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 ($\Sigma^2 \Pi \Sigma$); 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 ($\Sigma^k \Pi \Sigma$ for constant $k$) can be computed by a polynomial-size algebraic branching program (ABP). It was left open to show an exponential separation between the class of ABPs and $\Sigma^{k+1} \Pi \Sigma$.

In this article, we show a strongly-exponential separation between any two consecutive border classes, $\Sigma^k \Pi \Sigma$ and $\Sigma^{k+1} \Pi \Sigma$, 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.
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 $\Rightarrow$ 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 $f \in \mathbb{F}[x_1, \ldots, x_n]$, 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 $\ell_{ij}$ 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 $P_n \in \mathbb{F}[x_1, \ldots, x_n]$, such that $P_n$ can be ‘approximated’ by a small $\Sigma^k \Pi \Sigma$-circuit but not ‘approximated’ by a small $\Sigma^{k+1} \Pi \Sigma$-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 $\Sigma^{k+1} \Pi \Sigma$-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 \(\text{IP}_{k+1}^c\); 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 VBP \(\not\subseteq\) VNP (respectively VP \(\not\subseteq\) 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: \(\text{VNP} \not\subseteq \text{VBP}\), i.e. \(\ell^m-n \cdot \text{perm}_n(X) \not\in \text{GL}_m(\mathbb{C}) \cdot \text{det}_m(X)\), for a linear form \(\ell\) 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 \(f(x) \in \mathbb{F}[x_1, \ldots, x_n]\) 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 \(\lim_{\epsilon \to 0} g = f\). The class \(\mathcal{C}\), the approximative closure of a complexity class \(\mathcal{C}\), can be defined analogously. Note that, arbitrary \(\epsilon\)-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 \(\text{size}(f)\), is the minimum size of the circuit computing \(g\) over \(\mathbb{F}(\epsilon)\); evidently, \(\text{size}(f) \leq \text{size}(g)\). 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’ \(\epsilon\)-values from \(\mathbb{F}\). However, the trivial bound on the circuit size of \(f\) would linearly depend on the degree \(M\) of \(\epsilon\) (\& \(1/\epsilon\)); which is known to be exponential in the size of the circuit computing \(g\) ! Therefore, the following relation is the best one known: \(\text{size}(f) \leq \text{size}(f) \leq \exp(\text{size}(f))\) [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. $\Pi\Sigma = \Pi\Sigma$ and $\Sigma\Pi = \Sigma\Pi$, it is natural to study border of depth-3 circuits. Again, it is not hard to show that $\Pi\Sigma\Pi = \Pi\Sigma\Pi$ which leaves us to understand $\Sigma\Pi\Sigma$. Kumar [Kum20] showed that border depth-3 fanin-2 circuits are ‘universal’, i.e. $\Sigma[2]\Pi[3] \Sigma$ over $C(e)$ can approximate any $d$-degree, $n$-variante polynomial $1$; though this expression requires an exceedingly large $D = \exp(n,d)$. In the case of ‘smaller’ $D$, [DDS21] proved that $\Sigma[k]\Pi\Sigma \subseteq VBP$; formally any $f \in \Sigma[k]\Pi\Sigma$, 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: $\Sigma[k]\Pi\Sigma \not\subseteq VBP$ (even, $\Sigma[k]\Pi\Sigma \not\subseteq VNP$ 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, $Ch_d(W) \subseteq \mathbb{P}S^dW$; usually $W = C^n$ (or, for $det_n/perm_n$, $C^{n^2}$). We denote by $S^dW$, the space of polynomials of degree $d$ on $W^*$, $PV$ denotes the projective space and we denote $[v]$ as a corresponding point. Then, formally,

$$Ch_d(W) := \{ [z] \in \mathbb{P}S^dW \mid z = w_1 \ldots w_d, \text{ for } w_j \in W \}.$$ 

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

Moving to the approximative setting, one defines Chow border rank, $\text{rank}_{ch}(f)$, as the border analogue of the Chow rank. In other words, $\text{rank}_{ch}(f) = k \iff f \in \sigma_k(Ch_d(W))$ is the Zariski closure in $\mathbb{P}S^dW$ of $\sigma_k^0(Ch_d(W))$; it is called the $k$-th secant variety 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 ($& W = C^{n+1}$), Kumar’s result translates into geometric terms: For any degree-$d$ homogeneous polynomial $f$, there exists a linear form $\ell$ such that $\ell^{m-d}f \in \sigma_2(Ch_m(W))$, for $m = \exp(n,d)$, or equivalently, $\text{rank}_{ch}(f^{m-d}) = 2$. On the other hand, if one restricts $m = \text{poly}(n,d)$, Dutta, Dwivedi and Saxena [DDS21] showed that $\sigma_2(Ch_m(W)) \subseteq VBP$; the same result holds if one replaces $\sigma_2$ by $\sigma_k$, 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,

---

1 [Kum20] states the result for homogeneous polynomials. But trivially, the proof also works for any non-homogeneous polynomial, by homogenizing it using a new variable $x_0$, and finally de-homogenizing it by setting $x_0 = 1$.

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

---

4
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 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, AF22]. 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$^{3}$ $P_{k+1} := (P_{k+1,d}^1)_d$, 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 $IP_{k+1} \in \Sigma^{[2]} \Pi^{O(1)} \Sigma$. Therefore, we cannot show a strong size lower

---

$^{3}$This term has been taken from [Lan17, Section 8.12.2].
bound against $IP_{k+1}$. This can be mitigated by working with general $d$-degree monomials (instead 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 $P_2$ has been studied by Shpilka [Shp02], in a different context. This polynomial is closely related to the Trace-Iterated-Matrix-Multiplication, 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 $x_{ij}^{(r)}$. 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 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 $\Sigma^{[k]}\Pi\Sigma$ 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 $\epsilon \to 0$); 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 $\Sigma^{[2]}\Pi\Sigma$, the universal expressive power!

2. Also, efficient PIT algorithms are known for $\Sigma^{[k]}\Pi\Sigma$ circuits [LST21, DDS21]. However, they are not known to imply strong lower bounds in the same model. This is primarily due to the fact that this model is computationally weak; and does not facilitate the known proof techniques for proving lower bounds from derandomization results. The latter techniques, in particular, are insensitive to fanin-$k$ [HS80, KI03, AGS19, KS19]; as they require the ability to interpolate, which blows-up the top-fanin of the model.

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

4. 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 $\epsilon$-power that hides in the expression.

5. 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, $\epsilon^{m-d} \cdot \det_d \not\in \sigma_r(Ch_m(W))$, for any $m \leq 2^{\Omega(d)}$ and constant $r$. Geometrically, Theorem 1 exponentially separates the padded $k$-th-secant varieties of the Chow variety of $W = C^{n+1}$, for all constants $k$.

**Comparing our parameters with [LST21] breakthrough.** Interestingly, our proof method works with both IMM$_{k+1,d}$ (where instead of trace, one is interested in (1,1)-th entry), and the trace version Tr-IMM$_{k+1,d}$, already defined before. The linear-rank-based lower bound method by Limaye, Srinivasan & Tavenas [LST21] shows (which extends to the border as well) ‘only’ a superpolynomial separation between depth-3 (unbounded fanin) circuits and IMM, with different parameters [thus showing a superpolynomial separation between VBP and $\Sigma^{[k]}\Pi\Sigma$]. Their dominating variable is $n$ (with $d = o(\log n)$) and they showed an $n^{\sqrt{d}}$ lower bound (which further weakens above depth-
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 $\Sigma \land \Sigma$; they compute polynomials of the form $\sum_i \ell_i^a$, for linear polynomials $\ell_i$. 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. $P_{k+1}$, which after substituting constantly ($\leq k$)-many variables to 0, can not be approximated by a ratio of two sub-exponential size $\Sigma \land \Sigma$ circuits.

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 $\Sigma \land \Sigma$; they compute polynomials of the form $\sum_i \ell_i^a$, for linear polynomials $\ell_i$. 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. $P_{k+1}$, which after substituting constantly ($\leq k$)-many variables to 0, can not be approximated by a ratio of two sub-exponential size $\Sigma \land \Sigma$ 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 $R[[x_1, \ldots, x_n]]$ (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 $\log_y(f) = (\partial_y f) / f$ is another key tool which linearizes the product gate, since

$$dlog_y(f \cdot g) = \partial_y(fg) / (fg) = (f \cdot \partial_y g + g \cdot \partial_y f) / (fg) = \log_y(f) + \log_y(g).$$

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

**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 $f(0, x_2, \ldots, x_n)$. For an arbitrary $\ell$, we apply a suitable isomorphism $\phi$, on the space of linear forms, which sends $\ell$ 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 $f(\ell_1, \ldots, \ell_n)$, for some linear forms $\ell_i$ which are $x_1$-free, and rank$_F(\ell_1, \ldots, \ell_n) = n - 1$.

Similarly, this can be extended to looking at $f(x) \bmod \langle L_1, \ldots, L_t \rangle$, for some suitable $t$, where $\langle L_1, \ldots, L_t \rangle$ is the ideal generated by $t$-many linear forms $L_i$, for $i \in [t]$, i.e., any $g \in \langle L_1, \ldots, L_t \rangle$ 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 $f(\ell_1, \ldots, \ell_n)$ with rank$_F(\ell_1, \ldots, \ell_n) \geq n - t$; one could think of $\langle \ell_1, \ldots, \ell_n \rangle$ as containing at least $\{x_{t+1}, \ldots, x_n\}$. This perspective will play a crucial role in the proofs; see Lemma 2.
Lower bound for \(\Sigma[k]\Pi\Sigma\) circuits: The overall plan

Let \(x := \{x_1, \ldots, x_n\}\), \(n := (k+1)d\), and polynomial \(f := P_{k+1,d} \in \mathbb{F}[x]\). Suppose that \(f \in \Sigma[k]\Pi\Sigma\) of size \(s\), i.e. \(g = f + \epsilon \cdot S\), where 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 \text{ case.} \quad \text{[BIZ18, Prop. A.12] showed that } \Pi\Sigma = \Pi\Sigma. \text{ Eventually, the proof relies on the fact that the product distributes the } \epsilon\text{-powers showing } (\Pi\Sigma) \subseteq \Pi(\mathcal{C}), \text{ for any reasonable class } \mathcal{C} \text{ (see Lemma 14); here by } f \in \Pi\Sigma \text{ we mean } f =: \prod f_i, \text{ where } f_i \in \mathcal{C}. \text{ Thus, it suffices to show that } P_{2,d} \text{ does not have any } \Pi\Sigma \text{ circuit (i.e. impossibility). It is evident that } P_{2,d} \text{ is an irreducible polynomial. One way to argue is: since, } P_{2,d} \text{ is a homogeneous polynomial, all its factors must be homogeneous. Thus, take a linear form } \ell \text{ that divides } P_{2,d}. \text{ As argued in Lemma 2, there are linear forms } \ell_1, \ldots, \ell_n \text{ with } n = 2d \text{ such that}

\[ P_{2,d} \mod \ell = P_{2,d}(\ell_1, \ldots, \ell_n), \quad \text{and } \mathrm{rank}(\ell_1, \ldots, \ell_n) = n - 1. \]

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

1. (Easy case). \textit{Both } T_i \text{ have at least one factor whose } \epsilon\text{-free term is a homogeneous linear form over } \mathbb{F};

2. (Intermediate case). \textit{Exactly one of } T_i, \text{ say wlog, } T_1, \text{ has at least one factor whose } \epsilon\text{-free term is a homogeneous linear form;}

3. (Hard case). \textit{None of the factors of } T_i \text{ has } \epsilon\text{-free term as a homogeneous linear form.}

\bullet \text{ Handling the first case. The first case is almost similar to (but not exactly) } k = 1. \text{ Suppose, a linear form } \ell_1(x, \epsilon) \mid T_1 \text{ over } \mathbb{F}(\epsilon). \text{ Ideally, we would like to look at } g \mod \ell_1 \text{ which reduces fanin to } 1. \text{ Understanding } g \mod \ell_1. \text{ Wlog, } \ell_1 := x_1 + \epsilon \tilde{l}_1, \text{ where } \tilde{l}_1 \text{ is } x_1\text{-free; if } \ell_1(\epsilon = 0) \neq x_1, \text{ 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 \mod \ell_1 \text{ is equivalent to looking at } g\big|_{x_1=-\epsilon \cdot \tilde{l}_1}. \text{ The immediate question would be what happens to the limit, after this substitution. Interestingly,}

\[\lim_{\epsilon \to 0} g(-\epsilon \cdot \tilde{l}_1, x_2, \ldots, x_n) = \lim_{\epsilon \to 0} f(-\epsilon \cdot \tilde{l}_1, x_2, \ldots, x_n) = f(0, x_2, \ldots, x_n).\]

So, it is like looking at \(f\big|_{x_1=0}\).

Since, \(f\) is homogeneous, so is \(f(0, x_2, \ldots, x_n)\). By the reduced fanin, we have \(f(0, x_2, \ldots) \in \Pi\Sigma\), 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 a homogeneous linear form $\ell_1 \mid T_1$ and the above argument shows that $f(0, x_2, \ldots) \in \Pi_{\Sigma}$, a non-homogeneous circuit; since each linear factor of $T_2$ is non-homogeneous by assumption. But this cannot happen as $f(0, x_2, \ldots)$ 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 $\ell_1$ 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 $\ell_i|_{c=0} = 1$, 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 $\ell \in F[x]$. 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 $\epsilon$-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 poly($s$)-sized $\Sigma \wedge \Sigma$ circuit, (ii) $f = P_{3, d}$, requires $2^{\Omega(d)}$-sized $\Sigma \wedge \Sigma$ 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 $\Phi : F(\epsilon)[x] \to F(\epsilon)[x, z]$ 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 $\epsilon$ 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

$$\frac{\Phi(g)}{\tilde{T}_2} = \epsilon^{a_2} \cdot \frac{\Phi(T_1) / \tilde{T}_2}{\frac{\partial \Phi(g)}{\tilde{T}_2}} \Rightarrow \frac{\partial \Phi(g)}{\tilde{T}_2} = \frac{\partial \Phi(T_1)}{\tilde{T}_2},$$

where $\partial_z : \epsilon \mapsto \frac{d(\epsilon)}{dz}$. 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 $\epsilon \to 0$.

What is $\lim_{\epsilon \to 0} \frac{\partial \Phi(g)}{\tilde{T}_2}$? Note: (i) $\Phi$ 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 $\epsilon$, and $\lim_{\epsilon \to 0} \tilde{T}_2 = : c \mod z$, for $c \in F \setminus \{0\}$. Using these, it can be seen that $\frac{\partial \Phi(g)}{\tilde{T}_2} \equiv c' \cdot z^{d-1} \cdot f(x) \mod \langle z^d, c \rangle$, for nonzero constant $c'$ (see Claim 5). Formally, the following equation is well-defined:

$$c' \cdot z^{d-1} \cdot f(x) \equiv \frac{\partial \Phi(g)}{\tilde{T}_2} \equiv \frac{\partial \Phi(T_1)}{\tilde{T}_2} \mod \langle \epsilon, z^d \rangle. \tag{1}$$

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

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

Note that the $\text{dlog}$ operator distributes the product gate into summation giving $\text{dlog}(\Pi_{\Sigma}) = \Sigma \text{dlog}(\Sigma)$, where $\Sigma$ 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 $\mathcal{R} := F[z] / \langle z^d \rangle$. To achieve this, analytically expand $1/\ell$, where $\ell$ is a linear polynomial dividing $\Phi(T_i), i \in [2]$, as sum of powers of
linear functions; using the inverse identity:

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

Since $\Sigma \land \Sigma$ is ‘closed’ under taking product and addition (Appendix B), we obtain a final $\Sigma \land \Sigma$ circuit for $\text{dlog} \ (\Phi(T_1) / \hat{T}_2)$. Details of this step can be found in Claim 7. Therefore, $\partial_z (\Phi(T_1) / \hat{T}_2)$ is actually in a blookted class — $(\Pi \Sigma / \Pi \Sigma) \cdot (\Sigma \land \Sigma)$, over $R(\epsilon, x)$ — which computes elements of the form $(A/B) \cdot C$, where $A, B \in \Pi \Sigma$ and $C \in \Sigma \land \Sigma$. Moreover, from Equation 1, computing $\lim_{\epsilon \to 0} \partial_z (\Phi(T_1) / \hat{T}_2) \mod z^d$ suffices.

The crucial thing about these $\Pi \Sigma$ circuits, that appeared, is that their $|_{x=0=\epsilon}$ 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 $\lim_{\epsilon \to 0} \partial_z (\Phi(T_1) / \hat{T}_2)$, has a poly($s$)-size $\Sigma \land \Sigma$ circuit. In particular, coefficient of $z^{d-1}$ has also an $\Sigma \land \Sigma$-circuit of 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 poly($s$)-size $\Sigma \land \Sigma$. We point out that $\Sigma \land \Sigma$ 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 $\Sigma \land \Sigma$, it should have an upper bound of poly($s$). This implies the desired lower bound of $s \geq 2^{\Omega(d)}$; while trivially $f = P_{k+1,d}$ has $\Sigma^{[k+1]} \Pi \Sigma$-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):

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

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 $P_i, Q_i \in \Sigma \land \Sigma$, 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 $\Sigma^{[k]} \Pi \Sigma$-circuit, then there are linear forms $\ell_1, \ldots, \ell_n$, with $n := (k + 1)d$, such that rank($\ell_1, \ldots, \ell_n$) $\geq n - k$, and $P_{k+1,d}(\ell_1, \ldots, \ell_n)$ can be computed by an ‘all-non-homogeneous’ $\Sigma^{[m]} \Pi \Sigma$-circuit, for some $m \leq k$ (Lemma 2),

(ii) $P_{k+1,d}(\ell_1, \ldots, \ell_n) \cdot (\Sigma \land \Sigma / \Sigma \land \Sigma)$, 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 $\lim_{\epsilon \to 0} \text{Gen}(1, \cdot)$ is a ratio of two ‘small size’ $\Sigma \land \Sigma$ 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 $\Pi \Sigma |_{x=0=\epsilon}$ 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 \( P_{k+1,d} \cdot (\Sigma \wedge \Sigma / \Sigma \wedge \Sigma) \). 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 dog and finally taking limit, we will eventually get that

\[
\partial_z (P_{k+1,d}(zx + a) / \Sigma \wedge \Sigma) \equiv \Sigma \wedge \Sigma \cdot (\Sigma \wedge \Sigma) \mod z^d,
\]

for poly(s)-sized \( \Sigma \wedge \Sigma \) and \( \Sigma \wedge \Sigma \)-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 \( \Sigma \wedge \Sigma / \Sigma \wedge \Sigma \) (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 \( \Sigma^{[k]} \Sigma \), introduced in [DDS21]. DiDIL is an acronym for the steps: Divide, Derive, Induct, with Limit. 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-derivative 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 \( \Sigma^{[k]} \Sigma \)-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, GFKS16], 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 $\Sigma[2] \Pi[2] \Sigma$ circuits.

Even for showing lower bound for $\Sigma[2] \Pi[2] \Sigma$ circuits, we can not probably expect to use factoring or chinese remaindering (CRT) based ideas, since mod $\langle e^M \rangle$, we get non-unique, usually exponentially many, factors. For e.g., $x^2 \equiv (x - a \cdot e^{M/2}) \cdot (x + a \cdot e^{M/2}) \mod \langle e^M \rangle$; for all $a \in \mathbb{F}$. In this case, there are, in fact, infinitely many factorizations. Moreover, $\lim_{\epsilon \to 0} 1/e^M \cdot (x^2 - (x - a \cdot e^{M/2}) \cdot (x + a \cdot e^{M/2})) = a^2$. 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 $\Sigma[k] \Pi[k] \Sigma$ 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 $\lim_{\epsilon \to 0}$. For e.g. let $f_1 = x_1 + e^M \cdot x_2$, and $f_2 = x_1$, where $M$ is arbitrary large. Then, the underlying Jacobian $J(f_1, f_2) = e^M$ 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: $\nabla VBP_2 = \nabla VBP_3 = \cdots = \nabla VBP_k$, for any constant $k$, whereas it is also known that $\nabla VBP_2 \not\subseteq \nabla VBP_3 = \cdots = \nabla VBP_k$ [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 $[n] = \{1, \ldots, n\}$, and $x = (x_1, \ldots, x_n)$. 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 $\ell =: a_1 x_1 + \ldots + a_n x_n$, a linear polynomial without the constant term, as a *linear form*.

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

**Explicit.** A family $\{P_n \in \mathbb{F}[x_1, \ldots, x_n] \mid n \geq 1\}$, is said to be *explicit* if there is a deterministic algorithm that given as input $1^n$ and a monomial $m$ over the variables $x_1, \ldots, x_n$, 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 $\partial_y$ denotes the partial derivative wrt variable $y$. One important property of $\text{dlog}$ 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 ΠΣ, compute polynomials of the form \( \prod \ell_i \), where \( \ell_i \) are affine linear polynomials. Depth-3 circuits with top-fanin \( k \) are denoted as \( \Sigma^k \Pi \Sigma \); they compute polynomials of the form \( \sum_{i \in [k]} \prod \ell_{ij} \), where \( \ell_{ij} \) are affine linear functions. Also, depth-3 diagonal circuits are denoted as \( \Sigma^3 \Pi \Sigma \); they compute polynomials of the form \( \sum_{i=1}^k \ell_i^\mathsf{d} \), for linear polynomials \( \ell_i \). When we write \( \Sigma \land \Sigma \), 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 \( L_1, \ldots, L_r \), we denote \( \langle L_1, \ldots, L_r \rangle \), 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 \( C \) and \( D \) defined over ring \( R \), our bloated model is any combination of sum, product, and division of polynomials from respective classes. For instance, \( C / D = \{ f \cdot g : f \in C, 0 \neq g \in D \} \), similarly \( C \cdot D \) for products, \( C + D \) for sum, and other possible combinations. Also we use \( C_R \) to denote the basic ring \( R \) over which \( C \) is being computed.

Valuation. Valuation is a map \( \text{val}_y : R[y] \to \mathbb{Z}_{\geq 0} \), 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. the field of rationals \( \mathbb{Q} \), the field of reals \( \mathbb{R} \), the field of \( p \)-adics \( \mathbb{Q}_p \) etc.). All our results hold for other fields (eg. \( \mathbb{F}_p^\ast \) etc.) of large characteristic \( p \). We also denote \( \mathbb{F}^\ast \) as the multiplicative group of the field \( \mathbb{F} \).

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

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

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 \( a = (a_1, \ldots, a_n) \). 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 \( \Sigma^k \Pi [d] \Sigma \); these circuits compute polynomials \( f \), where \( e^M \cdot f(x) + e^{M+1} \cdot S(x, e) = \sum_{i \in [k]} T_i \), where each \( T_i = \prod_{j \in [d]} \ell_{ij} \), with \( \ell_{ij} \in \mathbb{F}[e][x] \), are affine linear functions. Since, in the homogeneous setting, exponential lower bound is easy-to-show, we somehow want to ‘exclude’ homogeneous polynomials \( \ell_i \) 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 \( \Sigma^{[k]} \Pi \Sigma \)-circuit, then there exists \( 0 \leq m \leq k \), and (homogeneous) linear forms \( \ell_1, \ldots, \ell_n \) such that

1. (large rank). \( \text{Rank}(\ell_1, \ldots, \ell_n) = n - t \), for some \( 0 \leq t \leq k - m \), and each \( \ell_i \) is \( \{x_1, \ldots, x_i\} \)-free;

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

\[
\epsilon^M \cdot f(\ell_1, \ldots, \ell_n) + \epsilon^{M+1} \cdot H(x, e) = T_1 + \ldots + T_m,
\]

where \( \forall i \in [m] \),

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

**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 \( \epsilon^M \cdot f(\ell_1, \ldots, \ell_n) = 0 \mod \epsilon^{M+1} \implies f(\ell_1, \ldots, \ell_n) = 0 \). 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, \( \epsilon^M \cdot f + \epsilon^{M+1} \cdot S(x, e) = T_1 + \ldots + T_k \), where \( T_i \in \mathbb{F}[e, x] \), product of linear polynomials. Now if already each \( T_i \) is such that after taking out the maximum \( \epsilon \)-power, \( \epsilon \)-free term \( \mod \langle x_1, \ldots, x_n \rangle \in \mathbb{F} - \{0\} \), then we are already done. Otherwise, it must happen that at least one of the \( T_i \) has a linear factor whose \( \epsilon \)-free term is a linear form.

Formally, after possible relabeling, \( T_1 := g_1 \cdot h_1 \), in \( \mathbb{F}[e, x] \), where \( g_1|_{e=0} \) 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 \( S(x, e) := S_0(x, e) \) (since this is the base case or the 0-th step). Also, let \( \mathcal{R} := \mathbb{F}[e] / \langle \epsilon^{M+1} \rangle \). We work in \( \mathbb{F}[x, e] \cap \mathcal{R}[x] \); we would not keep mentioning the underlying 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 \( \epsilon \)-free, and \( g_{11} \in \mathbb{F}[e, x] \). Since \( g_{10} \) is a nonzero linear form by assumption, extend \( \{1, g_{10}\} \) to a basis say \( \{1, L_1 := g_{10}, L_2, \ldots, L_n\} \) of the linear space \( V_n := \{\sum_{i=1}^n a_i x_i + a_0 | \forall a_i \in \mathbb{F}\} \), which has dimension \( n + 1 \). Define an isomorphism \( \phi_1 \) on \( V_n \), which maps \( L_i \mapsto x_i \), for all \( i \).

Let us write \( \phi_1(\epsilon) := (1 + \epsilon \cdot p_1(\epsilon)) \cdot x_1 + \epsilon \cdot R_1 \), 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 \( \phi_1(\epsilon) = x_1 \) and other terms of \( x_1 \) must come from multiple of \( \epsilon \) terms.

By assumption, \( \epsilon^M \cdot f + \epsilon^{M+1} S_0 = T_{10} + \ldots + T_{k0} \); apply \( \phi_1 \) 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 \( \psi_1 \). We want to stress: since \( 1/(1 + \epsilon \cdot p_1(\epsilon)) \mod \langle \epsilon^{M+1} \rangle \) exists, and belongs to \( \mathbb{F}[e] \), we can think of substituting \( x_1 := -\epsilon \cdot q_1(\epsilon) \cdot R_1 \), where \( q_1 := 1/(1 + \epsilon \cdot p_1(\epsilon)) \mod \epsilon^{M+1} \). By doing this, we get the following equation:

\[
\epsilon^M \cdot \psi_1 \circ \phi_1(f(x)) + \epsilon^{M+1} \cdot \psi_1 \circ \phi_1(S_0) = \sum_{i=k_1+1}^{k} \psi_1 \circ \phi_1(T_{i0}) .
\]
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 \( i \in \{k_1 + 1, k\} \):

\[
T_{i1} := \psi_1 \circ \phi_1(T_{i0}) \text{ and } v_{i1} := \text{val}_e(T_{i1}).
\]

Note that \( T_{i1} \) are elements in \( \mathbb{F}[e, x] \cap \mathcal{R}[x] \) which are \( x_1 \)-free. Just to state the obvious, if one of them becomes 0 after applying \( \psi_1 \), 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 \( \lim_{e \to 0} \psi_1 \circ \phi_1(f) =: f(\ell^{(1)}_1, \ldots, \ell^{(1)}_n) =: D_1(x) \) (which is \( x_1 \)-free), for linear forms \( \ell^{(1)}_i \in \mathbb{F}[x] \), where \( \text{rank}(\ell^{(1)}_1, \ldots, \ell^{(1)}_n) = n - 1 \); since \( \psi_1 \) 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):

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

Now if, for all \( i \in \{k_1 + 1, k\} \), \( (T_{ij}/e^{v_{ij}}) |_{e=0} \equiv 0 \mod (x_1, \ldots, x_n) \in \mathbb{F} - \{0\} \), 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 \( (T_{i1}/e^{v_{i1}}) \) has a linear factor (say \( g_{j+1} \)) with \( e \)-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 \( (k - k_j - 1) \). For \( j \geq 1 \), let us say we have

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

where

1. \( v_{ij} \) is a linear factor (say \( g_{j+1} \)) with \( e \)-free term being a linear form, along with \( \deg(T_{ij}) \leq s \). Moreover, \( k_j \geq j \).

2. \( D_j = \psi_j \circ \phi_j(D_{j-1}) = f(\ell^{(j)}_1, \ldots, \ell^{(j)}_n) \), where linear map \( \phi_j \) is invertible, and linear map \( \psi_j \) has the property that it sets \( x_1, \ldots, x_j \) to 0, but keeps the other \( n - j \) variables unchanged. Equivalently, \( \ell^{(j)}_i \in \mathbb{F}[x] \) are linear forms, which are \( \{x_1, \ldots, x_j\} \)-free, with \( \text{rank}(\ell^{(j)}_1, \ldots, \ell^{(j)}_n) = n - j \).

In the base case, of course, we used the notation: \( D_0 = f(x) \) and \( \ell^{(0)}_i = x_i \). 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 \( \mathbb{F}[e, x] \) where \( g_{j+1} |_{e=0} \neq 0 \) is a nonzero linear form in \( x_{j+1}, \ldots, x_n \). Let \( g_{j+1} := g_{j+1,0} + e \cdot g_{j+1,1} \). Since \( g_{j+1,0} \) is a linear form, extend \( \{1, x_1, \ldots, x_j, g_{j+1,0}\} \) to a basis say \( \{1, L_1 := x_1, L_2 := x_2, \ldots, L_j := x_j, \ell_{j+1} := g_{j+1,0}, L_{j+2}, \ldots, L_n\} \).

Define an isomorphism \( \phi_{j+1} \) on \( V_n \) which maps \( L_i \mapsto x_i \) for \( i \in [n] \). Let us write \( \phi_{j+1}(g_{j+1}) = (1 + e \cdot p_{j+1}(e)) \cdot x_{j+1} + e \cdot R_{j+1} \), where \( p_{j+1}(e) \in \mathbb{F}[e] \), and \( R_{j+1} \) is a polynomial which is \( \{x_1, \ldots, x_j, x_{j+1}\} \)-free; this can be achieved similarly as before, by clubbing all the \( x_{j+1} \) coefficients together which must look like \( 1 + e \cdot p_{j+1}(e) \).
Apply \( \phi_{j+1} \) both side and substitute \( x_{j+1} := -e \cdot R_{j+1}/(1 + e \cdot p_{j+1}(e)) \), which essentially sends \( g_{j+1} \to 0 \). We denote the substitution map by \( \psi_{j+1} \). Here again we want to stress: since 
\[
e/(1 + e \cdot p_{j+1}(e)) \in \mathcal{R},
\]
we can think of substituting \( x_{j+1} := -e \cdot q_{j+1}(e) \cdot R_{j+1} \), where \( q_{j+1} := 1/(1 + e \cdot p_{j+1}(e)) \mod e^{M+1} \). By doing this, we get the following equation:

\[
e^M \cdot \psi_{j+1} \circ \phi_{j+1}(D_j(x)) + e^{M+1} \cdot \psi_{j+1} \circ \phi_{j+1}(S_{j+1}) = \sum_{i=k_{j+1}}^k \psi_{j+1} \circ \phi_{j+1}(T_{ij}).
\]

In Equation 3 we used the fact that \( \psi_{j+1} \circ \phi_{j+1}(g_{j+1}) = 0 \); moreover, it kills many such \( T_{ij} \)'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 \( k_{j+1} > k_j \). This proves hypothesis-1-part-2.

Denote \( \lim_{e \to 0} \psi_{j+1} \circ \phi_{j+1}(D_j) =: D_{j+1} \), 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 \( T_{ij+1} := \psi_{j+1} \circ \phi_{j+1}(T_{ij}) \) and \( \nu_{ij+1} := \text{val}_e(T_{ij+1}) \). Note that, \( \deg(T_{ij+1}) \leq s \) remains an invariant, due to linear transformations. Therefore, one can write Equation 3 as follows:

\[
e^M \cdot D_{j+1}(x) + e^{M+1} \cdot \text{blah} = \sum_{i=k_{j+1}}^k T_{ij+1}.
\]

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 case, 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.

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 \( \Sigma^k \Pi \Sigma \) circuits.

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

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

By definition, \( P_{k+1,d}(\ell_1, \ldots, \ell_n) = \sum_{i \in [k+1]} \prod_{j \in [d]} \ell_{i-1}d+j \). It is easy to see that \( \prod_{j \in [d]} \ell_{(i-1)d+j} \) must be of the form \( m_t \cdot (\text{product of } t_i \text{ non-variable linear forms}) \), for some \( t_i \leq n-r \), since there are \( n-r \) many non-variable linear forms available and \( m_t \) is a multilinear monomial of degree \( d-t \); 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 \( m_{t_i} \), 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 \( m_i \) in at most \( n - r \leq k < d \) many variables (very few!) in its support.

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

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 \( 1/e^M \) in RHS representation, but this does not change the fundamental structure of polynomials \( T_i \).

**Problem 1** (Reduced lb problem). For a constant \( k \), and some \( m \in [k] \), with \( 1 \leq k < d \), say

\[ P_{k+1,d}(\ell_1, \ldots, \ell_n) + \epsilon \cdot H(x, \epsilon) = T_1 + \ldots + T_m \]

where –

1. (large rank). \( \ell_i \) are homogeneous linear forms (which are \( \{x_1, \ldots, x_{k-m}\} \)-free) such that \( \text{rank}(\ell_1, \ldots, \ell_n) = n - (k - m) \), and,

2. (all-non-hom.). \( T_i \) are product of linear polynomials such that \( (T_i/e^{v_i})(|x=\epsilon=0) \in F \setminus \{0\} \) (where \( v_i := \text{val}_k(T_i) \)).

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 \( \sum[k] \Pi \Sigma \)-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, \( P_{k+1,d}(\ell_1, \ldots, \ell_n) \) 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 \( P_{k+1,d}(\ell_1, \ldots, \ell_n) = 0 \), a contradiction!

2. Interestingly, the same robustness also holds for \( \text{IMM}_{k+1,d}, \text{det}_d, \text{perm}_d \) etc.

3. \( P_{k+1,d}(\ell_1, \ldots, \ell_n) \) 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 \( \ell_i \) are linear forms, the polynomial \( P_{k+1,d}(\ell_1, \ldots, \ell_n) \) is still homogeneous.

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

Assume that \( P_{k+1,d}(\ell_1, \ldots, \ell_n) =: f(\ell_1, \ldots, \ell_n) \) can be computed by the special \( \sum[m] \Pi \Sigma \), 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 say that a circuit $C$ is in the class $\text{Gen}(m, s)$, over the fraction-ring $R(x)$, with parameter $m$ and size $s$, if it computes polynomials $f \in R(x)$, of the form $f = \sum_{i \in [k]} T_i$, 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 $U_i \mod \langle x_1, \ldots, x_n \rangle$ and $V_i \mod \langle x_1, \ldots, x_n \rangle$ invertible in $R$, and
2. $P_i, Q_i \in \Sigma \land \Sigma$.

Further, size($C$) := $\sum_{i \in [m]} \text{size}(T_i)$, and size($T_i$) := size($U_i$) + size($V_i$) + size($P_i$) + 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 $T_i \mod \langle x \rangle$ has nonzero constant-term (after extracting the $e$-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 $f(x) := P_{k+1,d}(x_1, \ldots, x_n)$, of degree $d$ with $n := (k + 1)d$.

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

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

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

**Bloat out:** Reducing $\Sigma^{[m]} \Pi \Sigma$ to $\text{Gen}(m - 1, \cdot)$. Let $g_0 := f_0 + e \cdot S_0$, where $f_0 := f(\ell_1, \ldots, \ell_n)$ 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 $\mathbb{F}(e)$ with the property: after dividing by the $e$-power of $e$-valuation, it is a nonzero constant mod $\langle x_1, \ldots, x_n \rangle$. 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 $d_0 := d + 1$.

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

$$\Phi : \mathbb{F}(e)[x] \to \mathbb{F}(e)[x, z]$$

such that $x_i \mapsto z \cdot x_i$. 

18
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 $\Phi(T_{i,0}) := e^{a_{a_0}} \cdot \tilde{T}_{i,0}$, where $\tilde{T}_{i,0} := t_{i,0} + e \cdot \tilde{\epsilon}_{i,0}(x, z, \epsilon)$ (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 $\Phi(\tilde{g}_0)$ by $\tilde{T}_{m,0}$ and derive wrt $z$:  

$$
\Phi(f_0) \cdot \tilde{T}_{m,0} / 0 = e^{a_{a_0}} + \sum_{i \in [m-1]} \Phi(T_{i,0}) / \tilde{T}_{m,0} \quad \text{[Divide]}
$$

\[ \implies \partial_z (\Phi(f_0) / \tilde{T}_{m,0}) + e \cdot \partial_z \left( \Phi(S_0) / \tilde{T}_{m,0} \right) = \sum_{i \in [m-1]} \partial_z (\Phi(T_{i,0}) / \tilde{T}_{m,0}) \quad \text{[Derive]}
\]

\[ = \sum_{i \in [m-1]} (\Phi(T_{i,0}) / \tilde{T}_{m,0}) \cdot \text{dlog}(\Phi(T_{i,0}) / \tilde{T}_{m,0}) \quad (4)
\]

\[ =: \tilde{g}_1.
\]

**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. $\lim_{\epsilon \to 0} g_1 = f_1$, where $g_1$ (resp. $f_1$) is well-defined over $\mathcal{R}_1(e, x)$ (resp. $\mathcal{R}_1(x)$). 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  

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

We already remarked that $f_0$ is a homogeneous degree-$d$ polynomial; hence $\Phi(f_0) = z^d \cdot f_0$, and $t_{m,0} \in \mathbb{F}[x]$; 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, $\Phi(f_0) / \tilde{T}_{m,0} \in \mathbb{F}(x, \epsilon)[[z]]$. Moreover, $(\Phi(f_0) / \tilde{T}_{m,0})|_{\epsilon=0} = \Phi(f_0) / t_{m,0} \in \mathbb{F}(x, z)$. Combining these it follows that  

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

Since $t_{m,0} \equiv c \mod z$, 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 $z^{d_1+1}$. Thus after dividing by the 0-valuation element (wrt $z$), it still holds mod $z^{d_1+1}$; finally after differentiation it must hold mod $z^d$.

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

However, we stress that we also think of these intermediate expressions as elements over $\mathbb{F}(x, z, \epsilon)$, 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 $(\Pi\Sigma / \Pi\Sigma) \cdot (\Sigma \land \Sigma / \Sigma \land \Sigma)$-circuit over $\mathcal{R}_1(x, \epsilon)$.
**Inductive step (j-th step):** Reducing Gen($m - j, \cdot$) to 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 $\mathcal{R}_j(x, \epsilon)$, such that it approximates $f_j$ correctly, where $f_j \in \mathcal{R}_j(x)$, where $\mathcal{R}_j := \mathbb{F}[z]/(z^d)$.

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 $\mathcal{R}_j(\epsilon)[x]$. Each can be thought as an element in $\mathbb{F}(x, z, \epsilon) \cap \mathbb{F}(x, \epsilon)[[z]]$ 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 $\mathbb{F}(x, \epsilon)[[z]]$), then $\left(\frac{g_j}{z^b}\right)|_{z=0} = f_0 \cdot (\Sigma \wedge \Sigma/\Sigma \wedge \Sigma)$, where $\Sigma \wedge \Sigma$ are z-free (in $\mathbb{F}(\epsilon)[x]$). Equivalently, the coefficient of the minimum z-power in $f_j$, is of the form $f_0 \cdot (\Sigma \wedge \Sigma/\Sigma \wedge \Sigma)$, where both $\Sigma \wedge \Sigma$ 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 $T_{m-j,j} := e^{a_{m-j}} \cdot \bar{T}_{m-j,j}$, where $\bar{T}_{m-j,j} := (t_{m-j,j} + e \cdot \bar{t}_{m-j,j})$ is not divisible by $\epsilon$. Divide $g_j := f_j + e \cdot S_j$, by $\bar{T}_{m-j,j}$, to get:

$$
\frac{f_j}{\bar{T}_{m-j,j}} + e \cdot S_j / \bar{T}_{m-j,j} = e^{a_{m-j}} + \sum_{i \in [m-j-1]} T_{i,j} / \bar{T}_{m-j,j}
$$

$$
\implies \partial_z \left( \frac{f_j}{\bar{T}_{m-j,j}} \right) + e \cdot \partial_z \left( S_j / \bar{T}_{m-j,j} \right) = \sum_{i \in [m-j-1]} \partial_z \left( T_{i,j} / \bar{T}_{m-j,j} \right)
$$

$$
= \sum_{i \in [m-j-1]} \left( T_{i,j} / \bar{T}_{m-j,j} \right) \cdot \log \left( T_{i,j} / \bar{T}_{m-j,j} \right)
$$

$$
= g_{j+1}.
$$

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

$$
T_{i,j+1} := \left( T_{i,j} / \bar{T}_{m-j,j} \right) \cdot \log \left( T_{i,j} / \bar{T}_{m-j,j} \right), \text{ and } f_{j+1} := \partial_z(f_j / t_{m-j,j}).
$$

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

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

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

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 $R_{j+1}(e, x)$, or equivalently mod $z^{d_{j+1}}$; this is because of the division by $z$-valuation of $v_{m-j}$ and then differentiation, showing induction-hypothesis-(1). So, Eqn. (5) being computed mod $z^{d_{j+1}}$ is indeed valid. Further,
\[
\text{val}_z(f_j + e \cdot S_j) = \text{val}_z(\sum_{i \in [m-j]} T_{i,j}) \geq v_{m-j} \overset{\text{setting } e=0}{\Rightarrow} \text{val}_z(f_j) \geq v_{m-j}.
\]
Hence, $f_j/\tilde{T}_{m-j}$ is $\in \mathbb{F}(x, e)[[z]]$ (by Lemma 9). Further, since $\lim_{e \to 0} \tilde{T}_{m-j} = t_{m-j}$, we must have that $f_j/t_{m-j} \in \mathbb{F}(x)[[z]]$ and thus, $f_{j+1}$ exists. Finally, it is also easy to see that $\text{val}_z(S_j) \geq v_{m-j}$ if there is an $e$-power in $e \cdot S_j$ such that coefficient of $v_{m-j}$ power of $z$ exists, then the valuation of $f_j + e \cdot S_j$ must be $< v_{m-j}$ as well, a contradiction!

Thus, $\lim_{e \to 0} \partial_z(g_j/\tilde{T}_{m-j}) = \lim_{e \to 0} g_{j+1} = f_{j+1}$. This finishes Claim 6.

**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 $\text{dlog}(\Sigma \cdot \Sigma) \in \Sigma \cdot \Sigma / \Sigma \cdot \Sigma$, 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 $\text{dlog}(\Pi \Sigma) \in \Sigma \cdot \Sigma$ is of poly-size. For the time being, we assume these hold. Then, one can simplify to get the following:
\[
T_{i,j}/\tilde{T}_{m-j} = e^{-a_{m-j}} \cdot ((U_{i,j} \cdot V_{m-j})/(V_{i,j} \cdot U_{m-j})),
\]
and its dlog. Let $U_{i,j+1} := U_{i,j} \cdot V_{m-j}$; similarly $V_{i,j+1} := V_{i,j} \cdot U_{m-j}$. Essentially, dlog computation will produce $(\Sigma \cdot \Sigma/\Sigma \cdot \Sigma)$-circuits, which will further multiply with $P$’s and $Q$’s and we multiply $e^{-a_{m-j}}$ there; for details see Claim 7. This directly means: $U_{i,j+1}|_{z=0}, V_{i,j+1}|_{z=0} \in \mathbb{F}(e) \setminus \{0\}$. This proves the second part of induction-hypothesis-(3).

**Induction-hypothesis-(4).** By induction hypothesis, we have $g_j = z^{b_j} \cdot f_0 \cdot (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma) + z^{b_j+1} \cdot (\cdot)$. Since, $\tilde{T}_{m-j} \in (\Pi \Sigma/\Pi \Sigma) \cdot (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma)$, and that its $\Pi \Sigma|_{z=0} \in \mathbb{F}(e) \setminus \{0\}$, so, $g_j/\tilde{T}_{m-j}$ must look like
\[
g_j/\tilde{T}_{m-j} = z^{b_j} \cdot f_0 \cdot (\Pi \Sigma/\Pi \Sigma) \cdot (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma) + z^{b_j+1} \cdot (\text{blah}).
\]
In the first summand, we already used the fact that $\Sigma \cdot \Sigma$ is *closed* under multiplication (Lemma 10). Now of course, $\text{val}_z(\Sigma \cdot \Sigma/\Sigma \cdot \Sigma)$ in the first summand can be *negative*; however, it is $\geq -b_j$, as otherwise, $\text{val}_z(g_j/\tilde{T}_{m-j}) < 0$, a contradiction.

Since the degree of $z$ is polynomially bounded, we can extract the coefficient of the minimum $z$-power of $\Sigma \cdot \Sigma/\Sigma \cdot \Sigma$ (the $z$-power gets subtracted from $b_j$ accordingly) and we still get $\Sigma \cdot \Sigma/\Sigma \cdot \Sigma$ (this is $z$-free) of size polynomially bounded; since $\Pi \Sigma|_{z=0} \in \mathbb{F}(e) \setminus \{0\}$, it does not contribute anything (except nonzero multiples from $\mathbb{F}(e)$). By definition $\text{val}_z(g_j/\tilde{T}_{m-j}) = b_{j+1} + 1 \geq 1$, since its derivative has valuation $b_{j+1} \geq 1$. Hence, the coefficient of the minimum $z$-power must be of the form $f_0 \cdot \text{coef}_{z^{b_{j+1}}} (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma)$. Since $\text{deg}_{z}$ is polynomially-bounded, extracting the minimum $z$-power individually from numerator and denominator gives a ‘small’ $\Sigma \cdot \Sigma/\Sigma \cdot \Sigma$ 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
\[
f_{j+1} = \lim_{e \to 0} g_{j+1} = \lim_{e \to 0} \left(z^{b_{j+1}} \cdot f_0 \cdot \Sigma \cdot \Sigma/\Sigma \cdot \Sigma + z^{b_{j+1}+1} \cdot (\cdot)\right) = z^{b_{j+1}} \cdot f_0 \cdot (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma) + z^{b_{j+1}+1} \cdot (\cdot).
\]
Therefore, the coefficient of the minimum $z$-power in the above expression is $f_0 \cdot (\Sigma \cdot \Sigma/\Sigma \cdot \Sigma)$, because of the distributive property Property 14.
The overall size blowup. Finally, we show the main step: how to control \( \log \) 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 DiDL). \( T_{i,m-1} \in (\Pi\Sigma/\Pi\Sigma) \) (\( \Sigma \land \Sigma/\Sigma \land \Sigma \)) over \( R_{m-1}(x, \varepsilon) \) of size \( s^{O(m^2)} \). It is computed as an element in \( \mathbb{F}(e, x, z) \), with syntactic degree (in \( x, z \)) \( d^{O(m)} \).

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 \( \Phi \). However the main idea of using \( \log \) and expand it as a power-series is the same, which eventually shows that \( \log(\Pi\Sigma) \in \Sigma \land \Sigma \) with a controlled blowup.

d\log’s effect in \( j = 0 \) case. In this case, we would like to understand \( \log \)’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 \( \log \) distributes over product, it suffices to study \( \log(\ell) \), where \( \ell \in R(e)[x] \) is linear.

However, since \( \Pi\Sigma \) circuits are already non-homogeneous, the scaling \( \Phi \) makes each \( \ell \) of the form \( \ell = A - zB \), where \( A \in \mathbb{F}(e) \setminus \{0\} \) and \( B \in \mathbb{F}(e)[x] \). Using the power series expansion, we have the following, over \( R_1(x, \varepsilon) \):

\[
d\log(\ell) = -\frac{\partial_z (z \cdot B)}{A (1 - z \cdot B / A)} = -\frac{B}{A} \cdot \sum_{j=0}^{d_1-1} \left( \frac{z \cdot B}{A} \right)^j.
\]

Note, \( (B/A) \) and \( (-z \cdot B/A)^j \) have trivial powering circuits (\( \land \Sigma \) over \( R_1 \)), each of size \( O(dn) \). By Lemma 10, we get the final \( \Sigma \land \Sigma \) circuit for \( \log(\Pi\Sigma) \) of size \( O(d^2 \cdot s) \). We use the fact that \( d_1 < d_0 = d + 1 \). Here the syntactic degree blows up to \( O(d) \). This settles \( j = 0 \) case.

d\log’s effect in \( j \geq 1 \) case. For \( j > 0 \), the above equation holds over \( R_j(x) \). However, the degree could be \( D_j \) (possibly > \( d_j \)) of the corresponding \( \ell \) (nonlinear), and after exponentiation further increase to \( D_j \cdot D_j \). This is 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, \( \Sigma \land \Sigma \) is closed under differentiation (Lemma 13), effect of \( \log \) on \( \Sigma \land \Sigma \) is straightforward. Using Lemma 13, we obtain \( \Sigma \land \Sigma / \Sigma \land \Sigma \) circuit, computing \( \log(P_{i,j}) \) (similarly \( \log(Q_{i,j}) \)) of size \( O(D_j^2 \cdot s_j) \). Also, \( \log(U_{i,j} \cdot V_{m-j,i}) \in \Sigma \land \Sigma \), which could be computed using the above Equation. Thus,

\[
d\log(T_{i,j}/\tilde{T}_{m-j,i}) \in \log(\Pi\Sigma/\Pi\Sigma) \pm \Sigma^{[4]} \log(\Sigma \land \Sigma) \\
\subseteq \Sigma \land \Sigma + \Sigma^{[4]} (\Sigma \land \Sigma / \Sigma \land \Sigma) = (\Sigma \land \Sigma / \Sigma \land \Sigma).
\]

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

Size analysis of \( d\log(T_{i,j}/\tilde{T}_{m-j,i}) \). The \( \Sigma \land \Sigma \) expression, obtained from \( d\log(\Pi\Sigma/\Pi\Sigma) \) is of size \( O(D_j^2 d_j s_j) \). Next, there are 4-many \( \Sigma \land \Sigma / \Sigma \land \Sigma \) 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 \( \Sigma \land \Sigma / \Sigma \land \Sigma \) grows up to \( O(D_j) \). Finally, we club \( \Sigma \land \Sigma / \Sigma \land \Sigma \) expressions (4 of them) to express it as a single \( \Sigma \land \Sigma / \Sigma \land \Sigma \) expression using Lemma 13, with size blowup of \( O(D_j^4 s_j^4) \). Finally, add the single \( \Sigma \land \Sigma \) expression of size \( O(D_j^2 s_j) \), and degree \( O(D_j) \), to get \( O(s_j^5 D_j^4 d) \) size representation.

Overall size blowup from \( j \) to \( j + 1 \). Also, we need to multiply with \( T_{i,j}/\tilde{T}_{m-j,i} \) which is of the form \( (\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma \land \Sigma / \Sigma \land \Sigma) \), where each \( \Sigma \land \Sigma \) is basically product of two \( \Sigma \land \Sigma \) expressions.

22
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 $(\Pi \Sigma / \Pi \Sigma) \cdot (\Sigma \land \Sigma / \Sigma \land \Sigma)$-expression with the $\Sigma \land \Sigma / \Sigma \land \Sigma$ expression obtained from dlog-computation, gives a size blowup of $s_{j+1} := s_j^2 D_j^{O(1)} d$.

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

Hence successively doing this $m - 1$ times, we get that $g = g_{m-1} \in Gen(1, s^{O(m m^m)}) \mod z^{d_{m-1}}$ (first point of Lemma 4), such that $\lim_{z \to 0} g_{m-1} = f_{m-1} \in F(x, z)$ (second point of Lemma 4) and moreover, the coefficient of the minimum $z$-power in $g_{m-1}$ is nothing but $f_0 \cdot (\Sigma \land \Sigma / \Sigma \land \Sigma)$. These circuits $\Sigma \land \Sigma$ also have size $s^{O(m m^m)}$; since the degree bound on $\Sigma \land \Sigma$ circuits is $s^{O(m m^m)}$, 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).

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

Now it remains to establish the lower bound. Recall, we have shown that the minimum $z$-power in $g_{m-1}$ gives $f(\ell_1, \ldots, \ell_n) \cdot \Sigma \land \Sigma / \Sigma \land \Sigma$, where $\Sigma \land \Sigma$ are of size $s^{O(m m^m)}$. Moreover, since $g_{m-1} \in Gen(1, s^{O(m m^m)})$, over $F(e)$, the limit of $g_{m-1}$ itself is of the form $\Sigma \land \Sigma / \Sigma \land \Sigma$ (recall: the other factors satisfy $\Pi \Sigma |_{x=0} \in F^*$). Therefore, comparing the coefficients of the minimum $z$-power both sides, we have eventually shown: $f(\ell_1, \ldots, \ell_n) \cdot \Sigma \land \Sigma / \Sigma \land \Sigma = \Sigma \land \Sigma / \Sigma \land \Sigma \implies f(\ell_1, \ldots, \ell_n) = \Sigma \land \Sigma / \Sigma \land \Sigma$. Here $\Sigma \land \Sigma$ circuits have size $s^{O(m m^m)}$, 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 $f(\ell_1, \ldots, \ell_n) = \lim_{z \to 0} (\Sigma \land \Sigma / \Sigma \land \Sigma)$, such that

1. size of the $\Sigma \land \Sigma$ circuits, over $F(e)$, is $s^{O(k \ell^k)}$, and
2. rank($\ell_1, \ldots, \ell_n$) = $n - (k - m)$ $\geq n - k$,

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

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

$$V_{e,i} := \left\langle \frac{\partial (\Sigma \land \Sigma_i)}{\partial x^e} \mid e < \infty \right\rangle_{F(e)}$$ and $V_i := \left\langle \frac{\partial (\Sigma \land \Sigma_i)}{\partial x^e} \mid e < \infty \right\rangle_F$.

Since, size($\Sigma \land \Sigma_i$) $\leq s^{O(k \ell^k)}$, the partial derivative space of $\Sigma \land \Sigma_i$, over $F(e)$, is also bounded by $s^{O(k \ell^k)}$, i.e. dim($V_{e,i}$) $\leq s^{O(k \ell^k)}$ (see [CKW11, Lemma 10.2]). Consider the partial-derivative matrix $M_{e,i}$, where we index the rows by $\partial x^e$, while columns are indexed by monomials supporting $\Sigma \land \Sigma_i$, and each row expresses the operator-values $\partial x^e$ $(\Sigma \land \Sigma_i)$. We have, $q_i := \dim(V_{e,i}) \leq s^{O(k \ell^k)}$ (because of $\Sigma \land \Sigma_i$). So, any $(q_i + 1)$-many polynomials $\frac{\partial (\Sigma \land \Sigma_i)}{\partial x^e}$ are F-linearly dependent. In other words, determinant of any $(q_i + 1) \times (q_i + 1)$ minor of $M_{e,i}$ is 0. Note that $\lim_{e \to 0} M_{e,i} = M_i$, the corresponding partial-derivative matrix for $\Sigma \land \Sigma_i$. 23
Since, \( \det \) is a continuous function, the zeroness of the determinant of any \((q_i + 1) \times (q_i + 1)\) minor of \( M_{e,j} \) translates to the corresponding \((q_i + 1) \times (q_i + 1)\) submatrix of \( M_i \) as well. In particular,

\[
\dim(V_i) \leq q_i \leq s^{O(k^2)}.
\]

From this, it follows that leading monomial in \( \Sigma \wedge \Sigma_i \) (denoted \( \text{LM}(\Sigma \wedge \Sigma_i) \)) has cone-size at most \( s^{O(k^2)} \); see Lemma 15.

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

**Exploiting ‘large rank’**. As argued in Corollary 3, \( f(\ell_1, \ldots, \ell_n) = P_{k+1,d}(\ell_1, \ldots, \ell_n) \) 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: \( \text{size}(\Sigma \wedge \Sigma_i) \geq \text{cone-size}(\text{LM}(\Sigma \wedge \Sigma_i)) \geq \text{cone-size}(\text{LM}(f(\ell_1, \ldots, \ell_n) \cdot \Sigma \wedge \Sigma_2)) \geq 2^d \) (also, see the remark below). Consequently, \( s^{k^2} \geq 2^{\Omega(d)} \). This finishes the lower bound proof.

**Remark.** In the above, we only work with \( \Sigma \wedge \Sigma \). Since, \( \Sigma \wedge \Sigma \) 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 \( \Sigma[k] \Pi \Sigma \wedge \) and \( \Sigma[k] \Pi \Sigma \Pi \), since the cone-size-based argument no longer works.

### 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 \( \Sigma[k] \Pi \Sigma \wedge \)-circuit of size \( s \). We now reduce top-fan-in 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: \( P_{k+1,d}(\ell_1, \ldots, \ell_n) \neq 0 \) for rank\( (\ell_1, \ldots, \ell_n) \geq n - k \). 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-fan-in \( m \) 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.

### 5 Conclusion

In this work, we show a strong top-fan-in-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 \( \Sigma^{[o(n)]} \Pi \Sigma \)-circuits? The current method gives subexponential lower bound only as long as \( k = o(\log n) \).

2. Can we show exponential lower bound for \( \Sigma[k] \Pi \Sigma \wedge \)-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 $\delta$, is $\sum_1^{\delta} \Pi \Sigma \Pi \delta \subseteq \sum_2^{\delta} \Pi \Sigma \Pi \delta \subset \sum_3^{\delta} \Pi \Sigma \Pi \delta \cdots$, 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


[HS80] Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In STOC 1980, pages 262–272, 1980. 6

27


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

$$f = \sum_{\text{path } \gamma, s \rightarrow t} \text{wt} (\gamma)$$

where $\text{wt} (\gamma)$ is product of labels over the edges in path $\gamma$. Number of layers ($\Delta$) defines the depth and the maximum number of vertices in any layer ($\omega$) 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 $s \leq d \omega^2 \Delta$; 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 $(x_{v(1)}, \ldots, x_{v(n)})$ for some permutation $\sigma : [n] \rightarrow [n]$, if edges of $i$-th layer of ABP are univariate polynomials in $x_{v(i)}$.

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 $\Sigma \wedge \Sigma$. Since $\Sigma \wedge \Sigma \subseteq \overline{\text{ARO}}$, using duality trick (Lemma 16) and $\overline{\text{ARO}} = \text{ARO}$, from Nisan' characterization (Lemma 17), it follows that $\Sigma \wedge \Sigma \subset \text{ARO}$. Note that, $\Sigma \wedge \Sigma$ is a strict subset of ARO since $\prod_{i=1}^{n} x_i$ has a small ARO, but it requires $\exp (n)$-size $\Sigma \wedge \Sigma$-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 warping-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 $O((\prod_{i} d_i) \cdot \prod_{i} s_i)$. 

30
Using a simple interpolation, the coefficient of $y^e$ can be extracted from $f(x, y) \in \Sigma \land \Sigma$ again as a small $\Sigma \land \Sigma$ representation.

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

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

**Lemma 13** ($\Sigma \land \Sigma$ differentiation, [DDS21, Lemma A.13]). Let $f(x, y) \in \mathbb{F}[x][y]$ be computed by a $\Sigma \land \Sigma$ circuit of size $s$ and degree $d$. Then, $\partial_y(f)$ is a $\Sigma \land \Sigma$ 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 $(\hat{g}_1/\hat{g}_2) \cdot (h_1/h_2)$, where $g_i \in \mathcal{C}$ and $h_i \in \mathcal{D}$ ($g_2h_2 \neq 0$). 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]). $(\mathcal{C}/\mathcal{C}) \cdot (\mathcal{D}/\mathcal{D}) \subseteq (\overline{\mathcal{C}}/\mathcal{C}) \cdot (\overline{\mathcal{D}}/\mathcal{D})$.

**Proof.** Suppose $(\hat{g}_1/\hat{g}_2) \cdot (h_1/h_2) = f + \epsilon \cdot Q$, where $Q \in \mathbb{F}(x, \epsilon)$ and $f \in \mathbb{F}(x)$. Let $\text{val}_\epsilon(g_1) =: a_i$ and $\text{val}_\epsilon(h_1) =: b_i$. Denote $\hat{g}_i =: e^{a_i} \cdot \hat{g}_i$, similarly $\hat{h}_i$. Further, assume $\hat{g}_i =: \hat{g}_i + \epsilon \cdot \hat{g}_i'$; similarly for $\hat{h}_i$, 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 $= e^{a_1-b_1+1} \cdot (\hat{g}_1/\hat{g}_2) \cdot (\hat{h}_1/\hat{h}_2)$. This has a limit $\lim_{\epsilon \to 0}$, so $a_1 + b_1 - a_2 - b_2 \geq 0$. If it is $\geq 1$, 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 $f = (x_1 + \ldots + x_n)^d$ can be written as
$$f = \sum_{i \in [t]} f_{ij}(x_1) \cdots f_{in}(x_n),$$
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 $\Sigma \land \Sigma \subseteq \text{ARO}$.

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 $\mathbb{F}(e)[x]$. Then, there exists an ARO of size $s$ which exactly computes $f(x)$. 

31