

Separated borders: Exponential-gap fanin-hierarchy theorem for approximative depth-3 circuits

Pranjal Dutta *

Nitin Saxena †

Abstract

Mulmuley and Sohoni (2001) proposed an ambitious program, the Geometric Complexity Theory (GCT), to prove $P \neq NP$ and related conjectures using algebraic geometry and representation theory. Gradually, GCT has introduced new structures and questions in complexity. GCT tries to capture the algebraic/geometric notion of ‘approximation’ by defining *border* classes. Surprisingly, (Kumar ToCT’20) proved the universal power of the border of top-fanin-2 depth-3 circuits ($\overline{\Sigma^{[2]}\Pi\Sigma}$); which is in complete contrast to its classical model. Recently, (Dutta,Dwivedi,Saxena, FOCS’21) put an upper bound, by showing that bounded-top-fanin border depth-3 circuits ($\overline{\Sigma^{[k]}\Pi\Sigma}$ for constant k) can be computed by a polynomial-size algebraic branching program (ABP). It was left open to separate the class of ABPs from $\overline{\Sigma^{[k]}\Pi\Sigma}$.

In this article, we show a strongly-exponential separation between any two consecutive border classes, $\overline{\Sigma^{[k]}\Pi\Sigma}$ and $\overline{\Sigma^{[k+1]}\Pi\Sigma}$, establishing an *optimal hierarchy* of constant top-fanin border depth-3 circuits. Put in GCT language: we prove an exponential-hierarchy for *padded- k -th-secant-varieties* of the Chow variety of \mathbb{F}^{n+1} . This positively answers [Open question 2 of Dutta,Dwivedi,Saxena FOCS’21] and [Problem 8.10 with constant r , of Landsberg, Annal.Ferrara’15].

Keywords. approximative, border, depth-3, hierarchy, formula, GCT, secant variety, padded, ABP, ROABP, ARO, VF, inhomogeneous.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 1.1 | The Chow variety and lower bounds in GCT | 4 |
| 1.2 | Our results: The fanin-hierarchy theorem | 5 |
| 1.3 | An overview of the proof | 7 |
| 1.4 | Known depth-3 lower bounds and their limitations | 11 |
| 2 | Notation and preliminaries | 12 |
| 3 | Hardness lies in all-non-homogeneity | 13 |

*CSE, IIT Kanpur, and, Chennai Mathematical Institute, India. Email: pranjal@cmi.ac.in.

†Department of Computer Science & Engineering, IIT Kanpur, India. Email: nitin@cse.iitk.ac.in

| | |
|---|-----------|
| 4 Proving lower bound for the core case: Problem 1 | 17 |
| 4.1 Reducing to bloated fanin-1 model $\overline{\text{Gen}(1, \cdot)}$ | 18 |
| 4.2 Proving lower bound for bloated fanin-1 model $\overline{\text{Gen}(1, \cdot)}$ | 23 |
| 4.3 Tying the pieces together: Proof of Theorem 1 | 24 |
| 5 Conclusion | 24 |
| A Basics of algebraic complexity | 29 |
| B Basic tools | 29 |

1 Introduction

The main aim of Computational Complexity is to fathom, as meticulously as possible, the amount of computational resources required to perform computational tasks. These resources could be of various kinds depending on the computational model under consideration— e.g., time/ space for Turing machines; size/ depth/ fanin for boolean and algebraic circuits; and so on. A fundamental question in this context is “Does more (of the same) resource \implies more power?”. Classical theorems in Computational Complexity such as the Time Hierarchy [HS65] and Space Hierarchy [SHL65] answer this question (affirmatively) for the resources of time respectively space on multitape Turing machines. In this paper, we consider an analogous question for algebraic circuits.

A polynomial $f \in \mathbb{F}[x_1, \dots, x_n]$, over a field \mathbb{F} , is computable by an algebraic circuit of size s and depth d , if there exists a directed acyclic graph whose size (number of nodes and edges) is $\leq s$, and depth is $\leq d$, such that its leaf nodes are labeled by variables or field-constants, internal nodes are labeled with field-operators ($+$ and \times), and the polynomial computed at the root is f . For polynomial f , the size of the *smallest circuit* computing it, is denoted as $\text{size}(f)$. Another important complexity parameter is the *depth* – the length of the longest path in the circuit (from leaf to root). In this paper, we are interested in *depth-3 circuits* $\Sigma^{[k]}\Pi\Sigma$; they compute polynomials of the form $\sum_{i \in [k]} \prod_j \ell_{ij}$, where ℓ_{ij} are affine *linear* functions.

We consider the question of proving a *top-fanin-hierarchy* theorem for algebraic circuits (in the border/ approximative sense). Informally, we ask the following.

Question 1 (The fanin-hierarchy question). *For fixed $k \geq 1$, are there ‘explicit’ families of polynomials $P_n \in \mathbb{F}[x_1, \dots, x_n]$, such that P_n can be ‘approximated’ by a small $\Sigma^{[k+1]}\Pi\Sigma$ -circuit but not ‘approximated’ by a small $\Sigma^{[k]}\Pi\Sigma$ -circuit?*

See [section 2](#) for the definition of terms ‘explicit’ and ‘approximation’. Explicitness, and approximation by a small $\Sigma^{[k]}\Pi\Sigma$, are natural constructive hypotheses; as without these, a simple geometric-dimension argument lower-bounds top-fanin in a *non-constructive* way. Interestingly, if one considers the above question in the classical ‘exact’ setting (i.e. without approximation), there is an easy *impossibility* result known, which shows that the inner product polynomial, $IP_{k+1} := \sum_{i \in [k+1]} x_i \cdot y_i$, for $k \geq 1$, *cannot* be computed by $\Sigma^{[k]}\Pi\Sigma$ -circuit regardless of the size; see [[Kum20](#), [CGJ⁺18](#)].

However, the unexpected *universality* of border depth-3 fanin-2 circuits [[Kum20](#)] means that the classical impossibility result breaks down in the border for $k \geq 2$; for a detailed statement, see [subsection 1.1](#). Therefore, an exponential separation between the two classes becomes a nontrivial & intriguing question to study. In this work, we affirmatively and optimally answer [Question 1](#);

see [Theorem 1](#) and its remarks. In particular, this is related to fundamental variety constructions in geometric complexity theory (GCT). As one can see, [Question 1](#) is a simply-phrased question, and our candidate polynomial will be simple (namely, degree d version of IP_{k+1} ; see [subsection 1.2](#)). But, solving [Question 1](#) is far from obvious and we will employ quite powerful models like ABP, ARO, $\text{Gen}(k, s)$ in the proofs; for respective definitions, see [Appendix A](#) and [section 4](#).

So, here we remind a few important models of computation which appear in our proofs. For e.g., if in a circuit, out-degree of internal nodes is 1, then it is a *formula*. Any formula can be converted into a layered graph called *algebraic branching program* (ABP) with polynomial blowup in size. With different models, comes different complexity classes which accordingly classify polynomials. For e.g. VP is the class of polynomials of polynomial degree, computable by polynomial-sized circuits. Similarly, one can define VBP and VF, for ABPs respectively formulas. Finally, polynomials in VNP, can be expressed as an exponential-sum of projection of a VP circuit. Valiant [[Val79](#)] conjectured that $\text{VBP} \neq \text{VNP}$ (respectively $\text{VP} \neq \text{VNP}$), as an algebraic analog of the P vs. NP problem. For details, see [[SY10](#), [Mah13](#)].

Inception of GCT. In [[MS01](#)], Ketan Mulmuley and Milind Sohoni developed the Geometric Complexity Theory (GCT) program and strengthened Valiant’s conjecture to: $\text{VNP} \not\subseteq \overline{\text{VBP}}$, i.e. $\ell^{m-n} \cdot \text{perm}_n(X) \notin \overline{\text{GL}_{m^2}(\mathbb{C}) \cdot \det_m(X)}$, for a linear form ℓ and $m = \text{poly}(n)$. In words: padded permanent does not lie in the orbit closure of ‘small’ determinants. This padding was required to tackle *non-homogeneity*. Of course, the hope in the GCT program, is to convert complexity lower bound questions to algebraic geometry terms; and then translate algebro-geometric questions to those in representation theory, and finally solve them using rich mathematics. In a way, it advances the notion of *border complexity* which has intimate connections with diversified topics including – designing matrix multiplication algorithms [[Str74](#), [Bin80](#), [BCRL79](#), [CW90](#), [LO15](#)], computational invariant theory [[FS13](#), [Mul12b](#), [GGOW16](#), [BGO⁺18](#), [IQS18](#)], algebraic natural proofs [[GKSS17](#), [BIL⁺21](#), [CKR⁺20](#), [KRST20](#)], lower bounds [[BI13](#), [Gro15](#), [LO15](#)], optimization [[AZGL⁺18](#), [BFG⁺19](#)], derandomization [[Mul12a](#), [Muk16](#), [DDS21](#)], and many more. We refer to [[BLMW11](#), [Mul12b](#), [Mul12a](#)] for expository references.

Algebraic approximations. Perhaps the simplest notion of the approximative closure comes from the following definition [[Bür04](#), [Bür20](#)]: Polynomial $f(x) \in \mathbb{F}[x_1, \dots, x_n]$ is *approximated* by rational-function $g(x, \epsilon) \in \mathbb{F}(\epsilon)[x]$ if there exists polynomial $S(x, \epsilon) \in \mathbb{F}[\epsilon][x]$ such that $g = f + \epsilon S$. In $\mathbb{F} = \mathbb{R}$, it mimics the Euclidean topology: we can think analytically that $\lim_{\epsilon \rightarrow 0} g = f$. The class $\overline{\mathcal{C}}$, the *approximative closure* of a complexity class \mathcal{C} , can be defined analogously. Note that, arbitrary ϵ -power is allowed as ‘cost-free’ constants when computing $g \in \overline{\mathcal{C}}$. Further, algebraically one can define the closure as *Zariski closure*, over any field \mathbb{F} , i.e. take the closure of the set of polynomials, considered as points, of \mathcal{C} . Interestingly, these notions are known to be equivalent over the algebraically closed field \mathbb{C} [[Mum95](#), S2.C]; which is usually the source of all geometry ideas.

Complexity measure. The *approximative* (or *border*) complexity of f , denoted $\overline{\text{size}}(f)$, is the minimum size of the circuit computing g , over $\mathbb{F}(\epsilon)$; evidently, $\overline{\text{size}}(f) \leq \text{size}(f)$. Due to the possible $(1/\epsilon)$ -power terms in the circuit computing g , evaluation at $\epsilon = 0$ is not necessarily valid, although the ‘limit’ exists. Hence, any relation between the border and the *exact* complexity of f is not at all clear. Since $g = f + \epsilon \cdot S$, we could interpolate by setting ‘random’ ϵ -values from \mathbb{F} . However, the trivial bound on the circuit size of f would *linearly* depend on the degree M of ϵ (& $1/\epsilon$); which is known to be *exponential* in the size of the circuit computing g ! Therefore, the following relation is the best one known: $\overline{\text{size}}(f) \leq \text{size}(f) \leq \exp(\overline{\text{size}}(f))$ [[Bür04](#), Theorem 5.7].

1.1 The Chow variety and lower bounds in GCT

Border depth-3 circuits: An algebraic view. Since, depth-2 circuits are *closed* under taking limit, i.e. $\overline{\Pi\Sigma} = \Pi\Sigma$ and $\overline{\Sigma\Pi} = \Sigma\Pi$, it is natural to study border of depth-3 circuits. Again, it is not hard to show that $\overline{\Pi\Sigma\Pi} = \Pi\Sigma\Pi$ which leaves us to understand $\overline{\Sigma\Pi\Sigma}$. Kumar [Kum20] showed that border depth-3 fanin-2 circuits are ‘universal’; i.e. $\Sigma^{[2]}\Pi^{[D]}\Sigma$ over $\mathbb{C}(\epsilon)$ can approximate *any* homogeneous d -degree, n -variate polynomial; though this expression requires an exceedingly large $D = \exp(n, d)$. In the case of ‘smaller’ D , [DDS21] proved that $\overline{\Sigma^{[k]}\Pi\Sigma} \subseteq \text{VBP}$; formally any $f \in \overline{\Sigma^{[k]}\Pi\Sigma}$, of approximative circuit size s , can be exactly computed by (affine projection of) determinant of size $s^{\exp(k)}$. This raises a basic open question: $\overline{\Sigma^{[k]}\Pi\Sigma} \stackrel{?}{\neq} \text{VBP}$ (even, $\overline{\Sigma^{[k]}\Pi\Sigma} \stackrel{?}{\neq} \text{VNP}$ is left open!).

Border depth-3 circuits: A geometric view. Theoretical computer scientists are interested in proving ‘robust’ lower bounds, i.e. lower bound results & techniques which would also work under limit. This could be roughly translated into asymptotic geometry terms as follows: *Given a sequence of some ‘nice’ vector spaces V_n , and sequences of points and groups, does the inclusion (by inclusion, we mean the points under the group action in V_n) fail for every $n \geq n_0$, for some n_0 ?* The *Chow variety* is one of the simplest varieties studied in the field of algebraic geometry; it is believed to be a good testing ground for GCT [Lan15, Section 2]. Interestingly, this goes back to 19th century mathematics, studied independently by Hermite (1854) and Hadamard (1897).

Informally, if one specializes to group of diagonal matrices and takes the orbit closure, one obtains the famous *Chow variety*, $\text{Ch}_d(W) \subset \mathbb{P}S^d W$; usually $W = \mathbb{C}^n$ (or, for $\det_n / \text{perm}_n, \mathbb{C}^{n^2}$). We denote by $S^d W$, the space of polynomials of degree d on W^* , $\mathbb{P}V$ denotes the projective space and we denote $[v]$ as a corresponding point. Then, formally,

$$\text{Ch}_d(W) := \{[z] \in \mathbb{P}S^d W \mid z = w_1 \dots w_d, \text{ for } w_j \in W\}.$$

Therefore, one can define the *Chow rank* of a homogeneous polynomial f of degree d , denoted $\text{rank}_{\text{Ch}}(f)$, to be the minimum k such that $f = \sum_{i=1}^k \prod_{j=1}^d \ell_{ij}$, where ℓ_{ij} are linear forms. Often in the literature, $\text{rank}_{\text{Ch}}(f) = k$ is equivalently expressed as: the smallest k such that f (as a point) is in $\sigma_k^0(\text{Ch}_d(W))$ (= set of points with $\text{Ch}_d(W)$ -rank at most k).

Moving to the approximative setting, one defines *Chow border rank*, $\overline{\text{rank}}_{\text{Ch}}(f)$, as the border analogue of the Chow rank. In other words, $\overline{\text{rank}}_{\text{Ch}}(f) = k \iff f \in \sigma_k(\text{Ch}_d(W)) = \text{the Zariski closure in } \mathbb{P}S^d W \text{ of } \sigma_k^0(\text{Ch}_d(W))$; it is called the k -th *secant variety*¹ of the Chow variety of W . For details, refer to [Lan15, Lan17]. These two ranks happen to exactly coincide with the depth-3 respectively border depth-3 homogeneous circuits of f , with the smallest fanin k .

In general, from algebraic complexity perspective, we are interested in the *non-homogeneous* setting. For instance, Kumar’s expression [Kum20, Section 3.1] is non-homogeneous. However, with suitable padding (& $W = \mathbb{C}^{n+1}$), Kumar’s result translates into geometric terms: *For any degree- d homogeneous polynomial f , there exists a linear form ℓ such that $\ell^{m-d} f \in \sigma_2(\text{Ch}_m(W))$, for $m = \exp(n, d)$, or equivalently, $\overline{\text{rank}}_{\text{Ch}}(\ell^{m-d} f) = 2$.* On the other hand, if one restricts $m = \text{poly}(n, d)$, Dutta, Dwivedi and Saxena [DDS21] showed that $\sigma_2(\text{Ch}_m(W)) \subseteq \text{VBP}$; the same result holds if one replaces σ_2 by σ_k , as long as k is a constant. They called this phenomenon *de-bordering*.

Quest for lower bounds. Perhaps, the upper bound (de-bordering) results are must-to-understand, to demystify the limitations and power of computations/approximations in different models. However, identifying explicit polynomials which are hard to compute/approximate, and proving it

¹For a given variety X , s -th secant variety of X is Zariski closure of the union of all secant $(s - 1)$ -planes to X .

remains a major template in algebraic complexity. A significant result of Baur and Strassen [Str73, BS83] shows existence of an explicit n -variate degree- d polynomial which requires circuits of size at least $\Omega(n \cdot \log d)$. Though for $d = \text{poly}(n)$, it gives *superlinear* lower bound, ideally one would hope for at least a super-polynomial, optimistically even exponential, lower bound. Since it can be proved that ‘most’ polynomials require exponential size circuits to compute/approximate.

Similarly, if one restricts the model of computation, we have better lower bounds; for eg. formulas. Kalorkoti [Kal85] showed $\Omega(n^2 / \log n)$ formula lower bound for an explicit n -variate polynomial, which could be easily extended to the approximative setting. We refer the readers to [Sap19] for a comprehensive survey of lower bounds in algebraic complexity.

The situation in GCT, is no way better and near to what was expected at its inception. The linchpin of the GCT program was that the permanent and determinant are both *uniquely* characterized (up to a constant factor) by their symmetries; and moreover, lower bounds are equivalent to orbit closure containment [Gro12, Section 3.3.2]. Therefore, the ‘simplest’ way of proving lower bound would be to find *occurrence obstructions*, i.e. finding an irreducible representation with multiplicity for permanent larger than that of determinant. However a strongly ‘negative’ result by Bürgisser, Ikenmeyer, and Panova [BIP19] almost refuted this approach.

On the contrary, a recent breakthrough result by Limaye, Srinivasan & Tavenas [LST21] showed the first superpolynomial lower bound against general algebraic circuits/formulas of *constant*-depth, over all fields of characteristic 0 or large. Since their method is linear-rank-based, the proofs can be lifted in the border classes analogously [Gro15]. This gives us a stronger urge to continue the quest to show *exponential* separation, in the constant-depth regime using non-linear/non-geometric approaches. The current work is a step in that pursuit.

1.2 Our results: The fanin-hierarchy theorem

We state our result formally now. Our result holds for any field of characteristic 0/large characteristic.

Theorem 1 (Fanin-hierarchy exp-gap). *Fix any constant $k \geq 1$. There is an explicit n -variate, degree $< n$ polynomial f with $\Sigma^{[k+1]}\Pi\Sigma$ size $O(n)$, while f requires an exponential ($=2^{\Omega(n)}$) size $\Sigma^{[k]}\Pi\Sigma$ -circuit.*

Remarks. 1. Interestingly, **Theorem 1** is *optimally* exponential for $k \geq 2$, as $\binom{2n}{n} = 2^{\Theta(n)}$ is also an upper bound in Kumar’s [Kum20] fanin-2 representation. Thus, we have completely characterized the gap between two constant-top-fanin border depth-3 circuits.

2. Consider the following degree- d (> 2) polynomial on $3d$ -variables $f(x) := x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + x_{2d+1} \cdots x_{3d}$. Note that $f(x)$ has a trivial $\Sigma^{[3]}\Pi\Sigma$ circuit of size $O(d)$. Surprisingly, our proof method actually shows that f has a $\Sigma^{[2]}\Pi\Sigma$ -border complexity of $2^{\Theta(d)}$.

A word on the polynomial family. Our candidate polynomial family, for a fixed k , is the *sum-product polynomial*² $\mathcal{P}_{k+1} := (P_{k+1,d})_d$, where

$$P_{k+1,d} := \sum_{i \in [k+1]} \prod_{j=1}^d x_{(i-1)d+j}.$$

This is a strict generalization of the inner product polynomial IP_{k+1} , defined before. Using [Kum20, Theorem 1.3], it follows that $IP_{k+1} \in \Sigma^{[2]}\Pi^{[O(k)]}\Sigma$. Therefore, we cannot show a strong size lower bound against IP_{k+1} . This can be mitigated by working with general d -degree monomials (instead

²This term has been taken from [Lan17, Section 8.12.2].

of quadratic ones); though the technique to prove an exponential lower bound is far from obvious due to the ‘limit’ operator.

Clearly, $P_{k+1,d}$ is a multilinear degree- d polynomial on $(k+1) \cdot d$ -variables. Lower bounds for \mathcal{P}_2 has been studied by Shpilka [Shp02], in a different context. This polynomial is closely related to the Trace-Iterated-Matrix-Multiplication, $\text{Tr-IMM}_{k+1,d}$, which is the trace of the product of d -many $(k+1) \times (k+1)$ symbolic matrices X_r , $\text{tr}(\prod_{r \in [d]} X_r)$, where the (i,j) -th entry of the matrix X_r is $x_{i,j}^{(r)}$. In particular, take diagonal matrices X_r with (i,i) -th entry being the variable $x_{(i-1)d+r}$. Clearly, $P_{k+1,d} = \text{tr}(X_1 \cdots X_d)$. Trivially, this is a restriction on X_r , implying that our proof holds for $\text{Tr-IMM}_{k+1,d}$ as well.

Non-triviality and implications. 1. If we restrict ourselves to only *homogeneous* setting, then it is easy to argue that $P_{k+1,d}$, *cannot* be computed by a homogeneous $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuit. The proof directly extends to the border, where substituting k many linear forms to 0, one from each product, makes the circuit 0 while it cannot make $P_{k+1,d}$ vanish; for a proof see [Corollary 3](#). However, this impossibility result no longer holds when we allow *non-homogeneity* [Kum20]. The naive proof methods fail miserably, as setting many affine linear functions to zero leads to *inconsistency*. Eg. $x_1 - \epsilon x_2$ and $x_1 - \epsilon x_3 - 1$ are never both zero in the border (as $\epsilon \rightarrow 0$); while at the same time x_2, x_3 contribute in the computation due to the use of the powers of $(1/\epsilon)$. This is what lends the, seemingly innocuous, border computing model $\overline{\Sigma^{[2]}\Pi\Sigma}$, the universal expressive power!

2. Moreover, it was asked in [DDS21] whether $\overline{\Sigma^{[k]}\Pi\Sigma} \neq \text{VBP}$ or not. Our proof of [Theorem 1](#), in fact, shows an exponential separation between $\overline{\Sigma^{[k]}\Pi\Sigma}$ and $\Sigma^{[k+1]}\Pi\Sigma$. Consequently, we have exponentially-separated $\overline{\Sigma^{[k]}\Pi\Sigma}$ from VF (respectively VBP, VNP).

3. Further, ours is probably the first strong depth-3 lower bound in the border setting which is *non-rank*-based, non-geometric (absence of algebraic geometry abstraction), and does not really care about the upper bound on the ϵ -power that hides in the expression.

4. The same proof can be adapted for \det_d and perm_d analogously. Interestingly, [Theorem 1](#) answers a restricted version of a question asked in [Lan15, Problem 8.10]; namely, $\ell^{m-d} \cdot \det_d \notin \sigma_r(\text{Ch}_m(W))$, for any $m \leq 2^{o(d)}$ and constant r . Geometrically, [Theorem 1](#) exponentially separates the padded k -th-secant varieties of the Chow variety of $W = \mathbb{C}^{n+1}$, for *all* constants k .

Comparing our parameters with [LST21] breakthrough. Interestingly, our proof method works with both $\text{IMM}_{k+1,d}$ (where instead of trace, one is interested in $(1,1)$ -th entry), and the trace version $\text{Tr-IMM}_{k+1,d}$, already defined before. The linear-rank-based lower bound method by Limaye, Srinivasan & Tavenas [LST21] shows ‘only’ a superpolynomial separation between depth-3 (unbounded fanin) circuits and IMM, with different parameters. Their dominating variable is n (with $d = o(\log n)$) and they showed an $n^{\sqrt{d}}$ lower bound (which further weakens above depth-4); which is not optimal (as optimal would be n^d). While, we show an optimal lower bound of $\approx \binom{(k+1)d+d}{d} \approx 2^{\Omega(d)}$, since we take $k = O(1)$ and our dominating variable is d .

Moreover, [LST21] shows a *superpolynomial-gap* depth-hierarchy; while our result is about an exponential-gap fanin-hierarchy inside the depth-3 regime. In summary, the parameters and the results in [LST21] are incomparable with ours, and their method is very different since our method is *non-linear* at heart.

1.3 An overview of the proof

At a very high level, the idea behind our result is to show that if a ‘robust’ polynomial family can be computed by small bounded-top-fanin border depth-3 circuits, then after some suitable transformation, the polynomials can also be approximated by a ratio of two small depth-3 diagonal circuits. Recall the definition of *depth-3 diagonal* circuits, denoted as $\Sigma \wedge \Sigma$; they compute polynomials of the form $\sum_i \ell_i^{e_i}$, for linear polynomials ℓ_i . Informally, the robustness means that after substituting constantly many variables to 0, the polynomials still continue to carry its core properties. The proof concludes by showing that there are well-known hard polynomial families, for eg. \mathcal{P}_{k+1} , which after substituting constantly ($\leq k$)-many variables to 0, can not be approximated by a ratio of two sub-exponential size $\Sigma \wedge \Sigma$ circuits.

Till now it sounds uncomplicated, but the proof soon stumbles upon technical difficulties which need serious care & machinery. We will try to point out a few of them without obscuring the main idea; while asking the reader for a leap of faith and referring to the correct sections if needed.

Power series and dlog. In the proof, we will use the ring of *formal power series* $R[[x_1, \dots, x_n]]$ (in short $R[[x]]$), for some suitable ring R , see [Niv69, DSS18, Sin19]. One of the key benefit of this ring comes from the *inverse* identity: $(1 - x)^{-1} = \sum_{i \geq 0} x^i$.

The logarithmic derivative operator $\text{dlog}_z(f) = (\partial_z f) / f$ is another key tool which *linearizes* the product gate, since

$$\text{dlog}_y(f \cdot g) = \partial_y(fg) / (fg) = (f \cdot \partial_y g + g \cdot \partial_y f) / (fg) = \text{dlog}_y(f) + \text{dlog}_y(g).$$

This operator enables us to use power-series expansion, and converts the \prod -gate to \wedge .

Looking modulo an ideal generated by linear forms. Given a polynomial $f(x)$ and a linear form $\ell(x)$, let us try to understand what $f(x) \bmod \ell$ means; since a generalization of this will play a key role to define the ‘robustness’. If $\ell := x_1$, it is just looking at $f(0, x_2, \dots, x_n)$. For an arbitrary ℓ , we apply a suitable isomorphism ϕ , on the space of linear forms, which sends ℓ to, let us say, x_1 . Since, $\phi(f) = f(Ax)$, for some invertible matrix A , $f \bmod \ell$ essentially translates to looking at $f(Ax) \bmod x_1$. This is nothing but equal to $f(\ell_1, \dots, \ell_n)$, for some linear forms ℓ_i which are x_1 -free, and $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n) = n - 1$.

Similarly, this can be extended to looking at $f(x) \bmod \langle L_1, \dots, L_t \rangle$, for some suitable t , where $\langle L_1, \dots, L_t \rangle$ is the ideal generated by t -many linear forms L_i , for $i \in [t]$, i.e., any $g \in \langle L_1, \dots, L_t \rangle$ must be of the form $g =: \sum_{i \in [t]} a_i \cdot L_i$, for some $a_i \in \mathbb{F}[x]$. In that case, we are looking at $f(\ell_1, \dots, \ell_n)$ with $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n) \geq n - t$; one could think of $\{\ell_1, \dots, \ell_n\}$ as containing at least $\{x_{t+1}, \dots, x_n\}$. This perspective will play a crucial role in the proofs; see [Lemma 2](#).

Lower bound for $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits: The overall plan

Let $x := \{x_1, \dots, x_n\}$, $n := (k + 1)d$, and polynomial $f := P_{k+1,d} \in \mathbb{F}[x]$. Suppose that $f \in \overline{\Sigma^{[k]}\Pi\Sigma}$ of size s , i.e. $g = f + \epsilon \cdot S$, where $\text{size}_{\mathbb{F}(\epsilon)}(g) \leq s$ (as a $\Sigma^{[k]}\Pi\Sigma$ -circuit), $S \in \mathbb{F}[\epsilon, x]$. We want to understand the parameter s as a function of k and d . Assume that $1 \leq k < d$.

► **$k = 1$ case.** [BIZ18, Prop. A.12] showed that $\overline{\Pi\Sigma} = \Pi\Sigma$. Eventually, the proof relies on the fact that the product distributes the ϵ -powers showing $\overline{(\Pi\mathcal{C})} \subseteq \Pi(\overline{\mathcal{C}})$, for any reasonable class \mathcal{C} (see [Lemma 14](#)); here by $f \in \Pi\mathcal{C}$ we mean $f =: \prod f_i$, where $f_i \in \mathcal{C}$. Thus, it suffices to show that $P_{2,d}$ does not have *any* $\Pi\Sigma$ circuit (i.e. impossibility). It is evident that $P_{2,d}$ is an irreducible polynomial. One way to argue is: since, $P_{2,d}$ is a homogeneous polynomial, all its factors must be

homogeneous. Thus, take a linear form ℓ that divides $P_{2,d}$. As argued in [Lemma 2](#), there are linear forms ℓ_1, \dots, ℓ_n with $n = 2d$ such that

$$P_{2,d} \bmod \ell = P_{2,d}(\ell_1, \dots, \ell_n), \text{ and } \text{rank}(\ell_1, \dots, \ell_n) = n - 1.$$

Thus, it is easy to deduce: $\text{RHS} = \prod_{j \in [d]} \ell_j + \prod_{j \in [d]} \ell_{d+j} \neq 0$. This contradicts our assumption: $\ell \mid P_{2,d}$.

Unfortunately, there is no analogous result to eliminate ϵ -powers in $\overline{\Sigma^{[2]}\Pi\Sigma}$. So the above simple idea does not extend to $k = 2$; hence we cannot deduce impossibility of $P_{3,d}$. However, we can continue with a part of the basic proof template, and improvise the machinery.

► **$k = 2$ case: A synopsis.** Recall from the definition, $g := T_1 + T_2 = f + \epsilon \cdot S$ where T_1, T_2 are multiplication terms ($\Pi\Sigma$ -circuits over $\mathbb{F}(\epsilon)[x]$) and $f := P_{3,d}$. The sum gate makes it hard to give any relevant information. However, if we can somehow reduce it to $k = 1$ case ‘carefully’, our job gets done. To do that, let us focus on the structure of the polynomials T_i . By taking out ϵ , we could assume that each T_i is a product of an ϵ -power, and s -many linear polynomials in $\mathbb{F}[\epsilon, x]$, none of which is divisible by ϵ . One of the three things can happen.

- 1 (Easy case). Both T_i have at least one factor whose ϵ -free term is a linear form over \mathbb{F} ;
- 2 (Intermediate case). Exactly one of T_i , say wlog, T_1 , has at least one factor whose ϵ -free term is a linear form;
- 3 (Hard case). None of the factors of T_i , has ϵ -free term as a linear form.

• Handling the first case. The first case is almost similar to (but not exactly) $k = 1$. Suppose, a linear form $\ell_1(x, \epsilon) \mid T_1$, over $\mathbb{F}(\epsilon)$. Ideally, we would like to look at $g \bmod \ell_1$ which reduces fanin to 1.

Understanding $g \bmod \ell_1$. Wlog, $\ell_1 =: x_1 + \epsilon \tilde{\ell}_1$, where $\tilde{\ell}_1$ is x_1 -free; if $\ell_1(\epsilon = 0) \neq x_1$, we can always apply a suitable isomorphism & relabel. Interestingly, isomorphisms would not really change the proof structure; for details see [Lemma 2](#). Note that, $g \bmod \ell_1$ is *equivalent* to looking at $g|_{x_1 = -\epsilon \cdot \tilde{\ell}_1}$. The immediate question would be what happens to the limit, after this substitution. Interestingly,

$$\lim_{\epsilon \rightarrow 0} g(-\epsilon \cdot \tilde{\ell}_1, x_2, \dots, x_n) = \lim_{\epsilon \rightarrow 0} f(-\epsilon \cdot \tilde{\ell}_1, x_2, \dots, x_n) = f(0, x_2, \dots, x_n).$$

So, it is like looking at $f|_{x_1=0}$.

Since, f is homogeneous, so is $f(0, x_2, \dots, x_n)$. By the reduced fanin, we have $f(0, x_2, \dots) \in \overline{\Pi\Sigma}$, which looks similar to studying $P_{2,d}$. Thus, an argument similar to the above ($k = 1$) finishes the proof of the first case. We call this case ‘easy’ because we just had to work with an ideal of appropriate linear polynomials and do substitutions.

• Handling the second case. In the second case, we again work with linear form $\ell_1 \mid T_1$ and the above argument shows that $f(0, x_2, \dots) \in \overline{\Pi\Sigma}$, a *non-homogeneous* circuit; since each linear factor of T_2 is non-homogeneous by assumption. But this cannot happen as $f(0, x_2, \dots)$ is homogeneous and from $k = 1$ case, $\Pi\Sigma$ circuit must be homogeneous (as we have simply projected from $P_{3,d}$). This again leads to impossibility.

However, we point out that for general k , this strategy will just reduce this case to border of $k - 1$ sum of non-homogeneous product $\Pi\Sigma$ -circuits (assuming ℓ_1 was the easy case instance, otherwise we take modulo more linear forms, i.e. a non-rank-preserving projection of f via [Lemma 2](#)) which really *is* the third case. It is thus justified to call this the ‘intermediate’ step.

• Handling the third case. This is the hardest among the three, since the modulo idea becomes unreliable due to inhomogeneous ideals. Primary counter-example being: if each $\ell_i \mid T_1$ (similarly

T_2) is such that $\ell_i|_{\epsilon=0} = 1$, then the above substitution idea becomes nonsensical! For e.g., take $k = 2$, $\ell_1 = 1$ and $\ell_2 = 1 + \epsilon \cdot \ell$, for some non-zero linear form $\ell \in \mathbb{F}[x]$. Strikingly, this is exactly the format of expressing arbitrary f as border of $T_1 + T_2$ in [Kum20, Section 3.1] (T_1 was picked as an ϵ -power). Clearly, this implies that the *all-non-homogeneous* case should be the proof's core part. Our analysis in the above also shows that it suffices to show lower bound for this specific all-non-homogeneous case.

The basic idea now is to show that (i) assuming the all-non-homogeneous structure and after some operations, f can be almost-written as an $\text{poly}(s)$ -sized $\overline{\Sigma} \wedge \overline{\Sigma}$ circuit, (ii) $f = P_{3,d}$, requires $2^{\Omega(d)}$ -sized $\overline{\Sigma} \wedge \overline{\Sigma}$ to compute, implying $s \geq 2^{\Omega(d)}$, as desired!

- **Technical details: A bird's overview.** To start with, we apply a simple variable-scaling map $\Phi : \mathbb{F}(\epsilon)[x] \rightarrow \mathbb{F}(\epsilon)[x, z]$ that scales $x_i \mapsto z \cdot x_i$. This makes z the "degree counter" as it helps track the degree of the polynomial; but more importantly it allows us univariate derivation. Now, let the maximum power of ϵ dividing T_2 be a_2 (which could be < 0), i.e., $\Phi(T_2) = \epsilon^{a_2} \cdot \tilde{T}_2$. Dividing both sides by \tilde{T}_2 , we get

$$\Phi(g)/\tilde{T}_2 = \epsilon^{a_2} + \Phi(T_1)/\tilde{T}_2 \xrightarrow{\text{differentiate wrt } z} \partial_z(\Phi(g)/\tilde{T}_2) = \partial_z(\Phi(T_1)/\tilde{T}_2),$$

where $\partial_z(\cdot) := d(\cdot)/dz$. This has reduced the number of summands on the right hand side to 1, although the surviving summand has become more complicated now. Further, (seemingly) we have no control on what happens as $\epsilon \rightarrow 0$.

What is $\lim_{\epsilon \rightarrow 0} \partial_z(\Phi(g)/\tilde{T}_2)$? Note: (i) Φ is a scaling map and f is homogeneous, and (ii) \tilde{T}_2 is invertible *both* in the ring of power series in z and ϵ , and $\lim_{\epsilon \rightarrow 0} \tilde{T}_2 =: c \bmod z$, for $c \in \mathbb{F} \setminus \{0\}$. Using these, it can be seen that $\partial_z(\Phi(g)/\tilde{T}_2) \equiv c' \cdot z^{d-1} \cdot f(x) \bmod \langle z^d, \epsilon \rangle$, for nonzero constant c' (see [Claim 5](#)). Formally, the following equation is well-defined:

$$c' \cdot z^{d-1} \cdot f(x) \equiv \partial_z(\Phi(g)/\tilde{T}_2) \equiv \partial_z(\Phi(T_1)/\tilde{T}_2) \bmod \langle \epsilon, z^d \rangle. \quad (1)$$

Unfortunately, we have completely disfigured the model by introducing a division gate (and then taking a mod). This is exactly where *logarithmic derivative* (aka dlog) enters, with a bunch of helpful properties. In particular, the expression $\partial_z(\Phi(T_1)/\tilde{T}_2)$ can be re-written as

$$\Phi(T_1)/\tilde{T}_2 \cdot \text{dlog}(\Phi(T_1)/\tilde{T}_2) = (\Phi(T_1)/\tilde{T}_2) \cdot (\text{dlog}(\Phi(T_1)) - \text{dlog}(\tilde{T}_2)).$$

Note that the dlog operator distributes the product gate into summation giving $\text{dlog}(\Pi\Sigma) = \Sigma \text{dlog}(\Sigma)$, where Σ denotes linear polynomials. Observe that $\text{dlog}(\Sigma) = \Sigma/\Sigma \in \Sigma \wedge \Sigma$, the depth-3 powering circuits, over an appropriate function-ring $\mathcal{R}(\epsilon, x)$, where $\mathcal{R} := \mathbb{F}[z]/\langle z^d \rangle$. To achieve this, analytically expand $1/\ell$, where ℓ is a linear polynomial dividing $\Phi(T_i)$, $i \in [2]$, as sum of powers of linear functions; using the inverse identity:

$$1/(1 - a \cdot z) \equiv 1 + a \cdot z + \dots + a^{d-1} \cdot z^{d-1} \bmod z^d.$$

Since $\Sigma \wedge \Sigma$ is 'closed' under taking product and addition ([Appendix B](#)), we obtain a final $\Sigma \wedge \Sigma$ circuit for $\text{dlog}(\Phi(T_1)/\tilde{T}_2)$. Details of this step can be found in [Claim 7](#). Therefore, $\partial_z(\Phi(T_1)/\tilde{T}_2)$ is actually in a *bloated* class – $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma \wedge \Sigma)$, over $\mathcal{R}(\epsilon, x)$ – which computes elements of the form $(A/B) \cdot C$, where $A, B \in \Pi\Sigma$ and $C \in \Sigma \wedge \Sigma$. Moreover, from [Equation 1](#), computing $\lim_{\epsilon \rightarrow 0} \partial_z(\Phi(T_1)/\tilde{T}_2) \bmod z^d$ suffices.

The crucial thing about these $\Pi\Sigma$ circuits, that appeared, is that their $\Pi\Sigma|_{z=0=\epsilon}$ is a nonzero constant. Hence, the coefficient of the minimum z -power, does not have contribution from these $\Pi\Sigma$ circuits at all (in the limit). Further, de-bordering, for a product gate, is distributive ([Lemma 14](#)).

Consequently, we get that $\lim_{\epsilon \rightarrow 0} \partial_z(\Phi(T_1)/\tilde{T}_2) \bmod z^d$, has a $\text{poly}(s)$ -size $\overline{\Sigma} \wedge \overline{\Sigma}$ circuit. In partic-

ular, coefficient of z^{d-1} has also an $\overline{\Sigma\wedge\Sigma}$ -circuit of $\text{poly}(s)$ -size, by simple interpolation (Lemma 12). Comparing this with the LHS in Equation 1, we get that f can be computed by a $\text{poly}(s)$ -size $\overline{\Sigma\wedge\Sigma}$. We point out that $\overline{\Sigma\wedge\Sigma}$ circuit has an important property that it has *small* partial-derivative space (see [CKW11, Lemma 10.2] and Lemma 8), which will be crucial to show the lower bound below.

To show the lower bound, a simple cone-size argument (using partial-derivatives) shows: the *cone-size* [the number of monomials dividing the given monomial; for a formal definition see section 2] of the leading monomial (wrt some monomial ordering) in f is 2^d ; whereas for the $\overline{\Sigma\wedge\Sigma}$, it should have an upper bound of $\text{poly}(s)$. This implies the desired lower bound of $s \geq 2^{\Omega(d)}$; while trivially $f = P_{k+1,d}$ has $\overline{\Sigma^{[k+1]}\Pi\Sigma}$ -size $O(d)$. For details, see Lemma 8.

► **Extending the proof to general k .** The general constant k case is a bit more technical and it depends on a more general bloated class (it is in depth-5):

$$\text{Gen}(k, s) := \Sigma^{[k]} (\Pi\Sigma / \Pi\Sigma) (\Sigma\wedge\Sigma / \Sigma\wedge\Sigma) ,$$

they compute elements of the form $\sum_{i=1}^k (U_i / V_i) \cdot (P_i / Q_i)$, where $U_i, V_i \in \Pi\Sigma$, invertible in the ring, and $P_i, Q_i \in \Sigma\wedge\Sigma$, and the circuit (with division allowed) has size s ; see Definition 2. Moreover, we will show three important properties:

- (i) if $P_{k+1,d}$ was computed by a $\overline{\Sigma^{[k]}\Pi\Sigma}$ -circuit, then there are linear forms ℓ_1, \dots, ℓ_n , with $n := (k+1)d$, such that $\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$, and $P_{k+1,d}(\ell_1, \dots, \ell_n)$ can be computed by an ‘all-non-homogeneous’ $\overline{\Sigma^{[m]}\Pi\Sigma}$ -circuit, for some $m \leq k$ (Lemma 2),
- (ii) $P_{k+1,d}(\ell_1, \dots, \ell_n) \cdot (\overline{\Sigma\wedge\Sigma} / \overline{\Sigma\wedge\Sigma})$, is the coefficient of the *minimum* z -power in the limit of a small $\text{Gen}(1, \cdot)$ circuit, see induction-hypotheses-4 in subsection 4.1,
- (iii) the coefficient of the minimum z -power in $\lim_{\epsilon \rightarrow 0} \text{Gen}(1, \cdot)$ is a ratio of two ‘small size’ $\overline{\Sigma\wedge\Sigma}$ circuits (Claim 7).

Though this circuit ‘division’ along with the non-rank-preserving projection make things scarier, we keep calm and manage to adapt the above cone-size comparison, to argue an exponential lower bound on s ; see Lemma 8.

(im) Possibility of a ‘simpler’ proof, and comparison with [DDS21]

The crux of this lower bound approach is to — (1) ‘convert’ the general problem into all-non-homogeneous setting (to reap the advantages of $\Pi\Sigma|_{z=0=\epsilon}$ being a nonzero constant) and then, (2) do the standard dimension argument carefully. So, one would wonder whether we can just *shift* the x -variables randomly at first, achieving the desired non-homogeneity, and proceed without doing the case-analysis! However, if we work with $P_{k+1,d}(x + a)$, $a \in \mathbb{F}^n$, and variable-scale to get $P_{k+1,d}(zx + a)$, this becomes very hard to handle even after one division and derivation (subsection 4.1). Since $\partial_z(P_{k+1,d}(zx + a))$ ‘spreads’ the coefficients of $P_{k+1,d}$ across different z -powers; we cannot use anymore the last-step argument we gave above for $P_{k+1,d} \cdot (\overline{\Sigma\wedge\Sigma} / \overline{\Sigma\wedge\Sigma})$. This product gets stuck in the notorious polynomial multiplication *convolution*.

To elaborate, let us say, for $k = 2$, there is a size- s depth-3 fanin-2 circuit approximating $P_{k+1,d}$; after shifting-and-scaling, dividing, deriving and using dlog and finally taking limit, we will eventually get that

$$\partial_z (P_{k+1,d}(zx + a) / \Pi\Sigma) \equiv \overline{\Sigma\wedge\Sigma} \cdot (\Pi\Sigma) \pmod{z^d} ,$$

for $\text{poly}(s)$ -sized $\overline{\Sigma\wedge\Sigma}$ and $\Pi\Sigma$ -circuits. From this expression showing an exponential lower bound on s is not clear at all. The non-homogeneity kills the property (which holds in our proof sketched) that minimum z -power carries the full ‘hardness information’ of $P_{k+1,d}$ (see [subsection 4.1](#), induction-hypothesis-4). Moreover, for a larger k , it is even worse, since we do not understand how to ‘lift’ (back to f) the lower bound from the polynomial after doing $(k - 1)$ -many times ‘division and derivation’ to f . We remark that lifting requires interpolation, which is why [\[DDS21\]](#) moved to ABP/ABP, instead of our weaker model $\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$ (in fact, it was analyzed via ARO/ARO). We avoid lifting in our current proof case by exploiting the minimum z -power property.

Since the homogeneity of f is crucial for our proof to work, it requires us to reduce the border circuit to all-non-homogeneous setting (namely, the ‘hard case 3’ above). We do this carefully without any variable-shift ([Lemma 2](#)), so that the homogeneity of f is maintained and yet ‘all’ the linear functions are invertible! For a ‘coarser’ upper bound of ABP, [\[DDS21\]](#) did not require these innovations at all.

After reducing to the all-non-homogeneous case, we do use DiDIL technique to analyze $\overline{\Sigma^{[k]}\Pi\Sigma}$, introduced in [\[DDS21\]](#). DiDIL is an acronym for the steps: **D**ivide, **D**erive, **I**nduct, with **L**imit. In this paper, DiDIL process is applied in a new, somewhat simplified, setting; namely, to the all-non-homogeneous case spelled in [Lemma 2](#). Since, we do not use any shifting unlike in [\[DDS21\]](#), this really gives us the advantage to closely study the target polynomial $P_{k+1,d}$, and yields certain ‘bloated’ structures, which need an intricate analysis; for details see [subsection 4.1](#).

1.4 Known depth-3 lower bounds and their limitations

In this section, we briefly discuss about the well-known depth-3 circuit lower bounds (mostly in the classical setting), their techniques, and why they fail to yield our result in the border.

In a very influential piece of work, Nisan and Wigderson [\[NW96\]](#) showed that over any field \mathbb{F} , any *homogeneous* $\Sigma\Pi\Sigma$ circuit computing the determinant \det_d must be of size $2^{\Omega(d)}$. This uses partial-derivatives method and thus can be easily adapted in the border setting. We remark that the lower bound is actually on the top-fanin and thus for constant top-fanin k , \det_d can not be even computed by a homogeneous $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuit. For an alternative non-rank based proof, see [Lemma 2](#). Unfortunately, partial-derivative measure behaves very badly when we allow non-homogeneity, simply because the degree-bound gets lost, and thus the proof fails miserably. Moreover, if we focus on $P_{k+1,d}$ instead of \det_d or $\text{IMM}_{k+1,d}$, the exponential gap of k vs. $(k + 1)$, cannot be shown via rank methods since a naive calculation of the trivial upper bound and lower bounds do not yield anything useful. For similar reasons, techniques from Grigoriev and Karpinski [\[GK98\]](#), Grigoriev and Razborov [\[GR00\]](#), over *fixed* finite field, fails.

There have been significant work on restricted depth-3 circuit lower bounds where the restriction is on the bounded independence, bounded read/occur, bounded bottom-fanin etc. A major theme of the proofs depend on the following reductions: It reduces the (non-homogeneous) depth-3 circuits of top-fanin s to a subclass of depth-5 circuits where the top-fanin gets blown up to $s \cdot \exp(\sqrt{d})$ [\[SW01, GKKS16\]](#), use a random restriction to obtain a (homogeneous) depth-4 decomposition and then show lower bound using variant of shifted partial derivative measure [\[KLSS17, KS17, KS16\]](#). Although these proofs can be adapted in the border setting, they fail to give any meaningful hierarchy-theorem, or exponential-gap, for non-homogeneous $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits.

Even for showing lower bound for $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuits, we can not probably expect to use factoring or chinese remaindering (CRT) based ideas, since $\text{mod } \langle \epsilon^M \rangle$, we get non-unique, usu-

ally exponentially many, factors. For e.g., $x^2 \equiv (x - a \cdot \epsilon^{M/2}) \cdot (x + a \cdot \epsilon^{M/2}) \pmod{\langle \epsilon^M \rangle}$; for all $a \in \mathbb{F}$. In this case, there are, in fact, infinitely many factorizations. Moreover, $\lim_{\epsilon \rightarrow 0} 1/\epsilon^M \cdot (x^2 - (x - a \cdot \epsilon^{M/2}) \cdot (x + a \cdot \epsilon^{M/2})) = a^2$. Therefore, infinitely many factorizations may give infinitely many limits, and possibilities, and thus the analysis becomes much more intricate.

In the classical affine settings, it is known that computing immanant (which includes determinant and permanent) requires exponential size $\Sigma^{[k]}\Pi\Sigma$ circuits [ASSS16, Theorem 1.7]. Jacobian was used to show such strong lower bounds. However, it is not at all clear how it behaves (or defined?) wrt $\lim_{\epsilon \rightarrow 0}$. For eg. let $f_1 = x_1 + \epsilon^M \cdot x_2$, and $f_2 = x_1$, where M is arbitrary large. Then, the underlying Jacobian $J(f_1, f_2) = \epsilon^M$ is nonzero; but it becomes zero in the limit. Seemingly, this makes the whole Jacobian machinery collapse in the border setting; as it cannot possibly give a variable reduction for the border model. (E.g., one needs to keep faithful to both x_1 and x_2 above.)

Though not exactly comparable, but [BIZ18] showed a counter-intuitive ‘collapse’ in the border setting: $\overline{\text{VBP}}_2 = \overline{\text{VBP}}_3 = \dots = \overline{\text{VBP}}_k$, for any constant k , whereas it is also known that $\text{VBP}_2 \subsetneq \text{VF} = \text{VBP}_3 = \dots = \text{VBP}_k$ [BOC92, AW16]. The techniques used in these works (to show ‘collapse’) are quite different and do not help in achieving our result (to show strict ‘hierarchy’).

Lastly, in subsection 1.2, we have already compared and pointed out the similarities and dissimilarities with [LST21]. This concludes the comparisons, and obstacles, in the prior works.

2 Notation and preliminaries

In this section, we describe some of the assumptions and notations used throughout the paper.

Notation. Denote $[n] = \{1, \dots, n\}$, and $\mathbf{x} = (x_1, \dots, x_n)$. We use $\mathbb{F}[[x]]$, to denote the ring of formal power series over \mathbb{F} . Formally, $f = \sum_{i \geq 0} c_i x^i$, with $c_i \in \mathbb{F}$, is an element in $\mathbb{F}[[x]]$. Further, $\mathbb{F}(x)$ denotes the function field, where the elements are of the form f/g , where $f, g \in \mathbb{F}[x]$ ($g \neq 0$).

We call $\ell =: a_1 x_1 + \dots + a_n x_n$, a linear polynomial without the constant term, as a *linear form*.

Throughout the paper, by $\text{rank}(\ell_1, \dots, \ell_n)$, we mean $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n)$, i.e. the dimension of the linear space generated by the linear forms $\{\ell_1, \dots, \ell_n\}$.

Explicit. A family $\{P_n \in \mathbb{F}[x_1, \dots, x_n] \mid n \geq 1\}$, is said to be *explicit* if there is a deterministic algorithm that given as input 1^n and a monomial m over the variables x_1, \dots, x_n , computes the coefficient of the monomial m in P_n in time $\text{poly}(n)$.

Logarithmic derivative. Over a ring R and a variable y , the *logarithmic derivative* $\text{dlog}_y : R[y] \rightarrow R(y)$ is defined as $\text{dlog}_y(f) := \partial_y f / f$; here ∂_y denotes the partial derivative wrt variable y . One important property of dlog is that it is *additive* over a product as $\text{dlog}_y(f \cdot g) = \partial_y(fg) / (fg) = (f \cdot \partial_y g + g \cdot \partial_y f) / (fg) = \text{dlog}_y(f) + \text{dlog}_y(g)$. [*dlog linearizes product*]

Circuit size. Some of the complexity parameters of a circuit are – 1) *size*, the total number of nodes and edges, 2) *depth*, the number of layers, 3) *degree*, the maximum degree polynomial computed by any node, and 4) *fanin*, the maximum number of inputs to a node.

Depth-2 & Depth-3 circuits. Product depth-2 circuits, denoted as $\Pi\Sigma$, compute polynomials of the form $\prod_i \ell_i$, where ℓ_i are affine linear polynomials. Depth-3 circuits with top-fanin k are denoted as $\Sigma^{[k]}\Pi\Sigma$; they compute polynomials of the form $\sum_{i \in [k]} \prod_j \ell_{ij}$, where ℓ_{ij} are affine linear functions. Also, depth-3 diagonal circuits are denoted as $\Sigma^{[s]} \wedge \Sigma$; they compute polynomials of the form $\sum_{i=1}^s \ell_i^{\epsilon_i}$, for linear polynomials ℓ_i . When we write $\Sigma \wedge \Sigma$, it means the top-fanin is *unbounded* (and thus trivially bounded by the circuit size).

Ideal generated by linear forms. For given n -variate linear forms L_1, \dots, L_r , we denote $\langle L_1, \dots, L_r \rangle$, the ideal generated by L_i , for $i \in [r]$, which contains elements of the form $\sum_{i \in [r]} a_i \cdot L_i$, for $a_i \in \mathbb{F}[x]$.

Operation on Complexity Classes. For class \mathcal{C} and \mathcal{D} defined over ring R , our bloated model is any combination of sum, product, and division of polynomials from respective classes. For instance, $\mathcal{C}/\mathcal{D} = \{f/g : f \in \mathcal{C}, 0 \neq g \in \mathcal{D}\}$, similarly $\mathcal{C} \cdot \mathcal{D}$ for products, $\mathcal{C} + \mathcal{D}$ for sum, and other possible combinations. Also we use \mathcal{C}_R to denote the basic ring R on which \mathcal{C} is being computed over.

Valuation. Valuation is a map $\text{val}_y : R[y] \rightarrow \mathbb{Z}_{\geq 0}$, over a ring R , such that $\text{val}_y(\cdot)$ is defined to be the maximum power of y dividing the element. It can be easily extended to localized/fraction ring $R(y)$, by defining $\text{val}_y(p/q) := \text{val}_y(p) - \text{val}_y(q)$; where the integer value can be negative.

Field. We denote the underlying field as \mathbb{F} and assume that it is of characteristic 0 (eg. \mathbb{Q}, \mathbb{Q}_p etc.). All our results hold for other fields (eg. \mathbb{F}_{p^e} etc.) of large characteristic p .

Approximative closure. For an algebraic complexity class \mathcal{C} , the ‘approximation’ is formally modeled as follows [BIZ18, Definition 2.1].

Definition 1 (Approximative closure of a class). *Let $\mathcal{C}_{\mathbb{F}}$ be a class of polynomials defined over a field \mathbb{F} . Then, $f(x) \in \mathbb{F}[x_1, \dots, x_n]$ is said to be in approximative closure $\overline{\mathcal{C}}$ if and only if there exists polynomial $Q \in \mathbb{F}[\epsilon, x]$ such that $\mathcal{C}_{\mathbb{F}(\epsilon)} \ni g(x, \epsilon) = f(x) + \epsilon \cdot Q(x, \epsilon)$. In short, we denote $\lim_{\epsilon \rightarrow 0} g(x, \epsilon) := f(x)$.*

Cone-size of monomials. For a monomial x^a , the cone of x^a is the set of all sub-monomials of x^a . The cardinality of this set is called *cone-size* of x^a . It equals $\prod_{i \in [n]} (a_i + 1)$, where $a = (a_1, \dots, a_n)$. We will denote $\text{cs}(m)$, as the cone-size of the monomial m .

3 Hardness lies in all-non-homogeneity

We want to prove lower bounds on border depth-3 circuits $\overline{\Sigma^{[k]}\Pi^{[d]}\Sigma}$; these circuits compute polynomials f , where $\epsilon^M \cdot f(x) + \epsilon^{M+1} \cdot S(x, \epsilon) = \sum_{i \in [k]} T_i$, where each $T_i = \prod_{j \in [d]} \ell_{ij}$, where $\ell_{ij} \in \mathbb{F}[\epsilon][x]$, are affine linear functions. Since, in the homogeneous setting, exponential lower bound is easy-to-show, we somehow want to ‘exclude’ homogeneous polynomials ℓ_i that contribute to approximating the polynomial and focus *only* on all-non-homogeneity.

Intuitively speaking, since Kumar’s expression [Kum20] also involved the polynomials T_i with each linear factor being strictly affine (nonzero constant), these cases should be the ‘hard’ cases. We also briefly mentioned about this being the hardest to analyze, in subsection 1.3. Motivated thus, here is an important ‘reduction’ which shows that proving lower bound in general, reduces to proving for the ‘all-non-homogeneous’ case, i.e. every linear factor is affine with nonzero constant. The following lemma holds for a general n -variate polynomial $f(x)$.

Lemma 2 (Reduction to all-non-hom. setting). *If n -variate f is computed by an s -size $\overline{\Sigma^{[k]}\Pi\Sigma}$ -circuit, then there exists $0 \leq m \leq k$, and (homogeneous) linear forms ℓ_1, \dots, ℓ_n such that*

1 (large rank). *Rank* $(\ell_1, \dots, \ell_n) = n - t$, for some $0 \leq t \leq k - m$, and each ℓ_i is $\{x_1, \dots, x_t\}$ -free;

2 (all-non-hom.). *There exists $H(x, \epsilon) \in \mathbb{F}[x, \epsilon]$ and $M \geq 0$, such that*

$$\epsilon^M \cdot f(\ell_1, \dots, \ell_n) + \epsilon^{M+1} \cdot H(x, \epsilon) = T_1 + \dots + T_m,$$

where $\forall i \in [m]$,

$$\left(\frac{T_i}{\epsilon^{\text{val}_{\epsilon}(T_i)}} \right) \Big|_{x=\epsilon=0} \in \mathbb{F} - \{0\} \text{ and } \sum_i \text{size}_{\mathbb{F}(\epsilon)}(T_i) \leq O(s \cdot n).$$

Remarks. 1. $t = 0$ is the case where we do not need to reduce since it is already in all-non-homogeneous case; so, we simply take $\ell_i := x_i, \forall i \in [n]$. Cases $t \geq 1$ need more care and a longer proof.

2. $m = 0$ implies that $\epsilon^M \cdot f(\ell_1, \dots, \ell_n) = 0 \pmod{\epsilon^{M+1}} \implies f(\ell_1, \dots, \ell_n) = 0$. Though for an arbitrary polynomial f , this non-rank-preserving projection can make f zero, but for ‘robust’ polynomials, for e.g. $f := P_{k+1,d}$, one can argue that it *cannot* be zero. For details see the proof of [Corollary 3](#).

Proof of Lemma 2. Suppose, $\epsilon^M \cdot f + \epsilon^{M+1} \cdot S(x, \epsilon) = T_1 + \dots + T_k$, where $T_i \in \mathbb{F}[\epsilon, x]$, product of linear polynomials. Now if already each T_i is such that after taking out the maximum ϵ -power, ϵ -free term $\pmod{\langle x_1, \dots, x_n \rangle} \in \mathbb{F} - \{0\}$, then we are already done. Otherwise, it must happen that at least one of the T_i has a linear factor whose ϵ -free term is a linear form.

Formally, after possible relabeling, $T_1 =: g_1 \cdot h_1$, in $\mathbb{F}[\epsilon, x]$, where $g_1|_{\epsilon=0}$ is a nonzero linear form. Here is the first step which reduces the fanin from k to $k - 1$. We denote $T_i =: T_{i0}$, and $S(x, \epsilon) =: S_0(x, \epsilon)$ (since this is the base case or the 0-th step). Also, let $\mathcal{R} := \mathbb{F}[\epsilon] / \langle \epsilon^{M+1} \rangle$. We work in $\mathbb{F}[x, \epsilon] \cap \mathcal{R}[x]$; we would not keep mentioning the underline ring, unless it changes.

First step: Reduction from k to $(\leq k - 1)$. Let $g_1 =: g_{10} + \epsilon \cdot g_{11}$, where $g_{10} \in \mathbb{F}[x]$ is ϵ -free, and $g_{11} \in \mathbb{F}[\epsilon, x]$. Since g_{10} is a nonzero linear form by assumption, extend $\{1, g_{10}\}$ to a basis say $\{1, L_1 := g_{10}, L_2, \dots, L_n\}$ of the linear space $V_n := \{\sum_{i=1}^n a_i x_i + a_0 \mid \forall a_i \in \mathbb{F}\}$, which has dimension $n + 1$. Define an isomorphism ϕ_1 on V_n , which maps $L_i \mapsto x_i$, for all i .

Let us write $\phi_1(g_1) =: (1 + \epsilon \cdot p_1(\epsilon)) \cdot x_1 + \epsilon \cdot R_1$, where R_1 is a polynomial which is x_1 -free, and $p_1(\epsilon) \in \mathbb{F}[\epsilon]$; this can be achieved by clubbing all the x_1 coefficients together which must look like $1 + \epsilon \cdot p_1(\epsilon)$, since $\phi_1(g_{10}) = x_1$ and other terms of x_1 must come from multiple of ϵ terms.

By assumption, $\epsilon^M \cdot f + \epsilon^{M+1} S_0 = T_{10} + \dots + T_{k0}$; apply ϕ_1 both side and then substitute $x_1 := -\epsilon \cdot R_1 / (1 + \epsilon \cdot p_1(\epsilon))$, which essentially sends $g_1 \mapsto 0$. We denote the last substitution map by ψ_1 . We want to *stress*: since $1 / (1 + \epsilon \cdot p_1(\epsilon)) \pmod{\langle \epsilon^{M+1} \rangle}$ exists, and belongs to $\mathbb{F}[\epsilon]$, we can think of substituting $x_1 := -\epsilon \cdot q_1(\epsilon) \cdot R_1$, where $q_1 := 1 / (1 + \epsilon \cdot p_1(\epsilon)) \pmod{\epsilon^{M+1}}$. By doing this, we get the following equation:

$$\epsilon^M \cdot \psi_1 \circ \phi_1(f(x)) + \epsilon^{M+1} \cdot \psi_1 \circ \phi_1(S_0) = \sum_{i=k_1+1}^k \psi_1 \circ \phi_1(T_{i0}). \quad (2)$$

In the last equation, we used the fact that $\psi_1 \circ \phi_1(g_1) = 0 \implies \psi_1 \circ \phi_1(T_{10}) = 0$. In the above process, the fanin may reduce even further (and not just by 1), since the map can ‘kill’ some other T_{ij} as well. We may assume that the first k_1 -many T_{ij} vanish. Therefore, the reduced fanin becomes $k - k_1$. Denote, for $i \in [k_1 + 1, k]$:

$$T_{i1} := \psi_1 \circ \phi_1(T_{i0}) \text{ and } v_{i1} := \text{val}_\epsilon(T_{i1}).$$

Note that T_{i1} are elements in $\mathbb{F}[\epsilon, x] \cap \mathcal{R}[x]$ which are x_1 -free. Just to state the obvious, if one of them becomes 0 after applying ψ_1 , the summand further reduces. Also, size of T_{i1} may blowup by a factor of n , without changing its degree, since the linear-transformation $\psi_1 \circ \phi_1$ is acting on just n -variate linear functions.

Also, denote $\lim_{\epsilon \rightarrow 0} \psi_1 \circ \phi_1(f) =: f(\ell_1^{(1)}, \dots, \ell_n^{(1)}) =: D_1(x)$ (which is x_1 -free), for linear forms $\ell_i^{(1)} \in \mathbb{F}[x]$, where $\text{rank}(\ell_1^{(1)}, \dots, \ell_n^{(1)}) = n - 1$; since ψ_1 can be thought of as a single homogeneous

zero-constraint. Then, Equation 2 implies that (with abusing notation x , which now means x_1 -free):

$$\epsilon^M \cdot D_1(x) + \epsilon^{M+1} \cdot (\text{blah}) = \sum_{i=k_1+1}^k T_{i1}.$$

Now if, for all $i \in [k_1 + 1, k]$, $(T_{i1}/\epsilon^{v_{i1}})|_{\epsilon=0} \bmod \langle x_1, \dots, x_n \rangle \in \mathbb{F} - \{0\}$, that means we are done as promised! Otherwise, there must exist an $i \in [k_1 + 1, k]$, say w.l.o.g., $i = k_1 + 1$, such that $(T_{i1}/\epsilon^{v_{i1}})$ has a linear factor (say g_2) with ϵ -free term being a linear form, and we proceed similarly as before. Moreover, the inner structure of f changes in a bit more complicated way. To specify the details, we induct.

$(j + 1)$ -th step: Fanin reduction from $k - k_j$ to $\leq (k - k_j - 1)$. For $j \geq 1$, let us say we have

$$\epsilon^M \cdot D_j + \epsilon^{M+1} \cdot S_j(x, \epsilon) = \sum_{i=k_j+1}^k T_{ij},$$

where –

1. $v_{ij} := \text{val}_\epsilon(T_{ij})$ and $(T_{ij}/\epsilon^{v_{ij}})$ has a linear factor (say g_{j+1}) with ϵ -free term being a linear form, along with $\deg(T_{ij}) \leq s$. Moreover, $k_j \geq j$.
2. $D_j = \psi_j \circ \phi_j(D_{j-1}) = f(\ell_1^{(j)}, \dots, \ell_n^{(j)})$, where linear map ϕ_j is invertible, and linear map ψ_j has the property that it sets x_1, \dots, x_j to 0, but keeps the other $n - j$ variables unchanged. Equivalently, $\ell_i^{(j)} \in \mathbb{F}[x]$ are linear forms, which are $\{x_1, \dots, x_j\}$ -free, with $\text{rank}(\ell_1^{(j)}, \dots, \ell_n^{(j)}) = n - j$.

In the base case, of course, we used the notation: $D_0 = f(x)$ and $\ell_i^{(0)} = x_i$. Note that, we need Condition (1), otherwise we are already done and achieved the form we wanted!

We follow the similar strategy as in the base case. Let, after possible relabeling, $T_{k_j+1,j} =: g_{j+1} \cdot h_{j+1}$, over $\mathbb{F}[\epsilon, x]$ where $g_{j+1}|_{\epsilon=0}$ is a nonzero linear form in x_{j+1}, \dots, x_n . Let $g_{j+1} =: g_{j+1,0} + \epsilon \cdot g_{j+1,1}$. Since $g_{j+1,0}$ is a linear form, extend $\{1, x_1, \dots, x_j, g_{j+1,0}\}$ to a basis say $\{1, L_1 := x_1, L_2 := x_2, \dots, L_j := x_j, L_{j+1} := g_{j+1,0}, L_{j+2}, \dots, L_n\}$.

Define an isomorphism ϕ_{j+1} on V_n which maps $L_i \mapsto x_i$, for $i \in [n]$. Let us write $\phi_{j+1}(g_{j+1}) =: (1 + \epsilon \cdot p_{j+1}(\epsilon)) \cdot x_{j+1} + \epsilon \cdot R_{j+1}$, where $p_{j+1}(\epsilon) \in \mathbb{F}[\epsilon]$, and R_{j+1} is a polynomial which is $\{x_1, \dots, x_j, x_{j+1}\}$ -free; this can be achieved similarly as before, by clubbing all the x_{j+1} coefficients together which must look like $1 + \epsilon \cdot p_{j+1}(\epsilon)$.

Apply ϕ_{j+1} both side and substitute $x_{j+1} := -\epsilon \cdot R_{j+1}/(1 + \epsilon \cdot p_{j+1}(\epsilon))$, which essentially sends $g_{j+1} \mapsto 0$. We denote the substitution map by ψ_{j+1} . Here again we want to *stress*: since $\epsilon/(1 + \epsilon \cdot p_{j+1}(\epsilon)) \in \mathcal{R}$, we can think of substituting $x_{j+1} := -\epsilon \cdot q_{j+1}(\epsilon) \cdot R_{j+1}$, where $q_{j+1} := 1/(1 + \epsilon \cdot p_{j+1}(\epsilon)) \bmod \epsilon^{M+1}$. By doing this, we get the following equation:

$$\epsilon^M \cdot \psi_{j+1} \circ \phi_{j+1}(D_j(x)) + \epsilon^{M+1} \cdot \psi_{j+1} \circ \phi_{j+1}(S_{j+1}) = \sum_{i=k_j+1}^k \psi_{j+1} \circ \phi_{j+1}(T_{ij}). \quad (3)$$

In Equation 3 we used the fact that $\psi_{j+1} \circ \phi_{j+1}(g_{j+1}) = 0$; moreover, it kills many such $T_{i,j}$'s (say from $i = k_j + 1$ to k_{j+1}). This reduces the fanin to $k - k_{j+1} < k - k_j$; since we have $k_{j+1} > k_j \geq j$. This proves hypothesis-1-part-2.

Denote $\lim_{\epsilon \rightarrow 0} \psi_{j+1} \circ \phi_{j+1}(D_j) =: D_{j+1}$, which is $x_{\leq j+1}$ -free. Because of this new zero-constraint, the rank decreases by *exactly* 1; this proves hypothesis-2 (rank = $n - j - 1$).

Let $T_{i,j+1} := \psi_{j+1} \circ \phi_{j+1}(T_{i,j})$ and $v_{i,j+1} := \text{val}_\epsilon(T_{i,j+1})$. Note that, $\deg(T_{i,j+1}) \leq s$ remains an invariant, due to linear transformations. Therefore, one can write [Equation 3](#) as follows:

$$\epsilon^M \cdot D_{j+1}(\mathbf{x}) + \epsilon^{M+1} \cdot (\text{blah}) = \sum_{i=k_{j+1}+1}^k T_{i,j+1}.$$

Note that we have shown the desired hypothesis-1-part-2. If hypothesis-1-part-1 happens, then we proceed the induction; otherwise we have already achieved the desired form and we stop. In that case, let the number of reductions be t , therefore, $k_t \geq t$.

Finally we are in the all-non-homogeneous, where fanin becomes $m := k - k_t \leq k - t \implies t \leq k - m$. The size blowup is also trivial to see since all we have done is a linear transformation on n variables (and at the beginning, there are at most s many polynomials to consider). This finishes the inductive reduction process and the proof. \square

As a warmup application, we prove a ‘folklore’ impossibility result in *homogeneous* models.

Corollary 3 (Impossibility result). *For $1 \leq k < d$, the polynomial $P_{k+1,d}$ is uncomputable by homogeneous $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits.*

Proof. The proof is by contradiction. Suppose there is a $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuit computing $P_{k+1,d}$. From the above reduction process, we will eventually get $n := (k+1)d$ -many linear forms $\{\ell_1, \dots, \ell_n\}$, with $r := \text{rank}(\ell_1, \dots, \ell_n) \geq n - k$, such that $\epsilon^M \cdot P_{k+1,d}(\ell_1, \dots, \ell_n) = 0 \pmod{\epsilon^{M+1}}$; which implies it has to be *exactly* 0 (as after dividing by ϵ^M both side, LHS is ϵ -free). We show that this cannot happen.

To argue, $P_{k+1,d}(\ell_1, \dots, \ell_n) \neq 0$, from the above structure in [Lemma 2](#), ℓ_i are free of the variables x_1, \dots, x_{n-r} . Remaining variables have to occur in some ℓ_i , by the ‘large rank’ hypothesis. We could assume $\{x_{n-r+1}, \dots, x_n\}$ is a subset of $\{\ell_i \mid i \in [n]\} =: \mathcal{B}$; as we can apply an isomorphism that changes the basis of \mathcal{B} , without affecting the nonzeroness.

By definition, $P_{k+1,d}(\ell_1, \dots, \ell_n) = \sum_{i \in [k+1]} \prod_{j \in [d]} \ell_{(i-1)d+j}$. It is easy to see that $\prod_{j \in [d]} \ell_{(i-1)d+j}$ must be of the form $\mathbf{m}_i \cdot$ (product of t_i non-variable linear forms), for some $t_i \leq n - r$, since there are $n - r$ many non-variable linear forms available and \mathbf{m}_i is a *multilinear* monomial of degree $d - t_i$; here non-variable linear form means that there are at least two variables with nonzero coefficient in the linear form. Moreover, since $n - r < k + 1$, one of the $\prod_{j \in [d]} \ell_{(i-1)d+j}$, for some $i \in [k+1]$, remains ‘untouched’ i.e. it is just a pure *multilinear* monomial \mathbf{m}_i , a product of d -many distinct variables. We now claim that for any $[k+1] \ni i' \neq i$, the monomials generated by a single product $\prod_{j \in [d]} \ell_{(i'-1)d+j}$ can *overlap* with \mathbf{m}_i in at most $n - r \leq k < d$ many variables (very few!) in its support.

To argue, suppose, $\prod_{j \in [d]} \ell_{(i'-1)d+j} = \mathbf{m}_{i'} \cdot$ (product of $t_{i'}$ non-variable linear forms). Since the monomials in $P_{k+1,d}(\mathbf{x})$ have *disjoint* support and \mathcal{B} has $\leq (n - r)$ non-variables, \mathbf{m}_i can intersect with a monomial generated by $\prod_{j \in [d]} \ell_{(i'-1)d+j}$, in at most $t_{i'} \leq n - r \leq k$ many variables because \mathbf{m}_i and $\mathbf{m}_{i'}$ are variable disjoint. This proves the above claim. The ‘few overlapping variables’ phenomenon readily implies the nonzeroness of $P_{k+1,d}(\ell_1, \dots, \ell_n)$. This finishes the proof. \square

By [Lemma 2](#), it suffices to solve the following (restricted linear factors of T_i ’s) lower bound to prove [Theorem 1](#). Note: (1) $P_{k+1,d} =: f$, $n := (k+1)d$, with $d > k \geq 1$ and the hardest case is when $t := k - m$, and (2) We multiply $1/\epsilon^M$ in RHS representation, but this does not change the fundamental structure of polynomials T_i .

Problem 1 (Reduced lb problem). For a constant k , say $P_{k+1,d}(\ell_1, \dots, \ell_n) + \epsilon \cdot H(\mathbf{x}, \epsilon) = T_1 + \dots + T_m$, where –

1. (large rank). ℓ_i are linear forms (which are $\{x_1, \dots, x_{k-m}\}$ -free) such that $\text{rank}(\ell_1, \dots, \ell_n) = n - (k - m)$, for some $m \in [k]$, with $1 \leq k < d$ and,
2. (all-non-hom.). T_i are product of linear polynomials such that $(T_i / \epsilon^{v_i})|_{x=\epsilon=0} \in \mathbb{F} - \{0\}$ (where $v_i := \text{val}_\epsilon(T_i)$).

Then, $\sum_i \text{size}(T_i) \geq 2^{\Omega(d)}$.

Note that, solving the above problem suffices to prove **Theorem 1**. As **Lemma 2** shows that solving **Problem 1** implies: the size s of the approximative $\Sigma^{[k]}\Pi\Sigma$ -circuit for $P_{k+1,d}$ must be $\geq 2^{\Omega(d)}$, as we desired. A couple of important pointers before proceeding further:

1. As already argued in **Corollary 3**, $P_{k+1,d}(\ell_1, \dots, \ell_n)$ can not be a zero polynomial, i.e. $(P_{k+1,d})_d$ is a ‘robust’ family of polynomials. This is exactly why in **Problem 1**, $m \geq 1$, otherwise $m = 0$ readily implies $P_{k+1,d}(\ell_1, \dots, \ell_n) = 0$, a contradiction!
2. Interestingly, the same robustness also holds for $\text{IMM}_{k+1,d}$, det_d , perm_d etc. Moreover, our proofs (**subsection 4.1** and **Theorem 1**) hold for these polynomials as well. But we will only work with the polynomials $P_{k+1,d}$; as it suffices for our exponential-gap hierarchy-theorem.
3. $P_{k+1,d}(\ell_1, \dots, \ell_n)$ is a degree- d polynomial, with $n - t = (k + 1)d - (k - m)$ variables.
4. Since $P_{k+1,d}$ is a homogeneous polynomial and ℓ_i are linear forms, the polynomial $P_{k+1,d}(\ell_1, \dots, \ell_n)$ is still homogeneous.

4 Proving lower bound for the core case: **Problem 1**

Assume that $P_{k+1,d}(\ell_1, \dots, \ell_n) =: f(\ell_1, \dots, \ell_n)$ can be computed by the special $\overline{\Sigma^{[m]}\Pi\Sigma}$, as in the statement of **Problem 1**, of size s and we want to prove an exponential lower bound. Now, the proof will go in a different direction; we will induct over a more general circuit class. For $P_{k+1,d}$, we use $n := (k + 1)d$, throughout as the number of variables x .

Definition 2 (Bloated model, [DDS21]). We call a circuit $\mathcal{C} \in \text{Gen}(m, s)$, over the fraction-ring $R(x)$, with parameter m and size s , if it computes $f \in R(x)$ where $f = \sum_{i \in [k]} T_i$, such that $T_i = (U_i / V_i) \cdot (P_i / Q_i)$, with $U_i, V_i, P_i, Q_i \in R[x]$ such that

1. $U_i, V_i \in \Pi\Sigma$; with $U_i \bmod \langle x_1, \dots, x_n \rangle$ and $V_i \bmod \langle x_1, \dots, x_n \rangle$ invertible in R , and
2. $P_i, Q_i \in \Sigma\wedge\Sigma$.

Further, $\text{size}(\mathcal{C}) := \sum_{i \in [m]} \text{size}(T_i)$, and $\text{size}(T_i) := \text{size}(U_i) + \text{size}(V_i) + \text{size}(P_i) + \text{size}(Q_i)$.

It is evident that size- s circuit $\Sigma^{[m]}\Pi\Sigma$ lies in $\text{Gen}(m, s)$, which will be our general model of induction. The lower bound proof (of **Problem 1**) can be now divided into two parts:

1. Reducing **Problem 1** to proving lower-bound on $\text{Gen}(1, \cdot)$, in some appropriate ring; see **Lemma 4**.
2. Proving lower bound for $\overline{\text{Gen}(1, \cdot)}$, see **Lemma 8**.

4.1 Reducing to bloated fanin-1 model $\overline{\text{Gen}(1, \cdot)}$

In this section, we prove the first part which is to reduce it to fanin-1. Now we use the DiDIL-technique developed in [DDS21]. By hypothesis, each $T_i \bmod \langle x \rangle$ has nonzero constant-term (after extracting the ϵ -power). So, for the existence of $1/T_i$, DiDIL-technique *does not* need any additional shift: a mere scaling by a new variable z suffices (which is required for derivation and keeping a counter on the x -degree). Recall that our target polynomial is $f(x) := P_{k+1,d}(x_1, \dots, x_n)$, of degree d with $n := (k+1)d$.

Lemma 4 (Fanin-1 reduction). *If hypotheses of Problem 1 is true with $f(\ell_1, \dots, \ell_n)$ being approximated by $\Sigma^{[m]}\Pi\Sigma$ -circuits of size s , over $\mathbb{F}(\epsilon)$, then there is a rational function $g \in \mathbb{F}(x, \epsilon, z)$ such that*

1. g can be computed by a $\text{Gen}(1, s^{O(m7^m)})$ -circuit mod $z^{d'}$, over $\mathbb{F}(\epsilon)$, for some $d' \in [d+1]$,
2. $\lim_{\epsilon \rightarrow 0} g$ exists,
3. the coefficient of the minimum z -power in $\lim_{\epsilon \rightarrow 0} g$ is of the form $f(\ell_1, \dots, \ell_n) \cdot (\overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma})$, where $\overline{\Sigma \wedge \Sigma}$ are z -free circuits in $\mathbb{F}[x]$, of size and degree bounded by $s^{O(m7^m)}$.

Proof. We will prove the theorem by induction and reducing the top-fanin by 1 at each step.

Bloat out: Reducing $\overline{\Sigma^{[m]}\Pi\Sigma}$ to $\overline{\text{Gen}(m-1, \cdot)}$. Let $g_0 := f_0 + \epsilon \cdot S_0$, where $f_0 := f(\ell_1, \dots, \ell_n)$ and $g_0 := \sum_{i \in [m]} T_{i,0}$, such that $T_{i,0}$ is computable by a $\Pi\Sigma$ -circuit of size at most s over $\mathbb{F}(\epsilon)$ with the property: after dividing by the ϵ -power of ϵ -valuation, it is a nonzero constant mod $\langle x_1, \dots, x_n \rangle$. Moreover, define $U_{i,0} := T_{i,0}$ and $V_{i,0} := P_{i,0} := Q_{i,0} = 1$ as the base input case (of $\text{Gen}(1, \cdot)$). Also, let $d_0 := d+1$.

Φ : The scaling map. To ensure invertibility and facilitate derivation, we define a homomorphism (essentially a variable-scaling):

$$\Phi : \mathbb{F}(\epsilon)[x] \rightarrow \mathbb{F}(\epsilon)[x, z], \quad \text{such that } x_i \mapsto z \cdot x_i.$$

Note that the $\Pi\Sigma$ circuits (i.e. factors of $U_{i,0}$) are *invertible* mod z^{d_0} . We will be working with different ring $\mathcal{R}_i(x)$, at i -th step of induction, with $\mathcal{R}_0 := \mathbb{F}[z] / \langle z^{d_0} \rangle$, and think of the z -variables as ‘cost-free’. So, our target is to reduce the fanin in successive steps without changing the ‘core’ structure of f_0 .

Divide and Derive. Assume that $\Phi(T_{i,0}) := \epsilon^{a_{i,0}} \cdot \tilde{T}_{i,0}$, where $\tilde{T}_{i,0} := t_{i,0} + \epsilon \cdot \tilde{t}_{i,0}(x, z, \epsilon)$ (thus $t_{i,0} = \tilde{T}_{i,0}|_{\epsilon=0}$). Note that, $0 = v_{i,0} := \text{val}_z(\tilde{T}_{i,0})$. We divide $\Phi(g_0)$ by $\tilde{T}_{m,0}$ and derive wrt z :

$$\begin{aligned} \Phi(f_0)/\tilde{T}_{m,0} + \epsilon \cdot \Phi(S_0)/\tilde{T}_{m,0} &= \epsilon^{a_{m,0}} + \sum_{i \in [m-1]} \Phi(T_{i,0})/\tilde{T}_{m,0} \quad [\text{Divide}] \\ \implies \partial_z (\Phi(f_0)/\tilde{T}_{m,0}) + \epsilon \cdot \partial_z (\Phi(S_0)/\tilde{T}_{m,0}) &= \sum_{i \in [m-1]} \partial_z (\Phi(T_{i,0})/\tilde{T}_{m,0}) \quad [\text{Derive}] \\ &= \sum_{i \in [m-1]} (\Phi(T_{i,0})/\tilde{T}_{m,0}) \cdot \text{dlog} (\Phi(T_{i,0})/\tilde{T}_{m,0}) \quad (4) \\ &=: g_1. \end{aligned}$$

Definability. Let $\mathcal{R}_1 := \mathbb{F}[z] / \langle z^{d_1} \rangle$, and $d_1 := d_0 - 1 = d$. For $i \in [m-1]$, define

$$T_{i,1} := (\Phi(T_{i,0})/\tilde{T}_{m,0}) \cdot \text{dlog}(\Phi(T_{i,0})/\tilde{T}_{m,0}), \quad \text{and } f_1 := \partial_z (\Phi(f_0)/t_{m,0}).$$

Claim 5. Circuit g_1 approximates f_1 correctly, i.e. $\lim_{\epsilon \rightarrow 0} g_1 = f_1$, where g_1 (respec. f_1) is well-defined over $\mathcal{R}_1(\epsilon, \mathbf{x})$ (respec. $\mathcal{R}_1(\mathbf{x})$). Moreover, the coefficient of the minimum z -power in f_1 is a constant (in \mathbb{F}^*) multiple of f_0 .

Proof. As we divide by the minimum valuation, by **Lemma 9** we have

$$\text{val}_z(\Phi(T_{i,0})/\tilde{T}_{m,0}) \geq 0 \implies \Phi(T_{i,0})/\tilde{T}_{m,0} \in \mathbb{F}(\mathbf{x}, \epsilon)[[z]] \implies T_{i,1} \in \mathbb{F}(\mathbf{x}, \epsilon)[[z]].$$

We already remarked that f_0 is a homogeneous degree- d polynomial; hence $\Phi(f_0) = z^d \cdot f_0$, and $t_{m,0} \in \mathbb{F}[\mathbf{x}]$; actually by assumption, $t_{m,0}$ must be of the form $c \cdot \prod(1 + z \cdot A_j)$, for linear forms A_j . Hence, $\Phi(f_0)/\tilde{T}_{m,0} \in \mathbb{F}(\mathbf{x}, \epsilon)[[z]]$. Moreover, $(\Phi(f_0)/\tilde{T}_{m,0})|_{\epsilon=0} = \Phi(f_0)/t_{m,0} \in \mathbb{F}(\mathbf{x}, z)$. Combining these it follows that

$$\Phi(f_0)/t_{m,0} \in \mathbb{F}(\mathbf{x})[[z]] \implies f_1 = \partial_z(\Phi(f_0)/t_{m,0}) = \partial_z(z^d \cdot f_0/t_{m,0}) \in \mathbb{F}(\mathbf{x})[[z]].$$

Since $t_{m,0} \equiv c \pmod{z}$, for some nonzero constant c , the minimum power of z in f_1 is indeed a constant multiple of f_0 from the above equation.

Once we know that each $T_{i,1}$ and f_1 are well-defined power-series, we claim that Eqn. (4) holds mod z^{d_1} . Note that, $\Phi(f_0) + \epsilon \cdot \Phi(S_0) = \sum_{i \in [m]} T_{i,1}$, holds mod z^{d+1} . Thus after dividing by the 0-valuation element (wrt z), it still holds mod z^{d+1} ; finally after differentiation it must hold mod z^d .

Further, since $\lim_{\epsilon \rightarrow 0} \tilde{T}_{m,0}$ exists, we must have $f_1 = \partial_z(\Phi(f_0)/t_{m,0}) = \lim_{\epsilon \rightarrow 0} g_1$; i.e. g_1 approximates f_1 correctly, over $\mathcal{R}_1(\mathbf{x})$. This finishes **Claim 5**. \square

However, we stress that we also think of these intermediate expressions as elements over $\mathbb{F}(\mathbf{x}, z, \epsilon)$, with z -degree being ‘kept track of’ (which could be $> d$). All these different ‘lenses’ of looking and computing will be important later.

Quest for lower fanin. We need to induct using the structure of $T_{i,1}$; $T_{i,0}$ was rather ‘special’ and we managed easily. Now we have to work with general bloated models. We will eventually show that each $T_{i,1}$ has small $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ -circuit over $\mathcal{R}_1(\mathbf{x}, \epsilon)$.

Inductive step (j -th step): Reducing $\overline{\text{Gen}(m-j, \cdot)}$ to $\overline{\text{Gen}(m-j-1, \cdot)}$. Suppose, we are at the j -th ($j \geq 1$) step. Our induction hypothesis assumes–

1. $\sum_{i \in [m-j]} T_{i,j} =: g_j$, over $\mathcal{R}_j(\mathbf{x}, \epsilon)$, such that it approximates f_j correctly, where $f_j \in \mathcal{R}_j(\mathbf{x})$, where $\mathcal{R}_j := \mathbb{F}[z]/\langle z^{d_j} \rangle$.
2. Here, $T_{i,j} =: (U_{i,j}/V_{i,j}) \cdot (P_{i,j}/Q_{i,j})$, where $U_{i,j}, V_{i,j} \in \Pi\Sigma$ and $P_{i,j}, Q_{i,j} \in \Sigma\wedge\Sigma$, each in $\mathcal{R}_j(\epsilon)[\mathbf{x}]$. Each can be thought as an element in $\mathbb{F}(\mathbf{x}, z, \epsilon) \cap \mathbb{F}(\mathbf{x}, \epsilon)[[z]]$ as well.

Assume that the syntactic degree of each denominator and numerator of $T_{i,j}$ is bounded by D_j . This will be required in the size analysis.

3. Let $v_{i,j} := \text{val}_z(T_{i,j}) \geq 0$, for $i \in [m-j]$. Note that, unlike in $j = 0$, unfortunately we cannot always claim $v_{i,j}$ to be 0.

Therefore, without loss of generality, assume that $\min_i v_{i,j} = v_{m-j,j}$. Moreover, $U_{i,j}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$ (similarly for $V_{i,j}$), a unit in $\mathbb{F}(\epsilon)$.

4. If $\text{val}_z(g_j) = b_j$ (when viewed as element in $\mathbb{F}(\mathbf{x}, \epsilon)[[z]]$), then $(g_j/z^{b_j})|_{z=0} = f_0 \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, where $\Sigma\wedge\Sigma$ are z -free (in $\mathbb{F}(\epsilon)[\mathbf{x}]$). Equivalently, the coefficient of the minimum z -power in f_j , is of the form $f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$, where both $\overline{\Sigma\wedge\Sigma}$ circuits are nonzero.

Quick pointers. Eventually we will divide and derive like the $j = 0$ -th step done above, without applying any new homomorphism, to reduce the fanin further by 1.

Divide and Derive. Let $T_{m-j,j} =: \epsilon^{a_{m-j,j}} \cdot \tilde{T}_{m-j,j}$, where $\tilde{T}_{m-j,j} =: (t_{m-j,j} + \epsilon \cdot \tilde{t}_{m-j,j})$ is not divisible by ϵ . Divide $g_j =: f_j + \epsilon \cdot S_j$, by $\tilde{T}_{m-j,j}$, to get:

$$\begin{aligned} f_j / \tilde{T}_{m-j,j} + \epsilon \cdot S_j / \tilde{T}_{m-j,j} &= \epsilon^{a_{m-j,j}} + \sum_{i \in [m-j-1]} T_{i,j} / \tilde{T}_{m-j,j} \\ \implies \partial_z (f_j / \tilde{T}_{m-j,j}) + \epsilon \cdot \partial_z (S_j / \tilde{T}_{m-j,j}) &= \sum_{i \in [m-j-1]} \partial_z (T_{i,j} / \tilde{T}_{m-j,j}) \\ &= \sum_{i \in [m-j-1]} (T_{i,j} / \tilde{T}_{m-j,j}) \cdot \text{dlog} (T_{i,j} / \tilde{T}_{m-j,j}) \quad (5) \\ &=: g_{j+1}. \end{aligned}$$

Definability. Let $\mathcal{R}_{j+1} := \mathbb{F}[z] / \langle z^{d_{j+1}} \rangle$, where $d_{j+1} := d_j - v_{m-j,j} - 1$. For $i \in [m-j-1]$, define

$$T_{i,j+1} := (T_{i,j} / \tilde{T}_{m-j,j}) \cdot \text{dlog} (T_{i,j} / \tilde{T}_{m-j,j}), \text{ and } f_{j+1} := \partial_z (f_j / \tilde{T}_{m-j,j}).$$

Claim 6 (Induction hypotheses). Circuit g_{j+1} approximates f_{j+1} correctly, i.e. $\lim_{\epsilon \rightarrow 0} g_{j+1} = f_{j+1}$, where g_{j+1} (respec. f_{j+1}) are well-defined in the ring $\mathcal{R}_{j+1}(\mathbf{x}, \epsilon)$ (respec. $\mathcal{R}_{j+1}(\mathbf{x})$).

Proof. Remember, f_j and $T_{i,j}$'s are elements in $\mathbb{F}(\mathbf{x}, z, \epsilon)$ which also belong to $\mathbb{F}(\mathbf{x}, \epsilon)[[z]]$. Since, we divide by the minimum valuation w.r.t. z , by similar argument as in [Claim 5](#), it follows that

$$T_{i,j+1}, f_{j+1} \in \mathbb{F}(\mathbf{x}, z, \epsilon) \cap \mathbb{F}(\mathbf{x}, \epsilon)[[z]], \forall i \in [m-j-1],$$

proving the second part of induction-hypothesis-(2). In fact, trivially $v_{i,j+1} \geq 0$, for $i \in [m-j-1]$ proving induction-hypothesis-(3).

Similarly, Eqn. (5) holds over $\mathcal{R}_{j+1}(\epsilon, \mathbf{x})$, or equivalently mod $z^{d_{j+1}}$; this is because of the division by z -valuation of $v_{m-j,j}$ and then differentiation, showing induction-hypothesis-(1). So, Eqn. (5) being computed mod $z^{d_{j+1}}$ is indeed valid. Further,

$$\text{val}_z(f_j + \epsilon \cdot S_j) = \text{val}_z\left(\sum_{i \in [m-j]} T_{i,j}\right) \geq v_{m-j,j} \xrightarrow{\text{setting } \epsilon=0} \text{val}_z(f_j) \geq v_{m-j}.$$

Hence, $f_j / \tilde{T}_{m-j,j} \in \mathbb{F}(\mathbf{x}, \epsilon)[[z]]$ (by [Lemma 9](#)). Further, since $\lim_{\epsilon \rightarrow 0} \tilde{T}_{m-j,j} = t_{m-j,j}$, we must have that $f_j / t_{m-j,j} \in \mathbb{F}(\mathbf{x})[[z]]$ and thus, f_{j+1} exists.

Finally, it is also easy to see that $\text{val}_z(S_j) \geq v_{m-j,j}$; if there is an ϵ -power in $\epsilon \cdot S_j$ such that coefficient of $< v_{m-j,j}$ power of z exists, then the valuation of $f_j + \epsilon \cdot S_j$ must be $< v_{m-j,j}$ as well, a contradiction!

Thus, $\lim_{\epsilon \rightarrow 0} \partial_z (g_j / \tilde{T}_{m-j,j}) = \lim_{\epsilon \rightarrow 0} g_{j+1} = f_{j+1}$. This finishes [Claim 6](#). \square

Action of dlog : A teaser. Before going into the size analysis, we want to remark that the dlog computation plays a crucial role here. The action $\text{dlog}(\Sigma \wedge \Sigma) \in \Sigma \wedge \Sigma / \Sigma \wedge \Sigma$, is of poly-size ([Lemma 13](#)).

Since dlog distributes the product *additively*, so it suffices to work with $\text{dlog}(\Pi \Sigma)$; and we show that $\text{dlog}(\Pi \Sigma) \in \Sigma \wedge \Sigma$, is of poly-size. For the time being, we assume these hold. Then, one can simplify to get the following:

$$T_{i,j} / \tilde{T}_{m-j,j} = \epsilon^{-a_{m-j,j}} \cdot ((U_{i,j} \cdot V_{m-j,j}) / (V_{i,j} \cdot U_{m-j,j})) \cdot (P_{i,j} \cdot Q_{m-j,j}) / (Q_{i,j} \cdot P_{m-j,j}),$$

and its dlog. Let $U_{i,j+1} := U_{i,j} \cdot V_{m-j,j}$; similarly $V_{i,j+1} := V_{i,j} \cdot U_{m-j,j}$. Essentially, dlog computation will produce $(\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$ -circuits, which will further multiply with P 's and Q 's and we multiply $\epsilon^{-a_{m-j,j}}$ there; for details see [Claim 7](#). This directly means: $U_{i,j+1}|_{z=0}, V_{i,j+1}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$. This proves the second part of induction-hypothesis-(3).

Induction-hypothesis-(4). By induction hypothesis, we have $g_j = z^{b_j} \cdot f_0 \cdot (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma) + z^{b_j+1} \cdot (\cdot)$. Since, $\tilde{T}_{m-j,j} \in (\Pi \Sigma / \Pi \Sigma) \cdot (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$, and that its $\Pi \Sigma|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$; so, $g_j / \tilde{T}_{m-j,j}$ must look like

$$g_j / \tilde{T}_{m-j,j} = z^{b_j} \cdot f_0 \cdot (\Pi \Sigma / \Pi \Sigma) \cdot (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma) + z^{b_j+1} \cdot (\text{blah}) .$$

In the first summand, we already used the fact that $\Sigma \wedge \Sigma$ is *closed* under multiplication ([Lemma 10](#)). Now of course, $\text{val}_z(\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$ in the first summand can be *negative*; however, it is $\geq -b_j$, as otherwise, $\text{val}_z(g_j / \tilde{T}_{m-j,j}) < 0$, a contradiction.

Since the degree of z is polynomially bounded, we can extract the coefficient of the minimum z -power of $\Sigma \wedge \Sigma / \Sigma \wedge \Sigma$ (the z -power gets subtracted from b_j accordingly) and we still get $\Sigma \wedge \Sigma / \Sigma \wedge \Sigma$ (this is z -free) of size polynomially bounded; since $\Pi \Sigma|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$, it does not contribute anything (except nonzero multiples from $\mathbb{F}(\epsilon)$). By definition $\text{val}_z(g_j / \tilde{T}_{m-j,j}) = b_{j+1} + 1 \geq 1$, since its derivative has valuation $b_{j+1} \geq 0$. Hence, the coefficient of the minimum z -power must be of the form $f_0 \cdot \text{coef}_{z^{b_{j+1}+1}}(\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$. Since deg_z is polynomially-bounded, extracting the minimum z -power individually from numerator and denominator gives a 'small' $\Sigma \wedge \Sigma / \Sigma \wedge \Sigma$ which is now z -free. Finally, after differentiating, we get the desired form. Also, the nonzeroness is immediate.

To tackle the 'equivalent' part in the hypothesis-(4), note that

$$f_{j+1} = \lim_{\epsilon \rightarrow 0} g_{j+1} = \lim_{\epsilon \rightarrow 0} \left(z^{b_{j+1}} \cdot f_0 \cdot \Sigma \wedge \Sigma / \Sigma \wedge \Sigma + z^{b_{j+1}+1} \cdot (\cdot) \right) = z^{b_{j+1}} \cdot f_0 \cdot (\overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma}) + z^{b_{j+1}+1} \cdot (\cdot) .$$

Therefore, the coefficient of the minimum z -power in the above expression is $f_0 \cdot (\overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma})$, because of the distributive property [Lemma 14](#).

The overall size blowup. Finally, we show the main step: how to control dlog which is the crux of our reduction. We assume that at the j -th step, $\text{size}(T_{i,j}) \leq s_j$ and by assumption $s_0 \leq s$.

Claim 7 (Size blowup from DiDIL). $T_{1,m-1} \in (\Pi \Sigma / \Pi \Sigma) (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$ over $\mathcal{R}_{m-1}(\mathbf{x}, \epsilon)$ of size $s^{O(m7^m)}$. It is computed as an element in $\mathbb{F}(\epsilon, \mathbf{x}, z)$, with syntactic degree (in \mathbf{x}, z) $d^{O(m)}$.

Proof. We divide the proof into multiple titled paragraphs since it is a bit lengthy. We remark that steps $j = 0$ vs $j > 0$ are slightly different because of the homomorphism Φ . However the main idea of using dlog and expand it as a power-series is the same, which eventually shows that $\text{dlog}(\Pi \Sigma) \in \Sigma \wedge \Sigma$ with a controlled blowup.

dlog's effect in $j = 0$ case. In this case, we would like to understand dlog's effect on $\Phi(T_{i,0}) / \tilde{T}_{m,0}$ – (i) how it changes the structure and (ii) how the size gets blown up. As dlog distributes over product, it suffices to study $\text{dlog}(\ell)$, where $\ell \in \mathcal{R}(\epsilon)[\mathbf{x}]$ is linear.

However, since $\Pi \Sigma$ circuits are already non-homogeneous, the scaling Φ makes each ℓ of the form $\ell = A - zB$, where $A \in \mathbb{F}(\epsilon) \setminus \{0\}$ and $B \in \mathbb{F}(\epsilon)[\mathbf{x}]$. Using the power series expansion, we have the following, over $\mathcal{R}_1(\mathbf{x}, \epsilon)$:

$$\text{dlog}(\ell) = -\frac{\partial_z(z \cdot B)}{A(1 - z \cdot B/A)} = -\frac{B}{A} \cdot \sum_{j=0}^{d_1-1} \left(\frac{z \cdot B}{A} \right)^j . \quad (6)$$

Note, (B/A) and $(-z \cdot B/A)^j$ have trivial powering circuits ($\wedge\Sigma$ over \mathcal{R}_1), each of size $O(dn)$. By [Lemma 10](#), we get the final $\Sigma\wedge\Sigma$ circuit for $\text{dlog}(\Pi\Sigma)$ of size $O(d^2 \cdot s)$. We use the fact that $d_1 < d_0 = d + 1$. Here the syntactic degree blowsup to $O(d)$. This settles $j = 0$ case.

dlog's effect in $j \geq 1$ case. For $j > 0$, the above equation holds over $\mathcal{R}_j(x)$. However, the degree could be D_j (possibly $> d_j$) of the corresponding ℓ (nonlinear), and after exponentiation further increase to $d_j \cdot D_j$. This is exactly why we need to keep track of D_j , the syntactic degree as mentioned in induction-hypotheses-(2). In this calculation, one needs to use the fact that $d_j \leq d$.

Since, $\Sigma\wedge\Sigma$ is closed under differentiation ([Lemma 13](#)), effect of dlog on $\Sigma\wedge\Sigma$ is straightforward. Using [Lemma 13](#), we obtain $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ circuit, computing $\text{dlog}(P_{i,j})$ (similarly $\text{dlog}(Q_{i,j})$) of size $O(D_j^2 \cdot s_j)$. Also, $\text{dlog}(U_{i,j} \cdot V_{m-j,j}) \in \Sigma \text{dlog}(\Sigma)$, which could be computed using the above Equation. Thus,

$$\begin{aligned} \text{dlog}(T_{i,j}/\tilde{T}_{m-j,j}) &\in \text{dlog}(\Pi\Sigma/\Pi\Sigma) \pm \Sigma^{[4]}\text{dlog}(\Sigma\wedge\Sigma) \\ &\subseteq \Sigma\wedge\Sigma + \Sigma^{[4]}(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) = (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma). \end{aligned}$$

Here, $\Sigma^{[4]}(\cdot)$ means sum of 4-many expressions of the form (\cdot) . The first containment is by linearization. We can express $\text{dlog}(\Pi\Sigma/\Pi\Sigma)$ as a single $\Sigma\wedge\Sigma$ -expression since $\text{dlog}(\Sigma) \subset \Sigma\wedge\Sigma$. Similarly, 4-many $\text{dlog}(\Sigma\wedge\Sigma)$ expressions give 4-many $(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ expressions.

Size analysis of $\text{dlog}(T_{i,j}/\tilde{T}_{m-j,j})$. The $\Sigma\wedge\Sigma$ expression, obtained from $\text{dlog}(\Pi\Sigma/\Pi\Sigma)$ is of size $O(D_j^2 d_j s_j)$. Next, there are 4-many $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expressions of size $O(D_j^2 s_j)$ as there are 4-many P 's and Q 's. Additionally, the syntactic degree of each denominator and numerator of $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ grows up to $O(D_j)$. Finally, we club $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expressions (4 of them) to express it as a single $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expression using [Lemma 13](#), with size blowup of $O(D_j^{12} s_j^4)$. Finally, add the single $\Sigma\wedge\Sigma$ expression of size $O(D_j^3 s_j)$, and degree $O(dD_j)$, to get $O(s_j^5 D_j^{16} d)$ size representation.

Overall size blowup from j to $j + 1$. Also, we need to multiply with $T_{i,j}/\tilde{T}_{m-j,j}$ which is of the form $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, where each $\Sigma\wedge\Sigma$ is basically product of two $\Sigma\wedge\Sigma$ expressions of size s_j and syntactic degree D_j and clubbed together, owing a blowup of $O(D_j s_j^2)$. Hence, multiplying this $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ -expression with the $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expression obtained from dlog -computation, gives a size blowup of $s_{j+1} := s_j^7 D_j^{O(1)} d$.

As mentioned before, the main blowup of syntactic degree in the dlog computation could be $O(dD_j)$ and clearing expressions and multiplying the without- dlog expression increases the syntactic degree only by a constant multiple. Therefore, $D_{j+1} := O(dD_j) \implies D_j = d^{O(j)}$. Hence, $s_{j+1} = s_j^7 \cdot d^{O(j)} \implies s_j \leq (sd)^{O(j \cdot 7^j)}$. In particular, $s_{m-1} \leq s^{O(m \cdot 7^m)}$; here we used that $d \leq s$. This calculation quantitatively establishes induction-hypothesis-(2). This finishes [Claim 7](#). \square

Hence successively doing this $m - 1$ times, we get that $g = g_{m-1} \in \text{Gen}(1, s^{O(m7^m)}) \bmod z^{d_{m-1}}$ (first point of [Lemma 4](#)), such that $\lim_{\epsilon \rightarrow 0} g_{m-1} = f_{m-1} \in \mathbb{F}(x, z)$ (second point of [Lemma 4](#)) and moreover, the coefficient of the minimum z -power in g_{m-1} is nothing but $f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$. These circuits $\overline{\Sigma\wedge\Sigma}$ also have size $s^{O(m7^m)}$; since the degree bound on $\overline{\Sigma\wedge\Sigma}$ circuits is $s^{O(m7^m)}$, we had to extract the minimum coefficient from both numerator and denominator, as specified above (in induction-hypothesis-4). Finally, taking the limit finishes the proof of the third point (of [Lemma 4](#)). \square

4.2 Proving lower bound for bloated fanin-1 model $\overline{\text{Gen}(1, \cdot)}$

Now it remains to establish the lower bound. Recall, we have shown that the minimum z -power in g_{m-1} gives $f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma}$, where $\Sigma \wedge \Sigma$ are of size $s^{O(m7^m)}$. Moreover, since $g_{m-1} \in \text{Gen}(1, s^{O(m7^m)})$, over $\mathbb{F}(\epsilon)$, the limit of g_{m-1} itself is of the form $\overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma}$ (recall: the other factors satisfy $\Pi \Sigma|_{x=\epsilon=0} \in \mathbb{F}^*$). Therefore, comparing the coefficients of the minimum z -power both sides, we have eventually shown: $f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma} = \overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma} \implies f(\ell_1, \dots, \ell_n) = \overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma}$. Here $\overline{\Sigma \wedge \Sigma}$ circuits have size $s^{O(m7^m)}$, because of [Lemma 10](#). Next, we show that this implies: $s \geq 2^{\Omega(d)}$, when $m \leq k = O(1)$.

Lemma 8 (Lower bound for bloated fanin-1). *If $f(\ell_1, \dots, \ell_n) = \lim_{\epsilon \rightarrow 0} (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$, such that*

1. *size of the $\Sigma \wedge \Sigma$ circuits, over $\mathbb{F}(\epsilon)$, is $s^{O(k7^k)}$, and*
2. *$\text{rank}(\ell_1, \dots, \ell_n) = n - (k - m) \geq n - k$,*

then $s \geq 2^{\Omega(d/(k7^k))}$.

Proof. The proof is based on the cone-size measure. Note that, if $f(\ell_1, \dots, \ell_n) = \overline{\Sigma \wedge \Sigma_1} / \overline{\Sigma \wedge \Sigma_2} \implies f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma \wedge \Sigma_2} = \overline{\Sigma \wedge \Sigma_1}$. We will work with the partial derivatives spaces (defined below): Denote

$$V_{\epsilon, i} := \left\langle \frac{\partial (\Sigma \wedge \Sigma_i)}{\partial \mathbf{x}^e} \mid e < \infty \right\rangle_{\mathbb{F}(\epsilon)}, \quad \text{and} \quad V_i := \left\langle \frac{\partial (\overline{\Sigma \wedge \Sigma_i})}{\partial \mathbf{x}^e} \mid e < \infty \right\rangle_{\mathbb{F}}.$$

Since, $\text{size}(\Sigma \wedge \Sigma_i) \leq s^{O(k7^k)}$, the partial derivative space of $\Sigma \wedge \Sigma_i$, over $\mathbb{F}(\epsilon)$, is also bounded by $s^{O(k7^k)}$, i.e. $\dim(V_{\epsilon, i}) \leq s^{O(k7^k)}$ (see [[CKW11](#), Lemma 10.2]). Consider the partial-derivative matrix $M_{\epsilon, i}$, where we index the rows by $\partial_{\mathbf{x}^e}$, while columns are indexed by monomials supporting $\Sigma \wedge \Sigma_i$, and each row expresses the operator-values $\partial_{\mathbf{x}^e} (\Sigma \wedge \Sigma_i)$. We have, $q_i := \dim(V_{\epsilon, i}) \leq s^{O(k7^k)}$ (because of $\Sigma \wedge \Sigma_i$). So, any $(q_i + 1)$ -many polynomials $\frac{\partial (\Sigma \wedge \Sigma_i)}{\partial \mathbf{x}^e}$ are \mathbb{F} -linearly dependent. In other words, determinant of any $(q_i + 1) \times (q_i + 1)$ minor of $M_{\epsilon, i}$ is 0. Note that $\lim_{\epsilon \rightarrow 0} M_{\epsilon, i} = M_i$, the corresponding partial-derivative matrix for $\overline{\Sigma \wedge \Sigma_i}$.

Since, \det is a continuous function, the zeroness of the determinant of any $(q_i + 1) \times (q_i + 1)$ minor of $M_{\epsilon, i}$ translates to the corresponding $(q_i + 1) \times (q_i + 1)$ submatrix of M_i as well. In particular,

$$\dim(V_i) \leq q_i \leq s^{O(k7^k)}. \quad (7)$$

From this, it follows that leading monomial in $\overline{\Sigma \wedge \Sigma_i}$ (denoted $\text{LM}(\overline{\Sigma \wedge \Sigma_i})$) has cone-size at most $s^{O(k7^k)}$; see [Lemma 15](#).

Since, $\text{cone-size}(\text{LM}(f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma \wedge \Sigma_2})) \geq \text{cone-size}(\text{LM}(f(\ell_1, \dots, \ell_n)))$, it suffices to show that leading monomial of $f(\ell_1, \dots, \ell_n)$ has cone-size 2^d .

Exploiting ‘large rank’. As argued in [Corollary 3](#), $f(\ell_1, \dots, \ell_n) = P_{k+1, d}(\ell_1, \dots, \ell_n)$ is nonzero, and satisfies the property that after suitable isomorphism (which does not affect our proof), there is a pure multilinear monomial, i.e. product of distinct variables, that survives. This implies that with respect to a suitable monomial ordering, the leading term has cone-size $= 2^d$. Implying: $\text{size}(\overline{\Sigma \wedge \Sigma_1}) \geq \text{cone-size}(\text{LM}(\overline{\Sigma \wedge \Sigma_1})) = \text{cone-size}(\text{LM}(f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma \wedge \Sigma_2})) \geq 2^d$ (also, see the remark below). Consequently, $s^{k7^k} \geq 2^{\Omega(d)}$. This finishes the lower bound proof. \square

Remark. In the above, we only work with $\overline{\Sigma \wedge \Sigma}$. Since, $\overline{\Sigma \wedge \Sigma}$ has *low* partial derivative space, the cone-size-based proof goes through (Lemma 15). This is exactly why our proofs naively cannot give interesting lower bounds for $\overline{\Sigma^{[k]} \Pi \Sigma \wedge}$ and $\overline{\Sigma^{[k]} \Pi \Sigma \Pi}$, since the cone-size-based argument does no longer work.

4.3 Tying the pieces together: Proof of Theorem 1

Proof of Theorem 1. We have shown and proved all the necessary steps for Theorem 1. To summarize, we start with $P_{k+1,d}$ and assume that it can be computed by a $\overline{\Sigma^{[k]} \Pi \Sigma}$ -circuit of size s . We now reduce top-fanin in two totally different ways.

1. (*Setting linear forms zero*). If needed, use the reduction Lemma 2 to reduce to the setting of Problem 1. If we end up with $m = 0$, then a proof similar to Corollary 3 shows: $P_{k+1,d}(\ell_1, \dots, \ell_n) \neq 0$ for $\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$. Thus $m = 0$ is an impossibility.
2. (*DiDIL process*). If $m \geq 1$, then it suffices to solve Problem 1 as mentioned before. Use the reduction from subsection 4.1 which reduces the top-fanin m to 1 (at the cost of moving to the bloated model). Finally, use the lower bound for fanin-1 (Lemma 8) to conclude that $s \geq 2^{\Omega(d)}$ (for constant k). This finishes Theorem 1. \square

5 Conclusion

In this work, we show a strong top-fanin-hierarchy theorem for depth-3 class in the border setting. The methods used here, open a wide avenue of plausible questions, some of which may not be very hard to answer. We list a few of them below.

1. Can we show exponential lower bound for $\overline{\Sigma^{[o(n)]} \Pi \Sigma}$ -circuits? The current method gives subexponential lower bound only as long as $k = o(\log n)$.
2. Can we show exponential lower bound for $\overline{\Sigma^{[k]} \Pi \Sigma \wedge}$ -circuits (i.e. rather special depth-4)?
3. Can we extend the hierarchy theorem to bounded (top & bottom fanin) depth-4 circuits? i.e., for a *fixed* constant δ , is $\overline{\Sigma^{[1]} \Pi \Sigma \Pi^{[\delta]}} \subsetneq \overline{\Sigma^{[2]} \Pi \Sigma \Pi^{[\delta]}} \subsetneq \overline{\Sigma^{[3]} \Pi \Sigma \Pi^{[\delta]}} \dots$, where the respective gaps are exponential? Clearly, $\delta = 1$ holds, from this work.

Acknowledgement. Pranjal thanks Department of CSE, IIT Kanpur for the hospitality, and acknowledges the support of Google Ph. D. Fellowship. Nitin thanks the funding support from DST-SERB (CRG/2020/000045) and N. Rama Rao Chair. We thank Prateek Dwivedi and Amit Sinhababu for helpful comments on the draft.

References

- [ASSS16] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. [Jacobian hits circuits: Hitting sets, lower bounds for depth-D occur-k formulas and depth-3 transcendence degree-k circuits](#). *SIAM Journal on Computing*, 45(4):1533–1562, 2016. 12

- [AW16] Eric Allender and Fengming Wang. [On the power of algebraic branching programs of width two](#). *Computational Complexity*, 25(1):217–253, 2016. 12
- [AZGL⁺18] Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. [Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing](#). In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 172–181, 2018. 3
- [BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. [\$O\(n^{2.7799}\)\$ Complexity for \$n \times n\$ Approximate Matrix Multiplication](#). *Inf. Process. Lett.*, 8(5):234–235, 1979. 3
- [BFG⁺19] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. [Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes](#). In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 845–861. IEEE, 2019. 3
- [BGO⁺18] Peter Bürgisser, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. [Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory](#). *Proceedings of Innovations in Theoretical Computer Science (ITCS 2018)*, 2018. 3
- [BI13] Peter Bürgisser and Christian Ikenmeyer. [Explicit lower bounds via geometric complexity theory](#). In *Proceedings of the 45th Annual ACM symposium on Theory of Computing*, pages 141–150, 2013. 3
- [BIL⁺21] Markus Bläser, Christian Ikenmeyer, Vladimir Lysikov, Anurag Pandey, and Frank-Olaf Schreyer. [On the Orbit Closure Containment Problem and Slice Rank of Tensors](#). In Daniel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021*, pages 2565–2584. SIAM, 2021. 3
- [Bin80] Dario Bini. [Relations between exact and approximate bilinear algorithms. Applications](#). *Calcolo*, 17(1):87–97, 1980. 3
- [BIP19] Peter Bürgisser, Christian Ikenmeyer, and Greta Panova. [No occurrence obstructions in geometric complexity theory](#). *Journal of the American Mathematical Society*, 32(1):163–193, 2019. 5
- [BIZ18] Karl Bringmann, Christian Ikenmeyer, and Jeroen Zuiddam. [On Algebraic Branching Programs of Small Width](#). *J. ACM*, 65(5):32:1–32:29, 2018. 7, 12, 13
- [BLMW11] Peter Bürgisser, J. M. Landsberg, Laurent Manivel, and Jerzy Weyman. [An Overview of Mathematical Issues Arising in the Geometric Complexity Theory Approach to \$VP \neq VNP\$](#) . *SIAM J. Comput.*, 40(4):1179–1209, 2011. 3
- [BOC92] Michael Ben-Or and Richard Cleve. [Computing algebraic formulas using a constant number of registers](#). *SIAM Journal on Computing*, 21(1):54–58, 1992. 12
- [BS83] Walter Baur and Volker Strassen. [The complexity of partial derivatives](#). *Theoretical computer science*, 22(3):317–330, 1983. 5

- [Bür04] Peter Bürgisser. *The Complexity of Factors of Multivariate Polynomials*. *Found. Comput. Math.*, 4(4):369–396, 2004. 3
- [Bür20] Peter Bürgisser. *Correction To: The Complexity of Factors of Multivariate Polynomials*. *Found. Comput. Math.*, 20(6):1667–1668, 2020. 3
- [CGJ⁺18] Mahdi Cheraghchi, Elena Grigorescu, Brendan Juba, Karl Wimmer, and Ning Xie. *$AC^0 \circ MOD_2$ lower bounds for the Boolean Inner Product*. *Journal of Computer and System Sciences*, 97:45–59, 2018. 2
- [CKR⁺20] Prerona Chatterjee, Mrinal Kumar, C Ramya, Ramprasad Saptharishi, and Anamay Tengse. *On the Existence of Algebraically Natural Proofs*. *61st Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2020)*, 2020. 3
- [CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. *Partial derivatives in arithmetic complexity and beyond*. Now Publishers Inc, 2011. 10, 23
- [CW90] Don Coppersmith and Shmuel Winograd. *Matrix Multiplication via Arithmetic Progressions*. *J. Symb. Comput.*, 9(3):251–280, 1990. 3
- [DDS21] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. *Demystifying the border of depth-3 algebraic circuits*. *62nd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2021)*, 2021. 3, 4, 6, 10, 11, 17, 18, 29, 30, 31
- [DSS18] Pranjal Dutta, Nitin Saxena, and Amit Sinhababu. *Discovering the roots: uniform closure results for algebraic classes under factoring*. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *STOC 2018*, pages 1152–1165. ACM, 2018. 7
- [For16] Michael Forbes. *Some concrete questions on the border complexity of polynomials*. Presentation given at the Workshop on Algebraic Complexity Theory WACT 2016 in Tel Aviv, 2016. 31
- [FS13] Michael A Forbes and Amir Shpilka. *Explicit noether normalization for simultaneous conjugation via polynomial identity testing*. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 527–542. Springer, 2013. 3
- [GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. *A deterministic polynomial time algorithm for non-commutative rational identity testing*. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117. IEEE, 2016. 3
- [Gho19] Sumanta Ghosh. *Low Variate Polynomials: Hitting Set and Bootstrapping*. PhD thesis, PhD thesis. Indian Institute of Technology Kanpur, 2019. 30
- [GK98] Dima Grigoriev and Marek Karpinski. *An Exponential Lower Bound for Depth 3 Arithmetic Circuits*. In *STOC 1998*, pages 577–582, 1998. 11
- [GKKS16] Ankit Gupta, Prithish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. *Arithmetic Circuits: A Chasm at Depth 3*. *SIAM J. Comput.*, 45(3):1064–1079, 2016. FOCS 2013. 11

- [GKS16] Rohit Gurjar, Arpita Korwar, and Nitin Saxena. [Identity Testing for Constant-Width, and Commutative, Read-Once Oblivious ABPs](#). In Ran Raz, editor, *31st Conference on Computational Complexity, CCC 2016*, volume 50 of *LIPICs*, pages 29:1–29:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. 31
- [GKSS17] Joshua A Grochow, Mrinal Kumar, Michael Saks, and Shubhangi Saraf. [Towards an algebraic natural proofs barrier via polynomial identity testing](#). *arXiv preprint arXiv:1701.01717*, 2017. 3
- [GR00] 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. FOCS 1998. 11
- [Gro12] Joshua Abraham Grochow. [Symmetry and equivalence relations in classical and geometric complexity theory](#). The University of Chicago, 2012. 5
- [Gro15] Joshua A Grochow. [Unifying known lower bounds via geometric complexity theory](#). *Computational Complexity*, 24(2):393–475, 2015. 3, 5
- [HS65] Juris Hartmanis and Richard E Stearns. [On the computational complexity of algorithms](#). *Transactions of the American Mathematical Society*, 117:285–306, 1965. 2
- [IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. [Non-commutative Edmonds’ problem and matrix semi-invariants](#). *Comput. Complex.*, 26(3):717–763, 2018. 3
- [Kal85] Kyriakos Kalorkoti. [A Lower Bound for the Formula Size of Rational Functions](#). *SIAM Journal of Computing*, 14(3):678–687, 1985. 5
- [KLSS17] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. [An exponential lower bound for homogeneous depth four arithmetic formulas](#). *SIAM Journal on Computing*, 46(1):307–335, 2017. 11
- [KRST20] Mrinal Kumar, C Ramya, Ramprasad Saptharishi, and Anamay Tengse. [If VNP is hard, then so are equations for it](#). *arXiv preprint arXiv:2012.07056*, 2020. 3
- [KS16] Neeraj Kayal and Chandan Saha. [Lower bounds for depth-three arithmetic circuits with small bottom fanin](#). *computational complexity*, 25(2):419–454, 2016. 11
- [KS17] Mrinal Kumar and Shubhangi Saraf. [On the power of homogeneous depth 4 arithmetic circuits](#). *SIAM Journal on Computing*, 46(1):336–387, 2017. 11
- [Kum20] Mrinal Kumar. [On the Power of Border of Depth-3 Arithmetic Circuits](#). *ACM Trans. Comput. Theory*, 12(1):5:1–5:8, 2020. 2, 4, 5, 6, 9, 13
- [Lan15] Joseph M Landsberg. [Geometric complexity theory: an introduction for geometers](#). *Annali dell’universita’ di Ferrara*, 61(1):65–117, 2015. 4, 6
- [Lan17] J. M. Landsberg. [Geometry and Complexity Theory](#). Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017. 4, 5

- [LO15] Joseph M. Landsberg and Giorgio Ottaviani. [New Lower Bounds for the Border Rank of Matrix Multiplication](#). *Theory Comput.*, 11:285–298, 2015. 3
- [LST21] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. [Superpolynomial Lower Bounds Against Low-Depth Algebraic Circuits](#). *62nd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2021)*, page 81, 2021. 5, 6, 12
- [Mah13] Meena Mahajan. [Algebraic Complexity Classes](#). *CoRR*, abs/1307.3863, 2013. 3
- [MS01] Ketan Mulmuley and Milind A. Sohoni. [Geometric Complexity Theory I: An Approach to the P vs. NP and Related Problems](#). *SIAM J. Comput.*, 31(2):496–526, 2001. 3
- [Muk16] Partha Mukhopadhyay. [Depth-4 Identity Testing and Noether’s Normalization Lemma](#). In *CSR 2016*, 2016. 3
- [Mul12a] Ketan Mulmuley. [Geometric Complexity Theory V: Equivalence between Blackbox Derandomization of Polynomial Identity Testing and Derandomization of Noether’s Normalization Lemma](#). In *53rd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2012)*, pages 629–638, 2012. 3
- [Mul12b] Ketan D Mulmuley. [The GCT program toward the P vs. NP problem](#). *Communications of the ACM*, 55(6):98–107, 2012. 3
- [Mum95] David Mumford. *Algebraic geometry I: complex projective varieties*. Springer Science & Business Media, 1995. 3
- [Niv69] Ivan Niven. [Formal Power Series](#). *The American Mathematical Monthly*, 76(8):871–889, 1969. 7
- [NW96] Noam Nisan and Avi Wigderson. [Lower bounds on arithmetic circuits via partial derivatives](#). *Computational complexity*, 6(3):217–234, 1996. 11
- [Sap19] Ramprasad Saptharishi. [A survey of lower bounds in arithmetic circuit complexity](#). *Github Survey*, 2019. 5
- [Sax08] Nitin Saxena. [Diagonal Circuit Identity Testing and Lower Bounds](#). In *ICALP 2008*, pages 60–71, 2008. 30
- [SHL65] Richard Edwin Stearns, Juris Hartmanis, and Philip M Lewis. [Hierarchies of memory limited computations](#). In *6th annual symposium on switching circuit theory and logical design (SWCT 1965)*, pages 179–190. IEEE, 1965. 2
- [Shp02] Amir Shpilka. [Affine projections of symmetric polynomials](#). *Journal of Computer and System Sciences*, 65(4):639–659, 2002. 6
- [Sin19] Amit Kumar Sinhababu. [Power series in complexity: Algebraic Dependence, Factor Conjecture and Hitting Set for Closure of VP](#). PhD thesis, Indian Institute of Technology Kanpur, 2019. 7
- [Str73] Volker Strassen. [Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten](#). *Numerische Mathematik*, 20(3):238–251, 1973. 5

- [Str74] Volker Strassen. [Polynomials with Rational Coefficients Which are Hard to Compute](#). *SIAM J. Comput.*, 3(2):128–149, 1974. 3
- [SW01] Amir Shpilka and Avi Wigderson. [Depth-3 arithmetic circuits over fields of characteristic zero](#). *Computational Complexity*, 10(1):1–27, 2001. CCC 1999. 11
- [SY10] Amir Shpilka and Amir Yehudayoff. [Arithmetic Circuits: A survey of recent results and open questions](#). *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010. 3
- [Val79] Leslie G. Valiant. [Completeness Classes in Algebra](#). In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing*, pages 249–261, 1979. 3

A Basics of algebraic complexity

Definition 3 (Algebraic Branching Program (ABP)). *ABP is a computational model which is described using a layered graph with a source vertex s and a sink vertex t . All edges connect vertices from layer i to $i + 1$. Further, edges are labelled by univariate polynomials. The polynomial computed by the ABP is defined as*

$$f = \sum_{\text{path } \gamma: s \rightsquigarrow t} \text{wt}(\gamma)$$

where $\text{wt}(\gamma)$ is product of labels over the edges in path γ . Number of layers (Δ) defines the *depth* and the maximum number of vertices in any layer (w) defines the *width* of an ABP. The *size* (s) of an ABP is the sum of the graph-size and the degree of the univariate polynomials that label. If d is the maximum degree of univariates then $s \leq dw^2\Delta$; in fact, we will take the latter as the ABP-size bound in our calculations.

Our interest primarily is in the following two ABP-variants: ROABP and ARO.

Definition 4 (Read-once Oblivious Algebraic Branching Program (ROABP)). *An ABP is defined as Read-Once Oblivious Algebraic Branching Program (ROABP) in a variable order $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ for some permutation $\sigma : [n] \rightarrow [n]$, if edges of i -th layer of ABP are univariate polynomials in $x_{\sigma(i)}$.*

Definition 5 (Any-order ROABP (ARO)). *A polynomial $f \in \mathbb{F}[x]$ is computable by ARO of size s if for all possible permutation of variables there exists a ROABP of size at most s in that variable order.*

Remark. We can de-border $\overline{\Sigma \wedge \Sigma}$. Since $\overline{\Sigma \wedge \Sigma} \subseteq \overline{\text{ARO}}$, using duality trick ([Lemma 16](#)) and $\overline{\text{ARO}} = \text{ARO}$, from Nisan’ characterization ([Lemma 17](#)), it follows that $\overline{\Sigma \wedge \Sigma} \subsetneq \text{ARO}$. Note that, $\overline{\Sigma \wedge \Sigma}$ is a strict subset of ARO since $\prod_{i=1}^n x_i$ has a small ARO, but it requires $\exp(n)$ -size $\overline{\Sigma \wedge \Sigma}$ -circuits.

B Basic tools

Here is an important lemma to show that positive valuation with respect to y , lets us express a function as a power-series of y .

Lemma 9 (Valuation lemma, [[DDS21](#), Lemma A.17]). *Let $f \in \mathbb{F}(x, y)$ such that $\text{val}_y(f) \geq 0$. Then, $f \in \mathbb{F}(x)[[y]] \cap \mathbb{F}(x, y)$.*

In this section we will also discuss various properties of $\Sigma\wedge\Sigma$ circuits and basic waring-rank. The corresponding bloated model is $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$, that computes elements of the form f/g , where $f, g \in \Sigma\wedge\Sigma$. For the detailed proofs, we refer to [DDS21].

Firstly, it is known that $\Sigma\wedge\Sigma$ is *closed* under *constant-fold* multiplication.

Lemma 10 ($\Sigma\wedge\Sigma$ closed under multiplication, [DDS21, Lemma A.10]). *Let $f_i \in \mathbb{F}[x]$, of syntactic degree $\leq d_i$, be computed by a $\Sigma\wedge\Sigma$ circuit of size s_i , for $i \in [k]$. Then, $f_1 \cdots f_k$ has $\Sigma\wedge\Sigma$ circuit of size $O((d_2 + 1) \cdots (d_k + 1) \cdot s_1 \cdots s_k)$.*

Using the additive and multiplicative closure of $\Sigma\wedge\Sigma$, one can show that $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ is closed under constant-fold addition.

Lemma 11 ($\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ closed under addition, [DDS21, Lemma A.11]). *Let $f_i \in \mathbb{F}[x]$, of syntactic degree $< d_i$, be computable by $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ of size s_i , for $i \in [k]$. Then, $\sum_{i \in [k]} f_i$ has a $(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ representation of size $O(\prod_i d_i \cdot \prod_i s_i)$.*

Using a simple interpolation, the coefficient of y^e can be extracted from $f(x, y) \in \Sigma\wedge\Sigma$ again as a small $\Sigma\wedge\Sigma$ representation.

Lemma 12 ($\Sigma\wedge\Sigma$ coefficient extraction, [DDS21, Lemma A.12]). *Let $f(x, y) \in \mathbb{F}[x][y]$ be computed by a $\Sigma\wedge\Sigma$ circuit of size s and degree d . Then, $\text{coef}_{y^e}(f) \in \mathbb{F}[x]$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd)$, over $\mathbb{F}[x]$.*

Like coefficient extraction, differentiation of $\Sigma\wedge\Sigma$ circuit is easy too.

Lemma 13 ($\Sigma\wedge\Sigma$ differentiation, [DDS21, Lemma A.13]). *Let $f(x, y) \in \mathbb{F}[x][y]$ be computed by a $\Sigma\wedge\Sigma$ circuit of size s and degree d . Then, $\partial_y(f)$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd^2)$, over $\mathbb{F}[x][y]$.*

Let \mathcal{C} and \mathcal{D} be two classes over $\mathbb{F}[x]$. Consider the bloated-class $(\mathcal{C}/\mathcal{C}) \cdot (\mathcal{D}/\mathcal{D})$, which has elements of the form $(g_1/g_2) \cdot (h_1/h_2)$, where $g_i \in \mathcal{C}$ and $h_i \in \mathcal{D}$ ($g_2 h_2 \neq 0$). One can also similarly define its border (which will be an element in $\mathbb{F}(x)$). Here is an important observation.

Lemma 14 ([DDS21, Lemma A.19]). $\overline{(\mathcal{C}/\mathcal{C}) \cdot (\mathcal{D}/\mathcal{D})} \subseteq \overline{(\mathcal{C}/\mathcal{C})} \cdot \overline{(\mathcal{D}/\mathcal{D})}$.

Proof. Suppose $(g_1/g_2) \cdot h_1/h_2 = f + \epsilon \cdot Q$, where $Q \in \mathbb{F}(x, \epsilon)$ and $f \in \mathbb{F}(x)$. Let $\text{val}_\epsilon(g_i) =: a_i$ and $\text{val}_\epsilon(h_i) =: b_i$. Denote, $g_i =: \epsilon^{a_i} \cdot \tilde{g}_i$, similarly \tilde{h}_i . Further, assume $\tilde{g}_i =: \hat{g}_i + \epsilon \cdot \hat{g}'_i$; similarly for \tilde{h}_i , we define $\hat{h}_i \in \mathbb{F}[x]$. Note that $\hat{g}_i \in \overline{\mathcal{C}}$, similarly $\hat{h}_i \in \overline{\mathcal{D}}$.

So, LHS = $\epsilon^{a_1 - a_2 + b_1 - b_2} \cdot (\tilde{g}_1/\tilde{g}_2) \cdot (\tilde{h}_1/\tilde{h}_2)$. This has a limit $\lim_{\epsilon \rightarrow 0}$, so $a_1 + b_1 - a_2 - b_2 \geq 0$. If it is ≥ 1 , the limit in RHS is 0 and so $f = 0$. If $a_1 + b_1 - a_2 - b_2 = 0$, then

$$f = (\hat{g}_1/\hat{g}_2) \cdot (\hat{h}_1/\hat{h}_2) \in \overline{(\mathcal{C}/\mathcal{C})} \cdot \overline{(\mathcal{D}/\mathcal{D})}.$$

□

Lemma 15 ([Gho19, Lemma 2.3.15]). *Let \mathbb{F} be a field of characteristic 0 or greater than d . Let \mathcal{P} be a set of n -variate degree d polynomials over \mathbb{F} such that for all $P \in \mathcal{P}$, the dimension of the partial derivative space of P is at most k . Then, every nonzero $P \in \mathcal{P}$ has a $(\leq k)$ -cone-size leading-monomial.*

Lemma 16 (Duality trick [Sax08]). *The polynomial $f = (x_1 + \dots + x_n)^d$ can be written as*

$$f = \sum_{i \in [t]} f_{i1}(x_1) \cdots f_{in}(x_n),$$

where $t = O(nd)$, and f_{ij} is a univariate polynomial of degree at most d .

We remark that the above proof works for fields of characteristic $= 0$, or $> d$. This lemma eventually shows that $\Sigma \wedge \Sigma \subseteq \text{ARO}$.

Next we state that polynomials approximated by ARO can be easily de-bordered. To the best of our knowledge the following lemma was sketched in [For16]; also implicitly in [GKS16]. For a detailed proof, see [DDS21, Lemma A.21]

Lemma 17 (De-bordering ARO). *Consider a polynomial $f \in \mathbb{F}[x]$ which is approximated by ARO of size s over $\mathbb{F}(\epsilon)[x]$. Then, there exists an ARO of size s which exactly computes $f(x)$.*