

Demystifying Border Depth-3 Circuits via Differential Equations

Meghana Bhat*

Pranjal Dutta[†]Devansh Shringi[‡]

Abstract

Border complexity captures polynomials that can be approximated arbitrarily well by small algebraic circuits, and *debordering* asks how efficiently such an approximation can be converted into an exact computation. Debordering lies at the heart of the gap between Valiant’s determinant versus permanent conjecture and its strengthening by Mulmuley and Sohoni in Geometric Complexity Theory (GCT), which is phrased in terms of border determinantal complexity. However, efficient debordering results are known only for a few restricted models.

In this paper, we study border depth-3 circuits of top fan-in k , denoted $\overline{\Sigma^{[k]}\Pi\Sigma}$. Dutta, Dwivedi, and Saxena (FOCS 2021) introduced a recursive technique called DiDIL and used it to show that every polynomial in the border of size- s $\Sigma^{[k]}\Pi\Sigma$ circuits has an algebraic branching program (ABP) of size $s^{\exp(k)}$, and that this border class has an explicit hitting set of size $s^{\exp(k) \cdot \log \log s}$. Building on DiDIL, Dutta and Saxena (JACM 2026) proved an exponential-gap top-fan-in hierarchy theorem, via a lower bound of the form $2^{\Omega(d/\exp(k))}$. In all three settings, the bounds depend exponentially on k .

We bypass DiDIL entirely, replacing it with a single linear differential equation obtained from a Wronskian, and improve the dependence on k from exponential to polynomial in all three settings.

1. Every polynomial in the border of depth-3 circuits of size s and top fan-in k has an ABP of size $s^{O(k^2)}$; this is quasipolynomial even for $k = \text{poly}(\log s)$.
2. There is an explicit hitting set for this border class of size $s^{O(k^2)}$.
3. There is an explicit polynomial of degree d that is computed by a depth-3 circuit of size $O(kd)$ and top fan-in $k + 1$, but requires border depth-3 circuits of top fan-in k to have size $2^{\Omega(d/k^2)}$.

Finally, we show that the border of sums of k powers of δ -degree polynomials has depth-6 formulas of size $s^{\text{poly}(k\delta)}$. This resolves a special case of an open question (Open Question 9) of Dutta and Lysikov’s recent survey on debordering.

*Nanyang Technological University, meghana.bhat@ntu.edu.sg

[†]Nanyang Technological University, pranjal.dutta@ntu.edu.sg

[‡]University of Toronto, devansh@cs.toronto.edu

Contents

1	Introduction	2
1.1	Our results	5
2	Detailed proof ideas	8
2.1	Proof idea of Theorem 1.1	11
2.2	Proof idea of Theorem 1.2	12
2.3	Proof idea of Theorem 1.3.	13
2.4	Proof idea of Theorem 1.4	14
3	Preliminaries	17
3.1	Wronskian	17
3.2	Waring Rank	18
3.3	Arithmetic Branching Programs	19
3.4	Border Waring Rank PIT	19
3.5	Cramer's rule	20
4	Debordering Depth 3 circuits	20
4.1	Proof of Theorem 1.1	26
5	PIT for border depth-3	27
5.1	Proof of Theorem 1.3	30
6	Debordering $\overline{\Sigma \wedge \Sigma \Pi}(k, \delta, s)$	31
7	An improved lower bound for border depth-3 circuits	37
7.1	Structural Results for Border Depth-3 circuits	37
7.2	Nonhomogeneous reduction	43
7.3	Proof of Theorem 1.4	45
8	Conclusion	46

1 Introduction

Multivariate polynomials are fundamental objects in both mathematics and theoretical computer science: they are central to algebraic geometry, invariant theory and representation theory, and they play a crucial role in error-correcting codes, pseudorandomness and interactive proofs. Algebraic complexity theory studies the number of arithmetic operations, namely additions and multiplications over an underlying field, required to compute multivariate polynomials. Some of the main goals of the area are to understand the relative power of natural models of algebraic computation, to prove lower bounds for explicit polynomial families in these models, and to design efficient algorithms for computational algebraic problems such as polynomial identity testing and factorization.

Basic algebraic models. The most general model in algebraic complexity is the *algebraic circuit*: a directed acyclic graph whose leaves are labelled by variables or field constants and whose internal gates are labelled by $+$ or $\times$. The *size* of a circuit is the number of its gates and wires, and its *depth* is the length of a longest path from a leaf to the output gate. A *formula* is a circuit whose underlying graph is a tree, so that intermediate computations cannot be reused. An *algebraic branching program*

(ABP) is a layered directed acyclic graph with a source and a sink, whose edges are labelled by affine linear polynomials; it computes the sum, over all source-to-sink paths, of the product of the edge labels along the path. Polynomial-size formulas can be simulated by polynomial-size ABPs, which in turn can be simulated by polynomial-size circuits. Polynomial families of polynomially bounded degree that are computable by polynomial-size formulas, ABPs and circuits form the classes VF, VBP and VP, respectively. We refer the reader to the surveys [SY10, Mah13] for details.

The algebraic P versus NP problem. A central goal of the area is to resolve Valiant’s VP versus VNP problem. The class VNP consists of polynomial families $(f_n)_n$ of polynomially bounded degree and number of variables that can be expressed as

$$f_n(\mathbf{x}) = \sum_{\mathbf{b} \in \{0,1\}^{p(n)}} g_n(\mathbf{x}, \mathbf{b}),$$

where $\mathbf{x}$ denotes the variables $x_1, \dots, x_n$, and $p : \mathbb{N} \rightarrow \mathbb{N}$ is polynomially bounded and $(g_n)_n \in \text{VP}$. Valiant [Val79] conjectured that $\text{VP} \neq \text{VNP}$, as an algebraic analogue of the P versus NP problem. The two problems are in fact related: Bürgisser [Bür00] showed that the nonuniform version of $\text{P} \neq \text{NP}$, namely $\text{P/poly} \neq \text{NP/poly}$ implies $\text{VP} \neq \text{VNP}$ over every finite field and, further assuming the generalized Riemann hypothesis, over every field of characteristic zero. Over any field of characteristic different from two, the permanent family is VNP-complete, so the conjecture is equivalent to the statement that the permanent *cannot* be computed by polynomial-size algebraic circuits [Val79, Mah13].

Border complexity. In this paper, we study these models in a setting where a polynomial is only required to be approximated arbitrarily well, rather than computed exactly. This notion originated in the study of fast matrix multiplication [BCRL79, Bin80], and is formalized as follows. Throughout the introduction, we work over the field $\mathbb{C}$ of complex numbers. For a class $\mathcal{C}$ of circuits, let $\mathcal{C}(s)$ denote the set of polynomials that are computable by circuits in $\mathcal{C}$ of size at most s . Identifying a polynomial with its vector of coefficients, the *border* $\overline{\mathcal{C}(s)}$ is the closure of $\mathcal{C}(s)$, that is, the set of polynomials that are limits of polynomials in $\mathcal{C}(s)$; for the classes considered here, the Euclidean and Zariski closures coincide. A simple but well-known example already shows that taking limits might help:

$$x^{d-1}y = \lim_{\epsilon \rightarrow 0} \frac{(x + \epsilon y)^d - x^d}{d\epsilon},$$

and hence $x^{d-1}y$ is a limit of sums of two d -th powers of linear forms, whereas every exact expression of it as a sum of d -th powers of linear forms has at least d summands (see, e.g., [DL25]).

The border also has an equivalent algebraic description, which is often more convenient to work with: $f \in \overline{\mathcal{C}(s)}$ if and only if there is a polynomial

$$f_\epsilon(\mathbf{x}) = f(\mathbf{x}) + \epsilon \cdot g(\mathbf{x}, \epsilon), \quad g \in \mathbb{C}[\epsilon][\mathbf{x}],$$

that is computable by a circuit in $\mathcal{C}$ of size at most s over the field $\mathbb{C}(\epsilon)$ [Ald84, Bür04, DL25]. We emphasize that the constants in this circuit can be arbitrary rational functions of ϵ , and that their degrees and pole orders are not charged to the size. In particular, while the output has a defined limit as $\epsilon \rightarrow 0$, the intermediate computations need not (see the example above). At first sight, border complexity may appear to be a technical variant of exact complexity. It is, however, intrinsic to the study of algebraic lower bounds, for at least two reasons, which we now discuss.

Geometric complexity theory. Border complexity is central to geometric complexity theory (GCT), an approach to algebraic lower bounds via algebraic geometry and representation theory initiated by Mulmuley and Sohoni [MS01, MS08]. Its starting point is Valiant’s determinant versus permanent conjecture [Val79]: the permanent per_n of an $n \times n$ matrix of variables *cannot* be expressed as the determinant of an $m \times m$ matrix of affine linear polynomials unless m is superpolynomial in n . Since polynomial-size determinantal expressions and polynomial-size ABPs are equivalent in power, this is precisely the conjecture $\text{VBP} \neq \text{VNP}$. Mulmuley and Sohoni proposed a stronger conjecture, which asserts that the permanent is not even a limit of determinantal expressions of polynomial size! In geometric terms, one pads the permanent by a power of a new variable z to match the degree of the determinant, and the conjecture states that, for m polynomially bounded in n ,

$$z^{m-n} \text{per}_n \notin \overline{\text{GL}_{m^2} \cdot \det_m},$$

where the right-hand side is the closure of the orbit of the $m \times m$ determinant under invertible linear maps of variables [BLMW11]. The advantage of this formulation is that the orbit closure is an algebraic variety on which GL_{m^2} acts, so that the noncontainment can be studied through the representation theory of its coordinate ring. The closure is essential here: a lower bound obtained in this way is necessarily a lower bound on border complexity.

Lower bounds via polynomial equations. The relevance of border complexity is not confined to representation-theoretic methods. A common strategy for proving a lower bound against a circuit class is to exhibit a polynomial equation in the coefficients that is satisfied by every polynomial computed by the class, but not by the candidate hard polynomial; this template is formalized by the notion of algebraically natural proofs [FSV18, GKSS17]. A classical example is the *discriminant*: a binary quadratic form $ax^2 + bxy + cy^2$ is the square of a linear form, that is, has Waring rank at most one (i.e. it is a perfect square), if and only if $b^2 - 4ac = 0$. More generally, when the entries of a matrix are polynomial functions of the coefficients, as is the case for the matrix of partial derivatives, an upper bound on its rank is expressed by the vanishing of minors. By continuity, every such equation also holds on the closure of the class. Consequently, a lower bound proved via such equations is automatically a lower bound against the border of the class [Gro15, GMQ16]. Moreover, for classes that are closed under invertible linear changes of variables, after a linear change of variables these equations may be assumed to be a highest weight vectors, which ties this template back to the representation-theoretic approach of GCT [BDI21, vdBDG⁺25]. Understanding which polynomials are added by taking the closure is therefore part of understanding what these methods must separate.

VP versus $\overline{\text{VP}}$. For general circuits, the question is whether taking the closure adds anything at all. Let $\overline{\text{VP}}$ denote the border analogue of VP, defined using border circuit size in place of circuit size. Is $\overline{\text{VP}} = \text{VP}$, that is, can a polynomial-size exact one replace every polynomial-size border computation? This question remains open [GMQ16, DL25]. The geometric goal $\text{VNP} \not\subseteq \overline{\text{VP}}$ is stronger than $\text{VP} \neq \text{VNP}$ [GMQ16], and the two are equivalent if $\overline{\text{VP}} = \text{VP}$. This gap between exact versus approximation is of interest already for restricted models, where the question also has a quantitative form: what blow-up in size is incurred when a border computation is simulated by an exact one, and which lower bounds continue to hold under approximation? For several such models, researchers have exploited structural weaknesses to obtain both hitting sets and lower bounds for the corresponding border classes [BS21, LST21, AF22, DDS21b, DS26].

1.1 Our results

We study border depth-3 circuits of top fan-in k , and improve the dependence on k from exponential ([DDS21b, DS26]) to polynomial in three settings: debordering, hitting sets, and lower bounds. We also deborder sums of powers of bounded-degree polynomials, a restricted class of depth-4 circuits, thereby resolving open questions from [DDS21b, DL25]. In each case, we first discuss the context and the previously known results, and then state our results.

Beyond the quantitative improvements, the main takeaway of this work is a simpler Wronskian-based approach to border depth-3 and depth-4 circuits. Previous results in all three settings rely on the DiDIL technique of [DDS21b], which reduces the top fan-in recursively, one summand at a time; the exponential dependence on k is a consequence of this recursion. Our proofs bypass DiDIL entirely. They handle all k summands in one shot, via a linear differential equation of order at most k obtained from a Wronskian, and are, in our view, considerably simpler.

1.1.1 Debordering

A *debordering* result is an upper bound on the exact complexity of a polynomial in terms of its border complexity; the exact computation is allowed to be in a more general model than the approximating one. In this language, the question of whether $\overline{\text{VP}} = \text{VP}$ asks for a debordering of general circuits with only a polynomial blow-up in size. The best bound known there is exponential: eliminating ϵ by interpolation yields exact circuits of size $2^{\text{poly}(s)}$ from approximating circuits of size s [Bür04, Bür20]. Better bounds are known only for restricted models, and even there the gap between exact and border complexity is nontrivial.

Depth-3 circuits. A depth-3 circuit of top fan-in k computes a sum of k products of affine linear polynomials:

$$f(\mathbf{x}) = \sum_{i=1}^k T_i(\mathbf{x}), \quad T_i(\mathbf{x}) = \prod_{j=1}^{d_i} \ell_{ij}(\mathbf{x}).$$

We denote by $\Sigma\Pi\Sigma(k, s)$ the class of polynomials computed by such circuits of size at most s and top fan-in at most k . In particular, the formal degree is bounded by the size. This degree bound is essential in the border setting: Kumar [Kum20] showed that every homogeneous polynomial can be approximated by depth-3 circuits of top fan-in two, provided that the formal degree is allowed to be sufficiently large. Thus, a bound on the top fan-in alone does not restrict border complexity.

Dutta, Dwivedi, and Saxena [DDS21b] showed that every polynomial in $\overline{\Sigma\Pi\Sigma(k, s)}$ can be computed by an ABP of size $s^{O(k7^k)}$. They introduced a technique, called DiDIL (divide, derive, induct, with limit), that reduces the top fan-in one summand at a time, while controlling the divisions, derivatives, and limits introduced along the way. The resulting ABP has polynomial size for every constant k , but the exponent grows exponentially with k . Our first result improves this dependence to a quadratic one.

Theorem 1.1 (Improved Debordering depth-3 circuits). *Let $k \geq 1$ and let $f \in \overline{\Sigma\Pi\Sigma(k, s)}$. Then f can be computed by an algebraic branching program of size $s^{O(k^2)}$.*

We emphasize that k need not be a constant in Theorem 1.1. For instance, for $k = \log s$ the ABP has size $\exp(O(\log^3 s))$, so polynomial-size depth-3 circuits of logarithmic top fan-in can be debordered into quasipolynomial-size ABPs, whereas the bound of [DDS21b] yields an exponential (thus a trivial) upper bound.

Sums of powers of bounded-degree polynomials. We next consider a depth-4 model, in which a polynomial is expressed as a sum of k powers of polynomials of degree at most δ :

$$\sum_{i=1}^k P_i(\mathbf{x})^{e_i}, \quad \deg P_i \leq \delta.$$

Here, each P_i is given as a sum of monomials, and the size of the expression is $\sum_{i=1}^k (\text{size}(P_i) + e_i)$; in particular, the exponents contribute to the size. We denote by $\Sigma \wedge \Sigma \Pi(k, \delta, s)$ the class of polynomials that have such an expression of size at most s . We also mention that when $\delta = 1$, it is denoted as $\Sigma \wedge \Sigma$.

Dutta and Lysikov [DL25, Open Question 9] asked whether, for constant δ (but unbounded k), the border of this model is contained in VBP since this would imply that $\overline{\Sigma^{[k]} \Pi \Sigma \Pi^{[\delta]}} \subseteq \text{VBP}$ (implicit in [DDS21b] and explicitly mentioned in [DL25]). Our second debordering answers [DL25, Open Question 9] when the top fan-in k is constant.

Theorem 1.2 (Debordering powers of bounded-degree polynomials). *Let $k, \delta \geq 1$ and let $f \in \Sigma \wedge \Sigma \Pi(k, \delta, s)$. Then f can be computed by a depth-6 formula of size $s^{O(k^3 \delta)}$.*

Remark 1. In hindsight, DiDIL [DDS21b] can also be adapted to this model, giving formulas of size $s^{\delta \cdot \exp(k)}$ and depth $O(k)$ (see the Remark in Section 2.2). Theorem 1.2 improves the dependence on k in the exponent from exponential to polynomial, and also gives a constant depth which is independent of k .

Remark 2. The same argument gives ABPs of size $s^{\text{poly}(k)}$ for $\overline{\Sigma^k \wedge \Sigma \wedge \Sigma}$, and hence containment in VBP for constant k . We omit the details, as they follow from similar calculations; see Section 2.2.

1.1.2 Polynomial identity testing

Polynomial identity testing (PIT) is the problem of deciding whether a given algebraic circuit computes the identically zero polynomial. A nonzero polynomial of degree d vanishes on at most a $d/|S|$ fraction of the points of any finite grid S^n . This classical fact, often referred to as the Ore–DeMillo–Lipton–Schwartz–Zippel lemma or the Polynomial Identity Lemma [Ore22, DL78, Sch80, Zip79], immediately gives an efficient randomized algorithm for PIT, and obtaining an efficient deterministic algorithm is a central open problem in algebraic complexity. In the *whitebox* setting, the algorithm is allowed to inspect the circuit, whereas in the *blackbox* setting it only has query access to the polynomial. A deterministic blackbox algorithm for a class $\mathcal{C}$ amounts to an explicit *hitting set* for $\mathcal{C}$: a finite set $\mathcal{H} \subseteq \mathbb{C}^n$ such that every nonzero $f \in \mathcal{C}$ is nonzero at some point of $\mathcal{H}$. A further motivation for constructing such sets is the close connection between PIT and circuit lower bounds [KI03]. For an overview on PIT and its connections to other areas in theoretical computer science, we refer the reader to [SY10, Sax14, DG24].

Depth-3 circuits of bounded top fan-in have been a central test case for PIT [DS07, KS07, SS13]. In the border setting, one asks for a hitting set for the closure of the class, that is, a set that also hits every nonzero polynomial arising as a limit. A hitting set for the exact class *need not* have this property: the evaluations of an approximating polynomial at the points of the set may all be nonzero, and yet vanish in the limit. Dutta, Dwivedi, and Saxena [DDS21b] constructed a hitting set for $\overline{\Sigma \Pi \Sigma(k, s)}$ of size $s^{O(k 7^k \log \log s)}$, within the same time bound. Our next result significantly improves the dependence on the top fan-in, thus resolving [DDS21b, Question 3].

Theorem 1.3 (Improved Border depth-3 hitting set). *There is an explicit hitting set for $\overline{\Sigma\Pi\Sigma}(k, s)$ of size $s^{O(k^2)}$, constructible within the same time bound.*

In a very recent breakthrough by Bhardwaj and Saptharishi [BS26], a polynomial-size explicit hitting set was designed for all polynomials whose space of partial derivatives has small dimension. $\overline{\Sigma \wedge \Sigma}$ circuits fall into this class, implying a polynomial-size hitting set for the same¹. Before [BS26], only a quasipolynomial-size hitting set was known [FS13, AGKS15, GG20]. In particular, Theorem 1.3 gives a quasipolynomial size hitting set even when $k = \text{poly}(\log s)$, unlike [DDS21b] which would yield an exponential-size hitting set.

We remark that Theorem 1.3 does not follow from the ABP simulation of Theorem 1.1 alone, since efficient hitting sets for general ABPs are not known.

1.1.3 Lower bounds and a top-fan-in hierarchy

Our final result is a lower bound result. The debordering and hitting-set results above show that border depth-3 circuits of small top fan-in are a *weak* model; we now ask how their power grows with the top fan-in, which is a question of hierarchy type. Hierarchy theorems address a basic question: does more of a computational resource give strictly more computational power? The classical time hierarchy theorem [HS65] answers this question for time complexity. In algebraic complexity, Limaye, Srinivasan, and Tavenas [LST25] proved a depth hierarchy theorem: over fields of characteristic zero, for every fixed $\Delta \geq 2$ there are explicit polynomials that have depth- Δ circuits of size s , but require size $s^{\omega(1)}$ at depth $\Delta - 1$.

Thus, a finer question is whether such a hierarchy exists within a fixed depth. For depth-3 circuits, the natural resource is the top fan-in: does increasing the number of summands from k to $k + 1$ add power? In the exact setting, the answer is yes for a simple reason: it is folklore that the inner product polynomial $\sum_{i=1}^{k+1} x_i y_i$ cannot be computed by a depth-3 circuit of top fan-in k , regardless of its size (see [Kum20, DS26]). In the border setting, this impossibility breaks down: as mentioned above, circuits of top fan-in two can already approximate every homogeneous polynomial, at the cost of a large formal degree [Kum20]. A separation in the border setting must therefore be a lower bound on size. Dutta and Saxena [DS26] proved such a separation: for every constant k , there is an exponential gap between circuits of top fan-in $k + 1$ and border circuits of top fan-in k . Their explicit family is the sum of $k + 1$ monomials on disjoint sets of variables,

$$P_{k+1,d}(\mathbf{x}) := \sum_{i=1}^{k+1} \prod_{j=1}^d x_{i,j}.$$

This polynomial has degree d and $N = (k + 1)d$ variables, and is computed by a depth-3 circuit of size $O(kd)$ and top fan-in $k + 1$.

The lower bound of [DS26] yields a bound of the form $2^{\Omega(d/\exp(k))}$ on the size of border circuits of top fan-in k . This is exponential in d for constant k , but the exponential loss in k limits the separation when the top fan-in grows; in particular, it does not give a superpolynomial separation throughout the regime $k = \Theta(\log d)$. We improve this dependence, for the same explicit family.

Theorem 1.4 (Improved border top-fan-in hierarchy). *Let $d > k \geq 1$ be integers. The polynomial $P_{k+1,d}$ has a depth-3 circuit of size $O(kd)$ and top fan-in $k + 1$. If*

$$P_{k+1,d} \in \overline{\Sigma\Pi\Sigma}(k, s), \quad \text{then } s \geq 2^{\Omega(d/k^2)}.$$

¹applying [BS26] in [DDS21b] would yield an $s^{O(k7^k)}$ -size explicit hitting set

We stress that the approximating circuits in [Theorem 1.4](#) are not required to be homogeneous, and that their formal degree may exceed d ; it is only charged to their size. Thus, the theorem is a lower bound on size, and not merely an obstruction to approximating $P_{k+1,d}$ by k products of degree d .

For $k = \Theta(\log d)$, the lower bound is $2^{\Omega(d/(\log d)^2)}$, which is $2^{\Omega(N/(\log N)^3)}$ in terms of the number of variables (denoted as N), whereas the circuit of top fan-in $k + 1$ has size $O(N)$. More generally, the separation is superpolynomial whenever $d/k^2 = \omega(\log(kd))$. [Theorem 1.4](#) thus extends the exponential-gap hierarchy from constant top fan-in to growing top fan-in. Together with our upper bounds, it replaces the exponential dependence on k by a polynomial one in both directions: in simulating border circuits of small top fan-in by exact computations, and in showing that reducing the top fan-in requires a large blow-up in size.

2 Detailed proof ideas

Dutta, Dwivedi, and Saxena [[DDS21b](#)] introduced DiDIL (divide, derive, induct, with limit), which reduces the top fan-in one summand at a time through divisions, derivatives, and limits. Each step incurs a polynomial blow-up in size; repeating these steps leads to an exponential dependence on the top fan-in in the exponent of the final size bound. This technique gave efficient debordering and PIT results for border depth-3 circuits of bounded top fan-in, and was subsequently used to prove an exponential-gap top-fan-in hierarchy theorem [[DS26](#)]. Our approach captures the elimination underlying DiDIL in a *single differential equation*, avoiding the recursive reduction in top fan-in.

In a nutshell, we prove that every polynomial in the border of sums of k products of affine linear polynomials satisfies, after a change of variables, a linear differential equation of order at most k whose coefficients are simpler than the products themselves. The same holds when each summand is a power of a low-degree polynomial. We use the structure of these coefficients and the associated Wronskians to obtain our debordering, PIT, and lower-bound results. We explain these arguments primarily for $k = 2$. Throughout this overview, we work over $\mathbb{C}$.

A quick overview on DiDIL. Suppose,

$$f_\epsilon(\mathbf{x}) = T_1(\epsilon, \mathbf{x}) + T_2(\epsilon, \mathbf{x}) = f(\mathbf{x}) + O(\epsilon),$$

where each T_i is a product of at most d affine linear polynomials over $\mathbb{C}(\epsilon)$, and the total circuit size over $\mathbb{C}(\epsilon)$ is at most s . Here $O(\epsilon)$ denotes an error of the form $\epsilon g(\mathbf{x}, \epsilon)$ with $g \in \mathbb{C}[\mathbf{x}, \epsilon]$. Although their sum has a limit, the individual summands T_i need not. We focus on the case where f, T_1, T_2 are all nonzero.

Introduce a tuple $\mathbf{y}$ of shift variables and a new variable z , and set

$$\begin{aligned} \Phi_{\mathbf{y}}(g)(z, \mathbf{x}) &:= g(\mathbf{y} + z\mathbf{x}), & D &:= \partial_z, \\ S_i &:= \Phi_{\mathbf{y}}(T_i), & F_\epsilon &:= \Phi_{\mathbf{y}}(f_\epsilon), & F &:= \Phi_{\mathbf{y}}(f). \end{aligned}$$

Then $F_\epsilon = S_1 + S_2$ and $F = F_\epsilon|_{\epsilon=0}$. We recover f from F by setting $z = 1$ and replacing $\mathbf{x}$ by $\mathbf{x} - \mathbf{y}$. The variable z behaves like a *degree-counter*: the coefficient of z^j in F is homogeneous of degree j in $\mathbf{x}$. For debordering, it suffices to eventually set $\mathbf{y}$ to a random point $\alpha \in \mathbb{C}^n$. For PIT, however, such a point has to be found deterministically, so there we keep track of how the intrinsic computation depends on $\mathbf{y}$.

DiDIL [DDS21b] eliminates a summand by dividing and differentiating:

$$D\left(\frac{F_\epsilon}{S_2}\right) = D\left(\frac{S_1}{S_2}\right) = \frac{S_1}{S_2} \left(\frac{DS_1}{S_1} - \frac{DS_2}{S_2} \right). \quad (1)$$

Division turns the second summand into 1, which differentiation removes. Thus only one summand remains, although it now belongs to a more general model. The resulting expression involves logarithmic derivatives². These turn products into sums: if the shifted factors of S_i are denoted by $\ell_{i,j}$, then

$$\frac{DS_i}{S_i} = \sum_j \frac{D\ell_{i,j}}{\ell_{i,j}}. \quad (2)$$

Each shifted factor has the form $c + zL(\mathbf{x})$, where $c \in \mathbb{C}(\epsilon, \mathbf{y})$ is nonzero and L is homogeneous linear in $\mathbf{x}$; indeed, an affine factor ℓ becomes $\ell(\mathbf{y} + z\mathbf{x}) = \ell(\mathbf{y}) + zL(\mathbf{x})$, where L is the homogeneous linear part of ℓ . This is the reason for the shift: since the constant term $c = \ell(\mathbf{y})$ is nonzero, $1/(c + zL)$ is a power series in z . In particular, the logarithmic derivative has the expansion

$$\frac{L}{c + zL} \equiv \sum_{j=0}^d \frac{(-1)^j z^j L^{j+1}}{c^{j+1}} \pmod{z^{d+1}}. \quad (3)$$

Thus $D\ell_{i,j}/\ell_{i,j}$, and hence DS_i/S_i , lies in the power series ring $\mathbb{C}(\epsilon, \mathbf{y})[[\mathbf{x}][[z]]$, so truncation modulo z^{d+1} is well defined. Up to a scalar in $\mathbb{C}(\epsilon, \mathbf{y})$, each term in (3) is a product of powers of just two linear forms, z and L , and hence can be expressed as a small sum of powers of linear forms by Lemma 3.5. Consequently, the truncation of $DS_1/S_1 - DS_2/S_2$ modulo z^{d+1} has a polynomial-size $\Sigma \wedge \Sigma$ expression over $\mathbb{C}(\epsilon, \mathbf{y})$.

Now, multiply each S_i by the power of ϵ for which it has a nonzero limit, and denote this limit by s_i . Scaling (1) accordingly and taking the limit gives

$$D\left(\frac{F}{s_2}\right) \equiv \frac{s_1}{s_2} Q \pmod{z^{d+1}},$$

where Q is the limit of an appropriately rescaled truncation of the logarithmic-derivative difference. Thus Q has a polynomial-size $\overline{\Sigma} \wedge \overline{\Sigma}$ representation. Each s_i is a product of linear factors and therefore has a small ABP. Since small $\Sigma \wedge \Sigma$ circuits admit polynomial-size ABPs, the right-hand side becomes a quotient of polynomial-size ABPs. Finally, DiDIL reverses the differentiation by integrating and recovers F and therefore f .

For general $k \geq 3$, DiDIL repeats the above step $k - 1$ times, each time dividing by one summand and differentiating, until a single summand remains; f is then recovered by undoing these rounds one by one, taking limits and integrating. Each round incurs a polynomial blow-up, yielding the size to be $s^{\exp(k)}$.

From DiDIL to differential equations. Another way to view the DiDIL step is as finding a differential equation satisfied by both summands. Consider the matrix³

$$M := \begin{pmatrix} S_1 & DS_1 & D^2S_1 \\ S_2 & DS_2 & D^2S_2 \end{pmatrix}.$$

²The logarithmic derivative of a function g with respect to z is $(\partial_z g)/g$, defined wherever $g \neq 0$.

³The minor of M formed by its first two columns is $S_1 DS_2 - S_2 DS_1 = S_1 S_2 (DS_2/S_2 - DS_1/S_1)$; compare with (1).

Pick a nonzero vector $\mathbf{c} := (c_0, c_1, c_2)^\top$ in its kernel, where $c_i \in \mathbb{C}(\mathbf{x}, \mathbf{y}, z, \epsilon)$; such a vector exists by the rank-nullity theorem, since $\text{rk}(M) \leq 2$. That is, $c_0 S_i + c_1 D S_i + c_2 D^2 S_i = 0$ for $i = 1, 2$. How does this help? Adding these two equations gives

$$c_0 F_\epsilon + c_1 D(F_\epsilon) + c_2 D^2(F_\epsilon) = 0.$$

Note that $\lim_{\epsilon \rightarrow 0} F_\epsilon = F$, and further, $\lim_{\epsilon \rightarrow 0} D(F_\epsilon) = D(F)$ and $\lim_{\epsilon \rightarrow 0} D^2(F_\epsilon) = D^2 F$. Multiplying both sides by ϵ^{-m} , where m is the least order in ϵ among c_0, c_1, c_2 , and taking the limit gives

$$b_0 F + b_1 D(F) + b_2 D^2(F) = 0, \quad (\text{main differential equation}) \quad (4)$$

where the b_i are ϵ -free and not all zero (although some of them may be zero). A priori, the b_i are merely elements of $\mathbb{C}(\mathbf{x}, \mathbf{y}, z)$. The *hope* is that they have “small” complexity; if so, (4) may become useful for debordering, PIT, and lower bounds. A second crucial point is that F should be *uniquely* recoverable from (4). We first bound the complexity of the b_i , and return to uniqueness in the proof idea of [Theorem 1.1](#).

The first two columns already recover the expression from DiDIL. Their determinant is the Wronskian ([Definition 3.1](#))

$$W(S_1, S_2) := S_1 D S_2 - S_2 D S_1, \quad D\left(\frac{S_1}{S_2}\right) = -\frac{W(S_1, S_2)}{S_2^2}.$$

The third column lets us turn this elimination step into an equation for F_ϵ itself. To bound the complexity of the b_i , we factor S_i out of row i . This does not change the kernel, and gives

$$\widetilde{M} := \begin{pmatrix} 1 & a_1 & r_1 \\ 1 & a_2 & r_2 \end{pmatrix}, \quad a_i := \frac{D S_i}{S_i}, \quad r_i := \frac{D^2 S_i}{S_i}.$$

By (2) and (3), the truncation of a_i modulo z^{d+1} has a small $\Sigma \wedge \Sigma$ expression over $\mathbb{C}(\epsilon, \mathbf{y})$. For r_i , note that every shifted factor $\ell_{i,j}$ is affine in z , so $D^2 \ell_{i,j} = 0$, and the product rule gives

$$r_i = \sum_{j \neq j'} \frac{D \ell_{i,j}}{\ell_{i,j}} \cdot \frac{D \ell_{i,j'}}{\ell_{i,j'}}.$$

Expanding both factors by (3), each term becomes a product of powers of three linear forms (z and the linear parts of $\ell_{i,j}$ and $\ell_{i,j'}$). Hence, by [Lemma 3.5](#), the truncation of r_i modulo z^{d+1} also has a small $\Sigma \wedge \Sigma$ expression.

We now take the signed two-by-two minors of $\widetilde{M}$ as our kernel vector. The resulting equation is

$$\underbrace{(a_1 r_2 - a_2 r_1)}_{c_0} F_\epsilon + \underbrace{(r_1 - r_2)}_{c_1} D F_\epsilon + \underbrace{(a_2 - a_1)}_{c_2} D^2 F_\epsilon = 0. \quad (5)$$

Equivalently, the left-hand side of (5) equals $W(S_1, S_2, F_\epsilon)/(S_1 S_2)$, and this Wronskian vanishes because $F_\epsilon = S_1 + S_2$. If $a_1 \neq a_2$, this is a second-order equation. If $a_1 = a_2$, that is, if $D(S_1/S_2) = 0$ and hence S_1, S_2 are linearly dependent over $\mathbb{C}(\epsilon, \mathbf{y}, \mathbf{x})$, both summands satisfy $D S_i - a_1 S_i = 0$, and we use the first-order equation $D F_\epsilon - a_1 F_\epsilon = 0$ instead.

Each c_j is obtained from the a_i and r_i with at most one multiplication, and the product of two products of powers of a few linear forms is again such a product. Hence the truncation of each c_j modulo z^{d+1} has a small $\Sigma \wedge \Sigma$ expression over $\mathbb{C}(\epsilon, \mathbf{y})$, and the truncation of their scaled limits b_j modulo z^{d+1} has a small $\overline{\Sigma} \wedge \overline{\Sigma}$ representation over $\mathbb{C}(\mathbf{y})$. This is the desired low-complexity property of the b_j , as claimed in (4).

2.1 Proof idea of Theorem 1.1

Suppose that (4) has order two, that is, $b_2 \neq 0$ (otherwise, the same argument applies to the largest j with $b_j \neq 0$), and write $F = \sum_{i=0}^d F_i(\mathbf{y}, \mathbf{x})z^i$. The coefficient of z^0 determines F_2 from F_0, F_1 ; the coefficient of z^1 then determines F_3 , and so on. More precisely, the equation at z^j has the form

$$(j+2)(j+1)b_2(\mathbf{y}, 0, \mathbf{x})F_{j+2} + \text{a linear combination of } F_0, \dots, F_{j+1} = 0.$$

For this to determine F_{j+2} , the leading coefficient $b_2(\mathbf{y}, 0, \mathbf{x})$ must be nonzero; a priori, b_2 could be divisible by z . Here the shift helps once more. We argue it formally. Since, $D^a S_i$ is the a -th derivative of T_i in direction $\mathbf{x}$, evaluated at $\mathbf{y} + z\mathbf{x}$, every entry of $\widetilde{M}$ depends on $\mathbf{y}$ and z only through $\mathbf{y} + z\mathbf{x}$. Therefore so do the minors and their normalized limits, that is, $b_j(\mathbf{y}, z, \mathbf{x}) \equiv b_j(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}) \pmod{z^{d+1}}$. In particular, $b_2 \neq 0$ implies $b_2(\mathbf{y}, 0, \mathbf{x}) \neq 0$.

Thus F is determined by F_0, F_1 , and the coefficients b_0, b_1, b_2 . Moreover, $F_0 = f(\mathbf{y})$ and F_1 have degree at most one in $\mathbf{x}$, so they are trivially simple.

To compute the remaining coefficients *without* explicitly unrolling a recursion of length d , we view their equations as a *single linear system* in the unknowns $F_2, \dots, F_d$. This system is lower triangular, and its diagonal entries $(j+2)(j+1)b_2(\mathbf{y}, 0, \mathbf{x})$ are nonzero; hence it has a *unique* solution.

Cramer's rule expresses this solution as ratios of determinants whose entries are small ABPs, since $\overline{\Sigma \wedge \Sigma}$ circuits have small ABPs. Determinant computation and division elimination then give ABPs of size $\text{poly}(s)$. Finally, setting $\mathbf{y}$ to a point α at which all denominators remain nonzero, setting $z = 1$, and undoing the shift gives an ABP for f .

The strategy for general k is the same. Choose a *maximal* subset of summands that is linearly independent over $\mathbb{C}(\epsilon, \mathbf{y}, \mathbf{x})$, say $S_1, \dots, S_t$ with $t \leq k$, and form the analogous $t \times (t+1)$ matrix M whose (i, j) -th entry is $D^{j-1}S_i$. Since $S_1, \dots, S_t$ are linearly independent, the Wronskian criterion (Lemma 3.2) ensures that the minor obtained by deleting the last column, namely the Wronskian $W(S_1, \dots, S_t)$, is nonzero. Hence the maximal minors of M give a nonzero equation of order t satisfied by $S_1, \dots, S_t$. Every other summand is a linear combination of $S_1, \dots, S_t$ with coefficients that *do not* depend on z , so it satisfies the same equation, and hence so does F_ϵ . Taking limits as before, we get

$$\sum_{j=0}^r b_j(\mathbf{y}, z, \mathbf{x})D^j F \equiv 0 \pmod{z^{d+1}}, \quad b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0, \quad (6)$$

where $r \leq t \leq k$ is the largest index with $b_r \neq 0$; as above, $b_r \neq 0$ implies $b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0$. As in the case $k = 2$, the truncation of each b_j modulo z^{d+1} has a small $\overline{\Sigma \wedge \Sigma}$ representation, now of top fan-in $(kd)^{O(k^2)}$. The exponent k^2 can be explained informally as follows. A term of a $t \times t$ minor is a product of t entries $D^a S_i / S_i$ taken from distinct columns, and each $D^a S_i / S_i$ is a sum of products of a logarithmic derivatives of factors of S_i . Hence each term involves at most $1 + 2 + \dots + t = O(k^2)$ logarithmic derivatives, and after the expansion (3) it becomes a product of powers of $O(k^2)$ linear forms, which has a $\Sigma \wedge \Sigma$ expression of top fan-in $(kd)^{O(k^2)}$ by Lemma 3.5.

An important distinction from [DDS21b] in terms of complexity can be explained in a nutshell here. The single determinant explained above replaces the $k - 1$ nested rounds of DiDIL, which is why the exponent now depends polynomially, rather than exponentially on k .

The initial coefficients $F_0, \dots, F_{r-1}$ have small $\overline{\Sigma \wedge \Sigma}$ circuits as well. As before, the remaining F_j form the unique solution of a lower-triangular linear system whose diagonal entries are nonzero

multiples of $b_r(\mathbf{y}, 0, \mathbf{x})$, and Cramer's rule with division elimination gives an ABP of size $s^{O(k^2)}$. For a more precise statement regarding (6) and its proof, we refer to [Proposition 4.2](#).

2.2 Proof idea of [Theorem 1.2](#)

Now, instead of each T_i being a product of linear polynomials, let $T_i = P_i^{e_i}$ with $\deg P_i \leq \delta$. The argument above does not apply directly: P_i need not factor into affine linear polynomials, and [Theorem 1.2](#) asks for a formula of *constant depth*, whereas the reconstruction above computes large determinants. The key observation is that, for powers, the differential equation becomes *exact*, with coefficients of low degree. Put $G_i := \Phi_{\mathbf{y}}(P_i)$, so that $S_i = G_i^{e_i}$. Each entry $D^a S_i / S_i$ in row i of $\widetilde{M}$ is a rational function with denominator G_i^a , in which the possibly huge exponent e_i appears only through scalar coefficients; for instance,

$$\frac{DS_i}{S_i} = e_i \frac{DG_i}{G_i}, \quad \frac{D^2 S_i}{S_i} = e_i(e_i - 1) \frac{(DG_i)^2}{G_i^2} + e_i \frac{D^2 G_i}{G_i}.$$

Multiplying row i by G_i^k therefore clears all denominators, and since G_i has degree at most 2δ , the resulting minors have degree $O(k^2\delta)$. Taking limits as before, [Proposition 6.3](#) gives

$$\sum_{j=0}^r b_j D^j F = 0, \quad b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0, \quad \deg b_j \leq 2k^2\delta.$$

(In the depth-3 case, clearing denominators would cost degree about d , which is why we used truncated power series there.)

As in the proof idea of [Theorem 1.1](#), comparing the coefficients of powers of z gives a lower-triangular linear system for the Taylor coefficients F_j of F . The difference is that this system is now described by only $O(k^3\delta)$ polynomials, namely the z -coefficients of the b_j and $F_0, \dots, F_{r-1}$, which we can think of as *parameters* and collect in a tuple $\mathbf{b}$. Crucially, the number of parameters and their degrees in $\mathbf{x}$, which are $O(k^2\delta)$, *do not* depend on the exponents e_i , and hence not on s . Fix $\mathbf{y}$ to a random point α and put $\beta := b_r(\alpha, 0, \mathbf{x})$. Cramer's rule then gives, for every j ,

$$\beta^N \cdot F_j(\alpha, \mathbf{x}) = Q_j(\mathbf{b}(\alpha, \mathbf{x})),$$

where $N \leq \delta s$, and Q_j is a fixed polynomial in $O(k^3\delta)$ variables, of degree $\text{poly}(s)$, that does not depend on f (see [Proposition 6.3](#)). In other words: $\beta^N F_j$ is obtained by evaluating Q_j at the parameters.

Since there are only $p = O(k^3\delta)$ parameters, a polynomial of degree $\text{poly}(s)$ in them has at most $s^{O(p)}$ monomials, so we can afford to expand everything into monomials. As $f(\alpha + \mathbf{x}) = \sum_j F_j(\alpha, \mathbf{x})$, this gives a $\Sigma\Pi\Sigma\Pi$ formula of size $s^{O(k^3\delta)}$ for $\beta^N \cdot f(\alpha + \mathbf{x})$: a sum of monomials in the parameters, with each parameter written as a sum of monomials in $\mathbf{x}$. To divide by β^N , we shift $\mathbf{x}$ so that $\beta(\mathbf{0}) = c \neq 0$, write $\beta = c(1 - g)$, and use $\beta^{-N} = c^{-N} \sum_{j \geq 0} \binom{N+j-1}{j} g^j$; since $g(\mathbf{0}) = 0$, the terms with $j > \deg f$ do not affect the monomials of degree at most $\deg f$. Multiplying the two resulting $\Sigma\Pi\Sigma\Pi$ formulas, extracting the part of degree at most $\deg f$ by interpolation, and undoing the shifts gives a depth-six formula of size $s^{O(k^3\delta)}$ for f .

Remark Debordering $\overline{\Sigma^{[k]} \wedge \Sigma\Pi^{[\delta]}}$ DiDIL [[DDS21b](#)] can also be adapted to our model of sums of powers. The idea is to keep the logarithmic derivatives as rational functions during the elimination, instead of expanding them into power series. In each round, the degrees of their numerators and

denominators grow by at most a constant factor. Hence these rational functions have degree $\delta \cdot \exp(k)$ throughout, and can be written as sums of $s^{\delta \cdot \exp(k)}$ monomials. Reversing the rounds then gives formulas of size $s^{\delta \cdot \exp(k)}$. Each reversed round, an integration followed by a multiplication, adds $O(1)$ to the depth, so the final depth is $O(k)$. Thus the recursion causes both the exponential dependence on k and the growing depth. [Theorem 1.2](#) avoids the recursion and gives formulas of size $s^{O(k^3 \delta)}$ and depth six. Here, we omit the details on how to adapt DiDIL to get such a debordering result.

Remark (Debordering $\overline{\Sigma^k \wedge \Sigma \wedge \Sigma}$). The same comparison holds for $\overline{\Sigma^k \wedge \Sigma \wedge \Sigma}$: adapting DiDIL [[DDS21b](#)] gives ABPs of size $s^{\exp(k)}$; this follows by a similar modified argument as mentioned above. Whereas our Wronskian argument gives size $s^{\text{poly}(k)}$. After a generic shift, each z -coefficient of the Wronskian cofactor is a sum of $s^{O(k^2)}$ products of $O(k^2)$ powers of linear forms. One can show that the dimension of the derivative-space is at most $s^{O(k^2)}$, and this bound also holds after taking limit. Therefore, the resulting coefficients admit ABPs of this size; Cramer's rule and division elimination then give an ABP of size $s^{O(k^2)}$ for f .

2.3 Proof idea of [Theorem 1.3](#).

[Theorem 1.1](#) does not directly give a hitting set, since no efficient hitting sets are known for general ABPs. Instead, we use the differential equation (6) itself. Call a point $\alpha \in \mathbb{C}^n$ *good* if every $b_j(\alpha, z, \mathbf{x})$ is defined and $b_r(\alpha, 0, \mathbf{x}) \neq 0$. We show that at a good point, a nonzero f is already visible in its Taylor expansion up to degree $k - 1$, and that a good point can be found deterministically.

Nonzero solutions cannot vanish to high order. Let α be a good point, and write $F_\alpha := f(\alpha + z\mathbf{x}) = \sum_{i=0}^d F_{\alpha,i}(\mathbf{x})z^i$, so that $F_{\alpha,i}$ is the degree- i part of the Taylor expansion of f around α . By (6) at $\mathbf{y} = \alpha$, F_α satisfies a linear differential equation in z whose leading coefficient does not vanish at $z = 0$. Classically, a solution of such an equation that vanishes to order r at $z = 0$ is identically zero. The algebraic analogue is that if $f \neq 0$, then $F_{\alpha,i} \neq 0$ for some $i < r \leq k$; see [Lemma 5.1](#). Indeed, suppose instead that the lowest nonzero coefficient is $F_{\alpha,m}$ with $m \geq r$. Every term $b_j D^j F_\alpha$ with $j < r$ involves only the powers z^{m-j} and higher, so the coefficient of z^{m-r} in the equation is $\frac{m!}{(m-r)!} b_r(\alpha, 0, \mathbf{x}) F_{\alpha,m} \neq 0$. Since $m - r \leq d$, this coefficient survives the truncation modulo z^{d+1} , which contradicts the equation.

Thus, once a good point α is known, it suffices to hit a polynomial of low degree. Fix $i < k$ with $F_{\alpha,i} \neq 0$. Since $F_{\alpha,i}$ has degree less than k , it does not vanish at some point $\mathbf{v}$ of an explicit hitting set of size $(nk)^{O(k)}$. Then $f(\alpha + z\mathbf{v}) = \sum_i F_{\alpha,i}(\mathbf{v})z^i$ is a nonzero univariate polynomial of degree at most d , and hence it is nonzero at one of any $d + 1$ distinct values of z . It therefore remains to construct a small explicit set that contains a good point for every nonzero f in the class.

Finding a good point. The low complexity of b_r over $\mathbb{C}(\mathbf{y})$ does not help here, since it says nothing about how b_r depends on $\mathbf{y}$. Instead, we use the following structure from [Proposition 4.2](#):

$$b_r(\mathbf{y}, 0, \mathbf{x}) = \frac{p_r(\mathbf{y}, \mathbf{x})}{H(\mathbf{y})}, \quad \deg_{\mathbf{y}} p_r \leq kd,$$

where H is a nonzero product of at most kd affine linear polynomials. (At $z = 0$, the logarithmic derivative of a shifted factor $\ell(\mathbf{y} + z\mathbf{x})$ is $L(\mathbf{x})/\ell(\mathbf{y})$, so all denominators are products of the $\ell(\mathbf{y})$, suitably normalized in ϵ .) Thus α is good whenever $H(\alpha) \neq 0$ and $p_r(\alpha, \mathbf{x}) \neq 0$. A point $\mathbf{a}$ with $H(\mathbf{a}) \neq 0$ is easy to find, using a hitting set for products of affine linear polynomials. To make p_r nonzero as well, we move along a line $\mathbf{y} = \mathbf{a} + t\mathbf{u}$ whose direction $\mathbf{u}$ is yet to be chosen, and expand $b_r(\mathbf{a} + t\mathbf{u}, 0, \mathbf{x})$ as a power series in t . At this point, we need observations:

- some coefficient of t^v with $v \leq kd$ is nonzero, as otherwise the nonzero polynomial $p_r(\mathbf{a} + t\mathbf{u}, \mathbf{x})$, of degree at most kd in t , would be divisible by t^{kd+1} ;
- each such coefficient is a simple polynomial in $(\mathbf{u}, \mathbf{x})$: every denominator $\ell(\mathbf{a}) + tL(\mathbf{u})$ has a nonzero constant term, so expanding its reciprocal as in (3) shows, as for the b_j , that each coefficient lies in $\overline{\Sigma \wedge \Sigma}$, with top fan-in $(kd)^{O(k^2)}$ and degree $O(kd + k^2)$.

It remains to hit $\overline{\Sigma \wedge \Sigma}$ circuits, and for this we use a very recent breakthrough of Bhardwaj and Saptharishi [BS26]: explicit hitting sets of *polynomial* size for all polynomials whose space of partial derivatives has small dimension. $\overline{\Sigma \wedge \Sigma}$ circuits fall into this class, since a power L^e of a linear polynomial has at most $e + 1$ linearly independent partial derivatives, and this bound adds up under sums and can only decrease under limits. Applying their hitting set (Lemma 3.10) to the variables $(\mathbf{u}, \mathbf{x})$ yields a point $(\mathbf{u}, \boldsymbol{\xi})$ at which some coefficient from the first observation survives. Then $p_r(\mathbf{a} + t\mathbf{u}, \boldsymbol{\xi})$ and $H(\mathbf{a} + t\mathbf{u})$ are nonzero univariate polynomials of degree at most kd , so some τ among any $2kd + 1$ values makes both nonzero, and $\mathbf{a} + \tau\mathbf{u}$ is good. Ranging over all choices of $\mathbf{a}$, $(\mathbf{u}, \boldsymbol{\xi})$, and τ gives $s^{O(k^2)}$ candidate points (Lemma 5.2), and hence a hitting set of size $s^{O(k^2)}$. Since the hitting set of [BS26] has size polynomial in the top fan-in, which here is $(kd)^{O(k^2)}$, no $\log \log s$ factor appears in our exponent, unlike in the hitting set of [DDS21b].

2.4 Proof idea of Theorem 1.4

The lower bound uses the same Wronskians, but in a different way, and under one new constraint: we may not shift the variables. Our argument needs the entire hard polynomial to sit in the lowest power of z , and a shift would spread it over many powers of z ; this was crucial in [DS26]. Without a shift, however, the differential equation (4) itself carries no information about the hard polynomial. Indeed, under $\mathbf{x} \mapsto z\mathbf{x}$ a homogeneous polynomial f of degree d becomes $z^d f$, and every derivative $D^j(z^d f) = \frac{d!}{(d-j)!} z^{d-j} f$ is a multiple of f . Hence f factors out of $\sum_j b_j D^j(z^d f)$, and the equation only constrains the coefficients b_j . Instead of annihilating every summand, as in debordering, we will annihilate only the summands that *do not* carry f . We follow the template of Dutta and Saxena [DS26], which we recall first.

The approach of [DS26] Recall that $P_{k+1,d} = \sum_{i=1}^{k+1} \prod_{j=1}^d x_{i,j}$, and suppose that $\sum_{i=1}^k T_i = P_{k+1,d} + O(\epsilon)$, where each T_i is a product of affine linear polynomials over $\mathbb{C}(\epsilon)$. After scaling each affine factor ℓ by a power of ϵ , its ϵ -free part $\ell|_{\epsilon=0}$ is a nonzero affine linear polynomial of one of two kinds: either it is homogeneous, as for $\ell = x_{1,1} - x_{2,1} + \epsilon x_{1,2}$, or it has a nonzero constant term, as for $\ell = 1 + x_{1,1} + \epsilon x_{2,1}$. [DS26] first reduce to the case in which **every** factor is of the second kind, which they call *all-non-homogeneous*. The non-homogeneity seems to be at the heart of the power of border depth-3 circuits; for example, Kumar’s universal construction that *every* f admits a $\overline{\Sigma\Pi\Sigma}$ circuit [Kum20] is all-non-homogeneous.

Factors of the first kind are *killed* [DS26, Claims 3–4]. In the example above, the substitution $x_{1,1} \mapsto x_{2,1} - \epsilon x_{1,2}$, makes $\ell = 0$; therefore the product containing ℓ vanishes, while it changes $P_{k+1,d}$, up to $O(\epsilon)$, only into its restriction to the hyperplane $x_{1,1} = x_{2,1}$. As this hyperplane passes through the origin, the restricted hard polynomial $\tilde{P}$ is still homogeneous of degree d ; this homogeneity is extremely crucial. Moreover, one can argue that $\tilde{P}$ stays nonzero. At the end, every factor satisfies $\ell(0, \mathbf{0}) \neq 0$, so $1/\ell$ is a power series around the origin. This is exactly what the random shift provided in the debordering argument, but now obtained *without* a shift.

Next, [DS26] substitutes $\mathbf{x} \mapsto z\mathbf{x}$, which turns $\tilde{P}$ into $z^d \tilde{P}$, so that all of $\tilde{P}$ sits in a single power of z . They then run up to $k - 1$ rounds of DiDIL, keeping track of the coefficient of the lowest power

of z in the ϵ -free part; in every round, it is $\widetilde{P}$ times a ratio of border $\Sigma \wedge \Sigma$ circuits. After the last round, this expresses $\widetilde{P}$ as a ratio of two $\overline{\Sigma \wedge \Sigma}$ circuits of size $s^{\exp(k)}$. Finally, after clearing the denominator, a nonzero multiple of $\widetilde{P}$ has a small border $\Sigma \wedge \Sigma$ circuit, and comparing spaces of partial derivatives (via leading monomials) shows that this forces $s \geq 2^{\Omega(d/\exp(k))}$.

Our approach in one line after reduction. We follow the same three steps, but replace the $k-1$ rounds of DiDIL by a *single* Wronskian; this is what improves the lower bound from $2^{\Omega(d/\exp(k))}$ to $2^{\Omega(d/k^2)}$.

A new reduction lemma. Essentially, [Lemma 7.5](#) does the same reduction as [\[DS26\]](#), with one further simplification: instead of the whole restricted polynomial $\widetilde{P}$, we keep only a single monomial. Indeed, after a linear change of variables, one of the $k+1$ blocks survives the at most $k-1$ restrictions, and setting all other variables to zero turns $\widetilde{P}$ into $f(\mathbf{x}) := x_1 \cdots x_d$; as this substitution is homogeneous and does not involve ϵ , every factor keeps its nonzero constant term. We could equally keep $\widetilde{P}$, as [\[DS26\]](#) do, but the single monomial makes the final count of partial derivatives particularly simple. By doing this reduction, we obtain

$$\sum_{i=1}^m \epsilon^{a_i} T_i(\epsilon, \mathbf{x}) = \epsilon^M x_1 \cdots x_d + O(\epsilon^{M+1}), \quad m \leq k,$$

where $M, a_i \geq 0$, each $T_i(\epsilon, \mathbf{x}) \in \mathbb{C}[\epsilon, \mathbf{x}]$, and $T_i(0, 0) \neq 0$.

Removing ϵ -dependence. We consider the $\mathbb{C}$ -algebra homomorphism $\psi : \mathbb{C}[\epsilon, \mathbf{x}] \rightarrow \mathbb{C}[z, \mathbf{x}]$ defined by $\psi(\epsilon) = z^{d+1}$ and $\psi(\mathbf{x}) = z\mathbf{x}$. Write $V_i \in \mathbb{C}[z, \mathbf{x}]$ for the image of $\epsilon^{a_i} T_i$. Keeping ϵ in the coefficient field would not suffice for the lowest- z -coefficient argument: after $\mathbf{x} \mapsto z\mathbf{x}$, the approximation error could have z -degree smaller than d . Our substitution ensures the required separation. Since f is homogeneous of degree d and ϵ gets weight $d+1$, the term $\epsilon^M f$ becomes $z^\Delta f$ with $\Delta := M(d+1) + d$, while every error term, which carries ϵ^{M+1} , has order at least $(M+1)(d+1) = \Delta + 1$:

$$\sum_{i=1}^m V_i = z^\Delta f + O(z^{\Delta+1}).$$

Thus f is the lowest z -coefficient of the sum. Moreover, as every affine factor of T_i has a nonzero constant term at $\epsilon = 0$, the *lowest coefficient* of each V_i , that is, the coefficient of its lowest power of z , is a nonzero constant in $\mathbb{C}$.

Unlike [\[DS26\]](#), where ϵ and z stay separate and the limit $\epsilon \rightarrow 0$ is taken in every round of DiDIL before powers of z are compared, we compare powers of z directly in the summands, while the approximation error is still present (with ϵ kept separate, the error could contain powers z^j with $j < d$). Merging the two parameters makes this possible: the weight $d+1$ is just large enough to place f strictly below *every* error term, so a single valuation in z controls both the error and the Wronskians below, and at the end z also serves as the parameter of the border.

A Wronskian isolates f . Let $k = m = 2$ for simplicity. Modulo $z^{\Delta+1}$, the V_i span a space over $\mathbb{K} = \mathbb{C}(\mathbf{x})$ that contains z^Δ , and that is not spanned by z^Δ alone. We argue that $r \neq 1$, as every nonzero $V_i \bmod z^{\Delta+1}$ has $\text{val}_z(V_i) \cdot a_i(d+1) < \Delta$, since $\Delta = M(d+1) + d$ is not divisible by $d+1$. Such a V_i is therefore linearly independent of z^Δ , which already belongs to the span. Hence $r \geq 2$, and $m = 2$ forces $r = 2$. Pick $U \in \{V_1, V_2\}$ such that U and z^Δ form a basis, and write $U = cz^\lambda + O(z^{\lambda+1})$ with $c \in \mathbb{C}^\times$ and $\lambda < \Delta$. To isolate f , we use the map

$$L(F) := W(U, F) = U \cdot DF - F \cdot DU.$$

It has two simple properties: it is linear over $\mathbb{K}$, since elements of $\mathbb{K}$ do not depend on z , and it kills U , that is, $L(U) = 0$. Modulo $z^{\Delta+1}$, every V_i is a combination of the basis elements U and z^Δ :

$$V_i = \alpha_i U + \beta_i z^\Delta + E_i, \quad \alpha_i, \beta_i \in \mathbb{K}, \quad E_i = O(z^{\Delta+1}).$$

So L removes the U -part of each V_i and keeps its z^Δ -part: $L(V_i) = \beta_i L(z^\Delta) + L(E_i)$. (In the debordering argument, the differential equation killed every summand; here we keep the z^Δ -part, since it carries f .) We now compute the lowest terms. On the one hand,

$$L(z^\Delta) = U \cdot \Delta z^{\Delta-1} - z^\Delta \cdot DU = (\Delta - \lambda) c z^\nu + O(z^{\nu+1}), \quad \nu := \Delta + \lambda - 1,$$

and its lowest coefficient $h := (\Delta - \lambda) c$ is nonzero because $\lambda < \Delta$. On the other hand, if $E = O(z^{\Delta+1})$, then both terms of $L(E) = U \cdot DE - E \cdot DU$ have order at least $\lambda + \Delta = \nu + 1$. Thus errors of order $\Delta + 1$ do not affect the coefficient of z^ν ; in particular, every $L(V_i)$ has order at least ν . Now we apply L to both sides of $\sum_i V_i = z^\Delta f + O(z^{\Delta+1})$. As $f \in \mathbb{K}$ does not depend on z , we get

$$\sum_i L(V_i) = f L(z^\Delta) + O(z^{\nu+1}) = h f z^\nu + O(z^{\nu+1}).$$

Comparing the coefficients of z^ν gives

$$g = f h, \quad \text{where} \quad g := \sum_i \text{coeff}_{z^\nu}(L(V_i))$$

and $\text{coeff}_{z^\nu}(A)$ denotes the coefficient of z^ν in A . As every $L(V_i)$ has order at least ν , each $\text{coeff}_{z^\nu}(L(V_i))$ is either zero or the lowest coefficient of the Wronskian $W(U, V_i)$.

For general k , we use the operator $W(U_1, \dots, U_{r-1}, \cdot)$ for a basis $U_1, \dots, U_{r-1}, z^\Delta$ with $r \leq k$. After a change of basis, the U_j have distinct lowest powers of z , so the lowest term of $H := W(U_1, \dots, U_{r-1}, z^\Delta)$ is given by a nonzero Vandermonde determinant (Lemma 7.2), and again $g = f h$, where h is now the lowest coefficient of H (Proposition 7.1).

Lowest coefficients of Wronskians are simple. It remains to show that such lowest coefficients have small $\overline{\Sigma \wedge \Sigma}$ circuits (Lemma 7.3), without paying for ν , which grows with M . As in the debordering argument, we divide each Wronskian by the product of its columns, which changes its lowest coefficient only by a nonzero constant. By the product rule, every determinant term is then a combination, with coefficients in $\mathbb{C}((z))$, of products of $O(r^2)$ terms $D^j \ell / \ell$, where ℓ ranges over the factors of the columns; as the columns of a nonzero Wronskian are distinct, there are at most ks such factors. Write $\ell = \ell_0(z)(1 + \hat{\ell})$, where $\hat{\ell}$ is a homogeneous linear form in $\mathbf{x}$ that vanishes at $z = 0$, and expand $1/(1 + \hat{\ell}) = \sum_{e \geq 0} (-\hat{\ell})^e$. Since ν is huge, we truncate not in z , as in debordering, but in the *degree in $\mathbf{x}$* , which is harmless because the lowest coefficient has degree at most ks in $\mathbf{x}$. Each term then becomes a sum of products of powers of $O(r^2)$ linear forms, so Lemma 3.5 gives a $\Sigma \wedge \Sigma$ expression over $\mathbb{C}((z))$ of top fan-in $(ks)^{O(r^2)}$. Its lowest coefficient in z is a normalized limit $z \rightarrow 0$; that is, z now plays the role of ϵ . Hence g and h have $\overline{\Sigma \wedge \Sigma}$ circuits of top fan-in $R := (ks)^{O(k^2)}$ and degree at most ks .

Partial derivatives. As in [DS26], partial derivatives finish the proof. A power of an affine linear polynomial of degree at most ks has at most $ks + 1$ linearly independent partial derivatives, and this bound adds up under sums and can only drop under limits; so g has at most $R(ks + 1)$. On the other hand, the nonzero multiple $g = (x_1 \cdots x_d) h$ of $x_1 \cdots x_d$ has at least 2^d : with respect to any monomial order, differentiating once in each variable of a subset of $\{x_1, \dots, x_d\}$ produces 2^d distinct leading monomials. Since $k \leq s$, we get $2^d \leq O(R \cdot ks) \leq s^{O(k^2)}$, that is, $s \geq 2^{\Omega(d/k^2)}$; for details, see Section 7.3).

3 Preliminaries

Notation.

- Throughout, $\mathbb{F}$ denotes a field of characteristic zero. We write $[n] := \{1, \dots, n\}$. Bold letters denote tuples of variables, $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$. For $\gamma \in \mathbb{Z}_{\geq 0}^n$, we write $\mathbf{x}^\gamma := \prod_i x_i^{\gamma_i}$, $|\gamma| := \sum_i \gamma_i$, $\gamma! := \prod_i \gamma_i!$ and $\partial^\gamma := \prod_i \partial_{x_i}^{\gamma_i}$.
- $\deg F$ denotes the total degree of F , and $\deg_{\mathbf{y}, \mathbf{x}} F$, $\deg_z F$ the degree in the indicated variables. We write $D := \partial_z$ and $D^i := \partial_z^i$.
- For a commutative ring R , $R[[z]]$ and $R((z))$ denote the rings of formal power series and formal Laurent series in z over R . For $A = \sum_j a_j z^j \in R((z))$, we write $\text{coeff}_{z^j}(A) := a_j$ and $[A]_{\leq d} := \sum_{j \leq d} a_j z^j$. For $A \neq 0$, let $\text{val}_z(A) := \min\{j : a_j \neq 0\}$ and $\text{lc}_z(A) := \text{coeff}_{z^{\text{val}_z(A)}}(A)$; we set $\text{val}_z(0) := \infty$, and write $A = O(z^j)$ if $\text{val}_z(A) \geq j$.
- We use the same notation with ϵ in place of z , regarding $K(\epsilon)$ as a subfield of $K((\epsilon))$ for a field K . If $\text{val}_\epsilon(A) \geq m$, we write $\lim_{\epsilon \rightarrow 0} \epsilon^{-m} A := \text{coeff}_{\epsilon^m}(A)$.
- For $f \in \mathbb{F}[\mathbf{x}]$, $\partial^=k f$ denotes the span of all partial derivatives of f of order exactly k , and $\partial^* f$ the span of all partial derivatives of all orders.

Circuit classes and borders. For a class $\mathcal{C}$ of circuits, $\mathcal{C}(s)$ denotes the polynomials computed by circuits in $\mathcal{C}$ of size at most s , and $\overline{\mathcal{C}(s)}$ its border: $f \in \overline{\mathcal{C}(s)}$ if there is $g \in \mathbb{F}[\epsilon][\mathbf{x}]$ such that $f + \epsilon g$ is computed by a circuit in $\mathcal{C}$ of size at most s over $\mathbb{F}(\epsilon)$. We write $\Sigma^{[w]} \wedge \Sigma$ for the polynomials of the form $\sum_{j=1}^w c_j \ell_j^{e_j}$ with $c_j \in \mathbb{F}$ and ℓ_j affine linear, and $\Sigma \wedge \Sigma(w, e)$ when moreover $e_j \leq e$ for all j . The classes $\Sigma\Pi\Sigma(k, s)$ and $\Sigma \wedge \Sigma\Pi(k, \delta, s)$ are same as defined in [Section 1](#).

3.1 Wronskian

Definition 3.1 (Wronskian). Let $D = \frac{\partial}{\partial z}$. The Wronskian of $f_1, \dots, f_r$, denoted by $W(f_1, \dots, f_r)$, is the determinant of the matrix

$$W(f_1, \dots, f_r) := \det \begin{pmatrix} f_1 & D(f_1) & D^2(f_1) & \dots & D^{r-1}(f_1) \\ f_2 & D(f_2) & D^2(f_2) & \dots & D^{r-1}(f_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_r & D(f_r) & D^2(f_r) & \dots & D^{r-1}(f_r) \end{pmatrix}$$

We use the following facts about Wronskians.

Lemma 3.2 (Lemma 2.2, [\[BS26\]](#)). *Let $\mathbb{F}$ be any field of characteristic 0. Let $f_1, \dots, f_r \in \mathbb{F}(\mathbf{x})[z]$ be non-zero polynomials. Then,*

$$W(f_1, \dots, f_r) = 0$$

if and only if $f_1, \dots, f_r$ are $\mathbb{F}(\mathbf{x})$ -linearly dependent, i.e. there exists $a_1, \dots, a_r \in \mathbb{F}(\mathbf{x})$ s.t. $\exists i \in [r], a_i \neq 0$ and $\sum_{i \in [r]} a_i \cdot f_i = 0$.

Lemma 3.3. *Let $\mathbb{K} = \mathbb{C}(\mathbf{x})$ and let $F_1, \dots, F_q \in \mathbb{K}((z))$ such that $F_b = c_b z^{\mu_b} + O(z^{\mu_b+1})$ with $c_b \in \mathbb{K}^\times$ for $b \in [q]$, and μ_b are pairwise distinct. Then*

$$\text{val}_z W(F_1, \dots, F_q) = \sum_b \mu_b - \binom{q}{2}$$

where $W(F_1, \dots, F_q)$ is the Wronskian wrt z .

Proof. Let $D = \frac{\partial}{\partial z}$. Then

$$\text{val}_z(D^j F) \geq \text{val}_z(F) - j,$$

and hence every term in the expansion of the Wronskian of $F_1, \dots, F_q$ wrt z , $W(F_1, \dots, F_q)$ has valuation wrt z is at least $\sum_b \text{val}_z(F_b) - \binom{q}{2}$, and therefore

$$\text{val}_z W(F_1, \dots, F_q) \geq \sum_{b=1}^q \text{val}_z(F_b) - \binom{q}{2}. \quad (7)$$

Since the Wronskian is linear in each column, Equation (7) shows that $W(F_1, \dots, F_q)$ and $W(c_1 z^{\mu_1}, \dots, c_q z^{\mu_q})$ agree up to terms of valuation larger than $\sum_b \mu_b - \binom{q}{2}$. By [BD10, Lemma 1],

$$W(c_1 z^{\mu_1}, \dots, c_q z^{\mu_q}) = \left(\prod_{b=1}^q c_b \right) V(\mu_1, \dots, \mu_q) z^{\sum_b \mu_b - \binom{q}{2}},$$

where $V(\mu_1, \dots, \mu_q) = \prod_{a < b} (\mu_b - \mu_a)$ denotes the Vandermonde determinant, which is nonzero since $\text{char } \mathbb{K} = 0$. Hence $\text{val}_z W(F_1, \dots, F_q) = \sum_b \mu_b - \binom{q}{2}$. $\square$

3.2 Waring Rank

Let f be a homogeneous polynomial of degree d in $\mathbb{F}[x_1, \dots, x_n]$ where $\mathbb{F}$ is a field of characteristic zero. Then there exist homogeneous linear polynomials $L_1, \dots, L_r$ and $\alpha_1, \dots, \alpha_r \in \mathbb{F}$ such that

$$f = \sum_{i=1}^r \alpha_i L_i^d.$$

The least integer r such that f admits such a decomposition with r -summands is called the Waring rank of f . We denote the Waring rank of f as $\text{WR}(f)$. By abusing the notation, we also use $\text{WR}(f)$ for non-homogeneous polynomials f , which denotes the minimum top fan-in k such that f can be computed by a $\Sigma^k \wedge \Sigma$ circuit.

We will use $\underline{\text{WR}}(f)$ to denote the border waring rank of f , or in other words the minimum k , such that f can be computed by a $\bar{\Sigma}^k \wedge \Sigma$ circuit.

Lemma 3.4. *Let $\mathbb{F}$ be a field of characteristic zero. Then the Waring rank of any degree d monomial in n variables over $\mathbb{F}$ is at most d^n .*

Proof. Let $M = \prod_{j=1}^n x_i^{a_i}$ with $\sum_{i=1}^n a_i = d$ and $1 \leq a_1 \leq \dots \leq a_n$. By [HM22, Remark 3.22] we have

$$\text{WR}_{\mathbb{Q}}(M) \leq \prod_{i=1}^n (a_i + 1),$$

where $\text{WR}_{\mathbb{Q}}(M)$ denotes the Waring rank of M over $\mathbb{Q}$. Since each $a_i \neq 0$, $a_i + 1 \leq d$ for $1 \leq i \leq n$. Hence we obtain $\text{WR}_{\mathbb{Q}}(M) \leq d^n$. Since any field of characteristic zero contains $\mathbb{Q}$, we have $\text{WR}_{\mathbb{F}}(M) \leq d^n$. $\square$

Throughout the paper, we write $\text{WR}(f)$ to denote the Waring rank of f over an arbitrary field $\mathbb{F}$ of characteristic zero unless otherwise specified.

Lemma 3.5. *Let f be a product of linear forms as $\prod_{j=1}^m \ell_j^{d_j}$ such that $\sum_{j=1}^m d_j = d$, then $\text{WR}(f) \leq d^m$.*

Proof. Consider the monomial $M = \prod_{j=1}^m y_j^{d_j}$ in $\mathbb{F}[y_1, \dots, y_m]$. From Lemma 3.4 any monomial over m variables of total degree d , has waring rank d^m . Therefore, $\text{WR}(M) \leq d^m$ and we have $M = \sum_{i=1}^{d^m} \alpha_i L_i(y_1, \dots, y_m)^d$ where L_i are linear forms in $y_1, \dots, y_m$, $\alpha_i \in \mathbb{F}$ for $i \in [d^m]$.

Now consider $y_j = \ell_j(x_1, \dots, x_n)$ for all $j \in [m]$. Therefore, we have $f = \sum_{i=1}^{d^m} \alpha_i L_i(\ell_1, \dots, \ell_m)^d$. Since L_i are linear forms, $L_i(\ell_1, \dots, \ell_m)$ are linear forms in $x_1, \dots, x_n$. $\square$

3.3 Arithmetic Branching Programs

Definition 3.6. [Arithmetic branching Program (ABP), [DDS21a]] 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 linear 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 γ . Number of layers (Δ) defines the depth and the maximum number of vertices in any layer (w) defines the width of an ABP. The size (s) of an ABP is the sum of the graph-size.

It is easy to see from the definition that ABP is closed under both addition and multiplication. We can also eliminate division in ABPs. Here is an important lemma stated below that we will use.

Lemma 3.7 (Lemma 2.6, [DDS21a]). *(Strassen's division elimination) Let $g(\mathbf{x}, y)$ and $h(\mathbf{x}, y)$ be computed by ABPs of size s and degree $< d$. Further, assume $h(\mathbf{x}, 0) \neq 0$. Then, $g/h \text{ mod } y^d$ can be written as $\sum_{i=0}^{d-1} C_i \cdot y^i$, where each C_i is of the form ABP/ABP of size $O(sd^2)$. Moreover, in case g/h is a polynomial, then it has an ABP of size $O(sd^2)$.*

ABPs can capture all polynomials that can be approximated by $\Sigma \wedge \Sigma$ circuits.

Lemma 3.8 (Lemma 52, [DL25]). *If f is a n -variate degree d polynomial such that $f \in \overline{\Sigma^s \wedge \Sigma}$, then it can be computed by an ABP of size $\text{poly}(n, d, s)$.*

3.4 Border Waring Rank PIT

In a recent breakthrough, [BS26], the authors gave a poly-sized hitting set for $\Sigma \wedge \Sigma$ circuits, by giving a hitting set for a polynomial f with the dimension of the space of partial derivatives $\partial^* f$ is small. We use $\partial^* f$ to denote the space spanned by the partial derivatives of f of any order. The argument works for $\overline{\Sigma \wedge \Sigma}$ circuits as well, as they have a low-dimensional partial derivative space as well.

Theorem 3.9 (Theorem 1.1, [BS26]). *Let $\mathbb{F}$ be a field of characteristic zero. For any $n, r \geq 0$, there is an explicit polynomial map $(G_1(t), \dots, G_n(t)) \in \mathbb{F}[t]^n$ with $\deg(G_i) \leq O(nr^2)$ such that if $f(x_1, \dots, x_n)$ is a polynomial such that $\dim(\partial^* f) \leq r$, then*

$$f(x_1, \dots, x_n) = 0 \iff f(G_1(t), \dots, G_n(t)) = 0.$$

Lemma 3.10. *Let $\mathbb{F}$ be a field of characteristic zero, and let $p \in \mathbb{F}[x_1, \dots, x_n]$ of degree Δ . If p is computable by a $\overline{\Sigma^w \wedge \Sigma}$ circuits, then there is an explicit hitting set $\mathcal{A}_{n, \Delta, w} \subseteq \mathbb{F}^n$ for all such polynomials, of size*

$$|\mathcal{A}_{n, \Delta, w}| \leq \text{poly}(n, \Delta, w), \tag{8}$$

constructible within the same asymptotic bound.

Proof. First, we argue that for any polynomial f computed by a $\overline{\Sigma^w \wedge \Sigma}$ circuit, $\dim(\partial^* f) \leq w \cdot (\Delta + 1)$. To see that, first observe that taking a limit cannot increase the dimension of the partial derivative space. For that, consider the partial derivative matrix for partial derivatives of order k of f_ϵ . If $\dim(\partial^{=k} f_\epsilon) = r$, then the rank of the matrix is r , and all $(r + 1) \times (r + 1)$ minors of the partial derivative matrix vanish. These continue to vanish as we take $\lim_{\epsilon \rightarrow 0}$, and hence the $\dim(\partial^{=k} f) \leq r$.

To bound $\dim(\partial^{=k} f_\epsilon)$, write $f_\epsilon = \sum_{i=1}^w \alpha_i \ell_i^{e_i}$ with $\alpha_i \in \mathbb{F}(\epsilon)$ and each ℓ_i affine linear over $\mathbb{F}(\epsilon)$; note that the exponents e_i may exceed Δ . Every partial derivative of order k of $\ell_i^{e_i}$ is a scalar multiple of $\ell_i^{e_i-k}$, or zero if $k > e_i$. Hence $\partial^{=k} f_\epsilon$ is spanned by at most w polynomials, and $\dim(\partial^{=k} f_\epsilon) \leq w$ for every $k \geq 0$. Since $\deg f \leq \Delta$, all partial derivatives of f of order larger than Δ vanish. Combining this with the previous paragraph, we get

$$\dim(\partial^* f) \leq \sum_{k=0}^{\Delta} \dim(\partial^{=k} f) \leq \sum_{k=0}^{\Delta} \dim(\partial^{=k} f_\epsilon) \leq w(\Delta + 1).$$

Using the map in [Theorem 3.9](#) on f , gives us $g = f(G_1(t), \dots, G_n(t))$ such that $g \in \mathbb{F}[t]$, $\deg(g) \leq O(n\Delta^3 w^2)$ and $g = 0$ iff $f = 0$. We can check non-zerosness of g by evaluating on $\deg(g) + 1$ points from $\mathbb{F}$. To obtain the hitting set, evaluate $G_1, \dots, G_n$ on all these points, and it is a hitting set for f . The size of hitting set is $\deg(g) + 1 = O(n\Delta^3 w^2)$. $\square$

3.5 Cramer's rule

Lemma 3.11 (Cramer's rule). *Let A be a matrix in $\mathbb{F}^{n \times n}$ of rank n , and a system of n -equations in n -variables $v = [v_1 \dots v_n]^T$ denoted by*

$$Av = b.$$

If $\det(A) \neq 0$, then we obtain

$$v_i = \frac{\det A_i}{\det(A)}$$

where A_i denotes the matrix obtained by replacing i^{th} column of A by the vector b .

Lemma 3.12. *Let A be a matrix in $\mathbb{F}^{n \times (n+1)}$ of rank n and A_j denote the $n \times n$ matrix obtained by removing the j^{th} column. Then, for $v \in \mathbb{F}^{n+1}$ s.t. $v_j := (-1)^{j+1} \det(A_j)$, we have $Av = 0$.*

Proof. Consider the $(n + 1) \times (n + 1)$ matrix $A^{(i)}$ where the first row is the i -th row of A and rest of the rows are same as A . Clearly $\det(A^{(i)}) = 0$ as it has two identical rows. We also know from definition of determinant that $\det(A^{(i)}) = \sum_{j=1}^{n+1} a_{ij} (-1)^{j+1} \det(A_j)$.

Now we consider the i -th element of Av (represented $(Av)_i$), which is equal to $\sum_{j=1}^{n+1} a_{ij} (-1)^{j+1} \det(A_j)$. Therefore, $(Av)_i = 0$ for any $i \in [n]$, and thus $Av = 0$. $\square$

4 Debordering Depth 3 circuits

Let $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{x} = (x_1, \dots, x_n)$ be disjoint tuples of variables, and let z be a new variable. Define

$$\Phi_{\mathbf{y}} : \mathbb{F}(\epsilon)[\mathbf{x}] \longrightarrow \mathbb{F}(\epsilon)[\mathbf{y}, \mathbf{x}, z], \quad \Phi_{\mathbf{y}}(g)(z, \mathbf{x}) := g(\mathbf{y} + z\mathbf{x}), \quad D := \frac{\partial}{\partial z}. \quad (9)$$

Let $D^i(f) := \frac{\partial^i f}{\partial z^i}$. The tuple $\mathbf{y}$ will be specialized later. Derivatives are always taken before reduction modulo z^{d+1} . This convention is necessary, since D does not induce a derivation on $\mathbb{F}[z]/\langle z^{d+1} \rangle$.

First, we will start by showing that if a polynomial T is a product of linear forms, then its $D^a(T)/T$ can be computed by a $\Sigma \wedge \Sigma$ circuit of size $d^{O(a)}$. This is the equivalent of higher-order dlog computation by $\Sigma \wedge \Sigma$ circuits that has been used in [DDS21a].

Lemma 4.1. *If $T := \prod_{j=1}^d \ell_j$ with $\ell_j \in \mathbb{F}[\mathbf{x}]$ and ℓ_j 's are linear polynomials, then $\frac{D^a(\Phi_{\mathbf{y}}(T))}{\Phi_{\mathbf{y}}(T)} \bmod \langle z^{d+1} \rangle$, a polynomial in $\mathbb{F}(\mathbf{y})[\mathbf{x}, z]$, can be computed by a $\Sigma \wedge \Sigma$ circuit of top fan-in $O(d^{3a+1})$ for $1 \leq a \leq d$.*

Proof. As $\Phi_{\mathbf{y}}(T(\mathbf{x})) = T(\mathbf{y} + z\mathbf{x})$, $\Phi_{\mathbf{y}}(T(\mathbf{x}))$ is a product of polynomials which are linear in z . As we take the a -th derivative of it, by the product rule of derivatives, we will have terms of product of derivatives of a linear polynomials along with the not differentiated linear polynomials, summed over all subsets of $[d]$ of size a . Therefore, we have that

$$D^a(\Phi_{\mathbf{y}}(T)) = a! \cdot \sum_{S \subset [d], |S|=a} \left(\prod_{j \in S} D(\Phi_{\mathbf{y}}(\ell_j)) \right) \cdot \left(\prod_{j \in ([d] \setminus S)} (\ell_j) \right)$$

Dividing the above by T gives us:

$$\frac{D^a(\Phi_{\mathbf{y}}(T))}{\Phi_{\mathbf{y}}(T)} = a! \cdot \sum_{S \subset [d], |S|=a} \prod_{j \in S} \frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)}$$

Let $\ell_j(\mathbf{x}) = c_j + \tilde{\ell}_j(\mathbf{x})$ where c_j is the constant part and $\tilde{\ell}_j$ is the homogeneous part of $\ell_j(\mathbf{x})$. We have $\Phi_{\mathbf{y}}(\ell_j) = \ell_j(\mathbf{y}) + z \cdot \tilde{\ell}_j(\mathbf{x})$. Therefore, we can write it as a power series as

$$\frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)} = \frac{\tilde{\ell}_j(\mathbf{x})}{\ell_j(\mathbf{y}) + z \cdot \tilde{\ell}_j(\mathbf{x})} = \left(\sum_{i \geq 0} (-1)^i z^i \frac{(\tilde{\ell}_j)^{i+1}}{\ell_j(\mathbf{y})^{i+1}} \right)$$

Therefore, we have

$$\prod_{j \in S} \frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)} = \prod_{j \in S} \left(\sum_{i \geq 0} (-1)^i z^i \frac{(\tilde{\ell}_j)^{i+1}}{\ell_j(\mathbf{y})^{i+1}} \right)$$

Taking $\bmod \langle z^{d+1} \rangle$ gives us

$$\prod_{j \in S} \frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)} = \prod_{j \in S} \left(\sum_{i=0}^d (-1)^i z^i \frac{(\tilde{\ell}_j)^{i+1}}{\ell_j(\mathbf{y})^{i+1}} \right) \bmod \langle z^{d+1} \rangle$$

Swapping the sum and products, we get

$$\prod_{j \in S} \frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)} = \sum_{i=0}^d (-1)^i z^i \sum_{d_j \geq 0, \sum_{j \in S} d_j = i} \prod_{j \in S} \frac{(\tilde{\ell}_j)^{d_j+1}}{\ell_j(\mathbf{y})^{d_j+1}} \mod \langle z^{d+1} \rangle \quad (10)$$

Also observe that we have following expression for $\frac{D^a(\Phi_{\mathbf{y}}(T))}{\Phi_{\mathbf{y}}(T)} \Big|_{z=0}$,

$$\frac{D^a(\Phi_{\mathbf{y}}(T))}{\Phi_{\mathbf{y}}(T)} \Big|_{z=0} = a! \cdot \sum_{S \subset [d], |S|=a} \prod_{j \in S} \frac{(\tilde{\ell}_j)}{\ell_j(\mathbf{y})} = a! \cdot \frac{\sum_{S \subset [d], |S|=a} \left(\prod_{j \in S} \tilde{\ell}_j(\mathbf{x}) \right) \cdot \left(\prod_{j \in [d] \setminus S} \ell_j(\mathbf{y}) \right)}{T(\mathbf{y})} \quad (11)$$

for fixed S, i , the product in Equation (10) is the product of $|S| + 1 = a + 1$ linear forms of degree $2i + a$. Using Lemma 3.5, we have that the expression has waring rank at most $(2i + a)^{a+1}$. The summation over possibilities of d_j adds a d^a many possibilities. Since we move terms of $\mathbf{y}$ into the function field, we ignore the complexity of $\frac{1}{\ell_j(\mathbf{y})^{i+1}}$ terms. As i goes upto d , we have

$$\text{WR} \left(\prod_{j \in S} \frac{D(\Phi_{\mathbf{y}}(\ell_j))}{\Phi_{\mathbf{y}}(\ell_j)} \mod \langle z^{d+1} \rangle \right) \leq O(d^{2a+1}).$$

From Equation (10), we know the degree of the formal expression is $2d + a$. Since the number of subsets of $[d]$ size a , is $\binom{d}{a} \approx d^a$, we have

$$\text{WR} \left(\frac{D^a(\Phi_{\mathbf{y}}(T))}{\Phi_{\mathbf{y}}(T)} \mod \langle z^{d+1} \rangle \right) \leq O(d^{3a+1}).$$

□

In the following key proposition, we develop an identity characterizing the computation performed by the border depth-3 circuit, which gives us computable polynomials $b_0, \dots, b_r$ such that for $F = f(\mathbf{y} + z\mathbf{x})$

$$\sum_{j=0}^r b_j D^j(F) = 0 \mod \langle z^{d+1} \rangle.$$

This is also key for the PIT algorithm in Theorem 1.3.

Proposition 4.2 (Truncated differential equation). *Let $\mathbb{F}$ be a field of characteristic zero, and suppose that*

$$f_\epsilon(\mathbf{x}) = \sum_{i=1}^k T_i(\epsilon, \mathbf{x}) = f(\mathbf{x}) + O(\epsilon), \quad T_i = \prod_{a=1}^{d_i} \ell_{ia}(\epsilon, \mathbf{x}), \quad d_i \leq d, \quad (12)$$

where $f \neq 0$, every T_i is nonzero, and the $\ell_{ia} \in \mathbb{F}(\epsilon)[\mathbf{x}]$ are affine linear. Let $F(\mathbf{y}, z, \mathbf{x}) := \Phi_{\mathbf{y}}(f)(z, \mathbf{x}) = f(\mathbf{y} + z\mathbf{x})$.

There exist $1 \leq r \leq \min\{k, d+1\}$ and polynomials $b_0, \dots, b_r \in \mathbb{F}(\mathbf{y})[z, \mathbf{x}]$, with $\deg_z(b_j) \leq d$, such that

$$\sum_{j=0}^r b_j(\mathbf{y}, z, \mathbf{x}) D^j F(\mathbf{y}, z, \mathbf{x}) \equiv 0 \mod z^{d+1}, \quad (13)$$

and

$$b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0. \quad (14)$$

Moreover, the coefficients satisfy the following properties.

1. Viewed as polynomials in $(z, \mathbf{x})$ over $\mathbb{F}(\mathbf{y})$, each b_j has an $\overline{\Sigma} \wedge \overline{\Sigma}$ circuit of top fan-in

$$(kd)^{O(k^2)} \quad (15)$$

and formal degree $O(d + k^2)$.

2. There is a nonzero product $H(\mathbf{y})$ of at most kd affine linear polynomials such that, for every $0 \leq j \leq r$,

$$\beta_j(\mathbf{y}, \mathbf{x}) := b_j(\mathbf{y}, 0, \mathbf{x}), \quad p_j(\mathbf{y}, \mathbf{x}) := H(\mathbf{y})\beta_j(\mathbf{y}, \mathbf{x})$$

satisfy $p_j \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$,

$$\deg_{\mathbf{y}}(p_j) \leq kd, \quad \deg_{\mathbf{x}}(p_j) = O(k^2), \quad (16)$$

and

$$b_j(\mathbf{y}, z, \mathbf{x}) = \left[\frac{p_j(\mathbf{y} + z\mathbf{x}, \mathbf{x})}{H(\mathbf{y} + z\mathbf{x})} \right]_{\leq d} = [\beta_j(\mathbf{y} + z\mathbf{x}, \mathbf{x})]_{\leq d}. \quad (17)$$

Here $[\cdot]_{\leq d}$ denotes Taylor truncation in z .

In particular, all $b_j(\boldsymbol{\alpha}, z, \mathbf{x})$ are defined whenever $H(\boldsymbol{\alpha}) \neq 0$, and

$$b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \neq 0 \iff p_r(\boldsymbol{\alpha}, \mathbf{x}) \neq 0.$$

Proof. For all $i \in [k]$, let $\Phi_{\mathbf{y}}(T_i) = S_i$, and $\Phi_{\mathbf{y}}(\ell_{ij}) = \ell'_{ij}$. Write $\ell'_{ij} = \ell_{ij}(\epsilon, \mathbf{y}) + z \cdot \tilde{\ell}_{ij}(\epsilon, \mathbf{x})$ where $\tilde{\ell}_{ij}$ is the homogeneous part of ℓ_{ij} . Let $\Phi_{\mathbf{y}}(f_{\epsilon}) := F_{\epsilon}$ be computed by the following circuit C .

$$C := S_1 + \dots + S_k = \sum_{i=1}^k \prod_{j=1}^d \ell'_{ij}.$$

Consider the maximal linearly independent subset $S_1, \dots, S_t$ of S_i 's (linearly independent over $\mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$) and consider the $t \times (t+1)$ matrix M defined as follows:

$$M := \begin{pmatrix} S_1 & D(S_1) & D^2(S_1) & \dots & D^t(S_1) \\ S_2 & D^1(S_2) & D^2(S_2) & \dots & D^t(S_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_t & D^1(S_t) & D^2(S_t) & \dots & D^t(S_t) \end{pmatrix}$$

As we know for all $i \in [k]$, $S_i \neq 0$, we can divide row i in M with S_i and consider $\widetilde{M}$ as

$$\widetilde{M} := \begin{pmatrix} 1 & \frac{D^1(S_1)}{S_1} & \frac{D^2(S_1)}{S_1} & \dots & \frac{D^t(S_1)}{S_1} \\ 1 & \frac{D^1(S_2)}{S_2} & \frac{D^2(S_2)}{S_2} & \dots & \frac{D^t(S_2)}{S_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{D^1(S_t)}{S_t} & \frac{D^2(S_t)}{S_t} & \dots & \frac{D^t(S_t)}{S_t} \end{pmatrix}$$

From [Lemma 3.12](#), we also know vector $v \in \mathbb{F}(\epsilon, \mathbf{y}, z, \mathbf{x})^{t+1}$ with $v_j = (-1)^{j+1} \det(M_j)$ is such that $\widetilde{M}v = 0$, where M_j is the matrix $\widetilde{M}$ after removing the j -th column. These v_i 's are rational functions in $\mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x}, z)$ and are determinants of $t \times t$ matrices whose entries are $\frac{D^j(S_i)}{S_i}$.

Note that these v_i 's are not all zero, if and only if $\text{rk}(M) = t$ as $\text{rk}(M) = \text{rk}(\widetilde{M})$. Note that

$$v_{t+1} = (-1)^{t+2} \det(M_{t+1}) = (-1)^{t+2} \frac{W(S_1, \dots, S_t)}{S_1 \cdots S_t}$$

where $W(S_1, \dots, S_t)$ is the Wronskian wrt z of $S_1, \dots, S_t$. If $\text{rk}(M) < t$, then we have

$$W(S_1, \dots, S_t) = 0.$$

This implies that S_i 's are not linearly independent, as their Wronskian vanishes, which contradicts our choice of $S_1, \dots, S_t$. Therefore, we have that $\text{rk}(M) = t$ and v_i 's are not all zero.

Therefore, for all $i \in [t]$, we have

$$v_1 S_i + \sum_{j=1}^t v_{j+1} \cdot D^j(S_i) = 0$$

For any $t < i \leq k$, we can write $S_i = \sum_{j=1}^t c_{ij} S_j$ for some $c_{ij} \in \mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$. As c_{ij} 's are independent of z , we have for any $t < i \leq k$,

$$\begin{aligned} \sum_{j=0}^t v_{j+1} D^j(S_i) &= \sum_{j=0}^t v_{j+1} \sum_{a=1}^t c_{ia} D^j(S_a) \\ &= \sum_{a=1}^t c_{ia} \cdot \left(\sum_{j=0}^t v_{j+1} D^j(S_a) \right) \\ &= 0 \end{aligned}$$

Set $v_{t+2} = \dots = v_{k+1} = 0$. Therefore, for all $i \in [k]$, we have

$$\sum_{j=0}^k v_{j+1} D^j(S_i) = 0$$

Summing the equation for all i , gives us

$$v_1 \cdot F_\epsilon + \sum_{j=1}^k v_{j+1} \cdot D^j(F_\epsilon) = 0$$

Now, we consider the equation $\text{mod } \langle z^{d+1} \rangle$. From [Lemma 4.1](#), we know each of the $v_j \text{ mod } \langle z^{d+1} \rangle \in \mathbb{F}(\epsilon, \mathbf{y})[\mathbf{x}, z]$ are determinants of $t \times t$ ($t \leq k$) matrices whose entries can be computed by $\Sigma \wedge \Sigma$ circuits of top fan-in d^{3k+1} and degree $2d + k$. In each monomial of this determinant, we multiply the respective $\Sigma \wedge \Sigma$ expressions, leaving us with a sum of d^{3k^2} terms, each of which is a product of powers of k linear forms of degree at most $k \cdot (k + 2d)$. From [Lemma 3.5](#), we know each of these

terms has a $(kd)^k$ top fan-in $\Sigma \wedge \Sigma$ expression. Analysing (10) shows us that after truncation, the resulting degree is at most $O(d+k^2)$. Therefore, each monomial of the determinant can be computed by a $\Sigma \wedge \Sigma$ circuit of top fan-in $(kd)^{O(k^2)}$. Since, there are $k!$ monomials of the determinant, we have that each of the v_j 's $\bmod \langle z^{d+1} \rangle$ can be computed by $\Sigma \wedge \Sigma$ circuits of top fan-in $(kd)^{O(k^2)}$.

Let m be the minimum value of $\text{val}_\epsilon(v_i)$ for all non-zero v_i 's. For all $0 \leq i \leq k$, define

$$b_i := \lim_{\epsilon \rightarrow 0} \epsilon^{-m} [v_{i+1}]_{\leq d}$$

where $[v_{i+1}]_{\leq d}$ denotes the truncation of v_{i+1} upto degree d terms of z . Since $[v_{i+1}]_{\leq d}$'s are computable by $\Sigma \wedge \Sigma$ circuits of top fan-in $(kd)^{O(k^2)}$, we have b_i 's are computable by $\overline{\Sigma \wedge \Sigma}$ circuits of top fan-in $(kd)^{O(k^2)}$.

Let $\mathbf{u}$ be a new tuple of variables $u_1, \dots, u_n$. By the chain rule, for every $a \geq 0$, we can translate the derivatives of S_i into derivatives of T_i as

$$\frac{D^a(S_i)}{S_i}(\mathbf{y}, z, \mathbf{x}) = \frac{1}{T_i(\epsilon, \mathbf{u})} \cdot \left(\left(\sum_{l=1}^n x_l \frac{\partial}{\partial u_l} \right)^a T_i(\epsilon, \mathbf{u}) \right) \Big|_{\mathbf{u}=\mathbf{y}+z\mathbf{x}}$$

Therefore, each entry of $M(\mathbf{y}, z, \mathbf{x})$ is obtained from the corresponding entry of $M(\mathbf{u}, 0, \mathbf{x})$ by substituting $\mathbf{u} = \mathbf{y} + z\mathbf{x}$. Since each v_i is a minor of M , we have

$$v_i(\mathbf{y}, z, \mathbf{x}) = v_i(\mathbf{u}, 0, \mathbf{x})|_{\mathbf{u}=\mathbf{y}+z\mathbf{x}} = v_i(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}).$$

As we discussed earlier, the v_i 's are not all zero and therefore, at least one of the $v_i(\mathbf{y}, 0, \mathbf{x})$'s is non-zero. Consider the set $E \subseteq [k+1]$ of all i such that $v_i(\mathbf{y}, 0, \mathbf{x})$ are non-zero and have $\text{val}_\epsilon(v_i) = m$. Divide all the terms in the equation by ϵ^m , and take the limit as $\epsilon \rightarrow 0$. We have that at least one of the $b_i(\mathbf{y}, 0, \mathbf{x})$'s is non-zero. Let the largest index in E be $r+1$. Therefore, we have $\sum_{j=0}^r b_j \cdot D^j(F) = 0 \bmod \langle z^{d+1} \rangle$ and $b_r(\mathbf{y}, 0, \mathbf{x})$ is non-zero.

This finishes proof of part 1.

For part 2, we first normalize the linear forms with appropriate powers of ϵ . For all $i \in [t]$ and $a \in [d_i]$, let

$$\widehat{\ell}_{ia} := \epsilon^{-\text{val}_\epsilon(\ell_{ia})} \ell_{ia} \quad \text{and} \quad \widehat{T}_i := \prod_{a=1}^{d_i} \widehat{\ell}_{ia}$$

Each $\widehat{\ell}_{ia}$ has a nonzero limit as $\epsilon \rightarrow 0$. Since this normalization only multiplies T_i by a scalar in $\mathbb{F}(\epsilon)$, the ratios $\frac{D^a(S_i)}{S_i}$ remain unchanged.

From Equation (11), we know that after setting $z = 0$, all entries in row i have common denominator $\widehat{T}_i(\epsilon, \mathbf{y})$. Since each term of a determinant uses one entry from every row, multiplying $v_{j+1}(\mathbf{y}, 0, \mathbf{x})$ by $\prod_{i=1}^t \widehat{T}_i(\epsilon, \mathbf{y})$ clears its denominators. We will think of $H(\mathbf{y})$ as the common denominator of all $v_{j+1}(\mathbf{y}, 0, \mathbf{x})$'s, and $p_j(\mathbf{y}, \mathbf{x})$ as the numerator of $v_{j+1}(\mathbf{y}, 0, \mathbf{x})$ after clearing denominators.

Therefore, define

$$H(\mathbf{y}) := \prod_{i=1}^t \prod_{a=1}^{d_i} \left(\lim_{\epsilon \rightarrow 0} \widehat{\ell}_{ia}(\epsilon, \mathbf{y}) \right)$$

and

$$p_j(\mathbf{y}, \mathbf{x}) := \lim_{\epsilon \rightarrow 0} \epsilon^{-m} \left(\prod_{i=1}^t \widehat{T}_i(\epsilon, \mathbf{y}) \right) v_{j+1}(\mathbf{y}, 0, \mathbf{x})$$

These limits exist by the normalization in part 1. The expression defining p_j is polynomial in $\mathbf{y}, \mathbf{x}$ before taking the limit, as its denominators have been cleared. Thus $p_j \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$, and we have

$$p_j(\mathbf{y}, \mathbf{x}) = H(\mathbf{y})\beta_j(\mathbf{y}, \mathbf{x})$$

Also, H is a nonzero product of at most $td \leq kd$ linear forms.

After clearing the denominators, each entry in row i has degree at most d_i in $\mathbf{y}$ and at most t in $\mathbf{x}$. This is easy to see from Equation (11). Therefore, taking the determinant and then the normalized limit gives

$$\deg_{\mathbf{y}}(p_j) \leq \sum_{i=1}^t d_i \leq kd, \quad \deg_{\mathbf{x}}(p_j) \leq t^2 \leq k^2$$

As we showed above,

$$v_{j+1}(\mathbf{y}, z, \mathbf{x}) = v_{j+1}(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x})$$

The common denominator on the right is $\prod_{i=1}^t \widehat{T}_i(\epsilon, \mathbf{y} + z\mathbf{x})$, whose value at $\epsilon = z = 0$ is $H(\mathbf{y}) \neq 0$. Therefore, over $\mathbb{F}(\mathbf{y})$, we can take the normalized limit and truncate in z to obtain

$$b_j(\mathbf{y}, z, \mathbf{x}) = [\beta_j(\mathbf{y} + z\mathbf{x}, \mathbf{x})]_{\leq d} = \left[\frac{p_j(\mathbf{y} + z\mathbf{x}, \mathbf{x})}{H(\mathbf{y} + z\mathbf{x})} \right]_{\leq d}.$$

For any α such that $H(\alpha) \neq 0$, these expressions are defined, and

$$b_r(\alpha, 0, \mathbf{x}) = \frac{p_r(\alpha, \mathbf{x})}{H(\alpha)} \neq 0 \iff p_r(\alpha, \mathbf{x}) \neq 0.$$

□

Now, we have all the tools required to finish the proof of [Theorem 1.1](#).

4.1 Proof of [Theorem 1.1](#)

We now show that any polynomial f in $\overline{\Sigma \wedge \Sigma(k, s)}$ can be computed by an ABP of size $s^{O(k^2)}$.

Proof of [Theorem 1.1](#). Let $f \in \overline{\Sigma \Pi \Sigma(k, s)}$ be computed by circuit C in $\overline{\Sigma \Pi \Sigma(k, s)}$. From [Proposition 4.2](#), there exists non-zero polynomials $b_0, \dots, b_r$ such that

$$\sum_{j=0}^r b_j(\mathbf{y}, z, \mathbf{x}) D^j F(\mathbf{y}, z, \mathbf{x}) \equiv 0 \pmod{z^{d+1}},$$

Define F_i as the coefficient of z^i in $f(\mathbf{y} + z\mathbf{x})$. Therefore, We have $F(\mathbf{y}, z, \mathbf{x}) := \sum_{i=0}^d z^i F_i$. Also, $D^j(F) = \sum_{i=0}^{d-j} \frac{(i+j)!}{i!} z^i F_{i+j}$.

Let $b_j(\mathbf{y}, z, \mathbf{x}) := \sum_{i=0}^d z^i b_{j,i}(\mathbf{y}, \mathbf{x})$ where $b_{j,i}(\mathbf{y}, \mathbf{x})$ are polynomials in $\mathbb{F}(\mathbf{y})[\mathbf{x}]$ corresponding to the coefficients of z^i in $b_j(\mathbf{y}, z, \mathbf{x})$. Therefore, we have for all $q \in [d]$,

$$\sum_{j=0}^r \sum_{i=0}^q \frac{(i+j)!}{i!} b_{j,(q-i)}(\mathbf{y}, \mathbf{x}) F_{i+j} = 0$$

We first obtain ABP circuits for $F_0, \dots, F_{r-1}$. We know $r \leq \min k, d+1$. For $i \geq 1$, taking the coefficient of z^i in C gives a sum of at most $k \binom{d}{i}$ terms, each being a scalar times a product of i linear forms. From [Lemma 3.5](#), we have

$$\text{WR}(F_i) \leq k \binom{d}{i} i^i$$

For $i < r \leq k$, this is $(kd)^{O(k)}$. Also, f_0 is a scalar over $\mathbb{F}(\mathbf{y})$. Therefore, these initial coefficients have ABPs of size $(nkd)^{O(k)} = s^{O(k)}$ from [Lemma 3.8](#). If $r = d+1$, we are done. Otherwise, assume $r \leq d$. As b_i 's can be computed by $\overline{\Sigma} \wedge \overline{\Sigma}$ circuits of top fan-in $(kd)^{O(k^2)}$, we have b_i and $b_{i,j}$ can be computed by ABPs over $\mathbb{F}(y)$ of size $(nkd)^{O(k^2)} = s^{O(k^2)}$ from [Lemma 3.8](#).

Now, we move the terms of $F_0, \dots, F_{r-1}$ to the right-hand side of the equation and consider the coefficient of z^d in the equations we obtained above. Therefore, we have for all $0 \leq q \leq d-r$,

$$\sum_{j=0}^r \sum_{i=0}^q \frac{(i+j)!}{i!} b_{j,q-i}(\mathbf{y}, \mathbf{x}) F_{i+j} = 0$$

This creates a square system of linear equations in $F_r, \dots, F_d$. Let this system be characterised by matrix A . Equation q contains no F_i with $i > q+r$, and the coefficient of F_{q+r} is $\frac{(q+r)!}{q!} b_{r,0}(\mathbf{y}, \mathbf{x})$. Therefore, A is lower triangular and

$$\det(A) = \left(\prod_{q=0}^{d-r} \frac{(q+r)!}{q!} \right) b_r(\mathbf{y}, 0, \mathbf{x})^{d-r+1} \neq 0$$

All entries of A and of the right-hand side have ABPs of size $s^{O(k^2)}$, since they are sums and products of the coefficients already computed. This gives us an invertible square system in $F_r, \dots, F_d$ with coefficients all of which are computable by ABPs of size $s^{O(k^2)}$. Since the linear system is invertible and full rank, it also gives a *unique* solution for $F_r, \dots, F_d$.

Therefore, we can solve this system by [Lemma 3.11](#) and obtain $F_r, \dots, F_d$ as ratios of determinants with entries of $b_{j,i}$'s and $F_0, \dots, F_{r-1}$. Using the ABPs for them and the determinant along with division elimination, we can obtain ABP circuits for $F_r, \dots, F_d$ of size $s^{O(k^2)}$. Thus, we combine these to get an ABP computing $F(\mathbf{y}, z, \mathbf{x})$ with coefficients over $\mathbb{F}(y)$. Recall $F(z, \mathbf{x}) = f(\mathbf{y} + z\mathbf{x})$. So, we substitute $z = 1$, and $\mathbf{y} = \alpha$ such that $H(\alpha) \neq 0$ and $b_r(\alpha, 0, \mathbf{x}) \neq 0$ which exists as $H(\mathbf{y}), b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0$ and ensures all the denominator terms of the ABP are non-zero. This gives us an ABP for $f(\alpha + \mathbf{x})$, from which we can obtain an ABP for f by substituting $\mathbf{x} - \alpha$ in all the occurrences of $\mathbf{x}$. Thus, we have an ABP computing f of size $s^{O(k^2)}$. $\square$

5 PIT for border depth-3

Next, we prove a crucial identity lemma that leads to an efficient hitting set.

Lemma 5.1 (Identity Lemma). *Let f and $b_0, \dots, b_r$ be as defined in Proposition 4.2. Let $\mathbf{y} = \boldsymbol{\alpha} \in \mathbb{F}^n$ such that each b_j is defined and further,*

$$b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \neq 0. \quad (18)$$

Write

$$F_{\boldsymbol{\alpha}}(z, \mathbf{x}) := f(\boldsymbol{\alpha} + z\mathbf{x}) = \sum_{i=0}^d F_{\boldsymbol{\alpha},i}(\mathbf{x})z^i. \quad (19)$$

For notational convenience, set $F_{\boldsymbol{\alpha},i} := 0$ for $i > d$. Then

$$F_{\boldsymbol{\alpha},0} = F_{\boldsymbol{\alpha},1} = \dots = F_{\boldsymbol{\alpha},r-1} = 0 \iff f = 0.$$

Proof. $\Leftarrow$ is obvious. For $\Rightarrow$, if $r > d$, the hypothesis already implies $F_{\boldsymbol{\alpha}} = 0$, and hence $f = 0$. Assume that $r \leq d$ and, towards a contradiction, that $f \neq 0$. Choose $z = 1$, then $F_{\boldsymbol{\alpha}}(1, \mathbf{x}) = f(\boldsymbol{\alpha} + \mathbf{x}) \neq 0$, so $F_{\boldsymbol{\alpha}} \neq 0$. Let

$$m := \min\{i : F_{\boldsymbol{\alpha},i} \neq 0\}.$$

By hypothesis, $m \geq r$. For every $j < r$, the z -valuation of $b_j(\boldsymbol{\alpha}, z, \mathbf{x})D^j F_{\boldsymbol{\alpha}}$ is at least $m - j > m - r$. On the other hand, the coefficient of z^{m-r} in $b_r(\boldsymbol{\alpha}, z, \mathbf{x})D^r F_{\boldsymbol{\alpha}}$ is

$$b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \frac{m!}{(m-r)!} F_{\boldsymbol{\alpha},m}(\mathbf{x}), \quad (20)$$

which is nonzero by Equation (18), the choice of m , and the characteristic-zero assumption. Thus the coefficient of z^{m-r} in the left-hand side of Equation (13) is nonzero. Since $m - r \leq d$, this contradicts the congruence modulo z^{d+1} . $\square$

In particular, if $f \neq 0$ and $\boldsymbol{\alpha}$ satisfies Equation (18), then for some $i \in \{0, \dots, r-1\}$, we have

$$F_{\boldsymbol{\alpha},i} \neq 0. \quad (21)$$

Moreover, $F_{\boldsymbol{\alpha},i}$ is homogeneous of degree i in $\mathbf{x}$; indeed, Taylor's formula gives

$$F_{\boldsymbol{\alpha},i}(\mathbf{x}) = \sum_{|\gamma|=i} \frac{\partial^\gamma f(\boldsymbol{\alpha})}{\gamma!} \mathbf{x}^\gamma.$$

Thus the only condition that must be enforced when choosing $\boldsymbol{\alpha}$ is that the leading constant coefficient $b_r(\boldsymbol{\alpha}, 0, \mathbf{x})$ is non-zero and the coefficients b_j 's are defined. Once this condition holds, Equation (21) follows automatically for every nonzero f .

In the lemma below, we construct a hitting set that simultaneously hits both $H(\mathbf{y})$ and $b_r(\mathbf{y}, 0, \mathbf{x})$.

Lemma 5.2. *Let $n, d, k \leq s$. There is an explicit set $\mathcal{S} \subseteq \mathbb{F}^n$ of size $s^{O(k^2)}$ such that, for every nonzero polynomial f satisfying the hypotheses of Proposition 4.2, we have the coefficients $b_0, \dots, b_r$ constructed in the proposition's proof and a point $\boldsymbol{\alpha} \in \mathcal{S}$ at which every $b_j(\boldsymbol{\alpha}, z, \mathbf{x})$ is defined and*

$$b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \neq 0.$$

The set depends only on n, d, k, s and is constructible within the same time bound.

Proof. By [Proposition 4.2](#), it suffices to find α such that

$$H(\alpha) \neq 0 \quad \text{and} \quad p_r(\alpha, \mathbf{x}) \neq 0.$$

Let $\mathcal{H}_1$ be an explicit hitting set of size $(nkd)^{O(1)}$ for products of at most kd affine linear polynomials. Choose $\mathbf{a} \in \mathcal{H}_1$ with $H(\mathbf{a}) \neq 0$, and write

$$\beta_r(\mathbf{a} + t\mathbf{u}, \mathbf{x}) = \sum_{v \geq 0} \beta_{r,v}(\mathbf{u}, \mathbf{x}) t^v, \quad \beta_r = \frac{p_r}{H}.$$

Since $p_r(\mathbf{a} + t\mathbf{u}, \mathbf{x})$ is nonzero and has t -degree at most kd , while $H(\mathbf{a} + t\mathbf{u})$ has nonzero constant term, some $\beta_{r,v}$ with $v \leq kd$ is nonzero.

We know $\beta_r = b_r(\mathbf{y}, 0, \mathbf{x})$ is equal to $\lim_{\epsilon \rightarrow 0}$ of a $t \times t$ determinant of matrix with entries $\frac{D^a(\Phi_{\mathbf{y}}(T_i))}{\Phi_{\mathbf{y}}(T_i)} \Big|_{z=0}$ as $a \in [t]$ and $i \in [t]$ for $t \leq k$ as defined in [Proposition 4.2](#). Before taking the limit, we also normalise it with ϵ^{-m} as done in [Proposition 4.2](#). As the complexity increases with increasing a , we focus on case being $a = k$. From (11), we know for $T_i = \prod_{j=1}^{d_i} \ell_{ij}$, and $\Phi_{\mathbf{y}}(\ell_{ij}) = \ell_{ij}(\mathbf{y}) + z \cdot \tilde{\ell}_{ij}(\mathbf{x})$

$$\frac{D^k(\Phi_{\mathbf{y}}(T_i))}{\Phi_{\mathbf{y}}(T_i)} \Big|_{z=0} = k! \cdot \sum_{S \subset [d], |S|=k} \prod_{j \in S} \frac{(\tilde{\ell}_{ij})}{\ell_{ij}(\mathbf{y})}$$

Substituting $\mathbf{y} = \mathbf{a} + t\mathbf{u}$ into the above expression, gives us

$$\frac{D^k(\Phi_{\mathbf{y}}(T_i))}{\Phi_{\mathbf{y}}(T_i)} \Big|_{z=0} \Big|_{\mathbf{y}=\mathbf{a}+t\mathbf{u}} = k! \cdot \sum_{S \subset [d], |S|=k} \prod_{j \in S} \frac{(\tilde{\ell}_{ij}(\mathbf{x}))}{\ell_{ij}(\mathbf{a} + t\mathbf{u})}$$

Consider the coefficient of t^w in the above expression. As $H(\mathbf{a}) \neq 0$, we know $\ell_{ij}(\mathbf{a}) \neq 0$. Therefore, similar to (10), and ignoring constants, we get

$$\frac{D^k(\Phi_{\mathbf{y}}(T_i))}{\Phi_{\mathbf{y}}(T_i)} \Big|_{z=0} \Big|_{\mathbf{y}=\mathbf{a}+t\mathbf{u}} = \sum_{w \geq 0} t^w \cdot \left(\sum_{S \subset [d], |S|=k} \sum_{d_j \geq 0, \sum_{j \in S} d_j = w} \prod_{j \in S} \tilde{\ell}_{ij}(\mathbf{x}) \cdot \tilde{\ell}_{ij}(\mathbf{u})^{d_j} \right)$$

In this the product is a product of $2k$ linear forms of degree $w + k$. From [Lemma 3.5](#), we have that they can be computed by $\Sigma \wedge \Sigma$ circuits with top fan-in $(w + k)^{2k}$. Therefore, we get that coefficient of t^w has waring rank at most $d^k \cdot \binom{w+k}{k} \cdot (w + k)^{2k} = (d \cdot (w + k))^{O(k)}$.

In $\beta_r(\mathbf{a} + t\mathbf{u}, \mathbf{x}) = b_r(\mathbf{a} + t\mathbf{u}, 0, \mathbf{x})$, we have limit of determinant of at most k sized matrix with entries as above. As done in [Proposition 4.2](#), we multiply these $\Sigma \wedge \Sigma$ expressions and then use [Lemma 3.5](#) to obtain that the coefficient of t^v (as v goes upto kd) in $\beta_r(\mathbf{a} + t\mathbf{u}, \mathbf{x})$ can be computed by a $\overline{\Sigma \wedge \Sigma}$ circuit of top fan-in $(dk)^{O(k^2)}$ and degree $O(kd + k^2)$. To observe the degree, observe that the expression will have degree of $\mathbf{u}$ as v and degree of $\mathbf{x}$ as at most $O(k^2)$.

We claim that every such $\beta_{r,v}$ has degree $O(kd + k^2)$ and is computable by a $\overline{\Sigma \wedge \Sigma}$ circuit of top fan-in $(dk)^{O(k^2)}$. As we have $v \leq kd$ and $d \leq s$, if we pick $\mathbf{u}$ from a hitting set that hits all $\overline{\Sigma \wedge \Sigma}$ circuits of top fan-in $(k \cdot s)^{O(k^2)}$, we have $\beta_{r,v} \neq 0$.

We use the explicit hitting set, as stated in [Lemma 3.10](#). Let $w = s^{O(k^2)}$. Let $\mathcal{H}_2 \subseteq \mathbb{F}^{2n}$ be an explicit hitting set for polynomials in $(\mathbf{u}, \mathbf{x})$ of total degree $O(kd + k^2)$ that are computable by $\overline{\Sigma^w \wedge \Sigma}$ circuits. Its size is

$$|\mathcal{H}_2| \leq \text{poly}(n, (kd + k^2), w) \leq \text{poly}(s, w).$$

Every $\beta_{r,v}$ with $v \leq kd$ belongs to this class. Since at least one such $\beta_{r,v}$ is nonzero, there are $i \leq kd$ and $(\mathbf{u}, \boldsymbol{\xi}) \in \mathcal{H}_2$ such that

$$\beta_{r,v}(\mathbf{u}, \boldsymbol{\xi}) \neq 0.$$

Consequently, $\beta_r(\mathbf{a} + t\mathbf{u}, \boldsymbol{\xi})$ is a nonzero rational function of t .

Since $H(\mathbf{a}) \neq 0$, the polynomial $H(\mathbf{a} + t\mathbf{u})$ is nonzero. As $\beta_r = p_r/H$, it follows that $p_r(\mathbf{a} + t\mathbf{u}, \boldsymbol{\xi})$ is also nonzero.

Let Γ be an explicit univariate hitting set for degree $O(kd)$, of size $O(kd)$. The product

$$H(\mathbf{a} + t\mathbf{u})p_r(\mathbf{a} + t\mathbf{u}, \boldsymbol{\xi})$$

is nonzero and has degree $O(kd)$. Thus some $\tau \in \Gamma$ makes both factors nonzero. For $\boldsymbol{\alpha} = \mathbf{a} + \tau\mathbf{u}$, all $b_j(\boldsymbol{\alpha}, z, \mathbf{x})$ are defined and $b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \neq 0$.

We may therefore take

$$\mathcal{S} := \{\mathbf{a} + \tau\mathbf{u} : \mathbf{a} \in \mathcal{H}_1, (\mathbf{u}, \boldsymbol{\xi}) \in \mathcal{H}_2, \tau \in \Gamma\}.$$

Its size is at most

$$s^{O(1)} \text{poly}(s, w) = \text{poly}(s, w).$$

All constituent hitting sets are explicit and constructible within the stated bounds. $\square$

5.1 Proof of [Theorem 1.3](#)

Finally, we combine the tools, to get a $s^{O(k^2)}$ time PIT algorithm for polynomials in $\overline{\Sigma \wedge \Sigma(k, s)}$.

Proof of [Theorem 1.3](#). Let $\mathcal{S}$ be the set supplied by [Lemma 5.2](#). Let $\mathcal{H}_3 \subseteq \mathbb{F}^n$ be an explicit hitting set of size $(nk)^{O(k)}$ for polynomials of total degree less than k , and let $\Delta \subseteq \mathbb{F}$ be a univariate hitting set of size $O(d)$ for degree at most d . We claim that

$$\mathcal{H} := \{\boldsymbol{\alpha} + \zeta\mathbf{v} : \boldsymbol{\alpha} \in \mathcal{S}, \mathbf{v} \in \mathcal{H}_3, \zeta \in \Delta\}$$

hits every nonzero polynomial admitting a degeneration as in [\(12\)](#).

Fix such an f . By [Lemma 5.2](#), some $\boldsymbol{\alpha} \in \mathcal{S}$ makes every $b_j(\boldsymbol{\alpha}, z, \mathbf{x})$ defined and satisfies $b_r(\boldsymbol{\alpha}, 0, \mathbf{x}) \neq 0$. Write

$$f(\boldsymbol{\alpha} + z\mathbf{x}) = \sum_{i=0}^d F_{\boldsymbol{\alpha},i}(\mathbf{x})z^i.$$

[Lemma 5.1](#) gives $F_{\boldsymbol{\alpha},i} \neq 0$ for some $i < r \leq k$. Since $F_{\boldsymbol{\alpha},i}$ is homogeneous of degree i , there is $\mathbf{v} \in \mathcal{H}_3$ such that $F_{\boldsymbol{\alpha},i}(\mathbf{v}) \neq 0$. Hence $f(\boldsymbol{\alpha} + z\mathbf{v})$ is a nonzero univariate polynomial of degree at most d , and some $\zeta \in \Delta$ satisfies

$$f(\boldsymbol{\alpha} + \zeta\mathbf{v}) \neq 0.$$

Finally, using $n, d, k \leq s$,

$$|\mathcal{H}| \leq |\mathcal{S}| |\mathcal{H}_3| |\Delta| \leq s^{O(k^2)} \leq s^{O(k^2)}.$$

This finishes the proof. $\square$

6 Debordering $\overline{\Sigma \wedge \Sigma \Pi}(k, \delta, s)$

In this section, we deborder the sum of powers of bounded-degree polynomials using the Wronskian approach. Recall that

$$\Sigma \wedge \Sigma \Pi(k, \delta, s) = \left\{ \sum_{i=1}^k P_i^{e_i} : \deg(P_i) \leq \delta, \text{ and } \sum_i (\text{size}(P_i) + e_i) \leq s \right\}.$$

For $i \in [k]$, let $T_i = P_i^{e_i}$. The proof is analogous to that in [Section 4](#), except that we do *not* expand the entries $D^a S_i / S_i$ as power series in z ; instead, we keep them as rational functions and clear denominators.

The following lemma shows that out of all the size- r submatrices of M , the Wronskian has the least valuation wrt ϵ .

Lemma 6.1. *Let $S_1, \dots, S_r \in \mathbb{F}(\epsilon)[\mathbf{y}, z, \mathbf{x}]$, and M be the matrix given by*

$$M := \begin{pmatrix} S_1 & D^1(S_1) & D^2(S_1) & \dots & D^r(S_1) \\ S_2 & D^1(S_2) & D^2(S_2) & \dots & D^r(S_2) \\ \vdots & \vdots & \vdots & & \vdots \\ S_r & D^1(S_r) & D^2(S_r) & \dots & D^r(S_r) \end{pmatrix}$$

where $D = \partial z$, partial derivative wrt z . Let M_i denote the matrix obtained by deleting the i^{th} column of M . Denote $A_i = (-1)^i \det(M_{i+1})$ for $0 \leq i \leq r$. If $S_1, \dots, S_r$ are linearly independent set over $\mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$, then

$$\text{val}_\epsilon(A_i) \geq \text{val}_\epsilon(A_r) = \text{val}_\epsilon(W(S_1, \dots, S_r)).$$

Proof. Let $q_i := \text{val}_\epsilon(S_i)$ and $q = \min\{q_i \mid 1 \leq i \leq r\}$. Since $\text{val}_\epsilon(D^j S_i) \geq \text{val}_\epsilon(S_i) \geq q$, we have valuation of every term of M is at least q . Hence the determinant of any $r \times r$ sub matrix of has valuation at least rq . Hence we have

$$\text{val}_\epsilon(A_i) \geq rq$$

for all $0 \leq i \leq r$. In particular,

$$\text{val}_\epsilon(A_r) = \text{val}_\epsilon(W(S_1, \dots, S_r)) \geq rq.$$

Let $m := \text{val}_\epsilon(W(S_1, \dots, S_r))$. If $m = rq$, then we are done.

Suppose $\text{val}_\epsilon(W(S_1, \dots, S_r)) = m > rq$. Say $m = rq + N$. Write $S_i = \epsilon^q(t_i + \epsilon r_i)$ where $t_i \in \mathbb{F}[\mathbf{y}, z, \mathbf{x}]$ for $1 \leq i \leq r$. Since determinant is linear in each row, we obtain

$$W(S_1, \dots, S_r) = \epsilon^{rq}(W(t_1, \dots, t_r) + \epsilon R). \quad (22)$$

Since $\text{val}_\epsilon(W(S_1, \dots, S_r)) > rq$ we obtain $W(t_1, \dots, t_r) = 0$. Note that $t_i \in \mathbb{F}[\mathbf{y}, z, \mathbf{x}]$ for $1 \leq i \leq r$, and hence $t_1, \dots, t_r$ are linearly dependent over $\mathbb{F}(\mathbf{y}, \mathbf{x})$. Therefore, there exists $\lambda_i \in \mathbb{F}(\mathbf{y}, \mathbf{x})$, not all zeros, such that

$$t_j = \sum_{i \neq j} \lambda_i t_i$$

for some $1 \leq j \leq r$. Now apply the row operation j^{th} row R_j changes to $\epsilon^{-1}(R_j - \sum_{i \neq j} \lambda_i R_i)$ in M , and denote it by M' . Note that

$$\text{val}_\epsilon(\epsilon^{-1}(S_j - \sum_{i \neq j} \lambda_i S_i)) \geq q.$$

Therefore val_ϵ of each entry of M' is at least q , hence $\text{val}_\epsilon(A'_i) \geq rq$. For every submatrix of size $r \times r$ obtained by deleting i^{th} column of M' ,

$$A'_i = (-1)^i \det(M'_{i+1}) = \epsilon^{-1} A_i$$

for $0 \leq i \leq r$. Therefore

$$rq \leq \text{val}_\epsilon(A'_i) \leq \text{val}_\epsilon(A_i) - 1. \quad (23)$$

By inductively repeating this process, we obtain

$$rq \leq \text{val}_\epsilon(A'_i) \leq \text{val}_\epsilon(A_i) - N.$$

Hence $\text{val}_\epsilon(W(S_1, \dots, S_r)) = m = rq + N \leq \text{val}_\epsilon(A_i)$ for every $0 \leq i \leq r$. $\square$

We now show that the least degree coefficient of ϵ in the Wronskian evaluated at $z = 0$ is non-zero. This is useful to reconstruct each degree piece of f .

Lemma 6.2. *Let $T_1, \dots, T_r \in \mathbb{F}(\epsilon)[\mathbf{x}]$ and $S_i(\mathbf{y}, z, \mathbf{x}) := \Phi_{\mathbf{y}}(T_i)$ for $i \in [r]$. Suppose that $W(S_1, \dots, S_r) \neq 0$, let $m := \text{val}_\epsilon(W(S_1, \dots, S_r))$, and write*

$$W(S_1, \dots, S_r) = \epsilon^m (W_0(\mathbf{y}, z, \mathbf{x}) + \epsilon W'),$$

where $W_0 \in \mathbb{F}[\mathbf{y}, z, \mathbf{x}]$ is ϵ -free and $W' \in \mathbb{F}(\epsilon)[\mathbf{y}, z, \mathbf{x}]$ satisfies $\text{val}_\epsilon(W') \geq 0$; in particular, $W_0 \neq 0$. Then $W_0(\mathbf{y}, 0, \mathbf{x}) \neq 0$.

Proof. Let θ be a new variable. For every $i \in [r]$,

$$S_i(\mathbf{y}, z + \theta, \mathbf{x}) = T_i(\mathbf{y} + (z + \theta)\mathbf{x}) = S_i(\mathbf{y} + z\mathbf{x}, \theta, \mathbf{x}).$$

Write $W(\mathbf{y}, z, \mathbf{x}) := W(S_1, \dots, S_r)$. Taking the Wronskian of both sides with respect to θ gives $W(\mathbf{y}, z + \theta, \mathbf{x}) = W(\mathbf{y} + z\mathbf{x}, \theta, \mathbf{x})$, since ∂_θ acts on the left-hand side as $D = \partial_z$ does. Setting $\theta = 0$, we obtain

$$W(\mathbf{y}, z, \mathbf{x}) = W(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}). \quad (24)$$

The substitution $(\mathbf{y}, z) \mapsto (\mathbf{y} + z\mathbf{x}, 0)$ does not involve ϵ , so the right-hand side of (24) equals $\epsilon^m (W_0(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}) + \epsilon W'(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}))$, and $W'(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x})$ still has nonnegative ϵ -valuation. Dividing (24) by ϵ^m and setting $\epsilon = 0$, we get

$$W_0(\mathbf{y}, z, \mathbf{x}) = W_0(\mathbf{y} + z\mathbf{x}, 0, \mathbf{x}). \quad (25)$$

If $W_0(\mathbf{y}, 0, \mathbf{x}) = 0$, then the right-hand side of (25) vanishes, and hence $W_0 = 0$, a contradiction. Therefore $W_0(\mathbf{y}, 0, \mathbf{x}) \neq 0$. $\square$

Together with the lemmas proved above, we now recover the polynomial from a differential equation. Moreover, we obtain each higher degree piece of f as a quotient of two polynomials in $O(k^3\delta)$ -variables.

Proposition 6.3. *Let $\mathbb{F}$ be a field of characteristic zero. Suppose that $0 \neq f \in \mathbb{F}[\mathbf{x}]$ such that $f \in \Sigma \wedge \Sigma \Pi(k, \delta, s)$. i.e.,*

$$f_\epsilon(\mathbf{x}) := \sum_{i=1}^k T_i(\epsilon, \mathbf{x}) = f(\mathbf{x}) + O(\epsilon), \quad (26)$$

where $0 \neq T_i$, with $T_i = P_i^{e_i}$ and $\deg P_i \leq \delta$, $0 \leq e_i \leq e$ for $1 \leq i \leq k$. Let $F(\mathbf{y}, z, \mathbf{x}) := \Phi_{\mathbf{y}}(f)(z, \mathbf{x}) = f(\mathbf{y} + z\mathbf{x})$. Then there exist $1 \leq r \leq k$ and polynomials $b_0, \dots, b_r \in \mathbb{F}[\mathbf{y}, z, \mathbf{x}]$ with

$$\deg b_j(\mathbf{y}, z, \mathbf{x}) \leq 2k^2\delta$$

for $0 \leq j \leq r$ such that

$$\sum_{j=0}^r b_j(\mathbf{y}, z, \mathbf{x}) D^j F(\mathbf{y}, z, \mathbf{x}) = 0,$$

and

$$b_r(\mathbf{y}, 0, \mathbf{x}) \neq 0.$$

Moreover, we have $b_j(\mathbf{y}, z, \mathbf{x}) = \sum_{i=0}^{2k^2\delta} b_{ji}(\mathbf{y}, \mathbf{x}) z^i$ where $b_{ji}(\mathbf{y}, \mathbf{x}) \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$. Write

$$F(\mathbf{y}, z, \mathbf{x}) = \sum_{j=0}^{\delta e} F_j(\mathbf{y}, \mathbf{x}) z^j.$$

By considering $F_0, \dots, F_{r-1}$ and $b_{ji}(\mathbf{y}, \mathbf{x})$ for $0 \leq j \leq r$, $0 \leq i \leq 2k^2\delta$ as parameters, there exist polynomials $Q_l(\mathbf{w})$ in $(r+1)(2k^2\delta+1) + r$ variables $\mathbf{w}$, one for each parameter, such that $\deg Q_l \leq \delta e - r + 2$ and

$$F_{r-1+l}(\mathbf{y}, \mathbf{x}) \cdot b_{r0}(\mathbf{y}, \mathbf{x})^{\delta e - r + 1} = Q_l(\mathbf{b}) \quad (27)$$

for $1 \leq l \leq \delta e - r + 1$, where $\mathbf{b}$ denotes the tuple of parameters

$$\mathbf{b} := (b_{00}, \dots, b_{0,2k^2\delta}, \dots, b_{r0}, \dots, b_{r,2k^2\delta}, F_0, \dots, F_{r-1}).$$

Proof. Let $\Phi_{\mathbf{y}}(P_i) = G_i$, and $\Phi_{\mathbf{y}}(T_i)(\mathbf{x}) = S_i$. Hence $\Phi_{\mathbf{y}}(P_i^{e_i}) = G_i^{e_i} = S_i$ for $i \in [k]$. Let $S_1, \dots, S_r$ be a maximal linearly independent set over $\mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$ for some $r \leq k$. Consider the following $r \times (r+1)$ matrix given by

$$M' := \begin{pmatrix} S_1 & D^1(S_1) & D^2(S_1) & \dots & D^r(S_1) \\ S_2 & D^1(S_2) & D^2(S_2) & \dots & D^r(S_2) \\ \vdots & \vdots & \vdots & & \vdots \\ S_r & D^1(S_r) & D^2(S_r) & \dots & D^r(S_r) \end{pmatrix}$$

Since $S_i = G_i^{e_i}(\mathbf{y}, z, \mathbf{x})$ and by applying chain rule of differentiation, we have

$$\begin{aligned} D^1(S_i) &= e_i G_i^{e_i-1} D^1(G_i(\mathbf{y}, z, \mathbf{x})) = G_i^{e_i-1} Q_i^{(1)}, \\ D^2(S_i) &= D^1(G_i^{e_i-1} Q_i^{(1)}) \\ &= (e_i - 1) G_i^{e_i-2} D^1(G_i(\mathbf{y}, z, \mathbf{x})) Q_i^{(1)} + G_i^{e_i-1} D^1(Q_i^{(1)}) \\ &= G_i^{e_i-2} Q_i^{(2)}, \end{aligned}$$

where $Q_i^{(1)}$ is a polynomial such that $\deg_z Q_i^{(1)} \leq \delta - 1$, and total degree in $\mathbf{y}, \mathbf{x}$, $\deg_{\mathbf{y}, \mathbf{x}} Q_i^{(1)} \leq \delta$ and $Q_i^{(2)}$ is a polynomial with $\deg_z Q_i^{(2)} \leq 2(\delta - 1)$, and $\deg_{\mathbf{y}, \mathbf{x}} Q_i^{(2)} \leq 2\delta$. Inductively, we write

$$D^j(S_i) = G_i^{e_i-j} Q_i^{(j)} \quad (28)$$

where $Q_i^{(j)}$ is a polynomial such that $\deg_z Q_i^{(j)} \leq j(\delta - 1)$, and $\deg_{\mathbf{y}, \mathbf{x}} Q_i^{(j)} \leq j\delta$ for $1 \leq j \leq r$. Hence we have

$$M' = P'M$$

where $P' = \text{diag}(G_1^{e_1-r}, G_2^{e_2-r}, \dots, G_r^{e_r-r})$ an invertible diagonal matrix of size r and, and

$$M = \begin{pmatrix} G_1^r & G_1^{r-1}Q_1^{(1)} & \cdots & G_1Q_1^{(r-1)} & Q_1^{(r)} \\ G_2^r & G_2^{r-1}Q_2^{(1)} & \cdots & G_2Q_2^{(r-1)} & Q_2^{(r)} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ G_r^r & G_r^{r-1}Q_r^{(1)} & \cdots & G_rQ_r^{(r-1)} & Q_r^{(r)} \end{pmatrix}$$

where the degree of each entry of M is bounded above by $2k\delta$. Moreover, for any vector A , we have $M'A = 0$ if and only if $MA = 0$, since P' is invertible. By [Lemma 3.12](#), applied over the field $\mathbb{F}(\epsilon, \mathbf{y}, z, \mathbf{x})$, we have $MA = 0$ where $A = (A_0, \dots, A_r) \in \mathbb{F}(\epsilon)[\mathbf{y}, z, \mathbf{x}]^{r+1}$ with $A_j = (-1)^j \det(M_{j+1})$ where M_j denotes the matrix obtained by removing j^{th} -column of the matrix M for $0 \leq j \leq r$. Note that

$$A_r = (-1)^r \det(M_{r+1}) = \frac{(-1)^r}{\det(P')} \cdot W(S_1, \dots, S_r).$$

Since $S_1, \dots, S_r$ are linearly independent set over $\mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$, we get $W(S_1, \dots, S_r) \neq 0$. Moreover, $\det(P') \neq 0$. Thus $A \neq 0$. Since each entry of M has degree at most $2k\delta$, each A_j is a polynomial of degree at most $2k^2\delta$. By considering $M'A = 0$, we obtain $\sum_{j=0}^r A_j D^j(S_i) = 0$ for $1 \leq i \leq r$. For $r < i \leq k$, we can write $S_i = \sum_{a=1}^r c_{ia} S_a$ with $c_{ia} \in \mathbb{F}(\epsilon, \mathbf{y}, \mathbf{x})$; as the c_{ia} do not depend on z , the polynomial S_i satisfies the same equation. Summing over all $i \in [k]$, we obtain

$$\sum_{j=0}^r A_j(\mathbf{y}, \epsilon, z, \mathbf{x}) D^j(\Phi_{\mathbf{y}}(f_{\epsilon})) = 0. \quad (29)$$

By [Lemma 6.1](#), applied to $S_1, \dots, S_r$, the minor $W(S_1, \dots, S_r)$ has the least ϵ -valuation among the maximal minors of M' . As $M' = P'M$, each maximal minor of M' is $\det(P')$ times the corresponding maximal minor of M . Hence

$$m := \min\{\text{val}_{\epsilon}(A_j) \mid 0 \leq j \leq r\} = \text{val}_{\epsilon}(A_r).$$

Hence $\lim_{\epsilon \rightarrow 0} \epsilon^{-m} A_r \neq 0$, and $\lim_{\epsilon \rightarrow 0} \epsilon^{-m} A_j$ exists for $0 \leq j \leq r$. Normalize Equation (29) by multiplying ϵ^{-m} and further taking $\lim_{\epsilon \rightarrow 0}$ we obtain

$$\sum_{j=0}^r b_j(\mathbf{y}, z, \mathbf{x}) D^j F(\mathbf{y}, z, \mathbf{x}) = 0, \quad (30)$$

where $b_j(\mathbf{y}, z, \mathbf{x}) = \lim_{\epsilon \rightarrow 0} \epsilon^{-m} A_j$ for $0 \leq j \leq r$. Write

$$W(S_1, \dots, S_r) = \epsilon^{m'} (W_0(\mathbf{y}, z, \mathbf{x}) + \epsilon W')$$

for some ϵ -free polynomial $W_0(\mathbf{y}, z, \mathbf{x}) \neq 0$ and some W' with $\text{val}_{\epsilon}(W') \geq 0$. Similarly, write $P_i = \epsilon^{v_i}(\bar{P}_i + O(\epsilon))$ with $0 \neq \bar{P}_i \in \mathbb{F}[\mathbf{x}]$. Since $A_r = (-1)^r W(S_1, \dots, S_r) \prod_{i=1}^r G_i^{r-e_i}$ and $G_i = \Phi_{\mathbf{y}}(P_i)$, comparing the lowest powers of ϵ gives

$$b_r(\mathbf{y}, z, \mathbf{x}) = (-1)^r W_0(\mathbf{y}, z, \mathbf{x}) \prod_{i=1}^r \Phi_{\mathbf{y}}(\bar{P}_i)^{r-e_i} \neq 0. \quad (31)$$

Since $\deg A_j(\mathbf{y}, \epsilon, z, \mathbf{x}) \leq 2k^2\delta$, we have $\deg b_j(\mathbf{y}, z, \mathbf{x}) \leq 2k^2\delta$. For $0 \leq j \leq r$, write

$$b_j(\mathbf{y}, z, \mathbf{x}) = \sum_{i=0}^{2k^2\delta} b_{ji}(\mathbf{y}, \mathbf{x}) z^i$$

as a polynomial in z , where $b_{ji}(\mathbf{y}, \mathbf{x}) \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$. Since $\deg f \leq \delta e$, we can write $F(\mathbf{y}, z, \mathbf{x}) = \sum_{j=0}^{\delta e} F_j(\mathbf{y}, \mathbf{x}) z^j$. If $\delta e < r$, there is nothing more to prove, so assume that $\delta e \geq r$. By comparing the coefficients of $z^0, \dots, z^{\delta e - r}$ in Equation (30) we obtain

$$B(\mathbf{y}, \mathbf{x}) \begin{bmatrix} F_r & F_{r+1} & \dots & F_{\delta e} \end{bmatrix}^T = R(\mathbf{y}, \mathbf{x})$$

where $B(\mathbf{y}, \mathbf{x}) = [B_{pq}]_{1 \leq p, q \leq \delta e - r + 1}$ is a matrix of size $\delta e - r + 1$ with

$$B_{pq} = \begin{cases} \frac{(r+p-1)!}{(p-1)!} b_{r0}(\mathbf{y}, \mathbf{x}) & \text{if } p = q \\ 0 & \text{if } p < q, \end{cases}$$

and if $p > q$, B_{pq} is a linear polynomial in variables $b_{ji}(\mathbf{y}, \mathbf{x})$, and each entry of $R(\mathbf{y}, \mathbf{x})$ is a quadratic polynomial in variables $F_0, \dots, F_{r-1}$ and $b_{ji}(\mathbf{y}, \mathbf{x})$ for $0 \leq j \leq r$ and $0 \leq i \leq 2k^2\delta$. By Equation (31) we have $W_0(\mathbf{y}, z, \mathbf{x}) \neq 0$, hence $W_0(\mathbf{y}, 0, \mathbf{x}) \neq 0$ by Lemma 6.2. As $\Phi_{\mathbf{y}}(\bar{P}_i)$ becomes $\bar{P}_i(\mathbf{y}) \neq 0$ at $z = 0$, Equation (31) gives

$$b_{r0} = b_r(\mathbf{y}, 0, \mathbf{x}) = (-1)^r W_0(\mathbf{y}, 0, \mathbf{x}) \prod_{i=1}^r \bar{P}_i(\mathbf{y})^{r-e_i} \neq 0.$$

Therefore

$$\det(B(\mathbf{y}, \mathbf{x})) = c b_{r0}(\mathbf{y}, \mathbf{x})^{\delta e - r + 1} \neq 0,$$

for some $c \in \mathbb{F} \setminus \{0\}$. Let H_l be the matrix obtained by replacing l^{th} -column of $B(\mathbf{y}, \mathbf{x})$ by $R(\mathbf{y}, \mathbf{x})$. By Cramer's rule Lemma 3.11 we obtain

$$F_{r-1+l}(\mathbf{y}, \mathbf{x}) = \frac{c^{-1} \det(H_l)}{b_{r0}(\mathbf{y}, \mathbf{x})^{\delta e - r + 1}}$$

for $1 \leq l \leq \delta e - r + 1$.

Moreover, if we consider b_{ji} for $0 \leq j \leq r$, and $0 \leq i \leq 2k^2\delta$ and $F_0, \dots, F_{r-1}$ as parameters, then $c^{-1} \det(H_l)$ is a polynomial of degree at most $\delta e - r + 2$ in them, since the entries of $B(\mathbf{y}, \mathbf{x})$ are linear and those of $R(\mathbf{y}, \mathbf{x})$ are quadratic in the parameters. Hence there exists a polynomial $Q_l(\mathbf{w})$ in $(r+1)(2k^2\delta+1) + r$ variables $\mathbf{w}$ such that, after substituting all b_{ji} for $0 \leq j \leq r$, and $0 \leq i \leq 2k^2\delta$ and $F_0, \dots, F_{r-1}$, we obtain

$$Q_l(\mathbf{b}) = c^{-1} \det(H_l),$$

where $\mathbf{b}$ denotes the tuple of parameters $\mathbf{b} := (b_{00}, \dots, b_{0,2k^2\delta}, \dots, b_{r0}, \dots, b_{r,2k^2\delta}, F_0, \dots, F_{r-1})$. Hence

$$F_{r-1+l}(\mathbf{y}, \mathbf{x}) \cdot b_{r0}(\mathbf{y}, \mathbf{x})^{\delta e - r + 1} = Q_l(\mathbf{b}),$$

for $1 \leq l \leq \delta e - r + 1$.

□

We now prove the main result of this section.

Theorem 6.4 (Debordering powers of bounded-degree polynomials). *Let $k, \delta \geq 1$ and let $f \in \Sigma \wedge \Sigma \Pi(k, \delta, s)$. Then f can be computed by a depth-6 formula of size $s^{O(k^3 \delta)}$.*

Proof. We may assume that $f \neq 0$. Write $f_\epsilon = \sum_{i=1}^k P_i^{e_i} = f + O(\epsilon)$ as in Proposition 6.3, with $\deg P_i \leq \delta$, total size at most s , and every $P_i^{e_i}$ nonzero. Let $e := \max_i e_i$ and $d := \deg f$; then $e \leq s$ and $d \leq \delta e$. We may assume that $n \leq s$. Recall that there are $\binom{a+b}{b}$ monomials of degree at most a in b variables, and that $\binom{a+b}{b} \leq (a+b)^b/b! \leq (3(a+b)/b)^b$ for $b \geq 1$. In particular, as $n \leq s$, every polynomial of degree at most $2k^2\delta$ in $\mathbf{x}$ has sparsity at most

$$\sigma := \binom{n + 2k^2\delta}{2k^2\delta} \leq s^{O(k^2\delta)}.$$

Let $r, F_0, \dots, F_{\delta e}, b_{00}, \dots, b_{r, 2k^2\delta}, \mathbf{b}$ and $Q_1, \dots, Q_{\delta e - r + 1}$ be as defined in Proposition 6.3. If $\delta e < r$, then $\deg f < k$, so f is a sum of at most σ monomials, and we are done. Hence assume that $\delta e \geq r$, and put $N := \delta e - r + 1$, so that $1 \leq N \leq \delta s$. Then Proposition 6.3 gives

$$F_{r-1+l} \cdot b_{r0}^N = Q_l(\mathbf{b}) \quad \text{for } 1 \leq l \leq N, \quad (32)$$

where each Q_l is a polynomial of degree at most $N + 1$ in p variables, one for each entry of $\mathbf{b}$, and

$$p := (r + 1)(2k^2\delta + 1) + r \leq O(k^3\delta).$$

Since $N + 1 \leq \delta e + 1 \leq ep$ (as $p > \delta$ and $e \geq 1$), the sparsity of each Q_l is at most

$$\binom{(N + 1) + p}{p} \leq (3(e + 1))^p \leq s^{O(p)}.$$

We will substitute sparse representations of the entries of $\mathbf{b}$ into these monomials, to obtain a $\Sigma \Pi \Sigma \Pi$ formula over $\mathbf{x}$. To obtain coefficients in $\mathbb{F}$ rather than in $\mathbb{F}[\mathbf{y}]$, we first fix $\mathbf{y}$.

Fixing $\mathbf{y}$. Since $b_{r0} = b_r(\mathbf{y}, 0, \mathbf{x})$ is nonzero and $\mathbb{F}$ is infinite, there are $\alpha, \mathbf{a} \in \mathbb{F}^n$ with $b_{r0}(\alpha, \mathbf{a}) \neq 0$. Put

$$\beta(\mathbf{x}) := b_{r0}(\alpha, \mathbf{x}) \quad \text{and} \quad c := \beta(\mathbf{a}) \neq 0.$$

As $F(\alpha, 1, \mathbf{x}) = f(\alpha + \mathbf{x})$, we have $f(\alpha + \mathbf{x}) = \sum_{j=0}^{\delta e} F_j(\alpha, \mathbf{x})$. Substituting $\mathbf{y} = \alpha$ in (32) therefore gives

$$\Gamma(\mathbf{x}) := \beta(\mathbf{x})^N f(\alpha + \mathbf{x}) = \sum_{j=0}^{r-1} \beta(\mathbf{x})^N F_j(\alpha, \mathbf{x}) + \sum_{l=1}^N Q_l(\mathbf{b}(\alpha, \mathbf{x})). \quad (33)$$

From Proposition 6.3, we have $\deg(b_{ji}) \leq 2k^2\delta$. We also have $f(\mathbf{y} + z\mathbf{x}) = \sum_{j=0}^{\delta e} F_j(\mathbf{y}, \mathbf{x})z^j$, and therefore F_j is homogeneous of degree j in $\mathbf{x}$; in particular, $F_0, \dots, F_{r-1}$ have degree less than k in $\mathbf{x}$. Hence every entry of $\mathbf{b}(\alpha, \mathbf{x})$ has degree at most $2k^2\delta$, and this remains true after any affine linear substitution $\mathbf{x} \mapsto \pi(\mathbf{x})$. So every entry of $\mathbf{b}(\alpha, \pi(\mathbf{x}))$ has sparsity at most σ , and thus a $\Sigma \Pi$ formula of size $O(k^2\delta\sigma)$. We combine these. Substituting $\pi(\mathbf{x})$ for $\mathbf{x}$ in (33) and writing each Q_l as a sum of monomials, we obtain a $\Sigma \Pi \Sigma \Pi$ formula for $\Gamma(\pi(\mathbf{x}))$ with top fan-in at most $r + Ns^{O(p)}$, in which every summand is a constant times a product of at most $N + 1$ entries of $\mathbf{b}(\alpha, \pi(\mathbf{x}))$. Its size is

$$O((r + Ns^{O(p)})(N + 1)k^2\delta\sigma).$$

Division by β^N . Let $g(\mathbf{x}) := 1 - c^{-1}\beta(\mathbf{x} + \mathbf{a})$, so that $\beta(\mathbf{x} + \mathbf{a}) = c(1 - g(\mathbf{x}))$ and $g(\mathbf{0}) = 0$. Truncating the binomial series $(1 - Y)^{-N} = \sum_{u \geq 0} \binom{N+u-1}{u} Y^u$ after Y^d gives a polynomial $\eta \in \mathbb{F}[Y]$ with

$$(1 - Y)^N \sum_{u=0}^d \binom{N+u-1}{u} Y^u = 1 + Y^{d+1}\eta(Y).$$

We substitute $Y = g$ and multiply by $f(\alpha + \mathbf{a} + \mathbf{x})$. Since $\Gamma(\mathbf{x} + \mathbf{a}) = c^N(1 - g)^N f(\alpha + \mathbf{a} + \mathbf{x})$, this gives

$$\Psi(\mathbf{x}) := c^{-N} \Gamma(\mathbf{x} + \mathbf{a}) \sum_{u=0}^d \binom{N+u-1}{u} g(\mathbf{x})^u = f(\alpha + \mathbf{a} + \mathbf{x}) \left(1 + g(\mathbf{x})^{d+1}\eta(g(\mathbf{x}))\right).$$

As $g(\mathbf{0}) = 0$, every monomial of $g^{d+1}\eta(g) f(\alpha + \mathbf{a} + \mathbf{x})$ has degree larger than d . Since $\deg f = d$, the polynomial $f(\alpha + \mathbf{a} + \mathbf{x})$ is therefore the sum of the homogeneous components of Ψ of degree at most d .

Interpolation. Let $\kappa := \deg \Psi$, let Ψ_i be the homogeneous component of Ψ of degree i , and fix distinct $\lambda_0, \dots, \lambda_\kappa \in \mathbb{F}$. Since $\Psi(\lambda_t \mathbf{x}) = \sum_{i=0}^\kappa \lambda_t^i \Psi_i(\mathbf{x})$ and the Vandermonde matrix $(\lambda_t^i)_{t,i}$ is invertible, there are $\gamma_0, \dots, \gamma_\kappa \in \mathbb{F}$ with $\sum_{i=0}^\kappa \gamma_i \Psi_i = \sum_{t=0}^\kappa \gamma_t \Psi(\lambda_t \mathbf{x})$. Substituting $\mathbf{x} - \alpha - \mathbf{a}$ for $\mathbf{x}$ undoes the shifts, and we obtain

$$f(\mathbf{x}) = \sum_{t=0}^\kappa \gamma_t c^{-N} \Gamma(\pi_t(\mathbf{x})) \cdot \sum_{u=0}^d \binom{N+u-1}{u} \left(1 - c^{-1}\beta(\pi_t(\mathbf{x}))\right)^u, \quad (34)$$

where $\pi_t(\mathbf{x}) := \lambda_t(\mathbf{x} - \alpha - \mathbf{a}) + \mathbf{a}$.

The formula. As shown above, $\Gamma(\pi_t(\mathbf{x}))$ has a $\Sigma\Pi\Sigma\Pi$ formula of size $O((r + Ns^{O(p)})(N+1)k^2\delta\sigma)$. As $\deg \beta \leq 2k^2\delta$, the second factor in (34) has a $\Sigma\Pi\Sigma\Pi$ formula of size $O((d+1)^2k^2\delta\sigma)$. Hence (34) is a $\Sigma\Pi\Sigma\Pi\Sigma\Pi$ formula for f , that is, a formula of depth six. Finally, $r \leq k$, $N, d \leq \delta s$, and $\kappa \leq \deg \Gamma + d \deg g \leq 2k^2\delta(N+d) + d \leq O(k^2\delta^2s)$. So the size of this formula is at most

$$\text{poly}(k, \delta, s) \cdot s^{O(p)} \sigma \leq \text{poly}(k, \delta, s) \cdot s^{O(k^3\delta) + O(k^2\delta)} = s^{O(k^3\delta)}. \quad \square$$

7 An improved lower bound for border depth-3 circuits

For concreteness, we work over $\mathbb{C}$ in this section; the argument only uses that the field has characteristic zero, so [Theorem 1.4](#) holds over every such field. Write $\Sigma \wedge \Sigma(r, e)$ for the polynomials of the form $\sum_{j=1}^r c_j \ell_j^{e_j}$, where the ℓ_j are affine linear, $c_j \in \mathbb{C}$, and $0 \leq e_j \leq e$. Thus r bounds the top fan-in and e the formal degree; in particular, such an expression has size $O(n \cdot e \cdot r)$. We write $\overline{\Sigma \wedge \Sigma(r, e)}$ for its border: a polynomial $g \in \mathbb{C}[\mathbf{x}]$ lies in $\overline{\Sigma \wedge \Sigma(r, e)}$ if, for some $S \in \mathbb{C}[\epsilon, \mathbf{x}]$, the polynomial $g + \epsilon S$ has an expression as above with $c_j \in \mathbb{C}(\epsilon)$ and ℓ_j affine linear over $\mathbb{C}(\epsilon)$.

7.1 Structural Results for Border Depth-3 circuits

In this section, we will develop the structural ideas for homogeneous polynomials in $\overline{\Sigma^m \Pi \Sigma}$. The next proposition shows that if a homogeneous polynomial f is written as a border presentation using the sum of products of linear polynomials, with each linear polynomial having a non-zero constant term, then the following result shows that such f is a quotient of two polynomials of small border circuits.

Proposition 7.1 (A quotient of border diagonal circuits). *Let $0 \neq f \in \mathbb{C}[\mathbf{x}]$ be homogeneous of degree $d \geq 1$. Suppose that, for integers $m \geq 1$ and $M, a_i, d_i \geq 0$,*

$$\sum_{i=1}^m \epsilon^{a_i} T_i(\epsilon, \mathbf{x}) = \epsilon^M f(\mathbf{x}) + \epsilon^{M+1} E(\epsilon, \mathbf{x}), \quad (35)$$

where $E \in \mathbb{C}[\epsilon, \mathbf{x}]$ and

$$T_i = \prod_{a=1}^{d_i} \ell_{ia}(\epsilon, \mathbf{x}).$$

Each $\ell_{ia} \in \mathbb{C}[\epsilon, \mathbf{x}]$ is affine linear in $\mathbf{x}$, and $T_i(0, \mathbf{0}) \neq 0$. Let $\hat{d} := \sum_i d_i$. Then there are nonzero $g, h \in \mathbb{C}[\mathbf{x}]$ such that

$$g = fh, \quad g, h \in \overline{\Sigma \wedge \Sigma(R, \hat{d})}, \quad R := m \hat{d}^{O(m^2)}.$$

Notation. We use $D := \frac{\partial}{\partial z}$ and

$$W(A_1, \dots, A_r) := \det(D^{a-1} A_b)_{a,b \in [r]}.$$

For a Laurent series A in z , with coefficients in a field or in $\mathbb{C}[\mathbf{x}]$, we write $\text{coeff}_{z^j}(A)$ for the coefficient of z^j in A . For $A \neq 0$, let $\text{val}_z(A)$ be its least exponent and let $\text{lc}_z(A) := \text{coeff}_{z^{\text{val}_z(A)}}(A)$. Set $\text{val}_z(0) := \infty$, and write $A = O(z^j)$ if $\text{val}_z(A) \geq j$.

Proof. Consider the $\mathbb{C}$ -algebra homomorphism $\psi : \mathbb{C}[\epsilon, \mathbf{x}] \rightarrow \mathbb{C}[z, \mathbf{x}]$ given by $\epsilon \mapsto z^{d+1}$ and $x_j \mapsto zx_j$. Define

$$V_i(z, \mathbf{x}) := \psi(\epsilon^{a_i} T_i(\epsilon, \mathbf{x})) = z^{a_i(d+1)} T_i(z^{d+1}, z\mathbf{x}), \quad \Delta := M(d+1) + d.$$

Applying ψ to Equation (35) gives

$$\sum_{i=1}^m V_i = z^\Delta f(\mathbf{x}) + z^{\Delta+1} E(z^{d+1}, z\mathbf{x}). \quad (36)$$

Indeed, $\psi(\epsilon^M f) = z^{M(d+1)} f(z\mathbf{x}) = z^\Delta f$, since f is homogeneous of degree d , and $\psi(\epsilon^{M+1} E(\epsilon, \mathbf{x})) = z^{(M+1)(d+1)} E(z^{d+1}, z\mathbf{x})$, where $(M+1)(d+1) = \Delta + 1$ and $E(z^{d+1}, z\mathbf{x}) \in \mathbb{C}[z, \mathbf{x}]$. Hence the error term in (36) has z -valuation at least $\Delta + 1$. Comparing the coefficients of ϵ^M in (35) also shows that $d = \deg f \leq \max_i d_i \leq \hat{d}$; in particular, $\hat{d} \geq 1$.

Moreover, the map ψ preserves the number of affine factors in $\mathbf{x}$. Indeed, a factor $b_0(\epsilon) + \sum_j b_j(\epsilon) x_j$ becomes

$$b_0(z^{d+1}) + \sum_j z b_j(z^{d+1}) x_j.$$

Its value at $z = 0$ is the scalar $b_0(0)$, which is nonzero since $T_i(0, \mathbf{0}) \neq 0$. Below, z plays two roles. In Lemma 7.2, we regard the V_i as polynomials in z over the field $\mathbb{K} := \mathbb{C}(\mathbf{x})$. In Lemma 7.3, we regard them as products of affine linear polynomials in $\mathbf{x}$ whose coefficients are polynomials in z , and z serves as the parameter of the border.

For $1 \leq i \leq m$, let $\widetilde{V}_i$ denote the image of V_i in $\mathbb{K}[z]/(z^{\Delta+1})$. By (36), $\sum_i \widetilde{V}_i = f z^\Delta$ with $f \in \mathbb{K}^\times$, so the span of $\widetilde{V}_1, \dots, \widetilde{V}_m$ contains $\widetilde{z}^\Delta$. Let r be the dimension of this span, so that $1 \leq r \leq$

m . Extending $\{\widetilde{z^\Delta}\}$ to a basis by elements of the spanning set, we can choose $\{U_1, \dots, U_{r-1}\} \subseteq \{V_1, \dots, V_m\}$ such that $\{\widetilde{U}_1, \dots, \widetilde{U}_{r-1}, \widetilde{z^\Delta}\}$ is a basis of the span of $\{\widetilde{V}_1, \dots, \widetilde{V}_m\}$. Define

$$H := W(U_1, \dots, U_{r-1}, z^\Delta), \quad W_i := W(U_1, \dots, U_{r-1}, V_i).$$

[Lemma 7.2](#) proves that $H \neq 0$ and, for $\nu := \text{val}_z(H)$, that $\text{val}_z(W_i) \geq \nu$ for all $i \in [m]$ and

$$\sum_{i=1}^m W_i = fH + O(z^{\nu+1}). \quad (37)$$

Since $V_i \in \mathbb{C}[\mathbf{x}][z]$, the Wronskians H and W_i lie in $\mathbb{C}[\mathbf{x}][z]$. Let

$$g_i := \text{coeff}_{z^\nu}(W_i), \quad g := \sum_{i=1}^m g_i, \quad h := \text{coeff}_{z^\nu}(H) = \text{lc}_z(H).$$

Then $g_i, h \in \mathbb{C}[\mathbf{x}]$ and $h \neq 0$. Comparing the coefficients of z^ν in (37), and using that f does not depend on z , we obtain

$$g = fh;$$

hence $g \neq 0$. Moreover, as $\text{val}_z(W_i) \geq \nu$, each g_i is either zero or equal to $\text{lc}_z(W_i)$.

It remains to bound the complexity of h and of the nonzero g_i . Each V_i has the form required in [Lemma 7.3](#), with $c_i(z) = z^{a_i(d+1)}$ and the factors $\ell_{ia}(z^{d+1}, z\mathbf{x})$; so does z^Δ , with no affine factors. If $W_i \neq 0$, its columns are pairwise distinct, since a repeated column makes the determinant vanish; hence they are r distinct products among $V_1, \dots, V_m$, which together contain at most $\hat{d}$ affine factors. The same bound applies to H , whose last column z^Δ has no affine factors. As $\hat{d} \geq 1$, [Lemma 7.3](#) now gives

$$h, \text{lc}_z(W_i) \in \overline{\Sigma \wedge \Sigma(\hat{d}^{O(r^2)}, \hat{d})} \quad (W_i \neq 0).$$

Since g is a sum of at most m such leading coefficients and $r \leq m$, we obtain $g, h \in \overline{\Sigma \wedge \Sigma(R, \hat{d})}$ with $R = m(\hat{d} + 1)^{O(m^2)}$. $\square$

We now justify the Wronskian arguments used in the preceding proof. The following is useful to trace back f as a leading coefficient in a sum of Wronskians.

Lemma 7.2 (Truncated independence and Wronskians). *Let $\mathbb{K}$ be a field of characteristic zero, let $\Delta \geq 1$, and let $V_1, \dots, V_m \in \mathbb{K}[z]$ satisfy*

$$\sum_i V_i = z^\Delta f + O(z^{\Delta+1}), \quad f \in \mathbb{K}^\times.$$

For $F \in \mathbb{K}[z]$, let $\widetilde{F}$ denote its image in $\mathbb{K}[z]/(z^{\Delta+1})$. Suppose that $\{U_1, \dots, U_{r-1}\} \subseteq \{V_1, \dots, V_m\}$ is such that $\{\widetilde{U}_1, \dots, \widetilde{U}_{r-1}, \widetilde{z^\Delta}\}$ is a basis of the span of $\{\widetilde{V}_1, \dots, \widetilde{V}_m\}$ in $\mathbb{K}[z]/(z^{\Delta+1})$. If $H := W(U_1, \dots, U_{r-1}, z^\Delta)$, then $H \neq 0$ in $\mathbb{K}[z]$. Moreover, if $W_i := W(U_1, \dots, U_{r-1}, V_i)$ and $\nu := \text{val}_z(H)$, then

$$\text{val}_z(W_i) \geq \nu \quad (i \in [m]), \quad \sum_i W_i = fH + O(z^{\nu+1}).$$

Consequently, (37) holds.

Proof. If $r = 1$, then $H = z^\Delta$, $\nu = \Delta$, and every $\widetilde{V}_i$ is a multiple of $\widetilde{z}^\Delta$, so all claims are immediate. Therefore, assume that $r \geq 2$. Since $\widetilde{U}_1, \dots, \widetilde{U}_{r-1}, \widetilde{z}^\Delta$ are linearly independent, so are the images of $U_1, \dots, U_{r-1}$ in $\mathbb{K}[z]/(z^\Delta)$: if $\sum_j \gamma_j U_j \equiv 0 \pmod{z^\Delta}$ with $\gamma_j \in \mathbb{K}$, then $\sum_j \gamma_j \widetilde{U}_j$ is a multiple of $\widetilde{z}^\Delta$, and hence all $\gamma_j = 0$. Let A be the $(r-1) \times \Delta$ matrix whose j -th row is the coefficient vector of U_j with respect to $z^0, \dots, z^{\Delta-1}$; thus $\text{rank}(A) = r-1$. Using Gaussian elimination, obtain the row echelon form $A' = BA$ of A , where $B = (b_{jl}) \in \text{GL}_{r-1}(\mathbb{K})$. All rows of A' are nonzero, and their *pivots*, that is, their leftmost nonzero entries, lie in strictly increasing columns, indexed by $\lambda_1 < \dots < \lambda_{r-1}$.

Let $A_j := \sum_l b_{jl} U_l \in \mathbb{K}[z]$; its coefficients of $z^0, \dots, z^{\Delta-1}$ form the j -th row of A' . Hence the span of $A_1, \dots, A_{r-1}$ equals the span of $U_1, \dots, U_{r-1}$, and

$$A_j = c_j z^{\lambda_j} + O(z^{\lambda_j+1}), \quad c_j \in \mathbb{K}^\times, \quad 0 \leq \lambda_1 < \dots < \lambda_{r-1} < \Delta.$$

Let $(\mu_1, \dots, \mu_r) := (\lambda_1, \dots, \lambda_{r-1}, \Delta)$; these are pairwise distinct. Since the Wronskian is linear in each column and the entries of B do not depend on z ,

$$W(A_1, \dots, A_{r-1}, F) = \det(B) W(U_1, \dots, U_{r-1}, F) \quad \text{for every } F \in \mathbb{K}[z].$$

[Lemma 3.3](#) applied to $A_1, \dots, A_{r-1}, z^\Delta$ we obtain

$$W(A_1, \dots, A_{r-1}, z^\Delta) = \left(\prod_{j=1}^{r-1} c_j \right) V(\mu_1, \dots, \mu_r) z^{\sum_b \mu_b - \binom{r}{2}} + O(z^{\sum_b \mu_b - \binom{r}{2} + 1}),$$

and the leading coefficient is nonzero. Thus $H = \det(B)^{-1} W(A_1, \dots, A_{r-1}, z^\Delta) \neq 0$, and $\nu = \text{val}_z(H)$ satisfies

$$\nu = \Delta + \sum_{j=1}^{r-1} \lambda_j - \binom{r}{2}. \quad (38)$$

Since $\widetilde{V}_i \in \text{span}(\widetilde{U}_1, \dots, \widetilde{U}_{r-1}, \widetilde{z}^\Delta)$ for all $1 \leq i \leq m$, we can write

$$\widetilde{V}_i = \sum_{j=1}^{r-1} \alpha_{ij} \widetilde{U}_j + \beta_i \widetilde{z}^\Delta$$

for some $\alpha_{ij}, \beta_i \in \mathbb{K}$. Hence

$$V_i = \sum_{j=1}^{r-1} \alpha_{ij} U_j + \beta_i z^\Delta + E_i, \quad \alpha_{ij}, \beta_i \in \mathbb{K}, \quad E_i = O(z^{\Delta+1}). \quad (39)$$

We now show that $\text{val}_z(W_i) \geq \nu$. Since the determinant is an alternating multilinear function of the columns, Equation (39) gives

$$\begin{aligned} W_i &= W(U_1, \dots, U_{r-1}, V_i) = W(U_1, \dots, U_{r-1}, \beta_i z^\Delta + E_i) \\ &= \beta_i H + W(U_1, \dots, U_{r-1}, E_i). \end{aligned}$$

By Equation (7), applied to $A_1, \dots, A_{r-1}, E_i$, and since $\text{val}_z(E_i) \geq \Delta + 1$,

$$\text{val}_z W(U_1, \dots, U_{r-1}, E_i) = \text{val}_z W(A_1, \dots, A_{r-1}, E_i) \geq \Delta + 1 + \sum_{j=1}^{r-1} \lambda_j - \binom{r}{2} = \nu + 1.$$

Hence $\text{val}_z(W_i) \geq \nu$ for $1 \leq i \leq m$.

Finally, write $\sum_i V_i = z^\Delta f + E'$ with $E' = O(z^{\Delta+1})$. By linearity in the last column,

$$\begin{aligned} \sum_{i=1}^m W_i &= W\left(U_1, \dots, U_{r-1}, \sum_{i=1}^m V_i\right) = W(U_1, \dots, U_{r-1}, z^\Delta f + E') \\ &= fW(U_1, \dots, U_{r-1}, z^\Delta) + W(U_1, \dots, U_{r-1}, E'). \end{aligned}$$

As above, $\text{val}_z W(U_1, \dots, U_{r-1}, E') \geq \nu + 1$. Therefore

$$\sum_{i=1}^m W_i = fH + O(z^{\nu+1}). \quad \square$$

After we obtain f as a leading coefficient of the sum of Wronskians in [Lemma 7.2](#), we now show that the leading coefficient of the Wronskian has a small border circuit.

Lemma 7.3 (Leading coefficients of Wronskians). *Let*

$$V_i = c_i(z) \prod_{p=1}^{d_i} \ell_{ip}(z, \mathbf{x}), \quad i \in [r],$$

where $c_i \in \mathbb{C}(z)^\times$. Each ℓ_{ip} is affine linear in $\mathbf{x}$, its coefficients are rational functions of z that are regular at $z = 0$, and $\ell_{ip}(0, \mathbf{x}) \in \mathbb{C}^\times$. Let $\hat{d} \geq 1$ be an integer with $\hat{d} \geq \sum_i d_i$. If $W(V_1, \dots, V_r) \neq 0$, then, regarding $W(V_1, \dots, V_r)$ as a Laurent series in z with coefficients in $\mathbb{C}[\mathbf{x}]$,

$$\text{lc}_z W(V_1, \dots, V_r) \in \overline{\Sigma \wedge \Sigma(\hat{d}^{O(r^2)}, \hat{d})},$$

where the border is taken with respect to the parameter z . The bound does not depend on the degrees or valuations in z of the c_i and of the coefficients of the ℓ_{ip} .

Proof. We expand rational functions of z at $z = 0$ and work in the ring $\mathbb{C}[\mathbf{x}]((z))$ of Laurent series in z with coefficients in $\mathbb{C}[\mathbf{x}]$, on which D acts termwise.

Normalization. Write $\ell_{ip}(z, \mathbf{x}) = \ell_{ip}(z, \mathbf{0})(1 + \widehat{\ell}_{ip})$. Since $\ell_{ip}(0, \mathbf{0}) \neq 0$, the scalar $\ell_{ip}(z, \mathbf{0})$ is a unit of $\mathbb{C}[[z]]$; and since $\ell_{ip}(0, \mathbf{x})$ is constant, $\widehat{\ell}_{ip}$ is a homogeneous linear form in $\mathbf{x}$ with coefficients in $z\mathbb{C}[[z]]$. Absorbing the scalars $\ell_{ip}(z, \mathbf{0})$ into c_i , we write $V_i = c_i(z)P_i$, where now $c_i \in \mathbb{C}((z))^\times$ and $P_i := \prod_p (1 + \widehat{\ell}_{ip})$. As the coefficients of $\widehat{\ell}_{ip}$ have positive z -valuation, $(1 + \widehat{\ell}_{ip})^{-1} = \sum_{e \geq 0} (-\widehat{\ell}_{ip})^e$ in $\mathbb{C}[\mathbf{x}][[z]]$. Consider

$$Q := \frac{W(V_1, \dots, V_r)}{\prod_i V_i} = \det \left(\frac{D^{a-1} V_b}{V_b} \right)_{a,b \in [r]} \in \mathbb{C}[\mathbf{x}]((z)).$$

Entries. By the Leibniz rule, for every $a \geq 0$,

$$\frac{D^a V_b}{V_b} = \sum_{j=0}^a \binom{a}{j} \frac{D^{a-j} c_b}{c_b} \cdot \frac{D^j P_b}{P_b} \quad \text{and} \quad \frac{D^j P_b}{P_b} = \sum_{j_1 + \dots + j_{d_b} = j} \binom{j}{j_1, \dots, j_{d_b}} \prod_{\{p: j_p \geq 1\}} \frac{D^{j_p} \widehat{\ell}_{bp}}{1 + \widehat{\ell}_{bp}}.$$

Here $D^{a-j} c_b / c_b \in \mathbb{C}((z))$, and each $D^{j_p} \widehat{\ell}_{bp}$ is a homogeneous linear form in $\mathbf{x}$ with coefficients in $\mathbb{C}[[z]]$. For $d_b \geq 1$, the number of tuples $(j_1, \dots, j_{d_b})$ with sum j is $\binom{j+d_b-1}{j} \leq d_b^j$. Hence, also when

$d_b = 0$, the entry $D^a V_b / V_b$ is a $\mathbb{C}((z))$ -linear combination of at most $\sum_{j \leq a} d_b^j \leq (d_b + 1)^a$ products of the form

$$\prod_{t=1}^q \frac{L_t}{1 + \widehat{\ell}_t}, \quad q \leq a, \quad (40)$$

where each L_t is a homogeneous linear form in $\mathbf{x}$ with coefficients in $\mathbb{C}[[z]]$, and each $\widehat{\ell}_t$ is one of the $\widehat{\ell}_{bp}$.

Determinant. Each of the $r!$ terms of the determinant Q is a product of entries $D^{a-1} V_b / V_b$, one from each column $b \in [r]$, whose derivative orders $a - 1$ are distinct and add up to $\binom{r}{2}$. Since $d_b \leq \hat{d}$, the previous step shows that Q is a $\mathbb{C}((z))$ -linear combination of at most $r! (\hat{d} + 1)^{\binom{r}{2}}$ products of the form (40) with $q \leq \binom{r}{2}$.

Truncation in $\mathbf{x}$. Expanding the factors $(1 + \widehat{\ell}_{bp})^{-1}$ produces monomials in $\mathbf{x}$ of arbitrarily high degree, so we have to truncate. In Proposition 4.2, we truncated in z . This is *not possible* here: we need the coefficient of z^w , where $w := \text{val}_z(Q)$, and w is not bounded in terms of $\hat{d}$ (in the proof of Proposition 7.1, it can grow with the exponent M in (35)), so a truncation in z would give a bound depending on w . Instead, we truncate in the degree in $\mathbf{x}$. For $A \in \mathbb{C}[\mathbf{x}]((z))$, let $\tau(A)$ be obtained from A by deleting, in every coefficient, the monomials in $\mathbf{x}$ of degree larger than $\hat{d}$. Thus $\tau(A)$ is a polynomial in $\mathbf{x}$ of degree at most $\hat{d}$ with coefficients in $\mathbb{C}((z))$, and τ is $\mathbb{C}((z))$ -linear.

This truncation does not change the coefficient we need. Indeed, since D does not increase the degree in $\mathbf{x}$, every coefficient of $W(V_1, \dots, V_r)$ has degree at most $\sum_i d_i \leq \hat{d}$ in $\mathbf{x}$. Moreover, $P_b = 1 + O(z)$, so the leading coefficient of $\prod_b V_b = \prod_b c_b \cdot \prod_b P_b$ is a nonzero constant. As $W(V_1, \dots, V_r) = Q \prod_b V_b$ is nonzero, $\text{lc}_z W(V_1, \dots, V_r)$ is a nonzero constant multiple of $\text{lc}_z(Q)$; in particular, $\text{lc}_z(Q)$ has degree at most $\hat{d}$ in $\mathbf{x}$. Since τ acts coefficientwise, $\text{val}_z(\tau(Q)) = w$ and $\text{lc}_z(\tau(Q)) = \text{lc}_z(Q)$.

Expansion. Fix a product of the form (40). Expanding each factor $(1 + \widehat{\ell}_t)^{-1}$ as a geometric series, we write the product as the sum, over all $(e_1, \dots, e_q) \in \mathbb{Z}_{\geq 0}^q$, of the products $\prod_{t=1}^q L_t (-\widehat{\ell}_t)^{e_t}$. Since the L_t and $\widehat{\ell}_t$ are homogeneous linear forms in $\mathbf{x}$, the product indexed by $(e_1, \dots, e_q)$ is homogeneous of degree $q + e_1 + \dots + e_q$ in $\mathbf{x}$. Hence τ keeps exactly the products of degree at most $\hat{d}$:

$$\tau\left(\prod_{t=1}^q \frac{L_t}{1 + \widehat{\ell}_t}\right) = \sum_{\substack{e_1, \dots, e_q \geq 0 \\ q + e_1 + \dots + e_q \leq \hat{d}}} \prod_{t=1}^q L_t (-\widehat{\ell}_t)^{e_t}.$$

This sum has at most $(\hat{d} + 1)^q$ terms, and each term is a product of powers of at most $2q$ linear forms, of total degree at most $\hat{d}$.

Sums of powers. All coefficients that occur so far are rational functions of z : they are obtained from the rational functions in the statement of the lemma by field operations and by D . Hence, by Lemma 3.5 over the field $\mathbb{C}(z)$, each term in the expansion above is a $\mathbb{C}(z)$ -linear combination of at most $(\hat{d} + 1)^{2q}$ powers of linear forms over $\mathbb{C}(z)$, with exponents at most $\hat{d}$. Since τ is $\mathbb{C}((z))$ -linear, combining this with the determinant step, and using $q \leq \binom{r}{2}$ and $r! \leq 2^{r^2}$, we find that $\tau(Q)$ is a $\mathbb{C}(z)$ -linear combination of $\hat{d}^{O(r^2)}$ powers of linear forms over $\mathbb{C}(z)$, with exponents at most $\hat{d}$.

Taking the limit. Now think of $\tau(Q)$ as a polynomial in $\mathbf{x}$ over $\mathbb{C}((z))$. By the previous step, its coefficients are rational functions of z , and by the truncation step, $\text{val}_z(\tau(Q)) = w$ and $\text{lc}_z(\tau(Q)) =$

$\text{lc}_z(Q)$. Hence $z^{-w}\tau(Q)$ is a polynomial in $\mathbf{x}$ whose coefficients have no pole at $z = 0$, and setting $z = 0$ in it gives $\text{lc}_z(Q)$. That is,

$$\text{lc}_z(Q) = \lim_{z \rightarrow 0} z^{-w} \tau(Q),$$

where $z^{-w}\tau(Q)$ is a $\mathbb{C}(z)$ -linear combination of $\hat{d}^{O(r^2)}$ powers of linear forms with exponents at most $\hat{d}$; the individual summands need not have a limit. This is exactly membership in the border, with z as the parameter. (The definition asks for $\text{lc}_z(Q) + zS$ with $S \in \mathbb{C}[z, \mathbf{x}]$; we obtain this form by multiplying $z^{-w}\tau(Q)$ by a common denominator of its coefficients that equals 1 at $z = 0$.) Hence $\text{lc}_z(Q)$, and therefore its constant multiple $\text{lc}_z W(V_1, \dots, V_r)$, lies in $\Sigma \wedge \Sigma(\hat{d}^{O(r^2)}, \hat{d})$. $\square$

7.2 Nonhomogeneous reduction

We now have the tools to prove the lower bound for the sum of $k + 1$ disjoint degree- d monomials.

Theorem 7.4. *Let $d > k \geq 1$, $N := (k + 1)d$ and*

$$P_{k+1,d}(\mathbf{x}) := \sum_{i=1}^{k+1} \prod_{j=1}^d x_{i,j}.$$

If $P_{k+1,d} \in \overline{\Sigma\Pi\Sigma(k, s)}$, then

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

For constant k , this was proved in [DS26], for a larger family of polynomials, with a bound of the form $2^{\Omega(d/\exp(k))}$. We give a self-contained proof, which is shorter and improves the dependence on k .

Before proving the lower bound, we reduce it to the setting of Proposition 7.1: at most k products whose affine factors all have nonzero constant terms, approximating a single monomial. This reduction is essentially that of [DS26, Claims 3–4]. The only simplification is that, instead of keeping the whole restricted hard polynomial, we keep a single block of variables and set all other variables to zero, which leaves the monomial $x_1 \cdots x_d$. This makes the final count of partial derivatives particularly simple.

Lemma 7.5 (Reduction Lemma). *Let $d > k \geq 1$ and*

$$P_{k+1,d} := \sum_{i=1}^{k+1} \prod_{j=1}^d x_{i,j}.$$

Suppose that $P_{k+1,d} \in \overline{\Sigma\Pi\Sigma(k, s)}$ and let $\mathbf{x} = (x_1, \dots, x_d)$ be new variables. Then there exist integers $1 \leq m \leq k$ and $M, a_1, \dots, a_m \geq 0$, and products

$$T_i(\epsilon, \mathbf{x}) = \prod_{j=1}^{d_i} \ell_{ij}(\epsilon, \mathbf{x}), \quad 1 \leq i \leq m,$$

where each $\ell_{ij} \in \mathbb{C}[\epsilon, \mathbf{x}]$ is affine linear in $\mathbf{x}$, such that $T_i(0, \mathbf{0}) \neq 0$ for every i , and

$$\sum_{i=1}^m \epsilon^{a_i} T_i(\epsilon, \mathbf{x}) = \epsilon^M x_1 \cdots x_d + \epsilon^{M+1} E(\epsilon, \mathbf{x})$$

for some $E \in \mathbb{C}[\epsilon, \mathbf{x}]$. Moreover, $\sum_{i=1}^m d_i \leq ks$.

Proof. Let $N = (k+1)d$. Throughout the proof, $O(\epsilon)$ denotes a polynomial in the current variables whose coefficients lie in $\epsilon \mathbb{C}[[\epsilon]]$.

Step 1: Linear restrictions of $P_{k+1,d}$. Let $0 \leq t \leq k$, and let $L : \mathbb{C}^{N-t} \rightarrow \mathbb{C}^N$ be an injective linear map. We claim that a suitable linear restriction turns $P_{k+1,d} \circ L$ into $x_1 \cdots x_d$; in particular, $P_{k+1,d} \circ L \neq 0$. Indeed, the N coordinates of L are linear forms on $\mathbb{C}^{N-t}$ and, since L is injective, they span the space of all such forms. Choose $N-t$ of them that form a basis, and use them as coordinates on $\mathbb{C}^{N-t}$. Exactly t coordinate forms are left out of the basis. Since there are $k+1 > t$ blocks, some block has all of its d forms in the basis. Rename these forms $x_1, \dots, x_d$, and set all other basis forms to zero. Every other block has at most t forms outside the basis, and hence at least $d-t > 0$ forms in the basis, all of which are set to zero; its product therefore vanishes. The product of the chosen block becomes $x_1 \cdots x_d$, which proves the claim.

Step 2: Normalization. By assumption, there are products $\widehat{T}_1, \dots, \widehat{T}_k$ of affine linear polynomials over $\mathbb{C}(\epsilon)$, each with at most s factors, such that

$$\sum_{i=1}^k \widehat{T}_i = P_{k+1,d} + O(\epsilon).$$

We discard the products that are zero. Viewing $\mathbb{C}(\epsilon)$ as a subfield of the field $\mathbb{C}((\epsilon))$ of formal Laurent series in ϵ , we multiply each affine factor by the power of ϵ for which its coefficients lie in $\mathbb{C}[[\epsilon]]$ and its reduction modulo ϵ is nonzero; we call such a factor *normalized*. For each product, the extracted powers of ϵ are collected into a single factor.

Step 3: Eliminating factors without a constant term. We maintain the following invariant. After t steps, the current variables are $\mathbf{y} = (y_1, \dots, y_{N-t})$, and there is an injective linear map $L_t : \mathbb{C}^{N-t} \rightarrow \mathbb{C}^N$ such that at most $k-t$ nonzero products of normalized affine factors, with their prefactors, sum to $P_{k+1,d} \circ L_t + O(\epsilon)$. For $t=0$ this holds with L_0 the identity.

Suppose that some normalized factor A has zero constant term at $\epsilon=0$. With $q = N-t$, let

$$A = a_0(\epsilon) + \sum_{j=1}^q a_j(\epsilon) y_j, \quad a_0, \dots, a_q \in \mathbb{C}[[\epsilon]].$$

Then $a_0(0) = 0$ and, since A is normalized, $a_j(0) \neq 0$ for some $j \geq 1$. We substitute

$$y_j = -\frac{a_0(\epsilon) + \sum_{h \neq j} a_h(\epsilon) y_h}{a_j(\epsilon)}.$$

As $a_j(\epsilon)$ is a *unit* of $\mathbb{C}[[\epsilon]]$, this substitution is valid in ϵ , and as $a_0(0) = 0$, it specializes at $\epsilon=0$ to an injective homogeneous linear map from $\mathbb{C}^{q-1}$ to $\mathbb{C}^q$. Hence the right-hand side becomes $P_{k+1,d} \circ L_{t+1} + O(\epsilon)$, where L_{t+1} is the composition of L_t with this map. On the left-hand side, the product containing A vanishes. We discard all products that have become zero and normalize the factors of the remaining ones again. This gives the invariant for $t+1$.

Step 4: Termination. Each step removes at least one product. If no product survived after t steps, then $t \leq k$ and $P_{k+1,d} \circ L_t = O(\epsilon)$, that is, $P_{k+1,d} \circ L_t = 0$, contradicting Step 1. Hence the procedure stops after $t \leq k-1$ steps with m products, where $1 \leq m \leq k-t$, and at that point every normalized factor has a nonzero constant term at $\epsilon=0$.

Step 5: Restriction to one block. We now apply to $P_{k+1,d} \circ L_t$ the restriction from Step 1. It is a homogeneous linear substitution that does not involve ϵ , so it preserves the constant term of every factor, and it turns $P_{k+1,d} \circ L_t$ into $x_1 \cdots x_d$. We obtain

$$\sum_{i=1}^m \epsilon^{v_i} \prod_{j=1}^{d_i} \widehat{\ell}_{ij}(\epsilon, \mathbf{x}) = x_1 \cdots x_d + O(\epsilon),$$

where $v_i \in \mathbb{Z}$, each $\widehat{\ell}_{ij} \in \mathbb{C}[[\epsilon]][\mathbf{x}]$ is affine linear in $\mathbf{x}$, and $\widehat{\ell}_{ij}(0, \mathbf{0}) \neq 0$.

Step 6: Passing to polynomials in ϵ . Choose $M \geq 0$ such that $a_i := M + v_i \geq 0$ for all i , and multiply the identity above by ϵ^M . Let $\ell_{ij} \in \mathbb{C}[\epsilon, \mathbf{x}]$ be the truncation of $\widehat{\ell}_{ij}$ modulo ϵ^{M+1} , and let $T_i = \prod_j \ell_{ij}$. Since $a_i \geq 0$, replacing each $\widehat{\ell}_{ij}$ by ℓ_{ij} changes $\epsilon^{a_i} \prod_j \widehat{\ell}_{ij}$ by a multiple of ϵ^{M+1} . Therefore,

$$\sum_{i=1}^m \epsilon^{a_i} T_i(\epsilon, \mathbf{x}) = \epsilon^M x_1 \cdots x_d + \epsilon^{M+1} E(\epsilon, \mathbf{x})$$

for some $E \in \mathbb{C}[[\epsilon]][\mathbf{x}]$; as the left-hand side is a polynomial in ϵ , so is E . Moreover, $T_i(0, \mathbf{0}) = \prod_j \widehat{\ell}_{ij}(0, \mathbf{0}) \neq 0$.

Size bound. None of the substitutions, normalizations and truncations increases the number of factors of a product. Each of the original products has at most s factors and $m \leq k$, so $\sum_{i=1}^m d_i \leq ks$. $\square$

Finally, we put together all the pieces and finish the proof of [Theorem 1.4](#) below.

7.3 Proof of [Theorem 1.4](#)

Proof of [Theorem 1.4](#). Apply [Lemma 7.5](#), and put $\hat{d} := \sum_i d_i \leq ks$. [Proposition 7.1](#), with $f = x_1 \cdots x_d$, gives nonzero $g, h \in \mathbb{C}[\mathbf{x}]$ such that

$$g = (x_1 \cdots x_d)h, \quad g \in \overline{\Sigma \wedge \Sigma(R, \hat{d})}, \quad R = m\hat{d}^{O(m^2)}.$$

Let ∂^*g denote the span of all partial derivatives of g , of all orders. All partial derivatives of a power ℓ^e of an affine linear polynomial are scalar multiples of $\ell^e, \ell^{e-1}, \dots, 1$; hence, for $e \leq \hat{d}$, they span a space of dimension at most $\hat{d} + 1$. Therefore every polynomial in $\Sigma \wedge \Sigma(R, \hat{d})$ has derivative-space dimension at most $R(\hat{d} + 1)$. This bound also holds in the border, since it is expressed by the vanishing of minors of the coefficient matrix of the partial derivatives. Hence

$$\dim_{\mathbb{C}}(\partial^*g) \leq R(\hat{d} + 1).$$

Fix a monomial order and write $\text{LM}(g) = \mathbf{x}^\gamma$. Since $x_1 \cdots x_d$ divides g , we have $\gamma_j \geq 1$ for every j . Differentiation with respect to x_j multiplies the coefficient of a monomial $\mathbf{x}^\beta$ by β_j , and it preserves the order among the monomials with $\beta_j \geq 1$. Hence, for each $\alpha \in \{0, 1\}^d$, the derivative $\partial^\alpha g$ has leading monomial $\mathbf{x}^{\gamma-\alpha}$. These leading monomials are distinct, so the corresponding 2^d derivatives are linearly independent. Consequently,

$$2^d \leq \dim_{\mathbb{C}}(\partial^*g) \leq R(\hat{d} + 1).$$

The formal degree is bounded by the circuit size, so $k < d \leq s$. Together with $\hat{d} \leq ks$ and $m \leq k$, this gives

$$2^d \leq R(\hat{d} + 1) \leq k(k + 1)^{O(k^2)} \leq s^{O(k^2)}.$$

Thus $s \geq 2^{\Omega(d/k^2)}$. □

8 Conclusion

We replaced the recursive DiDIL technique of [DDS21b] by a single linear differential equation obtained from a Wronskian. After a shift, every polynomial in the border of depth-3 circuits of top fan-in k satisfies an equation of order at most k whose coefficients have low complexity and whose leading coefficient does not vanish at the origin. This one equation drives all our results: it determines f from its first few Taylor coefficients (debordering), it forces a nonzero f to be visible in these coefficients (PIT), and its Wronskian structure yields the lower bound. Since the equation is obtained in one step, rather than by recursion on the top fan-in, all our bounds depend polynomially, rather than exponentially, on k . We conclude with some open questions.

1. *Large top fan-in.* [Theorem 1.1](#) gives quasipolynomial-size ABPs as long as $k = \text{poly}(\log s)$, but it becomes trivial when k is polynomially large in s . Is $\overline{\Sigma\Pi\Sigma} \subseteq \text{VBP}$, where the top fan-in may be polynomially large (and the formal degree is bounded by the size)? We do not even know whether $\overline{\Sigma\Pi\Sigma} \subseteq \text{VNP}$. Recall that some bound on the formal degree is necessary, since border depth-3 circuits of top fan-in two and unbounded formal degree can approximate every homogeneous polynomial [Kum20].
2. *Sums of powers.* [Theorem 1.2](#) answers the question of Dutta and Lysikov [DL25] for constants k and δ . What happens when k is polynomially large? For $\delta = 1$, border $\Sigma \wedge \Sigma$ circuits have small ABPs for every top fan-in [Lemma 3.8](#). Is the border of $\Sigma^{[k]} \wedge \Sigma\Pi^{[\delta]}$ contained in VBP for $\delta \geq 2$ (already for $\delta = 2$, that is, for sums of powers of quadratics) when k is polynomially large?
3. *Dependence on k .* The exponent $O(k^2)$ in [Theorem 1.1](#) and [Theorem 1.3](#) comes from the size of the Wronskian determinants. Can it be improved to $O(k)$? Similarly, can the lower bound of [Theorem 1.4](#) be improved to $2^{\Omega(d/k)}$?
4. *Other models.* The differential-equation viewpoint only uses the fact that each summand has a simple logarithmic derivative. Does it yield debordering, PIT, or lower bounds for other border models, for instance border depth-4 circuits $\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}$, whose factors are low-degree polynomials rather than affine linear polynomials?

Acknowledgements. PD is supported by the Nanyang Assistant Professorship (NAP) grant and an SUG Grant WBS 03INS002590C130, both funded by NTU. MB is supported by the SUG Grant WBS 03INS002590C130. This work gained momentum after PD and DS met at [WACT 2026](#) in Copenhagen when PD was funded by GREAT Grant WBS D820011000 (Global Research Excellence Award Travel). DS's Research is partially supported by an NSERC Discovery Grant and a McLean Award.

AI disclosure. The debordering and PIT results were found by the authors *without* the help of AI, along a rather non-linear path. PD and MB started by discussing debordering for sums of powers (depth 4), where they realized that Wronskians can be used to get [Theorem 1.2](#). Only afterwards did the authors realize, rather unexpectedly, that the same idea also applies to border

depth-3 circuits, the very setting for which DiDIL was designed; this yields our debordering result in [Theorem 1.1](#). Given PD’s earlier experience with [\[DDS21b\]](#), PIT ([Theorem 1.3](#)) then followed almost immediately.

For the lower bound ([Theorem 1.4](#)), however, the authors initially did not expect (4) to help. The approach of [\[DS26\]](#) reduces to the case where every affine factor has a nonzero constant term, applies DiDIL, and compares the lowest powers of z to write f as a ratio of two $\overline{\Sigma \wedge \Sigma}$ circuits. This keeps the homogeneous polynomial f intact and allows no shift, and then f simply cancels out of (4) (see the proof idea of the lower bound in [Section 2.4](#)). The authors first asked an AI model for a lower-bound proof without suggesting the Wronskian approach; it did not produce anything fruitful. After the authors supplied the Wronskian idea together with the reduction of [\[DS26, Claims 3–4\]](#), the model produced the key argument ([Lemma 7.2](#)), including the substitution $\epsilon \mapsto z^{d+1}$ and the use of Gaussian elimination to analyze the Wronskian. The authors then spent considerable time simplifying the argument before writing it down.

AI tools were also used to check mathematical claims and to improve the exposition and the L^AT_EX formatting. We stress that the paper is **written by humans, for humans**: every part of the text was written or carefully revised by the authors, who take full responsibility for the correctness of all the proofs.

Human disclosure. PD became interested in debordering while working on [\[DDS21b\]](#), and is grateful to his coauthors for many clarifications during that time. He grasped much more of the area while writing the survey [\[DL25\]](#), and thanks his coauthor Vladimir Lysikov for the many discussions along the way. PD also thanks Christian Ikenmeyer, Vladimir Lysikov, Fulvio Gesmundo, Sumanta Ghosh, Amit Sinhababu, Gorav Jindal, and Zeyu Guo for numerous discussions on debordering over the years.

References

- [AF22] Robert Andrews and Michael A. Forbes. [Ideals, Determinants, and Straightening: Proving and Using Lower Bounds for Polynomial Ideals](#). *The 54th ACM Symposium on Theory of Computing (STOC 2022)*, 2022.
- [AGKS15] Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Hitting-sets for ROABP and sum of set-multilinear circuits. *SIAM J. Comput.*, 44(3):669–697, 2015.
- [Ald84] A. Alder. *Grenzzrang und Grenzkomplexität aus algebraischer und topologischer Sicht*. PhD thesis, Universität Zürich, 1984.
- [BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. [\$O\(n^{2.7799}\)\$ Complexity for \$n \times n\$ Approximate Matrix Multiplication](#). *Inf. Process. Lett.*, 8(5):234–235, 1979.
- [BD10] Alin Bostan and Philippe Dumas. Wronskians and linear independence. *The American Mathematical Monthly*, 117(8):722–727, 2010.
- [BDI21] Markus Bläser, Julian Dörfler, and Christian Ikenmeyer. On the complexity of evaluating highest weight vectors. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *LIPICs*, pages 29:1–29:36, 2021.

- [Bin80] Dario Bini. **Relations between exact and approximate bilinear algorithms. Applications.** *Calcolo*, 17(1):87–97, 1980.
- [BLMW11] Peter Bürgisser, J. M. Landsberg, Laurent Manivel, and Jerzy Weyman. **An Overview of Mathematical Issues Arising in the Geometric Complexity Theory Approach to $VP \neq VNP$.** *SIAM J. Comput.*, 40(4):1179–1209, 2011.
- [BS21] Pranav Bisht and Nitin Saxena. **Blackbox identity testing for sum of special ROABPs and its border class.** *Journal of Comput. Compl.*, 2021.
- [BS26] Shubham Bhardwaj and Ramprasad Satharishi. Hitting sets for polynomials with small partial derivative spaces. Technical Report TR26-217, Electronic Colloquium on Computational Complexity, September 2026.
- [Bür00] Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*. Algorithms and Computation in Mathematics. Springer, 2000. Online version: <http://math-www.uni-paderborn.de/agpb/work/habil.ps>.
- [Bür04] Peter Bürgisser. **The Complexity of Factors of Multivariate Polynomials.** *Found. Comput. Math.*, 4(4):369–396, 2004.
- [Bür20] Peter Bürgisser. Correction to: The complexity of factors of multivariate polynomials. *Foundations of Computational Mathematics*, 20(6):1667–1668, 2020.
- [DDS21a] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. Deterministic identity testing paradigms for bounded top-fanin depth-4 circuits. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:27. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
- [DDS21b] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. **Demystifying the Border of Depth-3 Algebraic Circuits.** In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 92–103. IEEE, 2021. Full version of the FOCS 2021 paper.
- [DG24] Pranjal Dutta and Sumanta Ghosh. Advances in polynomial identity testing. *SIGACT News*, 55(2):53–88, June 2024.
- [DL78] Richard A DeMillo and Richard J Lipton. **A Probabilistic Remark on Algebraic Program Testing.** *Information Processing Letters*, 7(4):193–195, 1978.
- [DL25] Pranjal Dutta and Vladimir Lysikov. Recent advances in debordering methods. arXiv:2510.13049, 2025.
- [DS07] Zeev Dvir and Amir Shpilka. **Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits.** *SIAM Journal on Computing*, 36(5):1404–1434, 2007.
- [DS26] Pranjal Dutta and Nitin Saxena. Separated borders: Exponential-gap fanin-hierarchy theorem for approximative depth-3 circuits. *Journal of the ACM*, 2026. Accepted manuscript published online August 13, 2026; preliminary version in FOCS 2022.
- [FS13] Michael A Forbes and Amir Shpilka. **Explicit noether normalization for simultaneous conjugation via polynomial identity testing.** In *Approximation, Randomization, and*

- Combinatorial Optimization. Algorithms and Techniques*, pages 527–542. Springer, 2013.
- [FSV18] Michael A. Forbes, Amir Shpilka, and Ben Lee Volk. Succinct hitting sets and barriers to proving lower bounds for algebraic circuits. *Theory of Computing*, 14(18):1–45, 2018. Preliminary version in STOC 2017.
 - [GG20] Zeyu Guo and Rohit Gurjar. **Improved Explicit Hitting-Sets for ROABPs**. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.
 - [GKSS17] Joshua A. Grochow, Mrinal Kumar, Michael E. Saks, and Shubhangi Saraf. Towards an algebraic natural proofs barrier via polynomial identity testing. *CoRR*, abs/1701.01717, 2017.
 - [GMQ16] Joshua A. Grochow, Ketan D. Mulmuley, and Youming Qiao. Boundaries of VP and VNP. In *ICALP2016*, pages 34:1–34:14, 2016.
 - [Gro15] Joshua A. Grochow. **Unifying Known Lower Bounds via Geometric Complexity Theory**. 24(2):393–475, 2015.
 - [HM22] Kangjin Han and Hyunsuk Moon. A new bound for the waring rank of monomials. *SIAM Journal on Applied Algebra and Geometry*, 6(3):407–431, 2022.
 - [HS65] Juris Hartmanis and Richard E Stearns. **On the computational complexity of algorithms**. *Transactions of the American Mathematical Society*, 117:285–306, 1965.
 - [KI03] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. In Lawrence L. Larmore and Michel X. Goemans, editors, *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA*, pages 355–364. ACM, 2003.
 - [KS07] Neeraj Kayal and Nitin Saxena. **Polynomial identity testing for depth 3 circuits**. *computational complexity*, 16(2):115–138, 2007.
 - [Kum20] Mrinal Kumar. **On the Power of Border of Depth-3 Arithmetic Circuits**. *ACM Trans. Comput. Theory*, 12(1):5:1–5:8, 2020.
 - [LST21] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. **Superpolynomial Lower Bounds Against Low-Depth Algebraic Circuits**. 62nd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2021), page 81, 2021.
 - [LST25] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. *Journal of the ACM*, 72(4):26:1–26:35, 2025. Preliminary version in FOCS 2021.
 - [Mah13] Meena Mahajan. **Algebraic Complexity Classes**. *CoRR*, abs/1307.3863, 2013.
 - [MS01] Ketan Mulmuley and Milind A. Sohoni. **Geometric Complexity Theory I: An Approach to the P vs. NP and Related Problems**. *SIAM J. Comput.*, 31(2):496–526, 2001.
 - [MS08] K. D. Mulmuley and M. Sohoni. Geometric Complexity Theory II: Towards explicit obstructions for embeddings among class varieties. *SIAM J. Computing*, 38(3):1175–1206, 2008.
 - [Ore22] Øystein Ore. Über höhere kongruenzen. *Norsk Mat. Forenings Skrifter*, 1(7):15, 1922.

- [Sax14] Nitin Saxena. **Progress on polynomial identity testing-II**. In *Perspectives in Computational Complexity*, pages 131–146. Springer, 2014.
- [Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980.
- [SS13] Nitin Saxena and C. Seshadhri. From sylvester-gallai configurations to rank bounds: Improved blackbox identity test for depth-3 circuits. *J. ACM*, 60(5):33:1–33:33, 2013.
- [SY10] Amir Shpilka and Amir Yehudayoff. **Arithmetic Circuits: A survey of recent results and open questions**. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010.
- [Val79] Leslie G. Valiant. **Completeness Classes in Algebra**. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 249–261, 1979.
- [vdBDG⁺25] Maxim van den Berg, Pranjal Dutta, Fulvio Gesmundo, Christian Ikenmeyer, and Vladimir Lysikov. Algebraic metacomplexity and representation theory. In *40th Computational Complexity Conference (CCC 2025)*, volume 339 of *LIPICs*, pages 26:1–26:35, 2025.
- [Zip79] Richard Zippel. **Probabilistic algorithms for sparse polynomials**. In *International symposium on symbolic and algebraic manipulation*, pages 216–226. Springer, 1979.