On $\Sigma \wedge \Sigma \wedge \Sigma$ Circuits: The Role of Middle Σ Fan-in, Homogeneity and Bottom Degree.

Christian Engels[1], B. V. Raghavendra Rao[2], and Karteek Sreenivasaiah[3]*

[1] Kyoto University, Kyoto, Japan, christian.engels@gmail.com
[2] IIT Madras, Chennai, India, bvrr@cse.iitm.ac.in
[3] Saarland University, Saarbrücken, Germany, karteek@mpi-inf.mpg.de

Abstract. We study polynomials computed by depth five $\Sigma \wedge \Sigma \wedge \Sigma$ arithmetic circuits where 'Σ' and '$\wedge$' represent gates that compute sum and power of their inputs respectively. Such circuits compute polynomials of the form $\sum_{i=1}^{t} Q_i^{\alpha_i}$, where $Q_i = \sum_{j=1}^{r_i} \ell_{ij}^{d_{ij}}$ where ℓ_{ij} are linear forms and r_i, α_i, $t > 0$. These circuits are a natural generalization of the well known class of $\Sigma \wedge \Sigma$ circuits and received significant attention recently. We prove an exponential lower bound for the monomial $x_1 \cdots x_n$ against depth five $\Sigma \wedge \Sigma^{[\leq n]} \wedge^{[\geq 21]} \Sigma$ and $\Sigma \wedge \Sigma^{[\leq 2^{\sqrt{n}/1000}]} \wedge^{[\geq \sqrt{n}]} \Sigma$ arithmetic circuits where the bottom Σ gate is homogeneous.
Our results show that the fan-in of the middle Σ gates, the degree of the bottom powering gates and the homogeneity at the bottom Σ gates play a crucial role in the computational power of $\Sigma \wedge \Sigma \wedge \Sigma$ circuits.

1 Introduction

Arithmetic circuits were introduced by Valiant [19] as a natural model for algebraic computation and conjectured that the permanent polynomial, perm_n, does not have polynomial size arithmetic circuits. Following Valiant's work, there have been intensive research efforts towards the resolution of Valiant's hypothesis. Further, obtaining super polynomial size lower bounds for arithmetic circuits computing explicit polynomials is a pivotal problem in Algebraic Complexity Theory. However, for general classes of arithmetic circuits, the best known lower bound is barely superlinear [2].

Lack of progress on lower bounds against general arithmetic circuits lead researchers to explore restricted classes of circuits. Grigoriev and Karpinski [5] proved an exponential size lower bound for depth three circuits computing the permanent over finite fields of fixed size. However, extending these results to infinite fields or depth four arithmetic circuits remains elusive. Agrawal and Vinay [1] (see also [18,11]) explained this lack of progress by establishing that proving exponential lower bounds against depth four arithmetic circuits is enough to resolve Valiant's conjecture. This was strengthened further to depth three circuits over infinite fields by Gupta et al. [6].

* This work was done while the author was working at Max Planck Institute for Informatics, Saarbrücken supported by IMPECS post doctoral fellowship.

Gupta et al. [7] obtained a $2^{\Omega(\sqrt{n})}$ size lower bound for depth four homogeneous circuits computing perm_n where the fan-in of the bottom product gate is bounded by $O(\sqrt{n})$. Following this, Fournier et al. [4] obtained a super polynomial lower bound against depth four homogeneous circuits computing a polynomial in VP. Further, the techniques in [7,8] have been generalized and applied to prove lower bounds against various classes of constant depth arithmetic circuits for polynomials in VP as well as in VNP (see e.g., [16] and references therein).

Most of the lower bound proofs against arithmetic circuits follow a common framework: 1) define a measure for polynomials that is sub-additive and/or sub-multiplicative, 2) show that the circuit class of interest has small measure and 3) show that the target polynomial has high measure. See [16] for a detailed survey of these measures.

Apart from the complexity measure based framework mentioned above, there have been two other prominent approaches towards a resolution of Valiant's hypothesis: A geometric approach by Mulmuley and Sohoni [15] and an approach based on the real τ conjecture proposed by Shub and Smale [17].

The geometric approach to complexity theory [15] involves the study of class of varieties associated with each of the complexity classes and studying their representations.

The real τ conjecture of Koiran [9] states that the number of real roots of a univariate polynomial computed by an arithmetic circuit of size s is bounded by a polynomial in s. Koiran [10] showed that any resolution of the real τ-conjecture or an integer variant of it, would imply a positive resolution of Valiant's hypothesis. There has been several approaches towards the resolution of the real τ-conjecture and its variants by Koiran et al. [13,12].

Circuit Model We consider the class of depth five powering circuits, i.e., $\Sigma\wedge\Sigma\wedge\Sigma$ circuits. It was shown in [6] that any homogeneous polynomial f of degree d over a sufficiently large field computed by a circuit of size s can also be computed by a homogeneous $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuit of size $s^{\sqrt{d\log n\log(sd)}}$ for suitably chosen a. Here the superscript $[a]$ for a gate denotes the fan-in (degree in the case of $\wedge$ gates) at that level. This was an intermediary step in [6] which went on to obtain a depth three $\Sigma\Pi\Sigma$ circuit of size $2^{O(\sqrt{d\log n\log(sd)})}$ for f.

Thus, combined with the results in [18], to prove Valiant's hypotheses over infinite fields, it is enough to prove a $2^{\omega(\sqrt{n}\log n)}$ size lower bound against any one of the following classes of circuits: (1) homogeneous depth four $\Sigma\Pi^{[\sqrt{n}]}\Sigma\Pi^{[O(\sqrt{n})]}$ circuits, (2) homogeneous depth five $\Sigma\wedge^{[\sqrt{n}]}\Sigma\wedge^{[O(\sqrt{n})]}\Sigma$ circuits or (3) depth three $\Sigma\Pi\Sigma$ circuits .

Models (1) and (3) have received extensive attention in the literature compared to model (2). It follows that obtaining a $2^{\omega(\sqrt{n}\log n)}$ lower bound for any one of the models above would give a similar lower bound to the other. However, known lower bounds for model (1) so far do not even imply a super polynomial lower bound for model (2) which leaves obtaining super polynomial lower bounds against this model wide open.

In this article, we prove lower bounds against two restrictions of model (2) mentioned above: $\Sigma \wedge \Sigma^{[\leq n]} \wedge^{[\geq 21]} \Sigma$ circuits and $\Sigma \wedge \Sigma^{[\leq 2^{\sqrt{n}/1000}]} \wedge^{[\geq \sqrt{n}]} \Sigma$ circuits with bottom gates computing homogeneous linear forms. Since the transformation from depth four $\Sigma\Pi^{[\sqrt{n}]}\Sigma\Pi^{[O(\sqrt{n})]}$ to depth five $\Sigma \wedge^{[a]} \Sigma \wedge^{[d/a]} \Sigma$ in [6], in contrast to their result from general circuits, works against any chosen parameter $a < d$, the restrictions on the degree of the bottom $\wedge$ gates in the models we consider are general enough.

Throughout, it helps to interpret the polynomials computed by $\Sigma \wedge \Sigma \wedge \Sigma$ as sums of powers of projections of power symmetric polynomials where the n variate power symmetric polynomial of degree d is given by $p_d(x_1, \ldots, x_n) = x_1^d + \cdots + x_n^d$.

Our Results We prove lower bounds against the restrictions of depth five $\Sigma\wedge\Sigma\wedge\Sigma$. We show:

Theorem 1. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = p_{d_i}(\ell_{i_1}, \ldots, \ell_{i_n}) + \beta_i$ for some scalars β_i and for every i, either $d_i = 1$ or $d_i \geq 21$ and $\ell_{i_1}, \ldots, \ell_{i_n}$ are homogeneous linear forms. If $g = x_1 \cdot x_2 \cdots x_n$ then $s = 2^{\Omega(n)}$.*

The proof of Theorem 1 involves the dimension of the space of projected multilinear derivatives as a complexity measure for a polynomial f. It is computed by first projecting the partial derivative space of f to its multilinear subspace and then setting a subset of variables to 0. The dimension of the resulting space of polynomials is our measure of complexity for polynomials. Further, the method of projected multilinear derivatives also gives our second important result of the paper: An exponential lower bound against depth five powering circuits where the middle Σ layers have fan-in at most $2^{\sqrt{n}/1000}$ with the degree of the bottom $\wedge$ gates at least $\sqrt{n}$:

Theorem 2. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = p_{d_i}(\ell_{i_1}, \ldots, \ell_{i_{N_i}}) + \beta_i$, for some scalars β_i and $\sqrt{n} \leq d_i \leq n$, $N_i \leq 2^{\sqrt{n}/1000}$, and $\ell_{i_1}, \ldots, \ell_{i_{N_i}}$ are homogeneous linear forms. If $g = x_1 \cdot x_2 \cdots x_n$ then $s = 2^{\Omega(n)}$.*

It is not difficult to see that the polynomial $x_1 \cdots x_n$ has a homogeneous $\Sigma \wedge^{[\sqrt{n}]} \Sigma^{[O(2^{\sqrt{n}})]} \wedge^{[\sqrt{n}]} \Sigma$ circuit of size $2^{O(\sqrt{n})}$ (see Lemma 16). Theorem 2 shows that reducing the middle Σ gate fan-in by a constant factor in the exponent leads to an exponential lower bound.

The homogeneity condition on the lower Σ and $\wedge$ gates seems to be necessary to our proofs of Theorem 1 and Theorem 2. In fact, Saptharishi [16], in a result attributed to Forbes, showed that $x_1 \cdots x_n$ can be computed by $\Sigma \wedge \Sigma \wedge$ circuits of size $2^{O(\sqrt{n})}$ where the lower Σ gates are not necessarily homogeneous.

Thus, it is important to study depth five powering circuits where the bottom Σ gates are not necessarily homogeneous. Towards this, in Section 4, we consider the widely used measure of the dimension of the shifted partial derivatives of a polynomial. We show:

Theorem 3. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = p_{d_i}(x_{i_1}, \ldots, x_{i_{m_i}}, \ell_{i_1}, \ldots, \ell_{i_{r_i}})$, $m_i \leq \frac{1}{40}n$, $r_i \leq n^{\epsilon}$, $d \leq 2^{o(n)}$, $\alpha_i \leq 2^{o(n)}$ for all i where $0 < \epsilon < 1$. If $g = x_1 x_2 \ldots x_n$ then $s = 2^{\Omega(n)}$.*

It should be noted that Theorem 3 is much weaker than Theorems 1 and 2, however, it allows non-homogeneous Σ gates at the bottom. It seems that the restrictions on r_i in the above theorem are necessary if the lower bound argument uses the method of shifted partial derivatives. In particular, we show:

Lemma 4. *Let $k \leq \min\{l, d\}$ and $\alpha > 0$ be large enough. Then*

$$\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial^{=k}\left(p_d(x_1,\ldots,x_n)^{\alpha}\right)\right\}\right) = \Omega\left(\frac{\binom{n}{k}\binom{n+l}{l}}{l^{l/(d-1)}}\right).$$

In the cases where $l/(d-1) = O(1)$ and $l = n^{O(1)}$ the above bound is tight up to a polynomial factor since $\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial^{=k}\left(p_d(x_1,\ldots,x_n)^{\alpha}\right)\right\}\right) \leq \binom{n}{k}\binom{n+l}{l}$ and hence indicating that the restrictions on the r_is in Theorem 3 would be necessary if the dimension of shifted partial derivatives is used as the measure of complexity.

2 Preliminaries

An *arithmetic circuit* is a labelled directed acyclic graph. Vertices of zero in-degree are called *input* gates and are labelled by elements in $\mathbb{F} \cup \{x_1, \ldots, x_n\}$. Vertices of in-degree two or more are called *internal* gates and have their labels from $\{\times, +\}$. An arithmetic circuit has at least one vertex of zero out-degree called an *output* gate. We assume that an arithmetic circuit has exactly one output gate. A polynomial p_g in $\mathbb{F}[x_1, \ldots, x_n]$ can be associated with every gate g of an arithmetic circuit defined in an inductive fashion. Input gates compute their label. Let g be an internal gate with children $f_1, \ldots, f_m$ then $p_g = p_{f_1}\ \mathsf{op}\ \cdots \mathsf{op}\ p_{f_m}$ where $\mathsf{op} \in \{+, \times\}$ is the label of g. The polynomial computed by the circuit is the polynomial at one of the output gates and denoted by p_C. The size of an arithmetic circuit is the number of gates in it and is denoted by size(C). We will denote a fan-in/degree bound on a layer as a superscript to the corresponding gate e.g., $\Sigma \wedge \Sigma^{[\leq n]} \wedge^{[\geq 21]} \Sigma$ denotes the class of families of polynomials computed by depth five circuits with powering and sum gates, where the middle layer of sum gates have fan-in bounded from above by n and the bottom most powering gates have degree at least 21.

The following bound on the binomial coefficient is useful throughout the paper:

Proposition 5 ([14]). *Let $r \leq n$. Then $\log_2 \binom{n}{r} \approx nH(r/n)$, where H is the binary entropy function, $H(p) = -p\log_2(p) - (1-p)\log_2(1-p)$, and $\approx$ is equality up to an additive $o(n)$ error.*

We denote by $[n]$ the set $\{1, \ldots, n\}$. For a set of polynomials S, let $\mathcal{M}_{\leq d}(S)$ ($\mathcal{M}_{=d}(S)$) denote the set of all products of at most (exactly) d not necessarily distinct elements from S. Note that when S is a set of variables, $|\mathcal{M}_{\leq d}(S)| = \binom{|S|+d}{d}$. When the set S is clear from the context, we use $\mathcal{M}_{\leq d}$ ($\mathcal{M}_{=d}$) instead of $\mathcal{M}_{\leq d}(S)$ ($\mathcal{M}_{=d}(S)$).

For a subset S of variables, let $\mathcal{X}_a^b(S)$ denote the set of all multilinear monomials of degree $a \leq d \leq b$ in variables from the set S, i.e.,

$$\mathcal{X}_a^b(S) = \{ \prod_{x_i \in S} x_i^{\delta_i} \mid a \leq \sum_{i=1}^{n} \delta_i \leq b, \delta_i \in \{0,1\}\}.$$

For two sets A and B, define $A \odot B \triangleq \{a \cdot b \mid a \in A, b \in B\}$. Additionally, we define $A \cdot f$ for some polynomial f to be the set $\{a \cdot f \mid a \in A\}$.

The notion of *shifted partial derivatives* is given as follows: For $k \geq 0$ and $f \in \mathbb{F}[x_1, \ldots, x_n]$ let $\partial^{=k} f$ denote the set of all partial derivatives of f of order k. For $l \geq 0$, the (k,l) shifted partial derivative space of f, denoted by $\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial^{=k} f\right\}$, is defined as:

$$\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial^{=k} f\right\} \triangleq \mathbb{F}\text{-Span}\left\{\mathbf{m} \cdot \partial^{=k} f \;\mid\; \mathbf{m} \in \mathcal{M}_{\leq \ell}(x_1, \ldots, x_n)\right\}$$

where $\mathbb{F}\text{-Span}\{S\} \triangleq \{\alpha_1 f_1 + \cdots + \alpha_m f_m \mid f_i \in S \text{ and } \alpha_i \in \mathbb{F} \text{ for all } i \in [m]\}$. We restate the well known lower bound for the dimension of the space of shifted partial derivatives $x_1 \cdots x_n$:

Proposition 6 ([8]).

$$\begin{aligned}\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial_{\mathsf{ML}}^{=k} x_1 \cdots x_n\right\}\right) &= \dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial^{=k} x_1 \cdots x_n\right\}\right) \\ &\geq \binom{n}{k} \cdot \binom{n-k+l}{l}.\end{aligned}$$

In the above, $\partial_{\mathsf{ML}}^{=k} f$ denotes the set of kth order multilinear derivative space of f, i.e., $\partial_{\mathsf{ML}}^{=k} f \triangleq \{\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}} \mid i_1 < \ldots < i_k \in \{1, \ldots, n\}\}$.

3 Projected Multilinear Derivatives and Proof of Theorems 1 and 2

This section is devoted to the proof of Theorems 1 and 2. Our proof follows the standard two step approach for proving arithmetic circuit lower bounds: First, define a sub-additive measure that is low for every polynomial computed in the model. Second, show that the measure is exponentially larger for a specific polynomial p. Hence allowing us to conclude that any circuit in the model that computes p requires exponential size.

We consider a variant of the space of partial derivatives, viz., the *projected multilinear derivatives* as the complexity measure for polynomials.

The Complexity Measure Let $f \in \mathbb{F}[x_1, \ldots, x_n]$. For $S \subseteq \{1, \ldots, n\}$, let $\pi_S : \mathbb{F}[x_1, \ldots, x_n] \to \mathbb{F}[x_1, \ldots, x_n]$ be the projection map that sets all variables in S to zero, i.e., for every $f \in \mathbb{F}[x_1, \ldots, x_n]$, $\pi_S(f) = f(x_i = 0 \mid i \in S)$. Let $\pi_{\mathsf{m}}(f)$ denote the projection of f onto its multilinear monomials, i.e., if $f = \sum_{\alpha \in \mathbb{N}^n} c_\alpha \prod_{i=1}^n x_i^{\alpha_i}$ then $\pi_{\mathsf{m}}(f) = \sum_{\alpha \in \{0,1\}^n} c_\alpha \prod_{i=1}^n x_i^{\alpha_i}$.

For $S \subseteq \{1,\ldots,n\}$ and $0 < k \leq n$, the dimension of Projected Multilinear Derivatives (PMD) of a polynomial f is defined as:

$$\mathrm{PMD}_S^k(f) \triangleq \dim(\mathbb{F}\text{-Span}\left\{\pi_S(\pi_{\mathsf{m}}(\partial_{\mathsf{ML}}^{=k} f))\right\}).$$

We omit the subscript S when either S is clear from the context or when it refers to an unspecified set S. It is not hard to see that PMD_S^k is sub-additive.

Lemma 7. *For any $S \subseteq \{1\ldots,n\}$, $k \geq 1$, and polynomials f and g:*

$$\mathrm{PMD}_S^k(f+g) \leq \mathrm{PMD}_S^k(f) + \mathrm{PMD}_S^k(g).$$

Lower Bound for the Measure

We establish a lower bound on the dimension of projected multilinear derivatives of the polynomial $x_1 \cdots x_n$. This follows from a simple argument and is shown below:

Lemma 8. *For any $S \subseteq \{1,\ldots,n\}$ with $|S| = n/2+1$ and $k = 3n/4$ we have:*

$$\mathrm{PMD}_S^k(x_1 \cdots x_n) \geq \binom{n/2-1}{n/4} \geq 2^{n/2}/n^2.$$

Proof. Let $T \subseteq \{1,\ldots,n\}$ with $|T| = k$. Then $\frac{\partial^k}{\partial T}(x_1 \cdots x_n) = \prod_{i \notin T} x_i$. Note that if $S \cap \overline{T} = \emptyset$ then we have $\pi_S(\pi_{\mathsf{m}}(\frac{\partial^k}{\partial T}(x_1 \cdots x_n))) = \prod_{i \notin T} x_i$ since setting variables in S to zero does not affect the variables in $\overline{T}$. Otherwise, if $S \cap \overline{T} \neq \emptyset$ then $\pi_S(\pi_{\mathsf{m}}(\frac{\partial^k}{\partial T}(x_1 \cdots x_n))) = 0$. Thus, we have:

$$\mathbb{F}\text{-Span}\left\{\pi_S(\pi_{\mathsf{m}}(\partial_{\mathsf{ML}}^{=k}(x_1 \cdots x_n)))\right\} \supseteq \mathbb{F}\text{-Span}\left\{\prod_{i \in \overline{T}} x_i \;\mid\; \overline{T} \subseteq \overline{S}, |\overline{T}| \leq n/4\right\}.$$

Hence, $\mathrm{PMD}_S^k(x_1 \cdots x_n) \geq \binom{n/2-1}{n/4} \geq 2^{n/2}/n^2$ using Stirling's approximation of binomial coefficients. □

$\Sigma \wedge \Sigma\wedge$ Circuits: The Curse of Homogeneity

Firstly, we observe that homogeneous $\Sigma \wedge \Sigma\wedge$ circuits of polynomial size cannot compute the monomial $x_1 \cdot \cdots \cdot x_n$ by eliminating bottom $\wedge$ gates of degree at least 2:

Observation 9. *Let $f = f_1^{\alpha_1} + \cdots + f_s^{\alpha_s}$ where $f_i = \sum_{j=1}^n \beta_{ij} x_j^{d_i} + \beta_{i0}$, $\beta_{ij} \in \mathbb{F}$. If $f = x_1 \cdots x_n$ then $s = 2^{\Omega(n)}$.*

The homogeneity condition for the bottom power gates is necessary due to the following result in [16]. Let $\mathrm{Sym}_{n,d} = \sum_{S \subseteq [n], |S|=d} \prod_{i \in S} x_i$, the elementary symmetric polynomial of degree d.

Proposition 10. *[16, Corollary 17.16] For any $d > 0$, $\mathrm{Sym}_{n,d}$ can be computed by a $\Sigma \wedge \Sigma\wedge$ circuit of size $2^{O(\sqrt{d})}\mathsf{poly}(n)$.*[4]

Is it all about homogeneity at the bottom Σ gates? The answer is no. In fact, Observation 9 can also be generalized to the case of powers of polynomials in the span of the set $\{x_{i_j}^{\alpha_{i_j}} \mid 1 \le i_j \le n,\ \alpha_{i_j} \ge 2\}$:

Lemma 11. *For any $\beta_0, \beta_1, \ldots, \beta_r \in \mathbb{F}$, $\alpha, d \in \mathbb{N}$ and for any $S \subseteq \{1, \ldots, n\}$ with $|S| + k > n$, we have $\mathrm{PMD}_S^k((\sum_{j=1}^r \beta_j x_{i_j}^{d_j} + \beta_0)^\alpha) \le 1$ where $1 \le i_j \le n$ and either $\forall j\ d_j \ge 2$ or $\forall j\ d_j = 1$.*

We get the following generalization of Observation 9:

Corollary 12. *Let $f = f_1^{\alpha_1} + \cdots + f_s^{\alpha_s}$ where for every i, either f_i is a linear form or $f_i = \sum_{j=1}^n \beta_{i,l_j} x_{l_j}^{d_{i_j}} + \beta_{i0}$ for $d_{i_j} \ge 2$ and $\beta_{i,l_j} \in \mathbb{F}$. If $f = x_1 \cdots x_n$ then $s = 2^{\Omega(n)}$. Moreover, $|\{i \mid f_i \text{ is linear}\}| = 2^{\Omega(n)}$.*

Proof. Let $S \subset \{1, \ldots, n\}$ with $|S| = n/2 + 1$ and $k = 3n/4$. From Lemmas 11 and 7 we have $\mathrm{PMD}_S^k(f) \le \sum_{i=1}^s \mathrm{PMD}_S^k(f_i^{\alpha_i}) \le s$. Hence by Lemma 8 we have $s \ge 2^{n/2}/n^2$ as required. Further, $\mathrm{PMD}_S^k(f_i^{\alpha_i})$ is non-zero only if f_i is a linear form, and hence $|\{i \mid f_i \text{ is linear}\}| = 2^{\Omega(n)}$. □

$\Sigma \wedge \Sigma \wedge \Sigma$ Circuits: Middle Σ Fan-in versus the Bottom Degree

The argument above fails even when the degree of the power symmetric polynomial is two (i.e., $d = 2$). Let $f = \ell_1^2 + \cdots + \ell_n^2 + \beta$ where $\ell_1, \ldots, \ell_n$ are homogeneous linear functions such that each of the ℓ_i have all n variables with non-zero coefficients and $\beta \neq 0$. It is not hard to see that the space $\partial_{\mathsf{ML}}^k f^\alpha$ of the kth order derivatives of f^α is contained in the span of $\{f^{\alpha-k} \prod_{i=1}^n \ell_i^{\gamma_i} \mid \sum_i \gamma_i \le k\}$. Even after applying the projections π_{m} and π_S for any $S \subseteq \{1, \ldots, n\}$, with $|S| = (n/2) + 1$, obtaining a bound on PMD_S^k better than the lower bound in Lemma 8 seems to be difficult. The reason is that every multilinear monomial of degree $|n/2 - 1 - k|$ appears in at least one of the projected multilinear derivatives of f^α.

A natural approach to overcome the above difficulty could be to obtain a basis for the projected multilinear derivatives of f^α consisting of a small set of monomials and a small set of products of powers of the linear forms multiplied by suitable powers of f. Surprisingly, as shown below in Lemma 13, the approach works when the degree $d \ge 21$, although it requires an involved combinatorial argument.

Lemma 13. *Suppose that $f = (\ell_1^d + \ldots + \ell_n^d + \beta)$ for some scalar β, and ℓ_j homogeneous linear forms, $1 \le j \le n$. Let $Y = \{\ell_i^{d-j} \mid 1 \le i \le n, 1 \le j \le d\}$ and*

[4] In [16], Corollary 17.16, it is mentioned that the resulting $\Sigma \wedge \Sigma\wedge$ circuit is homogeneous. However, a closer look at the construction shows that the application of Fischer's identity produces sum gates that are not homogeneous.

$\lambda = 1/4 + \varepsilon$ for some $0 < \varepsilon < 1/4$. Then, for $k = 3n/4$ and any $S \subseteq \{1, \ldots, n\}$ with $|S| = n/2 + 1$, we have:

$$\pi_S(\pi_{\mathsf{m}}(\partial^{=k}_{\mathsf{ML}} f^{\alpha})) \subseteq \mathbb{F}\text{-Span}\left\{\pi_S(\pi_{\mathsf{m}}(\mathcal{F} \odot \left(\mathcal{X}^{n/2-1}_{\lambda n}(\overline{S}) \cup \mathcal{M}_{\leq (1+\varepsilon)n/d}(Y)\right)))\right\}$$

where $\mathcal{F} = \cup_{i=1}^{k} f^{\alpha - i}$ and $\overline{S} = \{1, \ldots, n\} \setminus S$.

Proof. Let $T \subseteq \{x_1, \ldots, x_n\}$ with $|T| = k$, let $f^{(k)}_T$ denote kth order partial derivative of f with respect to T. Note that $f^{(k)}_T \in \mathbb{F}\text{-Span}\left\{\ell_j^{d-k} \mid 1 \leq j \leq n\right\}$. Let L_i denote $\{\ell_j^{d-i} \mid 1 \leq j \leq n\}$ so that $f^{(k)}_T \in \mathbb{F}\text{-Span}\{L_k\}$. Then

$$\frac{\partial^k f^{\alpha}}{\partial T} \in \mathbb{F}\text{-Span}\left\{f^{\alpha - i} \odot D^T_i(f) \mid 1 \leq i \leq k\right\} \tag{1}$$

where $D^T_i(f) = \left\{\prod_{r=1}^{i} f^{(t_r)}_{T_r} \;\mid\; T_1 \uplus \ldots \uplus T_i = T, \;\; \text{where } t_r = |T_r| > 0, 1 \leq r \leq i\right\}$. Intuitively, the set D^T_i contains one polynomial for each possible partition of T into i many parts. The polynomial corresponding to a particular partition is the product of the derivatives of f with respect to each of the parts. Now, the following claim bounds the span of D^T_i:

Claim. For any $1 \leq i \leq k$, $D^T_i \subseteq \mathbb{F}\text{-Span}\left\{\bigodot_{r=1}^{k} L_r^{\odot j_r} \;\mid\; \sum_{r=1}^{k} r \cdot j_r = k\right\}$.

Proof. Let $T_1 \uplus \cdots \uplus T_i = T$ be a partition and let j_r denote the number of parts with cardinality r, i.e., $j_r = |\{j \;\mid\; |T_j| = r\}|$. Then

$$\prod_{|T_j|=r} f^{(r)}_{T_j} \in \mathbb{F}\text{-Span}\left\{\bigodot_{|T_j|=r} L_r\right\} = \mathbb{F}\text{-Span}\left\{L_r^{\odot j_r}\right\}.$$

Thus, $\prod_{r=1}^{i} f^{(t_r)}_{T_r} \in \mathbb{F}\text{-Span}\left\{\bigodot_{r=1}^{k} L_r^{\odot j_r}\right\}$. Since, $\sum_{r=1}^{k} r \cdot j_r = k$ for any partition $T_1 \uplus \cdots \uplus T_i$ of T, the claim follows. □

Continuing from (1), we have:

$$\begin{aligned}
\frac{\partial^k f^{\alpha}}{\partial T} &\in \mathbb{F}\text{-Span}\left\{f^{\alpha - i} \odot D^T_i(f) \;\mid\; 1 \leq i \leq k\right\} \\
&\subseteq \mathbb{F}\text{-Span}\left\{\mathcal{F} \odot \{D^T_i(f) \;\mid\; 1 \leq i \leq d\}\right\} \\
&\subseteq \mathbb{F}\text{-Span}\left\{\mathcal{F} \odot \left\{\bigodot_{r=1}^{d} L_r^{\odot j_r} \;\mid\; 1 \cdot j_1 + \cdots + d \cdot j_d = k\right\}\right\}.
\end{aligned} \tag{2}$$

It remains to show that the right side of (2) is spanned by a set of polynomials that satisfy the properties stated in the lemma. The next claim completes the proof of Lemma 13.

Claim.

$$\pi_S(\pi_{\mathsf{m}}\left(\left\{\bigodot_{r=1}^{d} L_r^{\odot j_r} \;\mid\; \sum_{i=1}^{d} i\cdot j_i = k\right\}\right) \subseteq \mathbb{F}\text{-Span}\left\{\mathcal{X}_{\lambda n}^{\frac{n}{2}-1}(\bar{S}) \cup \mathcal{M}_{\leq \frac{(1+\varepsilon)n}{d}}(Y)\right\}.$$

Proof. Note that the polynomials in L_j are homogeneous non-constant polynomials of degree $d-j$, and hence the set $\bigodot_{r=1}^{d} L_r^{\odot j_r}$ consists of homogeneous polynomials of degree $\sum_{r=1}^{d} j_r(d-r)$.

Let $\deg(\bigodot_{r=1}^{d} L_r^{\odot j_r})$ denote the degree of polynomials in the set $\bigodot_{r=1}^{d} L_r^{\odot j_r}$. The remaining argument is split into three cases depending on the value of $\deg(\bigodot_{r=1}^{d} L_r^{\odot j_r})$.

Case 1: $\deg(\bigodot_{r=1}^{d} L_r^{\odot j_r}) \geq n/2$ then $\pi_S(\pi_{\mathsf{m}}(\bigodot_{r=1}^{d} L_r^{\odot j_r})) = \{0\}$. Note that here we have crucially used the fact that the ℓ_j are homogeneous.

Case 2: $\lambda n \leq \deg(\bigodot_{r=1}^{d} L_r^{\odot j_r})) < n/2$. In this case $\pi_S(\pi_{\mathsf{m}}(\bigodot_{r=1}^{d} L_r^{\odot j_r}))$ is spanned by the set of all multilinear monomials in the set of variables $\{x_j \mid j \notin S\}$ of degree at least λn and at most $n/2-1$. Therefore we have, $\pi_S(\pi_{\mathsf{m}}(\bigodot_{r=1}^{d} L_r^{\odot j_r})) \subseteq \mathbb{F}\text{-Span}\left\{\mathcal{X}_{\lambda n}^{n/2-1}(\bar{S})\right\}$.

Case 3: $\deg(\bigodot_{r=1}^{d} L_r^{\odot j_r})) < \lambda n$. Recall that $\deg(\bigodot_{r=1}^{d} L_r^{\odot j_r})) = \sum_{r=1}^{d} j_r(d-r) \leq \lambda n$. Then,

$$\begin{aligned}\sum_{r=1}^{d} j_r \cdot d &\leq \sum_{r=1}^{d} j_r \cdot r + \lambda n = k + \lambda n \qquad (\text{since } \sum_{r=1}^{d} r\cdot j_r = k.)\\ &= (\lambda + 3/4)n = (1+\varepsilon)n.\end{aligned}$$

Hence, $\pi_S(\pi_{\mathsf{m}}(\bigodot_{r=1}^{d} L_r^{\odot j_r}))$ is spanned by the set of all products of at most $(1+\varepsilon)n/d$ polynomials of the form ℓ_i^{d-j}, i.e.,

$$\pi_S(\pi_{\mathsf{m}}(\bigodot_{r=1}^{d} L_r^{\odot j_r})) \subseteq \mathbb{F}\text{-Span}\left\{\mathcal{M}_{\leq (1+\varepsilon)n/d}(Y)\right\}.$$

□

This completes the proof. □

Using Lemma 13 above and choosing suitable parameters k and S we obtain the following upper bound on the dimension of projected multilinear derivatives:

Theorem 14. *Let $f = (\ell_1^d + \ldots + \ell_n^d + \beta)$ where ℓ_j are homogeneous linear forms. For $d \geq 21$ and any $S \subseteq \{1,\ldots,n\}$ where $|S| = n/2+1$. Then*

$$\mathrm{PMD}_S^k(f^\alpha) \leq 2^{(0.498+o(1))n}.$$

Proof. By Lemma 13,

$$\pi_S(\pi_{\mathsf{m}}(\partial_{\mathsf{ML}}^{=k} f^\alpha)) \subseteq \mathbb{F}\text{-Span}\left\{\pi_S(\pi_{\mathsf{m}}(\{f^{\alpha-i}\}_{i=1}^{k} \odot \left\{\mathcal{X}_{\lambda n}^{n/2-1}(\bar{S}) \cup \mathcal{M}_{\leq (1+\epsilon)n/d}(Y)\right\}))\right\}.$$

Recall that $\lambda = \frac{1}{4} + \varepsilon$. We choose $\varepsilon = 1/50$ and hence $\lambda = 0.27$. We have:

$$\mathrm{PMD}_S^k(f^\alpha) \le k \cdot (|\mathcal{X}_{\lambda n}^{n/2-1}(\bar{S})| + |\mathcal{M}_{\le(1+\epsilon)n/d}(Y)|).$$

Now, since $1/4 < \lambda < 1/2$, we have

$$\begin{aligned}|\mathcal{X}_{\lambda n}^{n/2-1}(\bar{S})| &\le (n/2 - 1 - \lambda n) \cdot \binom{n/2-1}{\lambda n} \le c(n/2) \cdot \binom{n/2}{\lambda n} \\ &\le (cn/2) \cdot 2^{\frac{n}{2}\cdot\mathcal{H}(2\lambda)} \le (cn/2) \cdot 2^{0.498n}.\end{aligned}$$

Where c is an absolute constant. We bound $|\mathcal{M}_{\le(1+\epsilon)n/d}(Y)|$ as follows:

$$\begin{aligned}|\mathcal{M}_{\le(1+\epsilon)n/d}(Y)| &= \binom{|Y| + (1+\varepsilon)n/d}{(1+\varepsilon)n/d} = \binom{dn + (1+\varepsilon)n/d}{(1+\varepsilon)n/d} \\ &\le 2^{(dn+(1+\varepsilon)n/d)\mathcal{H}\left(\frac{(1+\varepsilon)n/d}{dn+(1+\varepsilon)n/d}\right)} \\ &= 2^{n(d+(1+\varepsilon)/d)\mathcal{H}\left((1+\varepsilon)/(d^2+(1+\epsilon))\right)} \le 2^{0.4955n} \quad \text{for } d \ge 21.\end{aligned}$$

For the last inequality, note that for fixed n and ε, $(d + (1+\varepsilon)/d)\mathcal{H}((1+\varepsilon)/(d^2+(1+\epsilon))$ is a monotonically decreasing function of d, with $\lim_{d\to\infty}(d + (1+\varepsilon)/d)\mathcal{H}((1+\varepsilon)/(d^2+(1+\epsilon)) = 0$. Therefore, the bound holds for $d \ge 21$. This completes the proof. □

Corollary 15. *Let $f = (\ell_1^d + \ldots + \ell_N^d + \beta)$ where ℓ_j are homogeneous linear forms. If d is such that $N \le 2^{(d/1000)}$, $d \le n$, and $n/d = o(n)$ then for any $\alpha > 0$,*

$$\mathrm{PMD}_S^k(f^\alpha) \le 2^{(0.498+o(1))n}.$$

Proof of Theorem 1: Let $S = \{1, \ldots, n/2+1\}$ and $k = 3n/4$. Then by Theorem 14 we have $\mathsf{PMD}_S^k(f_i) \le 2^{0.498n+o(n)}$. By the sub-additivity of PMD_S^k (Lemma 7), we have $\mathsf{PMD}_S^k(\sum_{i=1}^s f_i^{\alpha_i}) \le s \cdot 2^{0.498n+o(n)}$. Since $\mathsf{PMD}_S^k(x_1 \cdots x_n) \ge 2^{n/2}/n^2$, we conclude $s \ge 2^{0.001n}$, as required. □

Proof of Theorem 2: Let $S = \{1, \ldots, n/2+1\}$ and $k = 3n/4$. Since $d_i \ge \sqrt{n}$, it holds that $N_i \le 2^{d/1000}$. Then, by Corollary 15, we have $\mathsf{PMD}_S^k(f_i^{\alpha_i}) \le 2^{0.498n+o(n)}$. By the sub-additivity of PMD_S^k (Lemma 7), we have $\mathsf{PMD}_S^k(\sum_{i=1}^s f_i^{\alpha_i}) \le s \cdot 2^{0.498n+o(n)}$. Since $\mathsf{PMD}_S^k(x_1 \cdots x_n) \ge 2^{n/2}/n^2$, we conclude $s \ge 2^{0.001n}$ for large enough n, as required. □

A separation within $\Sigma \wedge \Sigma \wedge \Sigma$ Circuits: An alert reader might have wondered if the restriction on the fan-in of the middle layer of Σ gates in Theorem 2 is a limitation of the method of projected multilinear derivatives. By a simple application of Fischer's identity [3], we get:

Lemma 16. *Over fields of characteristic zero or characteristic greater than n, the polynomial $x_1 \cdots x_n$ can be computed by a homogeneous $\Sigma \wedge^{[\sqrt{n}]} \Sigma^{[O(2^{\sqrt{n}})]} \wedge^{[\sqrt{n}]} \Sigma$ circuit of size $2^{O(\sqrt{n})}$.*

This immediately leads to the following separation of homogeneous $\Sigma \wedge^{[\sqrt{n}]} \Sigma \wedge^{[\sqrt{n}]} \Sigma$ circuits:

Corollary 17. *The class of polynomials computed by $\Sigma \wedge^{[\sqrt{n}]} \Sigma^{[2^{\sqrt{n}/1000}]} \wedge^{[\sqrt{n}]} \Sigma$ of size $2^{O(\sqrt{n})}$ is strictly contained in the class computed by $\Sigma \wedge^{[\sqrt{n}]} \Sigma^{[2^{\sqrt{n}}]} \wedge^{[\sqrt{n}]} \Sigma$ of size $2^{O(\sqrt{n})}$.*

4 Dimension of Shifted Partial Derivatives

This section is devoted to the study of shifted partial derivatives of polynomials that are computed by restricted $\Sigma \wedge \Sigma \wedge \Sigma$ circuits and proofs of Theorem 3 and Lemma 4.

We begin with a simple upper bound on the dimension of the derivatives of powers of projections of p_d onto low-dimensional sub-spaces:

Lemma 18. *Let $f = p_d(\ell_1, \ldots, \ell_t)$ where $\ell_1, \ldots, \ell_t$ are linear forms. Then for any $k > 0$, we have $\dim\left(\mathbb{F}\text{-Span}\left\{\partial_{\mathsf{ML}}^{\leq k} f^\alpha\right\}\right) \leq (k+1)(dk)^r$ where r is the dimension of the span of $\{\ell_1, \ldots, \ell_t\}$.*

Proof. Without loss of generality, assume that $\ell_1, \ldots, \ell_r$ is a basis for the space spanned by $\ell_1, \ldots, \ell_t$ $r \leq t$. Observe that:

$$\partial_{\mathsf{ML}}^{\leq k} f^\alpha \subseteq \mathbb{F}\text{-Span}\left\{ f^{\alpha - i} \cdot \ell_1^{\beta_1} \cdots \ell_r^{\beta_r} \mid \sum_{j=1}^{r} \beta_j \leq dk \right\}_{i \in \{1,\ldots,k\}}$$

and therefore, $\dim\left(\mathbb{F}\text{-Span}\left\{\partial_{\mathsf{ML}}^{\leq k} f^\alpha\right\}\right) \leq (k+1)(dk)^r$ as required. □

Now, we bound the dimension of shifted partial derivatives of powers of the power symmetric polynomial:

Lemma 19. *Let $f = p_d(x_{j_1}, \ldots, x_{j_m})$ for some $j_1, \ldots, j_m \in \{1, \ldots, n\}$. Then for any $\alpha, l, k \geq 1$*

$$\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l} \partial_{\mathsf{ML}}^{=k} f^\alpha\right\}\right) \leq (k+1)\binom{n+m+k+l}{k+l}.$$

Note that the straightforward bound of $\binom{m}{k}\binom{n+l}{l}$ is better than this bound if m is large. However, when m is small (say $m \leq n/40$), the bound shown above is better for suitable values of k and l. Combining Lemmas 18 and 19 with the sum and product rules for partial derivatives, we get:

Lemma 20. *Let $\ell_1, \ldots \ell_t \in \mathbb{F}[x_1, \ldots, x_n]$ be linear forms and let r denote their rank. Let $f = p_d(x_{j_1}, \ldots, x_{j_m}, \ell_1, \ldots, \ell_t)$. Then for any $d > k > 0$, we have*

$$\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l} \partial_{\mathsf{ML}}^{=k} f^\alpha\right\}\right) \leq (\alpha+1)(k+1)^3(dk)^r\binom{m+n+k+l}{k+l}.$$

Finally, using sub-additivity of shifted partial derivatives and Lemma 20 we obtain the following upper bound:

Theorem 21. *Let $d > k > 0$ and $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where each of the polynomials $f_i = p_{d_i}(x_{i_1}, \ldots, x_{i_{m_i}}, \ell_{i_1}, \ldots, \ell_{i_{r_i}})$ and $\ell_{i_1}, \ldots, \ell_{i_{m_i}}$ are linear forms in $x_1, \ldots, x_n$. Then for any $l > 0$ with $k + l > n + m$:*

$$\dim\left(\mathbb{F}\text{-Span}\left\{\boldsymbol{x}^{\leq l}\partial_{\mathsf{ML}}^{=k} g\right\}\right) \leq s(\alpha+1)(k+1)^3(dk)^r\binom{n+m+k+l}{k+l}$$

where $m = \max_i m_i$ and $r = \max_i\{\dim(\mathbb{F}\text{-Span}\left\{\ell_{i_1}, \ldots, \ell_{i_{r_i}}\right\})\}$.

Combining the previous theorem with the lower bound from Proposition 6 gives us the required size lower bound.

Theorem 3. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = p_{d_i}(x_{i_1}, \ldots, x_{i_{m_i}}, \ell_{i_1}, \ldots, \ell_{i_{r_i}})$, $m_i \leq \frac{1}{40}n$, $r_i \leq n^{\epsilon}$, $d \leq 2^{n^{1-\gamma}}$ and $\alpha_i \leq 2^{n^{\delta}}$ for all i, for some $0 < \delta, \epsilon, \gamma < 1$, $\varepsilon < \gamma$. If $g = x_1 x_2 \ldots x_n$ then $s = 2^{\Omega(n)}$.*

Proof. Let $d \geq 2$ and $m = \max_i m_i$. Using Proposition 6 and Theorem 21

$$s \geq \frac{\binom{n}{k}\binom{n-k+l}{l}}{(\alpha+1)(k+1)^3(dk)^r\binom{n+m+k+l}{k+l}}$$

where $\alpha = \max_i \alpha_i$. Taking the logarithm and using that $3\log(k+1) \leq 3\log dk$ since $d \geq 2$ gives us

$$\log s \geq \log\binom{n}{k} + \log\binom{n-k+l}{l} - \left(\log(\alpha+1) + \log\binom{n+m+k+l}{k+l} + (r+3)\log dk\right).$$

Note that $(r+3)\log dk \in o(n)$ if $d \leq 2^{n^{1-\gamma}}$. Now, using the approximation of binomial coefficients in Proposition 5 and setting $k = n/10$ and $l = 10n$ we get $\log s \geq 0.0165n$. This proves the required bound. □

References

1. M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, pages 67–75, 2008.
2. W. Baur and V. Strassen. The complexity of partial derivatives. *Theor. Comput. Sci.*, 22:317–330, 1983.
3. I. Fischer. Sums of like powers of multivariate linear forms. *Mathematics Magazine*, 67(1):59–61, 1994.
4. H. Fournier, N. Limaye, G. Malod, and S. Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, pages 128–135, 2014.
5. D. Grigoriev and M. Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *STOC*, pages 577–582, 1998.
6. A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Arithmetic circuits: A chasm at depth three. In *FOCS*, pages 578–587, 2013.

7. A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Approaching the chasm at depth four. *J. ACM*, 61(6):33:1–33:16, 2014.
8. N. Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *ECCC*, 19:81, 2012.
9. P. Koiran. Shallow circuits with high-powered inputs. In *ICS*, pages 309–320. Tsinghua University Press, 2011.
10. P. Koiran. Shallow circuits with high-powered inputs. In *ICS*, pages 309–320, 2011.
11. P. Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.
12. P. Koiran, N. Portier, and S. Tavenas. A wronskian approach to the real tau-conjecture. *J. Symb. Comput.*, 68:195–214, 2015.
13. P. Koiran, N. Portier, S. Tavenas, and S. Thomassé. A tau -conjecture for newton polygons. *Foundations of Computational Mathematics*, 15(1):185–197, 2015.
14. D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
15. K. Mulmuley and M. A. Sohoni. Geometric complexity theory I: an approach to the P vs. NP and related problems. *SIAM J. Comput.*, 31(2):496–526, 2001.
16. R. Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Version 3.1.0, `https://github.com/dasarpmar/lowerbounds-survey/releases`, 2016.
17. M. Shub and S. Smale. On the intractability of hilbert's nullstellensatz and an algebraic version of" NP != P ?". *Duke Mathematical Journal*, 81(1):47–54, 1995.
18. S. Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *MFCS*, pages 813–824, 2013.
19. L. G. Valiant. Completeness classes in algebra. In *STOC*, pages 249–261, 1979.