in of Intuition - (2): split the circuit into multiquadratic and non-multiquadratic part, and show that the non-multiquadratic part contributes no multilinear monomials. But to just put things in perspective, we shall be dealing with parameters $r = \sqrt{d}/100$, $k = \sqrt{d}$ and $\ell = \frac{n}{2}(1 - \varepsilon)$ for $\varepsilon = o(1)$. The above bound, by Lemma 5, can be seen to reduce to

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \leq \quad s \cdot \binom{n}{\ell} \cdot (1+\varepsilon)^{2rs} \cdot 2^{O(\sqrt{d})}$$

Sanity checks

Let us first check if this measure can at least in principle yield a lower bound for us. The best way to do this is to get some heuristic estimate of what we expect the measure to be for a random n-variate degree d polynomial R.

Heuristic Estimate. For a random n-variate degree d polynomial R, we expect the $\Gamma^{\mathrm{PSD}}_{k,\ell}(R)$ to be as large as it can be, i.e.

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(R) \quad \approx \quad \min\left(\binom{n}{k} \cdot \binom{n}{\ell}, \binom{n}{\ell + d - k}\right)$$

As a first step, one should first check that if we could indeed find a polynomial P for which the bound is as large as stated above, do we get a useful lower bound from Lemma 10? Turns out that if we were to choose our parameters carefully, we do indeed get the lower bound. Just to give a sense of how *careful* we need to be, here is some of the parameters that are chosen in [19, 24].

- The number of variables n is at least the cube of the degree d.
- The model we shall be working with is bottom-support r where $r = \sqrt{d}/1000$.
- The order of derivatives $k = \sqrt{d}$.
- The degree of the shift ℓ shall be chosen as $\ell = \frac{n}{2}(1-\varepsilon)$ where $\varepsilon = \frac{\log d}{c\sqrt{d}}$ for a suitable constant c.

The above choice of parameters might already seem pretty fragile but these are not the most delicate choices! While proving the lower bound on $\Gamma^{\mathrm{PSD}}_{k,\ell}$ for an explicit polynomial, the number of monomials etc. need to be tailored to perfection to make the proof work.

5.2 The surrogate rank approach of [19]

The goal is now to find an explicit polynomial P such that $\mathrm{PSD}_{k,\ell}(P)$ has large rank. One way to prove that a set of polynomials are linearly independent is to show that they have distinct leading monomials (as used [11] etc.) Another method is to show that these polynomials are *almost orthogonal*. An example of this phenomenon can be seen in the following fact.

Fact 2. *Let M be a square matrix such that the absolute value of the diagonal entry is larger than sum of the absolute values of the non-diagonal entries in that row or column, i.e. $|M_{ii}| \geq \sum_{j \neq i} |M_{ij}|$ for all i. Then the matrix M is full rank.*

Such matrices are also called *diagonally dominant matrices*, and captures the notion of *almost orthogonal* vectors alluded to earlier. For symmetric matrices M, the following bound of Alon [2].

Lemma 11 ([2])**.** *For any real symmetric matrix M,*

$$\mathrm{rank}(M) \quad \geq \quad \frac{(\mathrm{Tr}(M))^2}{\mathrm{Tr}(M^2)}$$

We'll see the proof of this shortly but it would shed some more intuition to see what the above lemma yields for a diagonally dominant matrix. Let M be a matrix of the form

$$M = \begin{bmatrix} D & d & \dots & d \\ d & D & \dots & d \\ \vdots & \vdots & \ddots & \vdots \\ d & d & \dots & D \end{bmatrix}_{r \times r}$$

Then, $\mathrm{Tr}(M) = D \cdot r$, and $\mathrm{Tr}(M^2) = (D^2 + (r-1)d^2)r = O(D^2 r + r^2 d^2)$. If $D > (r-1)d^2$, then $\mathrm{Tr}(M^2) = O(D^2 r)$. Thus, the above lemma gives that $\mathrm{rank}(M) = \Omega(r)$.

Proof. By the spectral theorem, any real symmetric matrix has a basis of eigen vectors with eigenvalues $\lambda_1, \dots, \lambda_n$ where n is the dimension of the matrix. If $\lambda_1, \dots, \lambda_r$ are the non-zero eigenvalues, then

$$\begin{aligned} \mathrm{Tr}(M) &= \sum_{i=1}^{r} \lambda_i \\ &\leq \sqrt{r} \cdot \left(\sum_{i=1}^{r} \lambda_i^2 \right) = \sqrt{r} \cdot \mathrm{Tr}(M^2) \\ \implies r &\geq \frac{(\mathrm{Tr}(M))^2}{\mathrm{Tr}(M^2)} \end{aligned}$$

□

The bound of [19] for an explicit polynomial P proceeds by considering the matrix B where each row is indexed by a pair of multilinear monomials (m_1, m_2) of degree k and ℓ respectively, and the row is just the coefficients of the monomials of $\mathrm{mult}(m_2 \partial_{m_1}(P))$ in a fixed order. Note that B is not even a square matrix, and certainly not symmetric. However, the matrix $M = BB^T$ is a symmetric square matrix such that $\mathrm{rank}(M) \leq \mathrm{rank}(B)$.

Let us spend some time understand the entries of M. The (i,j)-th entry of M is precisely the inner-product of row i and row j of B. If P is a polynomial with just zero-one coefficients, then the i-th diagonal entry is precisely the number of non-zero entries in row i of B. Thus,

$$\begin{aligned} \mathrm{Tr}(M) &= \text{number of non-zero entries in } B \\ &= (\#\text{ cols of } B) \cdot \mathop{\mathbb{E}}_{i}[\#\text{ non-zero entries in } i\text{-th col of } B] \end{aligned}$$

The calculation for $\mathrm{Tr}(M^2)$ requires a little more care. Let M_i refer to the i-th row of M and B_i refer to the i-th row of B. Then,

$$\begin{aligned}
\mathrm{Tr}(M^2) &= \sum_i \langle M_i, M_i \rangle \\
&= \sum_i \sum_j \langle B_i, B_j \rangle^2 \quad = \quad \sum_i \sum_j \left(\sum_m B_{im} B_{jm} \right)^2 \\
&= \sum_i \sum_j \sum_m B_{im}^2 B_{jm}^2 \quad + \quad \sum_i \sum_j \sum_{m \neq m'} B_{im} B_{im'} B_{jm} B_{jm'} \\
&= \sum_m \left(\sum_i \sum_j B_{im} B_{jm} \right) \quad + \quad \sum_i \sum_j \sum_{m \neq m'} B_{im} B_{im'} B_{jm} B_{jm'} \\
&= \qquad T_1 \qquad\qquad + \qquad\qquad T_2
\end{aligned}$$

The first term T_1 is easy to calculate:

$$\begin{aligned}
T_1 \quad &= \quad (\#\text{ cols of } B) \cdot \mathop{\mathbb{E}}_i [(\#\text{ non-zero entries in } i\text{-th col of } B)^2] \\
&\overset{\text{(hopefully)}}{\approx} \quad (\#\text{ cols of } B) \cdot \mathop{\mathbb{E}}_i [(\#\text{ non-zero entries in } i\text{-th col of } B)]^2
\end{aligned}$$

The term T_2 roughly corresponds to the number of 2×2 submatrices of B that is $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. If we could somehow show that there are not too many such submatrices, then $\mathrm{Tr}(M^2)$ is essentially dominated by T_1. That would then yield that $\mathrm{rank}(M) \gtrapprox (\#\text{ cols of } B)$.

Obtaining a bound on T_2:

$$T_2 \quad = \quad \sum_i \sum_j \sum_{m \neq m'} B_{im} B_{im'} B_{jm} B_{jm'}$$

Each term $B_{im} B_{im'} B_{jm} B_{jm'}$ that is non-zero corresponds to a 2×2 submatrix of B (indexed by rows i, j and columns m, m') that is $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

The columns of B are indexed by multilinear monomials of degree $\ell + d - k$, and the rows of B are indexed by a derivative and a shift. Let row i correspond to $\mathrm{mult}(\gamma_1 \cdot \partial_{\alpha_1}(P))$ and row j to $\mathrm{mult}(\gamma_1 \cdot \partial_{\alpha_1}(P))$. Thus, if the 2×2 minor indexed by rows i, j and columns m, m' equals $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then

there exists $\beta_1, \beta_2, \beta_3, \beta_4 \in P$ such that

$$\begin{aligned} m &= \frac{\beta_1}{\alpha_1} \cdot \gamma_1 = \frac{\beta_3}{\alpha_2} \cdot \gamma_2 \\ m' &= \frac{\beta_2}{\alpha_1} \cdot \gamma_1 = \frac{\beta_4}{\alpha_2} \cdot \gamma_2 \\ &\implies \frac{\beta_1}{\beta_3} = \frac{\beta_2}{\beta_4} \end{aligned}$$

Following notation used in [19], we shall call $\beta_1, \beta_2, \beta_3, \beta_4$ as the *label* of the 2×2 minor. Since $m \neq m'$, we also have that $\beta_1 \neq \beta_2$. What we'd like to say that the only way $\beta_1/\beta_3 = \beta_2/\beta_4$ is if $\beta_3 = \beta_1$ and $\beta_2 = \beta_4$. This need not be true in general of course, but this is where the choice of the polynomial comes in.

Claim 12. *If P is the $\mathrm{NW}_{d,d^3,e}$ polynomial for $e = \frac{d}{3}$ then any 2×2 minor of B (with the order of derivatives $k = o(d)$) that is $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has label $\beta_1, \beta_2, \beta_3, \beta_4$ where $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$, or $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$.*

Proof. Assume that $\beta_1 \neq \beta_3$. Then by Lemma 6 we know that they differ in at least $2d/3$ places. But then, $\beta_1/\beta_3 = \beta_2/\beta_4$ forces that β_1 and β_3 must agree at least $2d/3$ places forcing $\beta_1 = \beta_2$. □

Thus, for the NW-polynomial the number of such boxes is quite small. Using this, albeit with a reasonable amount of sweat, one can estimate T_2 to show that $T_2 = O(T_1)$. Thus, [19] obtain the following bound.

Lemma 13 ([19])**.** *For the polynomial $\mathrm{NW}_{d,d^3,e}$, for $e = \frac{d}{3}$, and $k = \sqrt{d}$ and $\ell = \frac{n}{2}\left(1 - \frac{\log d}{\sqrt{d}}\right)$ we have the bound*

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathrm{NW}_{d,d^3,e}) \geq \frac{1}{\mathsf{poly}(n,d)} \cdot \min\left(\binom{n}{\ell+d-k}, \binom{d}{k}^2 \cdot d^k \cdot k! \cdot \binom{n}{\ell}\right)$$

Note that the first term of the min in the RHS is the number of columns of B, as we had heuristically estimated. Simplifying the RHS using Lemma 5, we get

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathrm{NW}_{d,d^3,e}) \geq \frac{1}{\mathsf{poly}(n,d)} \cdot \binom{n}{\ell} \cdot \exp\left(c \cdot \varepsilon(d-k)\right)$$

for some constant $c > 0$. Since $\varepsilon = \frac{\log d}{\sqrt{d}}$, we get

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathrm{NW}_{d,d^3,e}) \geq \frac{1}{\mathsf{poly}(n,d)} \cdot \binom{n}{\ell} \cdot \exp\left(c \cdot \sqrt{d} \cdot \log d\right)$$

With the above bound and Lemma 10, we get the lower bound of [19].

Theorem 14 ([19]). *Any depth 4 homogeneous circuit of bottom support $r = \sqrt{d}/1000$ computing the polynomial $\mathrm{NW}_{d,d^3,d/3}$ over a characteristic zero field must have top fan-in $s = d^{\Omega(\sqrt{d})}$.*
In fact, more generally, any homogeneous depth 4 circuit of bottom support bounded by r computing $\mathrm{NW}_{d,m,e}$ for suitably chosen parameters must have top fanin $s = d^{\Omega(d/r)}$.

Coupling with Lemma 9, we obtain (a slight reformulation of) their main theorem.

Theorem 15 ([19]). *Any depth 4 homogeneous computing the polynomial $\mathrm{NW}_{d,d^3,d/3} \circ \mathrm{Lin}$ over a characteristic zero field must have size $s = d^{\Omega(\sqrt{d})}$.*

5.3 The leading monomial approach of [24]

Shortly after [19], a purely combinatorial proof of the result was presented by [24]. More over, they were able to prove the lower bound of $n^{\Omega(\sqrt{d})}$ for the size of any homogeneous depth 4 circuit computing $\mathrm{IMM}_{n,d}$ (for some suitable choices of n and d). This was a strengthening of [19] in two ways – (1) it worked over any field, and (2) the lower bound was for a polynomial that we know can be computed small arithmetic circuit.
The calculations of [24] are much more trickier than [19] but there are quite a few interesting ideas that would even have application in other areas.

The earlier lower bounds of [11, 25, 9] required a lower bound on the dimension of shifted partial derivatives of a polynomial P, and this was obtained by finding a *large* set of *distinct leading monomials*. In [24], 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 γ, we also have that $\beta \cdot \gamma$ is the leading monomial of γP.
>
> However, the leading monomial of $\mathrm{mult}(\gamma P)$ could be $\beta' \cdot \gamma$ for some $\beta' \neq \beta$ (as higher monomials could be made non-multilinear during the shift by γ).

The multilinear projection makes the task of counting leading monomials much harder and [24] come up with a clever method to estimate this.

Leading monomials after multilinear projections

Let P the polynomial 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

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

In other words, we want the number of distinct monomials that are contributed by β, which are also distinct leading monomials obtained from $\partial_\alpha(P)$ that are divisible by β. We then have

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \geq \quad \left| \bigcup_{\alpha,\beta} A(\alpha,\beta) \right| \tag{5}$$

The standard technique to obtain a lower bound on the union of sets is via the *Inclusion-Exclusion* principle.

Lemma 16 (Inclusion-Exclusion Principle)**.** *For any collection of sets $A_1, \ldots, A_r$,*

$$\left| \bigcup_i A_i \right| \quad \geq \quad \sum_i |A_i| \quad - \quad \sum_{i \neq j} |A_i \cap A_j|$$

If we were to somehow show that $\sum_{i \neq j} |A_i \cap A_j| \leq \frac{1}{2} \sum_i |A_i|$, then we obtain that $|\cup_i A_i| \geq \frac{1}{2} \cdot \sum_i |A_i|$. This is what shall be employed for the sets $A(\alpha,\beta)$, except that we quickly run into two immediate problems.

1. How do we even estimate $A(\alpha,\beta)$? The set of γ' such that $\gamma'\beta = \mathrm{LM}(\partial_\alpha(P))$ do not seem to have any nice combinatorial structure.
2. What if it so happens that $\sum |A(\alpha_1,\beta_1) \cap A(\alpha_2,\beta_2)| = 100 \sum |A(\alpha,\beta)|$? Inclusion-Exclusion does not yield anything in that case.

It so turns out that the second point actually is the case. In fact for $\mathrm{IMM}_{n,d}$, the second term turns out to be greater than the first term by a factor of $n^{\sqrt{d}/1000}$ or so! In [24], they prove a wonderful strengthened version of the Inclusion-Exclusion principle which allows them to handle the second hurdle.

Lemma 17 (Stronger Inclusion-Exclusion [24])**.** *Let $A_1, \ldots, A_r$ be sets such that there is some $\lambda > 1$ such that*

$$\sum_{i \neq j} |A_i \cap A_j| \quad \leq \quad \sum_i \lambda \cdot |A_i|$$

Then,

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

In other words, as long as the second term of the Inclusion-Exclusion principle is *not too much larger* than the first term, we still can get non-trivial bounds on the union.

Proof. Let $p = \frac{1}{2\lambda} < 1$. Define sets $A'_1, \ldots, A'_r$ such that $A'_i \subseteq A_i$ obtained by adding each element of A_i to A'_i independently with probability p. Since $A'_i \subseteq A_i$, we also have that $|\cup A_i| \geq |\cup A'_i|$. By linearity of expectation,

$$\mathbb{E}\left[\sum_i |A'_i|\right] \quad = \quad p \sum_i |A_i|$$

More importantly, by the sampling process,

$$\mathbb{E}\left[\left|A'_i \cap A'_j\right|\right] \quad = \quad p^2 \cdot |A_i \cap A_j|$$

as any common element must be added to both A'_i *and* A'_j, and either of these events happen independently with probability p each. Since $\sum_{i,j} \left|A'_i \cap A'_j\right|$ drops by a factor of p^2, we are now in a position to apply the Lemma 16 to the A'_is.

$$\begin{aligned}
\left|\bigcup A_i\right| &\geq \mathbb{E}\left[\left|\bigcup A'_i\right|\right] \\
&\geq \mathbb{E}\left[\sum_i |A'_i|\right] \quad - \quad \mathbb{E}\left[\left|A'_i \cap A'_j\right|\right] \\
&= p\left(\sum_i |A_i|\right) \quad - \quad p^2\left(\sum_{i \neq j} |A_i \cap A_j|\right) \\
&\geq p\left(\sum_i |A_i|\right) \quad - \quad p^2\lambda\left(\sum_i |A_i|\right) \\
&\geq \frac{p}{2}\left(\sum_i |A_i|\right) \quad = \quad \frac{1}{4\lambda}\left(\sum_i |A_i|\right)
\end{aligned}$$

□

We can now proceed to lower bound $|\bigcup A(\alpha, \beta)|$ via inclusion exclusion.

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

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

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 or IMM, for any reasonable monomial ordering. However, [24] 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.

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

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 18 ([24])**.** *For any α,*

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

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

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

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 19 ([24])**.**

$$\begin{aligned} T_1 &= (\#\ \textit{derivs}) \cdot (\#\ \textit{mons in a deriv}) \cdot \binom{n-d+k}{\ell} \\ &\approx (\#\ \textit{derivs}) \cdot (\#\ \textit{mons in a deriv}) \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{d-k} \end{aligned}$$

Let $T_1(\alpha) = \sum_\beta |S(\alpha, \beta)|$ for any choice of α. So far we have not used any property of the polynomial P. But this becomes crucial in the calculation of T_2 and T_3. To get a sense of how these calculations proceed in [24], we outline the calculation of T_2 for the case of $P = \mathrm{NW}_{d,m,e}$ for suitable choices of the parameters m, d, e.

Lemma 20 ([24])**.** *For the polynomial* $\mathrm{NW}_{d,m,e}$*, if* $n = md$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = o(1)$

$$T_2 \leq (\#\ \textit{derivs}) \cdot (\#\ \textit{mons per deriv})^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k}$$

Proof. For any fixed derivative α, define

$$T_2(\alpha) = \sum_{\beta \neq \beta'} |S(\alpha, \beta) \cap S(\alpha, \beta')| \,.$$

For any pair of multilinear degree $(d-k)$ monomials $\beta \neq \beta' \in \partial_\alpha(P)$ such that $\deg(\gcd(\beta, \beta')) = t$, we know that

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

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 $\mathrm{NW}_{d,m,e}$, any two $\beta, \beta' \in \partial_\alpha(\mathrm{NW}_{d,m,e})$ can agree on at most $e-k$ places. Further, the number of pairs that agree in exactly $0 \leq t \leq e-k$ places is at most

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

as there are m^{e-k} choices for β, and $\binom{d-k}{t}$ choices for places where they may agree, and $(m-1)^{e-t}$ choices for β' that agree with β on those t places. Thus,

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

Thus,

$$T_2 \quad \leq \quad (\#\text{ derivs}) \cdot (\#\text{ mons per deriv})^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k}$$

□

Combining this with Lemma 19 and using Lemma 17,

$$\sum_{\alpha,\beta} |A(\alpha,\beta)| \quad \geq \quad (\#\text{ derivs}) \cdot \frac{T_1(\alpha)}{\max(2, \frac{4T_2(\alpha)}{T_1(\alpha)})}$$

To maximize this, if we choose the parameters m, d, e such that $T_1(\alpha) = T_2(\alpha)$, we obtain the following corollary.

Corollary 21. *Consider the polynomial* $\mathrm{NW}_{d,m,e}$ *with* $n = md$ *and* $m = \Omega(d)$. *If* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = o(1)$ *and* e *chosen so that*

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

then

$$\sum_{\alpha,\beta} |A(\alpha,\beta)| \quad \geq \quad (\#\ \textit{derivs}) \cdot \binom{n}{\ell} \cdot 2^{\Theta(\sqrt{d})}$$

Proof. If $T_1(\alpha) = T_2(\alpha) \cdot 2^{-\Theta(\sqrt{d})}$ then

$$\begin{aligned}\sum_{\alpha,\beta} |A(\alpha,\beta)| &\geq (\#\text{ derivs}) \cdot \frac{T_1(\alpha)}{\max(2, \frac{4T_2(\alpha)}{T_1(\alpha)})} \\ &= (\#\text{ derivs}) \cdot \frac{T_1(\alpha)^2}{4T_2(\alpha)} \\ &= (\#\text{ derivs}) \cdot \binom{n}{\ell} \cdot 2^{\Theta(\sqrt{n})}\end{aligned}$$

Note that $T_1(\alpha) = T_2(\alpha) \cdot 2^{-\Theta(\sqrt{d})}$ forces

$$(\#\text{ mon per deriv}) \quad = \quad m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

□

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!
In the case of IMM this calculations gets a lot messier. The calculation would similarly force that the number of monomials must be in a very narrow range. This is achieved by instead looking at a random subgraph of the generic ABP of suitable sparsity to ensure the following two properties:

- The number of monomials in any derivative is exactly as demanded.
- 'Most' pairs of monomials (β, β') agree on 'few' places.

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

We are still left with the task of upper bounding

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

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 γ', γ'' of degree ℓ satisfying

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

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

Observation 22. *If* $\deg(\gcd(\beta,\beta')) = t$, *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

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

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 $\deg(\gcd(\beta,\beta')) = t$ and sum them up accordingly. We shall just state the bound of [24] here without proof.

Lemma 23 ([24])**.** *For the polynomial* $\mathrm{NW}_{d,m,e}$, *and* $n = md$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = o(1)$,

$$T_3 \quad \leq \quad (\#\ \textit{deriv})^2 (\#\ \textit{mons per deriv})^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1}{2}\right)^{2d-2k}$$

Recalling that we have chosen our parameters so that

$$(\#\ \text{mons per deriv}) = \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

the above equation reduces to

$$T_3 \quad \leq \quad (\#\ \text{deriv})^2 \left(\frac{1}{1+\varepsilon}\right)^{2(d-k)} \cdot \binom{n}{\ell}.$$

Combining with Corollary 21, we obtain the required bound for $|\bigcup A(\alpha,\beta)|$.

Lemma 24. *Consider polynomial* $\mathrm{NW}_{d,m,e}$ *where* $n = md$ *and* e *chosen so that*

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

If $\varepsilon = \frac{\log d}{c\sqrt{d}}$ *for a large enough constant* c, *and* $k = O(\sqrt{d})$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$, *then*

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

With Lemma 10, we obtain the lower bound for low-bottom-support homogeneous depth 4 circuits.

Theorem 25 ([24]). *Any homogeneous depth 4 circuit with bottom support bounded by $r = \sqrt{d}/1000$ computing, over any field $\mathbb{F}$, the polynomial $\mathrm{NW}_{d,m,e}$ with parameters as defined above must have top fan-in $s = d^{\Omega(\sqrt{d})}$.*
In fact, more generally, any homogeneous depth 4 circuit of bottom support bounded by r computing $\mathrm{NW}_{d,m,e}$ for suitably chosen parameters must have top fanin $s = d^{\Omega(d/r)}$.

Again, coupling with Lemma 9, we obtain (a slight reformulation of) their theorem.

Theorem 26 ([19]). *Any homogeneous depth 4 circuit computing, over any field $\mathbb{F}$, the polynomial $\mathrm{NW}_{d,m,e} \circ \mathrm{Lin}$ with parameters as defined above must have top fan-in $s = d^{\Omega(\sqrt{d})}$.*
A similar lower bound $d^{\Omega(\sqrt{d})}$ holds also for the polynomial $\mathrm{IMM}_{n,d} \circ \mathrm{Lin}$ for suitable choices of n and d.

6 Non-homogeneous depth 3 circuits

In a very recent result, [21] show that similar techniques can also be used to prove lower bounds for subclasses of non-homogeneous depth three circuits, namely depth three circuits with *bounded bottom fan-in*. We shall denote the class of depth three circuits of bottom fan-in bounded by r as $\Sigma\Pi\Sigma^{[r]}$ circuits.
But before we see this lower bound, let us first understand the computational power of depth three circuits, and the depth reduction of [12].

6.1 Computational power of depth three circuits

A $\Sigma\Pi\Sigma$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_{i=1}^{s} \ell_{i1} \dots \ell_{iD}$$

If the circuit is non-homogeneous, the degree of the circuit D could potentially be much larger than $\deg(f)$.

The class of depth three arithmetic circuits can compute polynomials in non-trivial ways. To illustrate a couple of examples, there is a homogeneous $\Sigma\Pi\Sigma$

circuit for Perm_n of size $2^{O(n)}$ called Ryser's Formula [31]

$$\mathsf{Perm}_n \quad = \quad \sum_{S \subseteq [n]} (-1)^{n-|S|} \prod_{i=1}^{n} \left(\sum_{j \in S} x_{ij} \right) \tag{6}$$

On the other hand, no $\Sigma\Pi\Sigma$ circuit for the Det_n significantly better than writing it as a sum of $n!$ monomials was known (until [12]). Further, the elementary symmetric polynomials $\mathrm{Esym}_k(x_1, \ldots, x_n)$ of degree k defined as

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\substack{S \subset \mathbf{x} \\ |S|=k}} \prod_{x_i \in S} x_i$$

can be computed by a non-homogeneous depth 3 circuit of size $O(n^2)$ over any characteristic zero field. In stark contrast, [27] showed that any homogeneous depth 3 circuit computing Esym_k requires size $n^{\Omega(k)}$. [27] also showed a $2^{\Omega(n)}$ lower bound for homogeneous depth 3 circuits computing Perm_n or Det_n. Also, the results of [13, 10] showed a $2^{\Omega(n)}$ lower bound for $\Sigma\Pi\Sigma$ circuits *over finite fields* that compute Det_n or Perm_n. All these results seemed to suggest that there perhaps is an $2^{\Omega(n)}$ lower bound for $\Sigma\Pi\Sigma$ circuits computing Det_n over characteristic zero fields as well. If it was true over finite fields, and for homogeneous $\Sigma\Pi\Sigma$ circuits, how much power can characteristic zero fields and non-homogeneity add? As it turns out, quite a lot!

Theorem 27 ([12]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s over any characteristic zero field. Then there is a $\Sigma\Pi\Sigma$ circuit of size $s' \leq s^{O(\sqrt{d})}$ that computes f.*

Corollary 28 ([12]). *There is a $\Sigma\Pi\Sigma$ circuit over $\mathbb{Q}$, the field of rational numbers, of size $n^{O(\sqrt{n})}$.*

The proof is quite short and comprises of two steps using known reductions, and going through a bizarre intermediate model of *depth* 5 *powering circuits*. Simply presenting the proof step-by-step would rob the readers of the intuition as to why one would study depth 5 powering circuits. This result was really a bi-product of an attempt to prove a stronger lower bound for depth 4 circuits. We believe this perspective, albeit lengthier, is more insightful than seeing the proof directly.

6.1.1 Towards proving better lower bounds for depth 4 circuits

From Theorem 3, it suffices to prove a better lower bound for explicit polynomials computed as

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \ldots Q_{ir} \quad \text{where} \quad \deg(Q_{ij}) \leq \sqrt{d}\,, \; r \leq O(\sqrt{d}) \tag{7}$$

The goal is to show a lower bound of $s = n^{\omega(\sqrt{d})}$. Perhaps a simpler question to ask is to prove a lower bound for expressions of the form

$$f \quad = \quad \sum_{i=1}^{s} Q_i^{\sqrt{d}} \quad \text{where } \deg(Q_i) \leq \sqrt{d} \tag{8}$$

Fortunately, if the goal is to prove lower bounds of $n^{\omega(\sqrt{d})}$, then without loss of generality we can focus on this equation instead!

Lemma 29. *Over any characteristic zero field, given an expression of the form*

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ir} \quad \textit{where} \quad \deg(Q_{ij}) \leq \sqrt{d}\ ,\ r \leq O(\sqrt{d})$$

there is an equivalent equation

$$f \quad = \quad \sum_{i=1}^{s'} Q_i^r \quad \textit{where} \quad \deg(Q_i) \leq \sqrt{d}$$

with $s' \leq s \cdot 2^{O(\sqrt{r})}$.

Proof. Consider Ryser's formula (6) applied for to the $r \times r$ matrix where each row is $[y_1, \dots, y_r]$.

$$\mathsf{Perm} \begin{bmatrix} y_1 & \dots & y_r \\ \vdots & \ddots & \vdots \\ y_1 & \dots & y_r \end{bmatrix} \quad = \quad r! \cdot y_1 \dots y_r \quad = \quad \sum_{S \subseteq [r]} \left(\sum_{j \in S} y_j \right)^n$$

This specific identity is often attributed to Fischer [8]. The lemma follows by applying this identity on each term $Q_{i1} \dots Q_{ir}$. □

Note that since we need to divide by $r!$, the above lemma fails over low characteristic fields, in particular finite fields. Thus, proving an $n^{\omega(\sqrt{d})}$ lower bound for expressions such as (8) implies an $n^{\omega(\sqrt{d})}$ lower bound for expressions such as (7). We shall call expressions such as (8) as $\Sigma\wedge\Sigma\Pi^{[\sqrt{d}]}$ circuits. Just as we converted the top Π layer into powering layers using Fischer's identity, the same can be done to the lower layer of Π gates as well.

Corollary 30. *If a homogeneous n-variate degree d polynomial f can be computed by a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ of size $s = n^{O(\sqrt{d})}$, then f can also be computed by an $\Sigma\wedge^{[O(\sqrt{d})]}\Sigma\wedge^{[\sqrt{d}]}\Sigma$ circuit of size $s' = s \cdot 2^{O(\sqrt{d})}$.*
Conversely, if f requires $\Sigma\wedge^{[O(\sqrt{d})]}\Sigma\wedge^{[\sqrt{d}]}\Sigma$ circuits of size $s' = n^{\omega(\sqrt{d})}$ to compute it, then f cannot be computed by polynomial sized arithmetic circuits.

We shall take a small detour to see if non-homogeneous depth 3 circuits can be converted to homogeneous shallow circuits without much blow-up in size.

6.1.2 Non-homogeneous depth 3 to homogeneous depth 5 circuits

Let f be a homogeneous degree d polynomial computed by a possibly non-homogeneous depth 3 circuit C of the form

$$f \quad = \quad \sum_{i=1}^{s} \ell_{i1} \dots \ell_{iD}$$

As a first step, let us extract the degree d homogeneous component of each summand on the RHS. Since f is a homogeneous degree d polynomial, f has to be sum of the degree d homogeneous components of each summand on the RHS. Consider a single term of the form

$$T \quad = \quad (\ell_1 + \alpha_1) \cdots (\ell_D + \alpha_D)$$

where each ℓ_i is a homogeneous linear polynomial, and α are elements from the field. Assuming that the first r of the α_i's are zero, we can write T in the form (with some reuse of symbols)

$$\begin{aligned} T &= \alpha \cdot \ell_1 \dots \ell_r \cdot (\ell_{r+1} + 1) \dots (\ell_D + 1) \\ \implies \quad \mathrm{Hom}_d(T) &= \ell_1 \dots \ell_r \cdot \mathrm{Esym}_{d-r}(\ell_{r+1}, \dots, \ell_D) \end{aligned}$$

where $\mathrm{Esym}_k(\mathbf{x})$, the elementary symmetric polynomial of degree k defined as

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\substack{S \subset \mathbf{x} \\ |S|=k}} \prod_{x_i \in S} x_i$$

Hence, if we can show that $\mathrm{Esym}_{d-r}(\mathbf{x})$ has a not-too-large homogeneous depth 4 circuit, then we can immediately infer that f can be computed by a not-too-large homogeneous depth 5 circuit. The following identities, attributed to Newton (cf. [26]), is exactly what we need. Define the *power symmetric polynomials*, denoted by $\mathrm{Pow}_k(\mathbf{x})$ as

$$\mathrm{Pow}_k(\mathbf{x}) \quad = \quad \sum_{x_i \in \mathbf{x}} x_i^k$$

Lemma 31 (Newton Identities)**.** *Let* $\mathrm{Esym}_k(x_1, \dots, x_m)$ *and* $\mathrm{Pow}_k(x_1, \dots, x_m)$ *denote the* elementary symmetric *and* power symmetric *polynomials of degree* k *respectively, as defined above. Then,*

$$\mathrm{Esym}_k \quad = \quad \frac{1}{k!} \cdot \begin{vmatrix} \mathrm{Pow}_1 & 1 & 0 & \cdots & \\ \mathrm{Pow}_2 & \mathrm{Pow}_1 & 2 & 0 & \cdots \\ \vdots & & \ddots & \ddots & \\ \mathrm{Pow}_{k-1} & \mathrm{Pow}_{k-2} & \cdots & \mathrm{Pow}_1 & k-1 \\ \mathrm{Pow}_k & \mathrm{Pow}_{k-1} & \cdots & \mathrm{Pow}_2 & \mathrm{Pow}_1 \end{vmatrix}.$$

Expanding the determinant on the RHS, we obtain a homogeneous expression

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\mathbf{a}\,:\,\sum_i ia_i = k} \alpha_{\mathbf{a}} \cdot (\mathrm{Pow}_1)^{a_1} \ldots (\mathrm{Pow}_k)^{a_k} \tag{9}$$

The number of summands bounded by the number of non-negative solutions to $\sum ia_i = k$, which is precisely the number of partitions of k. By the estimates of [14], we know that the number of partitions of k is bounded by $2^{\Theta(\sqrt{k})}$. Thus, (9) yields a homogeneous depth 4 circuit for $\mathrm{Esym}_k(x_1, \ldots, x_m)$ of size $2^{\Theta(\sqrt{k})} \cdot m$. In fact, the circuit is a homogeneous $\Sigma\Pi\Sigma\wedge$ circuit, i.e. a $\Sigma\Pi\Sigma\Pi$ circuit where the bottom layer of multiplication in fact just raises a single variable to a higher power.

Corollary 32. *Let T be a product of D linear polynomials over n variables, not necessarily homogeneous. Then, the degree d homogeneous component of T, denoted by $\mathrm{Hom}_d(T)$ can be computed by a homogeneous $\Sigma\Pi\Sigma\wedge$ circuit of size $nD \cdot 2^{O(\sqrt{d})}$.*
Hence, if f is a homogeneous degree d polynomial over n variables that is computed by a non-homogeneous depth 3 circuit C of size s, then f can be computed by a homogeneous $\Sigma\Pi\Sigma\wedge\Sigma$ circuit of size $\mathsf{poly}(ns) \cdot 2^{O(\sqrt{d})}$.

To convert the $\Sigma\Pi\Sigma\wedge\Sigma$ circuit to a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit, we could use Fischer's identity again. At first sight, it appears as though this would yield a blow up of 2^d as some of the product gates could have fan-in d. However, notice that the sum is over a_i's satisfying $\sum i \cdot a_i = d$. Hence, there can be at most $O(\sqrt{d})$ of the a_i's that are non-zero. By looking at Fischer's identity applied on $y_1^{a_1} \ldots y_d^{a_d}$ more carefully, we see that it uses at most $(1+a_1)\ldots(1+a_d) \leq d^{O(\sqrt{d})}$ distinct linear powers instead of the naïve bound of 2^d. This fact of expressing any degree d monomial over m variables as a $\Sigma\wedge\Sigma$ circuit of size $d^{O(m)}$ was also observed by Ellison [7].
Thus, if f admits a poly-sized depth three circuit, then f also admits a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $d^{O(\sqrt{d})} \cdot \mathsf{poly}(n)$. The following lemma summarizes this discussion.

Lemma 33. *Let f be an n-variate degree d polynomial that is computable by depth three circuit of size s over $\mathbb{Q}$. Then, f is equivalently computable by a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $d^{O(\sqrt{d})} \cdot \mathsf{poly}(s)$.*
Conversely, if f requires $\Sigma\wedge\Sigma\wedge\Sigma$ circuits of size $n^{\omega(\sqrt{d})}$ over $\mathbb{Q}$ to compute it, then f requires depth three circuits of size $n^{\omega(\sqrt{d})}$.

6.1.3 Completing the picture

We now have an interesting situation (Figure 1). On one hand, Corollary 30 states that a lower bound of $n^{\omega(\sqrt{d})}$ for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits would yield a

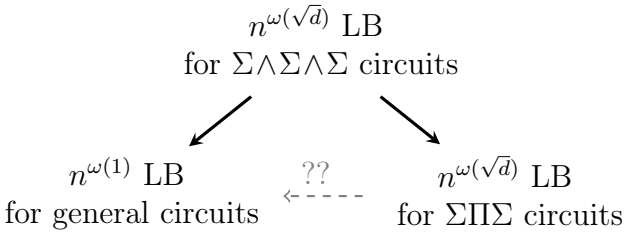

Figure 1: Power of $\Sigma\wedge\Sigma\wedge\Sigma$ ckts.

super-polynomial lower bound for general arithmetic circuits. On the other, Lemma 33 states that an $n^{\omega(\sqrt{d})}$ lower bound for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits would yield a lower bound of $n^{\omega(\sqrt{d})}$ for depth three circuits.
Could this just be a coincidence? Or, is it the case that any poly-sized arithmetic circuit can be equivalently expressed as a depth three circuit of size $n^{O(\sqrt{d})}$ over $\mathbb{Q}$? As it turns out, there is indeed a depth reduction to convert any arithmetic circuit to a not-too-large depth three circuit over $\mathbb{Q}$.

To complete the picture, it suffices to show that a $\wedge\Sigma\wedge$ circuit can be expressed as a $\Sigma\Pi\Sigma$ circuit. This would automatically imply a reduction from $\Sigma\wedge\Sigma\wedge\Sigma$ circuits to $\Sigma\Pi\Sigma$ circuits. The last step of the puzzle is the *duality trick* of [32].

Lemma 34 (The Duality Trick [32])**.** *There exists univariate polynomials f_{ij}'s of degree at most b such that*

$$(z_1 + \cdots + z_s)^b \quad = \quad \sum_{i=1}^{sb+1} f_{i1}(z_1) \ldots f_{is}(z_s).$$

It is worth noting that the degree of each term on the RHS is sb, whereas the LHS just has degree b. This is the place where non-homogeneity is introduced. Applying the above lemma for a $\wedge\Sigma\wedge$ circuit such as $(y_1^a + \cdots + y_s^a)^b$ gives

$$\begin{aligned}(y_1^a + \cdots + y_s^a)^b &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} f_{ij}(y_j^a) \\ &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} \tilde{f}_{ij}(y_j)\end{aligned}$$

where $\tilde{f}_{ij}(y) = f_{ij}(y^a)$. Since each $\tilde{f}_{ij}(y)$ is a univariate polynomial, it can be factorized completely over the $\mathbb{C}$, the field of complex numbers. Hence, if $f_{ij}(y) = \prod_k (y - \zeta_{ijk})$, then we get

$$\begin{aligned}(y_1^a + \cdots + y_s^a)^b &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} \tilde{f}_{ij}(y_j) \\ &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} \prod_{k=1}^{b} (y_j - \zeta_{ijk})\end{aligned}$$

which is a depth three circuit! Thus, $(y_1^a + \cdots + y_s^a)$ can be expressed as a depth three circuit of size $\mathsf{poly}(s, a, b)$ over $\mathbb{C}$. With a little more effort, one can construct a depth three circuit over $\mathbb{Q}$ as well. Summarizing this is a lemma, we have the following.

Lemma 35. *Any n-variate degree d polynomial f computed by a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ of size s over a characteristic zero field $\mathbb{F}$ can also be computed by a depth three circuit of size $\mathsf{poly}(s, n, d)$ over $\mathbb{F}$.*

Combining with Corollary 30 and Theorem 3, we obtain the main result of [12].

Theorem 27 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s over any characteristic zero field. Then there is a $\Sigma\Pi\Sigma$ circuit of size $s' \leq s^{O(\sqrt{d})}$ that computes f.*

Remark. Note that if we were to start with a degree d polynomial f and apply the above depth reduction, all the linear polynomials that we obtain at bottom are essentially from the application of Fischer's identity on the bottom Π layer of fanin $\sqrt{d}$ of the $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit. Hence, the each of the linear polynomials that appear in the final $\Sigma\Pi\Sigma$ circuit depend on at most $\sqrt{d}$ variables. In other words, the above Theorem yields a reduction to $\Sigma\Pi\Sigma^{[\sqrt{d}]}$ circuits.

6.2 Lower bounds for $\Sigma\Pi\Sigma$ circuits with small bottom fan-in

Now let us focus on $\Sigma\Pi\Sigma^{[r]}$ circuits, where all linear polynomials in the circuit depend on at most r variables. The following is the key observation of [21] and can be verified easily.

Observation 36 ([21]). *Starting with a $\Sigma\Pi\Sigma^{[r]}$ circuit C of size s computing a homogeneous n-variate polynomial of degree d, the resulting $\Sigma\Pi\Sigma\wedge\Sigma$ circuit C' obtained from Corollary 32 is in fact a $\Sigma\Pi\Sigma\wedge\Sigma^{[r]}$ circuit of size $s' = \mathsf{poly}(ns)\cdot 2^{O(\sqrt{d})}$.*
Thus, by expanding the all powers of linear polynomials computed in the bottom two layers of the $\Sigma\Pi\Sigma\wedge\Sigma$ circuit C', the circuit C' can be rewritten as a homogeneous depth 4 circuit of bottom support bounded by r and size $s'' = s'\cdot d^r$

This observation in combination with Theorem 14 immediately yields the main theorem of [21].

Theorem 37 ([21]). *Over any characteristic zero field $\mathbb{F}$, any $\Sigma\Pi\Sigma^{[r]}$ circuit C computing the polynomial $\mathrm{IMM}_{n,d}$, for suitably chosen parameters n and d with $n = d^{O(1)}$, must have size $s = n^{\Omega(d/r)}$.*

6.3 Extensions to low-bottom-fanin depth 5 circuits

[21] also prove lower bounds for depth 5 circuits where the bottom fan-in is bounded. The result proceeds by analysing the random restriction process carefully to decompose any $\Sigma\Pi\Sigma\Pi\Sigma^{[\tau]}$ circuit into a $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit and another circuit C' such that $\Gamma^{[\mathrm{PSD}]}_{k,\ell}(C') = 0$. We just state their theorem here without proof.

Theorem 38. *Let $\mathbb{F}$ be a field of characteristic zero, and let $0 \le \mu < 1$. If $\alpha = \frac{2\mu+1}{1-\mu}$ and $\tau = O(N^{\mu})$, then there is a family of n-variate degree d polynomials $\{f_n\}$ in VNP with $n \in [d^{2+\alpha}, 2d^{2+\alpha}]$ such that any homogeneous $\Sigma\Pi\Sigma\Pi\Sigma^{[\tau]}$ circuit computing this polynomial requires size $n^{\Omega(\sqrt{d})}$.*

7 Speculation about lower bounds for homogeneous formulas

In this section, we shall look at a possible approach to proving an $n^{\Omega(\log n)}$ lower bound for homogeneous formulas. It is conceivable that variants of the dimension of shifted partial derivatives would be able to give such a lower bound. We will not be presenting any candidate measures, but would instead present a normal form that could perhaps be useful.

7.1 A stronger(?) depth reduction for homogeneous formulas

If we were to prove a lower bound for homogeneous formulas via a depth reduction to $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits, the first step would be to answer the following question:

> Suppose we apply Theorem 3 to a polynomial sized circuit C to obtain a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit C', and also apply Theorem 3 to a polynomial sized homogeneous formulas $\tilde{C}$ to obtain a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit $\tilde{C}'$, how is $\tilde{C}'$ structurally different from C'?

Unless we are able to find a non-trivial structural difference between $\tilde{C}'$ and C', it does not make sense to attempt proving lower bounds for homogeneous formulas via $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits. The original proof of [35] of Theorem 3 does not seem to suggest any structural difference between the two. However, the alternate proof described in Section 3 allows one to understand this difference better.

Recall how the alternate proof proceeded. If g is the polynomial computed by any gate in the circuit C of size s, then g can be written as

$$g \quad = \quad \sum_{i=1}^{\mathsf{poly}(s)} g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5}$$

where $\sum_j \deg(g_{ij}) = \deg(g)$ for all i and $\deg(g_{ij}) \leq \deg(g)/2$. Further g_{ij}s are polynomials computed by the children/grandchildren of g. With an equation as above, we could recursively keep expanding the larger degree g_{ij}'s to eventually get all degrees to be less than $\sqrt{d}$.
For homogeneous formulas, we can start with a slightly more structured equation instead of the one above. This more structured decomposition was first observed by [15].

Lemma 39 ([15]). *Let Φ be a homogeneous formula of size s. If f is a polynomial computed at an arbitrary gate of f, then f can be written as*

$$f \quad = \quad \sum_{i=1}^{s} f_{i1} \cdot f_{i2} \cdot \cdots \cdot f_{i\ell} \tag{10}$$

where

- *Each f_{ij} is computable by a homogeneous formula of size at most s*

- $\sum_j \deg(f_{ij}) = \deg(f)$ *for all* i
- $\left(\frac{1}{3}\right)^i \deg(f) \leq \deg(f_{ij}) \leq \left(\frac{2}{3}\right)^i \deg(f)$ *for all* i, j.
- $\deg(f_{i\ell}) = 1$ *for all* i.

Proof. Assume that the Φ is a formula of fan-in 2 at each gate. This would only increase the depth by a polynomial factor. Starting from the root, walk down the tree by always picking the child of largest degree until we hit a node v of degree at most $\frac{2\deg(f)}{3}$ for the first time. Since the path always picked the child of largest degree, we must have that

$$\frac{\deg(f)}{3} \quad \leq \quad \deg(v) \quad \leq \quad \frac{2\deg(f)}{3}$$

Let Φ_v denote the sub-formula rooted at v, and let $\Phi_{v=0}$ refer to the formula obtained from Φ by replacing the sub-tree rooted at v by 0. Let s_1 and s_2 be the size of Φ_v and $\Phi_{v=0}$ respectively. (We shall abuse notation and also use Φ_v and $\Phi_{v=0}$ to refer to the polynomial computed by these formulas.) Then,

$$f \quad = \quad \Phi_v \cdot A \quad + \quad \Phi_{v=0}$$

for some polynomial A. Note that homogeneity implies that $\deg(A)+\deg(\Phi_v) = \deg(f)$ and hence $\frac{\deg(f)}{3} \leq \deg(A) \leq \frac{2\deg(f)}{3}$. ($A$ is going to play the role of f_{i1} for some of the i's.)
Observe that the formulas Φ_v and $\Phi_{v=0}$ 'partition' the formula Φ and hence $s_1 + s_2 \leq s$. By induction on these smaller formulas, we can write

$$\begin{aligned} \Phi_v &= \sum_{i=1}^{s_1} g_{i1} \cdots g_{i\ell} \\ \Phi_{v=0} &= \sum_{i=1}^{s_2} h_{i1} \cdots h_{i\ell} \end{aligned}$$

satisfying the necessary conditions. Since $\frac{\deg(f)}{3} \leq \deg(\Phi_v) \leq \frac{2\deg(f)}{3}$, we have that

$$f \quad = \quad \sum_{i=1}^{s_1} A \cdot g_{i1} \cdots g_{i\ell} \quad + \quad \sum_{i=1}^{s_2} h_{i1} \cdots h_{i\ell}$$

satisfies all the degree conditions with A as claimed. To complete the proof, it suffices to show that A can be computed by a homogeneous formula of size s. Indeed, the polynomial A is just the product of all siblings of multiplication gates encountered in the path from v to the root. Since each of the siblings are disjoint sub-formulas of Φ, the polynomial A is computable by a homogeneous formula of size at most s. □

With equation (10) instead, we can repeat the strategy we used to prove Theorem 3.

> Start with (10) for the root of the homogeneous formula.
>
> For each summand $f_{i1} \dots f_{ir}$ in the RHS, if the largest degree f_{ij} has degree more than $\sqrt{d}$, expand that f_{ij} with the its corresponding representation using Lemma 39.
>
> Repeat this process until all f_{ij}'s on the RHS have degree at most $\sqrt{d}$.

Again, in the expansion of f of degree d via Lemma 39, every term on the LHS has at least two factors of degree more than $d/9$. The same proof would then yield a $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit of top fan-in at most $s^{O(\sqrt{d})}$. Did we gain anything with this? We certainly did – observe that every expansion via Lemma 39 yields $O(\log d)$ more factors in each term. In the case of Theorem 3, we gained only constantly many factors at each term. Thus, in the resulting depth 4 circuit has the form

$$f \quad = \quad \sum_{i=1}^{s^{O(\sqrt{d})}} Q_{i1} \dots Q_{ir} \quad , \quad \text{where } 1 \leq \deg(Q_{ij}) \leq \sqrt{d}$$

and, most importantly, $r = O(\sqrt{d} \log d)$ as opposed to $O(\sqrt{d})$ in the case of Theorem 3. This seems to be a key structural difference between depth 4 circuits obtained from homogeneous formulas as opposed to depth 4 circuits obtained from general arithmetic circuits! We summarize this below.

Theorem 40. *If f is an n-variate degree d polynomial computed by a homogeneous formula of size s, then there is a homogeneous $\Sigma\Pi^{[O(\sqrt{d}\log d)]}\Sigma\Pi^{[\sqrt{d}]}$ circuit computing f with top fanin at most $s^{O(\sqrt{d})}$.*

Is this useful?

It is not clear if the above structural difference can be exploited to give a complexity measure. But it is very much possible that the some small modification of measure of dimension of shifted partials derivatives might be a measure that works. The reason we believe that is because the results of [11, 25, 23] give explicit n-variate degree d polynomials that admit a top fan-in lower bound of $n^{\Omega(d/t)}$ for depth 4 circuits with *maximum bottom degree* bounded by t.

Question. Could that be changed to give an $n^{\Omega(d/t)}$ lower bound for depth 4 circuits with *average bottom degree* bounded by t?

If this were true, then we obtain an $n^{\Omega(\log d)}$ lower bound for the size of homogeneous formulas computing an explicit n-variate degree d polynomial. Also, we need to keep in mind that any circuit of polynomial size has an equivalent homogeneous formula of size $n^{O(\log d)}$. Hence, if we are hoping to come up with a method that might prove a lower bound for homogeneous formulas but not for general circuits, then method should not be able to yield a lower bound better than $n^{\Omega(\log d)}$. This indeed seems to be the case in this approach. Maybe this is the right depth reduction to work with to prove lower bounds for homogeneous formulas, maybe not. Either way, we shall probably find out soon enough!

8 Conclusion

Quite a lot seems to be happening lately in arithmetic circuits. The last few results were on $n^{\Omega(\sqrt{d})}$ lower bounds for homogeneous depth 4, non-homogeneous depth 3 circuits with small bottom fanin, and homogeneous depth 5 with small bottom fanin. Perhaps in the near future, we would be able to obtain $n^{\Omega(\sqrt{d})}$ lower bounds for non-homogeneous depth 3 or homogeneous depth 5 circuits without any bottom fanin restrictions. These are all interesting problems to work on, and should be very much within reach of current techniques. However, it is to be noted that if we wish to separate VP and VNP, we need to break past $n^{\Omega(\sqrt{d})}$. It appears (at least to us) this task would require very different techniques and seems unlikely that small variants of shifted partial derivatives might just get us past $n^{\Omega(\sqrt{d})}$. Nevertheless, Open Problem 1 presents a concrete and extremely simple looking model to work with, for which we need to prove an $n^{\omega(\sqrt{d})}$ lower bound to separate VP and VNP. We conclude by stating the problem again to emphasize the point.

Open Problem. *Find an explicit n-variate degree d polynomial f such that any expression of the form*

$$f \quad = \quad (Q_1)^{\sqrt{d}} + \cdots + (Q_s)^{\sqrt{d}} \quad , \quad \deg(Q_i) \leq \sqrt{d} \text{ for all } i$$

must have $s = n^{\omega(\sqrt{d})}$.

References

[1] Manindra Agrawal. Proving Lower Bounds Via Pseudo-random Generators. In *Foundations of Software Technology and Theoretical Computer Science Science (FSTTCS)*, pages 92–105, 2005.

[2] Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability and Computing*, 18(1-2):3–15, March 2009.

[3] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Foundations of Computer Science (FOCS)*, pages 67–75, 2008.

[4] Peter Bürgisser, Michael Clausen, and Mohammad A. Shokrollahi. *Algebraic Complexity Theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1997.

[5] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[6] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial Derivatives in Arithmetic Complexity (and beyond). *Foundation and Trends in Theoretical Computer Science*, 2011.

[7] W. J. Ellison. A 'waring's problem' for homogeneous forms. *Proceedings of the Cambridge Philosophical Society*, 65:663–672, 1969.

[8] I. Fischer. Sums of like powers of multivariate linear forms. *Mathematics Magazine*, 67(1):59–61, 1994.

[9] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:100, 2013.

[10] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Symposium on Theory of Computing (STOC)*, pages 577–582, 1998.

[11] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Conference on Computational Complexity (CCC)*, 2013.

[12] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth Three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013.

[13] 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.

[14] G. H. Hardy and S. Ramanujan. Asymptotic formulaæ in combinatory analysis. *Proceedings of the London Mathematical Society*, s2-17(1):75–115, 1918.

[15] Pavel Hrubes and Amir Yehudayoff. Homogeneous formulas and symmetric polynomials. *Computational Complexity*, 20(3):559–578, 2011.

[16] Mark Jerrum and Marc Snir. Some exact complexity results for straight-line computations over semirings. *Journal of the ACM*, 29(3):874–897, 1982.

[17] Kyriakos Kalorkoti. A Lower Bound for the Formula Size of Rational Functions. *SIAM Journal of Computing*, 14(3):678–687, 1985.

[18] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

[19] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In *Foundations of Computer Science (FOCS)*, 2014.

[20] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65, 2012.

[21] Neeraj Kayal and Chandan Saha. Lower Bounds for Depth Three Arithmetic Circuits with small bottom fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 2014.

[22] Neeraj Kayal and Ramprasad Saptharishi. A selection of lower bounds for arithmetic circuits. In Manindra Agrawal and V Arvind, editors, *Perspectives in Computational Complexity*. "Birkhäuser", Basel, 2014.

[23] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 136–145, 2014.

[24] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Foundations of Computer Science (FOCS)*, 2014.

[25] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:91, 2013.

[26] D.E. Littlewood. *The Theory of Group Characters and Matrix Representations of Groups*. Ams Chelsea Publishing. AMS Chelsea Pub., 2nd edition, 1950.

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

[28] R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *Journal of the ACM*, 56(2), 2009.

[29] Ran Raz. How to fool people to work on circuit lower bounds. Invited talk at Microsoft Research, 2010.

[30] Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.

[31] H. J. Ryser. Combinatorial mathematics. *Math. Assoc. of America*, 14, 1963.

[32] Nitin Saxena. Diagonal Circuit Identity Testing and Lower Bounds. In *ICALP (1)*, pages 60–71, 2008.

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

[34] Amir Shpilka and Amir Yehudayoff. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[35] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Mathematical Foundations of Computer Science (MFCS)*, pages 813–824, 2013.

[36] Leslie G. Valiant. Completeness Classes in Algebra. In *Symposium on Theory of Computing (STOC)*, pages 249–261, 1979.

[37] 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.