Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 15 (2021)

Hitting Sets for Orbits of Circuit Classes and Polynomial Families

Chandan Saha
Indian Institute of Science
chandan@iisc.ac.in

Bhargav Thankey
Indian Institute of Science
thankeyd@iisc.ac.in

Abstract

The orbit of an n-variate polynomial $f(\mathbf{x})$ over a field $\mathbb{F}$ is the set $\mathrm{orb}(f) := \{f(A\mathbf{x}+\mathbf{b}) : A \in \mathrm{GL}(n,\mathbb{F}) \text{ and } \mathbf{b} \in \mathbb{F}^n\}$. The orbit of a polynomial f is a geometrically interesting subset of the set of affine projections of f. Affine projections of polynomials computable by seemingly weak circuit classes can be quite powerful. For example, the polynomial $\mathsf{IMM}_{3,d}$ – the $(1,1)$-th entry of a product of d generic 3×3 matrices – is computable by a constant-width read-once oblivious algebraic branching program (ROABP), yet every polynomial computable by a size-s general arithmetic formula is an affine projection of $\mathsf{IMM}_{3,\mathrm{poly}(s)}$ [BC92]. To our knowledge, no efficient hitting set construction was known for $\mathrm{orb}(\mathsf{IMM}_{3,d})$ before this work.

In this paper, we initiate the study of explicit hitting sets for the *orbits* of polynomials computable by several natural and well-studied circuit classes and polynomial families. In particular, we give quasi-polynomial time hitting sets for the orbits of:

1. Low-individual-degree polynomials computable by *commutative ROABPs*. This implies quasi-polynomial time hitting sets for the orbits of the *elementary symmetric polynomials* and the orbits of *multilinear sparse polynomials*.
2. Multilinear polynomials computable by *constant-width ROABPs*. This implies a quasi-polynomial time hitting set for the orbits of the family $\{\mathsf{IMM}_{3,d}\}_{d\in\mathbb{N}}$.
3. Polynomials computable by *constant-depth, constant-occur formulas*. This implies quasi-polynomial time hitting sets for the orbits of *multilinear depth-*4 *circuits with constant top fan-in*, and also polynomial-time hitting sets for the orbits of the *power symmetric polynomials* and the *sum-product polynomials*.
4. Polynomials computable by *occur-once formulas*.

We say a polynomial has low individual degree if the degree of every variable in the polynomial is at most $\mathrm{poly}(\log n)$, where n is the number of variables.

The first two results are obtained by building upon and strengthening the rank concentration by translation technique of [ASS13]; the second result additionally uses the merge-and-reduce idea from [FS13b, FSS14]. The proof of the third result applies the algebraic independence based technique of [ASSS16, BMS13] to reduce to the case of constructing hitting sets for the orbits of sparse polynomials. A similar reduction using the Shpilka-Volkovich (SV) generator based argument in [SV15] yields the fourth result. The SV generator plays an important role in all the four results.

ISSN 1433-8092

Contents

1 Introduction

Polynomial identity testing (PIT) is a fundamental problem in arithmetic circuit complexity. PIT is the problem of deciding if a given arithmetic circuit computes an identically zero polynomial. It is one of the few natural problems in BPP (in fact, in co-RP) for which we do not know of deterministic polynomial-time algorithms. A probabilistic polynomial-time algorithm for PIT follows from the DeMillo-Lipton-Schwartz–Zippel lemma [DL78, Zip79, Sch80]. There are several algorithms for other interesting problems that have PIT at their core. The fast parallel algorithms for the perfect matching problem [Lov79, KUW86, MVV87, FGT16, ST17], the linear matroid intersection problem [NSV94, GT20], and the maximum rank matrix completion problem [Mur93, GT20] are based on PIT. The deterministic primality testing algorithm in [AKS04] derandomizes a particular instance of PIT over a ring [AB03]. Also, multivariate polynomial factorization for general circuits can be efficiently reduced to PIT and factoring univariate polynomials [Kal89, KT90, KSS15].

Derandomizing PIT is closely connected to proving circuit lower bounds. A sub-exponential time derandomization of PIT implies either a super-polynomial Boolean circuit lower bound or a super-polynomial arithmetic circuit lower bound [KI04]. A sub-exponential time derandomization of *black-box*[1] PIT implies a super-polynomial arithmetic circuit lower bound [HS80, Agr05]. Conversely, a super-polynomial lower bound for arithmetic circuits implies a deterministic sub-exponential time algorithm for *low-degree*[2], black-box PIT [KI04, NW94][3]. Similar hardness versus randomness tradeoffs are also known for constant depth arithmetic circuits [DSY09, CKS18]. Thus, derandomizing black-box PIT is essentially equivalent to proving arithmetic circuit lower bounds. The black-box PIT problem for a circuit class $\mathcal{C}$ is also known as the problem of constructing *hitting sets* for $\mathcal{C}$ (see Definition 5).

Two restricted circuit classes. In the past two decades, PIT algorithms and hitting set constructions have been studied for various restricted classes/models of circuits. Bounding the read of every variable is a natural restriction that has received a lot of attention. In particular, two constant-read models have been intensely studied in the literature. These are *read-once oblivious algebraic branching programs* (ROABPs) and *constant-read* (more generally, *constant-occur*) *formulas* (see Definition 1 and 3) . The ROABP model is surprisingly rich and powerful. It captures several other interesting circuit classes such as sparse polynomials or depth-two circuits, depth-three powering circuits (symmetric tensors), set-multilinear depth-three circuits (tensors) and its generalization set-multilinear algebraic branching programs, and semi-diagonal circuits [FS13b]. Some notable polynomials such as the iterated matrix multiplication polynomial, the elementary and the power symmetric polynomials, and the sum-product polynomial can be computed by linear-size ROABPs. A polynomial-time PIT algorithm and a quasi-polynomial time hitting set construction for ROABPs are known [RS05, FS13b, AGKS15]. Hitting sets for ROABPs, which can be viewed as the algebraic analogue of pseudorandomness for randomized space-bounded computation [Nis92, INW94, FK18], have also led to the derandomization of an interesting case of the Noether Normalization Lemma [Mul17, FS13a], and to hitting sets for non-

[1] An algorithm for the black-box PIT problem takes as input black-box access to a circuit. The algorithm cannot "see" the circuit but can query it at any point.

[2] i.e., the input circuit computes a polynomial of degree poly(n), where n is the number of variables.

[3] A stronger lower bound yields a stronger derandomization result: an exponential lower bound for arithmetic circuits implies a quasi-polynomial time derandomization of low-degree, black-box PIT.

commutative algebraic branching programs [FS13b]. The constant-occur formula model is also reasonably natural; it captures other interesting circuit classes like multilinear depth-four circuits with bounded top fan-in [KMSV13, SV18] and sums of constantly many read-once formulas [SV15]. A quasi-polynomial time hitting set construction for multilinear constant-read formulas was given by [AvMV15]. [ASSS16] obtained polynomial-time constructible hitting sets for constant-depth, constant-occur formulas.

Hitting sets for orbits. In this paper, we study hitting set constructions for the *orbits* of ROABPs and constant-occur formulas. The orbit of a polynomial f is the set of polynomials obtained by applying invertible affine transformations on the variables of f, i.e., by replacing the variables of f with linearly independent affine forms. The orbit of a circuit class is the union of the orbits of the polynomials computable by the circuits in the class. Our reasons for studying hitting sets for the orbits of ROABPs and constant-occur formulas are threefold:

1. *The power of orbit closures:* The set of affine projections of an n-variate polynomial $f(\mathbf{x})$ over a field $\mathbb{F}$ is $\mathrm{aproj}(f) := \{f(A\mathbf{x}+\mathbf{b}) \;:\; A \in \mathbb{F}^{n\times n} \text{ and } \mathbf{b} \in \mathbb{F}^n\}$; the orbit of f is the set $\mathrm{orb}(f) = \{f(A\mathbf{x}+\mathbf{b}) \;:\; A \in \mathrm{GL}(n,\mathbb{F}) \text{ and } \mathbf{b} \in \mathbb{F}^n\} \subseteq \mathrm{aproj}(f)$.[4] Affine projections of polynomials computable by polynomial-size ROABPs or constant-occur formulas have great expressive power. For example, the iterated matrix multiplication polynomial $\mathsf{IMM}_{w,d}$ – the $(1,1)$-th entry of a product of d generic $w \times w$ matrices – is computable by a linear-size ROABP, yet every polynomial computable by a size-s *general* algebraic branching program[5] is in $\mathrm{aproj}(\mathsf{IMM}_{s,s})$. In fact, every polynomial computable by a size-s arithmetic formula is in $\mathrm{aproj}(\mathsf{IMM}_{3,\mathrm{poly}(s)})$ [BC92]. The sum-product polynomial $\mathsf{SP}_{s,d} := \sum_{i\in[s]}\prod_{j\in[d]} x_{i,j}$ is computable by a depth-2 read-once formula, but even so every polynomial computable by a general depth-3 circuit with top fan-in s and formal degree d is in $\mathrm{aproj}(\mathsf{SP}_{s,d})$. As demonstrated by the depth reduction results in [GKKS16, Tav15, Koi12, AV08, VSBR83], depth-3 circuits are incredibly powerful. Also, affine projections of read-once formulas capture general arithmetic formulas. The orbit of f being a mathematically interesting subset of $\mathrm{aproj}(f)$, it is natural to ask if we can give efficient hitting set constructions for the orbits of the above-mentioned polynomial families and circuit classes. Moreover, $\mathrm{orb}(f)$ is not 'much smaller' than $\mathrm{aproj}(f)$, as the latter is contained in the *orbit closure* of f if $\mathrm{char}(\mathbb{F}) = 0$ (see Appendix F). By identifying n-variate, degree-d polynomials with their respective coefficient vectors in $\mathbb{F}^{\binom{n+d}{d}}$, the orbit closure of f (denoted by $\overline{\mathrm{orb}(f)}$) is defined as the Zariski closure of $\mathrm{orb}(f)$. The polynomials in $\overline{\mathrm{orb}(f)}$, and hence also the polynomials in $\mathrm{aproj}(f)$, can be approximated infinitesimally closely by the polynomials in $\mathrm{orb}(f)$ over $\mathbb{C}$.[6] In this sense, $\mathrm{orb}(f)$ is a dense subset of $\mathrm{aproj}(f)$.

2. *Geometry of the circuit classes:* Consider an n-variate polynomial $f \in \mathbb{R}[\mathbf{x}]$ that is computable by a polynomial-size ROABP or a polynomial-size constant-occur formula. Let $\mathbb{V}(f)$ be the variety (i.e., the zero locus) of f. The geometry of $\mathbb{V}(f)$ is preserved by any rigid transfor-

[4] Ideally, we should use the notations $\mathrm{aproj}_{\mathbb{F}}$ and $\mathrm{orb}_{\mathbb{F}}$, but we are dropping the subscripts here for simplicity, and because we would be always working with the underlying field $\mathbb{F}$.

[5] Thanks to the depth reduction result in [VSBR83], low-degree polynomials computable by arithmetic circuits are also computable by quasi-polynomially large algebraic branching programs.

[6] However, $\overline{\mathrm{orb}(f)}$ can be strictly larger than $\mathrm{aproj}(f)$.

mation[7] on $\mathbb{R}^n$. Computation of a set $\mathcal{H} \subseteq \mathbb{R}^n$ that is not contained in $T(\mathbb{V}(f))$, for every rigid transformation T, would have to be "mindful" of the geometry of $\mathbb{V}(f)$ and oblivious to the choice of the coordinate system. Computing such an $\mathcal{H}$ is exactly the problem of constructing a hitting set for the polynomials $\{f(R\mathbf{x}+\mathbf{b}) \ : \ R \in O(n,\mathbb{R}) \text{ and } \mathbf{b} \in \mathbb{R}^n\}$. We can generalize the problem slightly by replacing $R \in O(n,\mathbb{R})$ with $A \in \mathrm{GL}(n,\mathbb{R})$.[8] A hitting set for ROABPs or constant-occur formulas does not immediately give a hitting set for $\{f(A\mathbf{x}+\mathbf{b}) \ : \ A \in \mathrm{GL}(n,\mathbb{R}) \text{ and } \mathbf{b} \in \mathbb{R}^n\}$, as the definitions of an ROABP and a constant-occur formula are tied to the choice of the coordinate system. In fact, we show in Appendix E.1 that there is an explicit polynomial g in the orbit of a sparse polynomial such that any ROABP computing g has exponential size. We also show in Appendix E.2 that there is an explicit polynomial $g \in \mathrm{orb}(x_1x_2\cdots x_n)$ such that any occur-once formula computing g has size at least 2^{n-1}. It is thus natural to ask if there is anything special about the geometry of $\mathbb{V}(f)$ which can facilitate efficient constructions of hitting sets for $\mathrm{orb}(f)$.

3. *Strengthening existing techniques:* Finally, it is worth investigating whether the techniques used to design hitting sets for ROABPs and constant-occur formulas can be applied or strengthened or combined to give hitting sets for the orbits of these circuit classes.

Indeed, the results in this paper are obtained by building upon, strengthening and combining several tools and techniques from the literature, in particular the rank concentration by translation technique from [ASS13], the merge-and-reduce idea from [FS13b, FSS14], the algebraic independence based technique from [ASSS16, BMS13], and the Shpilka-Volkovich generator from [SV15]. Our work here on hitting sets for the orbits of the above-mentioned circuit classes probes a line of research that – to our knowledge – has remained largely unexplored. In obtaining these results, we have highlighted the efficacy and the versatility of some of the existing tools and techniques. We describe the relevant circuit models in the next section and state our results in Section 1.2.

1.1 The models

Unless otherwise stated, we will assume that polynomials have coefficients that belong to a field $\mathbb{F}$.

Algebraic branching programs (ABPs) were defined by Nisan in [Nis91]. As the name suggests, read-once oblivious algebraic branching programs (ROABPs) are a read-once variant of ABPs. While Nisan defined ABPs using directed graphs, in this work we use the following equivalent and conventional definition of an ROABP.

Definition 1 (ROABP [FS13b])**.** An n-variate, width-w read-once oblivious algebraic branching program (ROABP) is a product of the form $\mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n)\cdot \mathbf{1}$, where $\mathbf{1}$ is the $w \times 1$ vector of all ones, and for every $i \in [n]$, $M_i(x_i)$ is a $w \times w$ matrix whose entries are in $\mathbb{F}[x_i]$.

Definition 2 (Commutative ROABP)**.** An n-variate, width-w *commutative* ROABP is an n-variate, width-w ROABP $\mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n)\cdot \mathbf{1}$, where for all $i,j \in [n]$, $M_i(x_i)$ and $M_j(x_j)$ commute with each other.

[7] A rigid transformation T is given by an orthogonal matrix $R \in O(n,\mathbb{R})$ (which stands for reflections and rotations) and a translation vector $\mathbf{b} \in \mathbb{R}^n$ such that every $\mathbf{x} \in \mathbb{R}^n$ maps to $T(\mathbf{x}) = R\mathbf{x}+\mathbf{b}$.

[8] An invertible transformation A is essentially an orthogonal transformation up to "scaling": from singular value decomposition, we have $A = UDV$, where U, V are orthogonal matrices and D is a diagonal matrix.

A polynomial f is *s-sparse* if it has at most s monomials with non-zero coefficients; these monomials will be referred to as the *monomials of f*. It is easy to see that an s-sparse polynomial of degree d can be computed by a depth-2 circuit of size at most sd. Also, observe that every s-sparse polynomial can be computed by a width-s commutative ROABP.

Definition 3 (Occur-k formula [ASSS16])**.** An occur-k formula is a rooted tree whose leaves are labelled by s-sparse polynomials and whose internal nodes are sum $(+)$ gates or product-power $(\times\curlywedge)$ gates. Each variable appears in at most k of the sparse polynomials that label the leaves. The edges feeding into a $+$ gate are labelled by field elements and have 1 as *edge weights*, whereas the edges feeding into a $\times\curlywedge$ gate have natural numbers as edge weights. A leaf node computes the s-sparse polynomial that labels it. A $+$ gate with inputs from nodes that compute $f_1, ..., f_m$ and with the corresponding input edge labels $\alpha_1, ..., \alpha_m$, computes $\alpha_1 f_1 + \cdots + \alpha_m f_m$. A $\times\curlywedge$ gate with inputs from nodes that compute $f_1, ..., f_m$ and with the corresponding input edge weights $e_1, ..., e_m$, computes $f_1^{e_1} \cdots f_m^{e_m}$. The formula computes the polynomial that is computed by the root node.

The *size* of an occur-k formula is the weighted sum of all the edges in the formula (i.e., an edge feeding into a $\times\curlywedge$ gate is counted as many times as its edge weight, whereas an edge feeding into a $+$ gate is counted once) plus the sizes of the depth-2 circuits computing the s-sparse polynomials at the leaves. The *depth* of an occur-k formula is equal to the depth of the underlying tree plus 2, to account for the depth of the circuits computing the sparse polynomials at the leaves.[9]

Read-k formulas have been studied intensely in the literature (see Section 1.4). Occur-k formulas generalize read-k formulas in two ways – the leaves are labelled by arbitrary sparse polynomials instead of just variables, and powering gates are included along with the usual sum and product gates. These generalizations help make the occur-k model complete[10], and capture other interesting circuit classes (such as multilinear depth-4 circuits with constant top fan-in [SV18, KMSV13]) and polynomial families (such as the power symmetric polynomials). Besides, there is no restriction of multilinearity on the model, unlike the case in some prior works [AvMV15, SV18, KMSV13].

We will identify the variable set $\mathbf{x} = \{x_1, \ldots, x_n\}$ with the column vector $(x_1 \; x_2 \; \cdots \; x_n)^T$.

Definition 4 (Orbits of polynomials)**.** Let $f(\mathbf{x})$ be an n-variate polynomial over a field $\mathbb{F}$. The *orbit* of f, denoted by $\text{orb}(f)$, is the set $\{f(A\mathbf{x}) : A \in \text{GL}(n, \mathbb{F})\}$. The orbit of a set of polynomials $\mathcal{C}$, denoted by $\text{orb}(\mathcal{C})$, is the union of the orbits of the polynomials in $\mathcal{C}$.

Remark. The results we present in this paper hold even if we define the orbit of an n-variate polynomial f as $\text{orb}(f) = \{f(A\mathbf{y} + \mathbf{b}) \; : \; |\mathbf{y}| = m \geq n, \; A \in \mathbb{F}^{n \times m}$ has rank n, and $\mathbf{b} \in \mathbb{F}^n\}$. However, we work with this slightly conventional definition of $\text{orb}(f)$ for simplicity of exposition, and because the proofs in the general setting are nearly the same as the proofs we present here.

By the 'orbit of a circuit class $\mathcal{C}$', we mean the union of the orbits of the polynomials computable by the circuits in the class $\mathcal{C}$. Our main results are efficient constructions of hitting sets for the orbits of commutative ROABPs and constant-width ROABPs (under low individual degree restriction), and the orbits of constant-depth constant-occur formulas and occur-once formulas.

[9] Observe that if f is computable by a size-s, depth-Δ, occur-k formula, then it is also computable by a size-s, depth-Δ circuit that has only $+$ and $\times$ gates.

[10] For example, the power symmetric polynomial $x_1^n + \ldots + x_n^n$ cannot be computed by a read-k formula for any $k < n$, but it can be computed by an occur-once formula.

1.2 Our results

Definition 5 (Hitting set)**.** Let $\mathcal{C}$ be a set of n-variate polynomials. A set of points $\mathcal{H} \subseteq \mathbb{F}^n$ is a hitting set for $\mathcal{C}$ if for every non-zero $f \in \mathcal{C}$, there is a point $\mathbf{a} \in \mathcal{H}$ such that $f(\mathbf{a}) \neq 0$.

By a 'T-time hitting set', we mean that the hitting set can be computed in T time. Typically, T is a function of the input parameters such as the number of variables, the size of the input circuit, and the degree or the individual degree of the input polynomial. The *individual degree* of a monomial is the largest of the exponents of the variables that appear in it. The individual degree of a polynomial is the largest of the individual degrees of its monomials. We are now ready to state our results.

Theorem 6 (Hitting sets for the orbits of commutative ROABPs with low individual degree)**.** *Let $\mathcal{C}$ be the set of n-variate polynomials with individual degree at most d that are computable by width-w commutative ROABPs. If $|\mathbb{F}| > n^2 d$, then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $(nd)^{O(d \log w)}$ time.*

An interesting subclass of commutative ROABPs is the class of *sums of products of univariates*. This model, which is a broad generalization of the class of sparse polynomials, has found important applications in several other works [Sax08, SSS13, GKKS16]. We say an n-variate polynomial $f(x_1, x_2, \ldots, x_n)$ can be expressed as a sum of s products of univariates if $f = \sum_{i \in [s]} \prod_{j \in [n]} f_{i,j}(x_j)$, where each $f_{i,j}(x_j)$ is a univariate polynomial in x_j. Theorem 7 below (which follows as a corollary from the above theorem) gives a quasi-polynomial time hitting set for the orbits of sums of products of *low degree* univariates.

Theorem 7 (Hitting sets for the orbits of sums of products of low degree univariates)**.** *Let $\mathcal{C}$ be the set of n-variate polynomials that can be expressed as sums of s products of univariates of degree at most d. If $|\mathbb{F}| > n^2 d$, then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $(nd)^{O(d \log s)}$ time.*

Remarks.

1. Even under the low individual degree restriction the above class remains reasonably natural and interesting. For example, the elementary symmetric polynomial $\mathsf{ESym}_{n,D} = \sum_{S \in \binom{[n]}{D}} \prod_{i \in S} x_i$ can be expressed as a sum of $n+1$ products of univariate affine forms. This is due to a nice interpolation trick attributed to Ben-Or in [NW97, Shp02]. The theorem then implies an $n^{O(\log n)}$-time hitting set for $\mathrm{orb}(\mathsf{ESym}_{n,D})$.

2. The theorem also implies a quasi-polynomial time hitting set for the orbits of multilinear sparse polynomials, and more generally, for the orbits of sparse polynomials with low individual degree. It is easy to see that the orbit of a multilinear sparse polynomial may contain a non-sparse polynomial. So, the existing hitting set constructions for sparse polynomials [KS01, LV03] (where the complexity depends polynomially on the sparsity parameter) may no longer remain efficient for the orbits of sparse polynomials.

3. It turns out though that for the particular case of sparse polynomials it is possible to remove the individual degree restriction from the above theorem. This is due to an independent and simultaneous work by [MS21]. We state their result next.

Theorem 8 (Hitting sets for the orbits of sparse polynomials [MS21])**.** *Let $\mathcal{C}$ be the set of n-variate, s-sparse polynomials of degree at most d. If $|\mathbb{F}| > nd$ and* $\mathrm{char}(\mathbb{F}) = 0$ *or $> d$, then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $(nd)^{O(\log s)}$ time.*

Remarks.

1. Hitting sets for the orbits of sparse polynomials play a basic role in our proofs of Theorem 10 and Theorem 11 (stated later). There, we apply the algebraic independence based analysis from [ASSS16, BMS13] and the Shpilka-Volkovich (SV) generator based argument from [SV15], respectively, to reduce to the case of constructing hitting sets for the orbits of sparse polynomials. While in the original version of our work [ST21] we applied Theorem 7 in the base case of the proofs of Theorem 10 and 11, here we plug-in Theorem 8 in the base case. This helps us forgo the low individual degree restriction that was present in these theorems in the original version.

2. It is worth noting though that the proof of Theorem 8, which is also based on the SV-generator, does not seem to scale to commutative ROABPs or even the sums of products of univariates model. For the sake of completeness, we provide [MS21]'s nice proof of Theorem 8 in Appendix C.

Theorem 9 (Hitting sets for the orbits of multilinear constant-width ROABPs)**.** *Let $\mathcal{C}$ be the set of n-variate multilinear polynomials that are computable by width-w ROABPs. If $|\mathbb{F}| > n^{O\left(w^4\right)}$, then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $n^{O\left(w^6 \cdot \log n\right)}$ time.*

Remarks.

1. The theorem gives a quasi-polynomial time hitting set for $\mathrm{orb}(\mathsf{IMM}_{3,d})$, as $\mathsf{IMM}_{3,d}$ is computable by a width-9 ROABP. As mentioned before, the family $\{\mathsf{IMM}_{3,d}\}_{d \in \mathbb{N}}$ is complete for the class of arithmetic formulas under affine projections (in fact, under p-projections) [BC92].

2. The set of affine projections of $\mathsf{IMM}_{2,d}$ is also quite rich, despite the fact that there are simple quadratic polynomials that are not in $\mathrm{aproj}(\mathsf{IMM}_{2,d})$ for *any* d [AW16, SSS09]. This is because hitting sets for $\mathrm{aproj}(\mathsf{IMM}_{2,d})$ give hitting sets for depth-3 circuits [SSS09]. Moreover, $\overline{\mathrm{orb}(\mathsf{IMM}_{2,d})}$ captures the orbit closures of arithmetic formulas [BIZ18]. The above theorem implies a quasi-polynomial time hitting set for $\mathrm{orb}(\mathsf{IMM}_{2,d})$.

Theorem 10 (Hitting sets for the orbits of constant-depth, constant-occur formulas)**.** *Let $\mathcal{C}$ be the set of n-variate, degree-D polynomials that are computable by depth-Δ, occur-k formulas of size s. Let $R := (2k)^{2\Delta \cdot 2^{\Delta}}$. If* $\mathrm{char}(\mathbb{F}) = 0$ *or* $> (2ks)^{\Delta^3 R}$, *then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $(nRD)^{O(R(\log R + \Delta \log k + \Delta \log s) + \Delta R)}$ time. If the leaves are labelled by b-variate polynomials, then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $(nRD)^{O(Rb + \Delta R)}$ time. In particular, if Δ and k are constants, then the hitting sets can be constructed in time $(nD)^{O(\log s)}$ and $(nD)^{O(b)}$, respectively.*

Remarks.

1. The above theorem gives hitting sets for the orbits of two other interesting models that have been studied in the literature: There is a polynomial-time constructible hitting set for multilinear depth-4 circuits with constant top fan-in [SV18, KMSV13]. Theorem 10 implies a quasi-polynomial time hitting set for the orbit of this model, as a multilinear depth-4 circuit with constant top fan-in can be viewed as a depth-4 constant-occur formula. [BMS13] gave a polynomial-time hitting set for $C(f_1, \ldots, f_m)$, where C is a low-degree circuit and $f_1, \ldots, f_m$ are sparse polynomials with bounded transcendence degree. The proof of the above theorem also implies a quasi-polynomial time hitting set for the orbit of this model.

2. The theorem yields polynomial-time hitting sets for the orbits of the power symmetric polynomial $\mathsf{PSym}_{n,D} = \sum_{i \in [n]} x_i^D$ and the sum-product polynomial $\mathsf{SP}_{n,D} = \sum_{i \in [n]} \prod_{j \in [D]} x_{i,j}$. This is because the polynomials PSym and SP are computable by constant-depth, occur-once formulas whose leaves are labelled by univariate polynomials. Prior to our work, [KS19] gave a polynomial-time hitting set for $\mathrm{orb}(\mathsf{PSym}_{n,D})$ using a different argument that involves the Hessian matrix.

Theorem 11 (Hitting sets for the orbits of occur-once formulas)**.** *Let $\mathcal{C}$ be the set of n-variate, degree-D polynomials that are computable by occur-once formulas whose leaves are labelled by s-sparse polynomials. If $|\mathbb{F}| > nD$ and $\mathrm{char}(\mathbb{F}) = 0$ or $> D$, then a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in $(nD)^{O(\log n + \log s)}$ time. If the leaves are labelled by b-variate polynomials, then a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in $(nD)^{O(\log n + b)}$ time.*

Remark. The independent and concurrent work by [MS21] gave (among other results) a quasi-polynomial time hitting set construction for the orbits of read-once formulas. We note that this result also follows from the second part of the above theorem which is already present in the original version of this work [ST21].

1.3 Proof techniques

Let us briefly discuss the techniques that go into proving the above results.

Commutative ROABPs with low individual degree. Theorem 6 is proved by adapting the rank concentration by translation technique of [ASS13][11] to work for the orbits of commutative ROABPs. Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$ be a commutative ROABP and $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For any $A \in \mathrm{GL}(n, \mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. Suppose that A maps x_i to a linear form $\ell_i(\mathbf{x})$ for every $i \in [n]$, and let $y_i = \ell_i(\mathbf{x})$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. We show that if $g \neq 0$, then there exist *explicit* "low" degree polynomials $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z})$, where $\mathbf{z}$ is a "small" set of variables, such that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has a "low" support[12] monomial. This is done by proving that $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has low support rank concentration over $\mathbb{F}(\mathbf{z})$ in the "$\mathbf{y}$-variables" (see Section 2.2 for the meaning of low support rank concentration.). That done, we use the assumption that f has low individual degree to argue that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ also has a low support $\mathbf{x}$-monomial. This and the fact that $|\mathbf{z}|$ is small imply that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$, when viewed as a polynomial in $\mathbb{F}[\mathbf{x}, \mathbf{z}]$, has a low support monomial. Finally, we use the SV generator to hit g.

Our analysis differs from that in [ASS13] at a crucial point: In [ASS13], it was shown that $F(\mathbf{x} + \mathbf{t}) = M_1(x_1 + t_1)M_2(x_2 + t_2)\cdots M_n(x_n + t_n)$ has low support rank concentration over $\mathbb{F}(\mathbf{t})$ if the nonzeroness of every polynomial in a certain collection of polynomials – each in a "small" set of $\mathbf{t}$-variables – is preserved. As each polynomial in the collection has "few" $\mathbf{t}$-variables, a substitution $t_i \leftarrow t_i(\mathbf{z})$ that preserves its nonzeroness is relatively easy to construct. But the collection of polynomials that we need to preserve to show low support rank concentration for $G(\mathbf{x} + \mathbf{t})$ is such that every polynomial in the collection has potentially all the $\mathbf{t}$-variables. However, we are able to argue that each of these polynomials still has a low support $\mathbf{t}$-monomial. This then helps

[11] [ASS13] proved their result for products of univariate polynomials over a Hadamard algebra which form a subclass of commutative ROABPs. However, their analysis also works for general commutative ROABPs.

[12] Support of a monomial is the number of variables with non-zero exponents in the monomial.

us construct a substitution $t_i \mapsto t_i(\mathbf{z})$ that preserves the nonzeroness of these polynomials.

Multilinear constant-width ROABPs. Theorem 9 is proved by combining the rank concentration by translation technique of [ASS13] with the merge-and-reduce idea from [FS13b] and [FSS14]. Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$ be a multilinear, width-w ROABP; here $M_i(x_i) \in \mathbb{F}^{w\times w}[x_i]$ for all $i \in [n]$. Also, let $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For any $A \in \mathrm{GL}(n,\mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. For $i \in [n]$, suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$, where ℓ_i is a linear form, and let $y_i = \ell_i(\mathbf{x})$ and $\mathbf{y} = \{y_1,\ldots,y_n\}$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. Much like in the case of commutative ROABPs, we show that if $g \neq 0$, then there exist explicit "low" degree polynomials $t_1(\mathbf{z}),\ldots,t_n(\mathbf{z})$, where $\mathbf{z}$ is a "small" set of variables such that $G(x_1+t_1(\mathbf{z}),\ldots,x_n+t_n(\mathbf{z}))$ has "low" support rank concentration in the "$\mathbf{y}$-variables". While in the rank concentration argument for commutative ROABPs the $\mathbf{x}$-variables were translated only once, here the translations can be thought of as happening sequentially and in stages. There will be $\lceil \log n \rceil$ stages with each stage also consisting of multiple translations. After the p-th stage, the product of any 2^p consecutive matrices in G will have low support rank concentration in the $\mathbf{y}$-variables. Thus, after $\lceil \log n \rceil$ stages, we will have low support rank concentration in the $\mathbf{y}$-variables for $G(x_1+t_1(\mathbf{z}),\ldots,x_n+t_n(\mathbf{z}))$.

As in the case of commutative ROABPs, we show that $G(\mathbf{x}+\mathbf{t})$ has low support rank concentration if each polynomial in a certain collection of non-zero polynomials in the $\mathbf{t}$-variables is kept non-zero by the substitution $t_i \mapsto t_i(\mathbf{z})$. However, in this case, it is trickier to show that these polynomials have low support $\mathbf{t}$-monomials. We do this by arguing that each such polynomial can be expressed as a ratio of a polynomial that contains a low support $\mathbf{t}$-monomial and a product of linear forms in the $\mathbf{t}$-variables.

Remark. A quasi-polynomial time hitting set for general ROABPs was given by [AGKS15] using an elegant generalization of the monomial isolation method [KS01], namely the *basis isolation method*. As shown in [GKST17, FGS18], designing a basis isolating weight assignment is a stronger objective than achieving rank concentration by translation. It is not immediately clear how to obtain efficient constructions of basis isolating weight assignments for the orbits of ROABPs, even under additional restrictions such as commutativity, constant-width or low individual degree. However, our work here shows that the weaker objective of rank concentration by translation can be achieved for the orbits of the above-mentioned subclasses of ROABPs.

Constant-depth, constant-occur formulas. We prove Theorem 10 by combining the algebraic independence based technique in [ASSS16] with Theorem 8. Let f be a constant-depth, constant-occur formula. We first show that it can be assumed without loss of generality that the top-most gate of f is a $+$ gate whose fan-in is upper bounded by the occur of f, say k. In [ASSS16], they were able to upper bound the top fan-in by simply translating a variable by 1 and subtracting the original formula. However, the same idea does not quite work here, because we have only access to a polynomial in the *orbit* of f. To upper bound the top fan-in, we show that there exists a variable x_i such that $\frac{\partial f}{\partial x_i}$ is a constant-depth, constant-occur formula with top fan-in bounded by k. Then, using the chain rule of differentiation, we show that one can construct a hitting set generator for $\mathrm{orb}(f)$ from a generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$; this means that we can shift our attention to $f' = \frac{\partial f}{\partial x_i}$, which we shall henceforth refer to as f.

Let $f = f_1 + \cdots + f_k$, $A \in \mathrm{GL}(n, \mathbb{F})$, $g = f(A\mathbf{x})$, $g = g_1 + \ldots + g_k$ where for all $i \in [k]$, $g_i = f_i(A\mathbf{x})$. It was shown in [ASSS16] that a homomorphism, which is faithful (see Definition 18) to $f_1, \ldots, f_k$, is a hitting set generator for f. In our case, this translates to 'a homomorphism that is faithful to $g_1, \ldots, g_k$ is a hitting set generator for g'. [ASSS16] also showed that the problem of constructing a homomorphism ϕ that is faithful to $f_1, \ldots, f_k$ reduces to constructing a homomorphism ψ that preserves the determinant of a certain matrix. This matrix is an appropriate sub-matrix of the Jacobian of $f_1, \ldots, f_k$. Also, it was argued that its determinant is a product of sparse polynomials and so ψ was obtained from [KS01]. We use a similar argument, along with the chain rule, to show that the problem of constructing a homomorphism ϕ that is faithful to $g_1, \ldots, g_k$ reduces to constructing a homomorphism ψ that preserves the determinant of a sub-matrix of the same Jacobian *evaluated at* $A\mathbf{x}$. As this determinant is a product of polynomials in the orbit of sparse polynomials, we can use Theorem 8 to construct such a ψ.

Occur-once formulas. We prove Theorem 11 by building upon the arguments in [SV15] and linking it with Theorem 8. At first, we show two structural results (Lemma 40 and 41) for occur-once formulas. These lemmas are generalizations of similar structural results for read-once formulas shown in [SV15]. Much like in [SV15], the structural results help us show that for a "typical" occur-once formula f with a $+$ gate as the root node, there exists a variable x_i such that $\frac{\partial f}{\partial x_i}$ is a product of occur-once formulas, each of which has at most half as many non-constant leaves as f. We then use this fact to show that a hitting-set generator for $\mathrm{orb}(f)$ can be constructed from a generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$. [SV15] uses the derivatives of f in a similar way to show that a generator for f can be constructed from that for $\frac{\partial f}{\partial x_i}$ using the SV generator (see Definition 13). However, in our case, we want a generator for $\mathrm{orb}(f)$ and not just for f. For this reason, we first use the chain rule for derivatives to relate the gradient of a $g \in \mathrm{orb}(f)$ with that of f, and then argue that there exists a x_j such that a generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$ is also a generator for $\frac{\partial g}{\partial x_j}$. Finally, we use this generator for $\frac{\partial g}{\partial x_j}$ to construct a generator for g. The argument then proceeds by induction on the number of non-constant leaves. In the base case, we need a hitting set generator for orbits of sparse polynomials which we get from Theorem 8.

1.4 Related work

We give a brief account of known results on PIT and hitting sets for arithmetic circuits. The results on hitting sets for the constant-read models are most relevant to our work here. However, for the sake of completeness, we mention a few other prominent results.

Constant-read models. [SV15] initiated the study of PIT for read-once formulas. They gave a polynomial-time PIT algorithm and a quasi-polynomial time hitting set construction for sums of constantly many *preprocessed* read-once formulas (PROFs). The leaves of a PROF are labelled by univariate polynomials and every variable appears in at most one leaf; PROFs form a subclass of occur-once formulas. Later, a polynomial-time hitting set construction for the same model was given by [MV18]. A sum of k ROFs is a special case of a multilinear read-k formula. [AvMV15] gave a quasi-polynomial time hitting set construction for multilinear read-k formulas. Their construction also works for multilinear *sparse-substituted* read-k formulas, wherein the leaves are replaced by sparse polynomials and every variable appears in at most k of the sparse polynomials.

Observe that a sparse-substituted read-k formula is an occur-k formula (without the powering gates), however the arguments in [AvMV15] additionally require the multilinearity assumption.

A polynomial-time PIT for ROABPs follows from the PIT algorithm for non-commutative formulas [RS05]. [FS13b] gave a quasi-polynomial time construction of hitting sets for ROABPs, when the order of the variables is known; prior to their work, a quasi-polynomial time hitting set for multilinear, constant-width, known-variable-order ROABPs was given by [JQS10]. Building on the rank concentration by translation technique from [ASS13] and the merge-and-reduce idea from [FS13b], [FSS14] gave a quasi-polynomial time hitting set construction for multilinear ROABPs (more generally, low individual degree ROABPs). Finally, [AGKS15] obtained a quasi-polynomial time constructible hitting set for ROABPs using a different and simpler method, namely *basis isolation*, which can be thought of as a generalization of the monomial isolation method in [KS01]. It was also shown later that translation by a basis isolating weight assignment leads to rank concentration [GKST17,FGS18], and so, constructing a basis isolating weight assignment is a stronger objective than showing rank concentration by translation. This fact was used effectively in [GKST17] to design hitting sets for sums of constantly many ROABPs in quasi-polynomial time; they also gave a polynomial-time PIT algorithm for the same model. A conjunction of the basis isolation and the rank concentration techniques have been used to give more efficient constructions of hitting sets for ROABPs [GG20], sometimes under additional restrictions on the model such as commutativity and constant-width [GKS17]. The latter work also gave a polynomial-time hitting set for constant-width ROABPs, when the order of the variables is known. For read-k oblivious algebraic branching programs, [AFS+18] obtained a subexponential-time PIT algorithm.

Orbits and orbit closures. A polynomial-time hitting set for the *orbit* of the power symmetric polynomial $\mathsf{PSym}_{n,d} = x_1^d + \ldots + x_n^d$ was given by [KS19]. Observe that PSym is computable by a constant-depth occur-once formula with univariate polynomials at the leaves. So, Theorem 10 subsumes this result. Our hitting-set construction is different from the one in [KS19] which involves second order derivatives (in particular, the Hessian matrix), whereas the proofs here work with first order derivatives. Very recently and independent of our work here, [MS21] gave quasi-polynomial time hitting sets for the orbits of sparse polynomials and read-once formulas. For the orbit closures of polynomials that are computable by low-degree, polynomial-size circuits (i.e., VP circuits), [FS18,GSS18] gave PSPACE constructions of hitting sets.

Constant-depth models. The polynomial-time hitting set construction for depth-2 circuits (i.e., sparse polynomials) in [KS01] is one of the widely used results in black-box PIT. Depth-3 circuit PIT has also received a lot of attention. [DS07] gave a quasi-polynomial time PIT algorithm for depth-3 circuits with constant top fan-in by showing a structural result on the rank[13] of a circuit. [KS07] improved the complexity to polynomial-time using a different method, which is based on a generalization of the Chinese Remaindering Theorem (CRT). The structural result of [DS07], along with the rank extractors of [GR08], played a central role in devising polynomial-time constructible hitting sets for depth-3 circuits with constant top fan-in over $\mathbb{Q}$ [KS11,KS09,SS13]. Ultimately, a combination of ideas from the CRT method and rank extractors led to a polynomial-time hitting set construction for the same model over any field [SS12,SS13]. Meanwhile, [Sax08,Kay10] gave polynomial-time PIT for depth-3 powering circuits. Using ideas from [KS07] and [Sax08], [SSS13]

[13] Rank of a depth-3 circuit is the number of linearly independent linear polynomials appearing in the circuit.

gave polynomial-time PIT for the sum of a depth-3 circuit with constant top fan-in and a *semi-diagonal* circuit (which is a special kind of a depth-4 circuit). [SSS09] showed that polynomial-time PIT (hitting sets) for $\mathrm{aproj}(\mathsf{IMM}_{2,d})$ implies polynomial-time PIT (hitting sets) for depth-3 circuits.

A quasi-polynomial time hitting set for set-multilinear depth-3 circuits with known variable-partition was given by [FS12]. Independently and simultaneously, [ASS13] gave a quasi-polynomial time hitting set for set-multilinear depth-3 circuits with *unknown* variable-partition (and more generally, for constant-depth *pure* formulas [NW97]) using a different technique, namely *rank concentration by translation*. Set-multilinear depth-3 circuits (in fact, pure formulas) form a subclass of ROABPs. [dOSlV16] gave subexponential-time hitting sets for multilinear depth-3 and depth-4 formulas (and more generally, for constant-depth multilinear regular formulas) by reducing the problem to constructing hitting sets for ROABPs. For multilinear depth-4 circuits with constant top fan-in, [KMSV13] gave a quasi-polynomial time hitting set. This was improved to a polynomial-time hitting set in [SV18]. Multilinear depth-4 circuits with constant top fan-in form a subclass of depth-4 constant-occur formulas. [ASSS16] gave a unifying method based on algebraic independence to design polynomial-time hitting sets for both depth-3 circuits with constant top fan-in and constant-depth, constant-occur formulas. A generalization of depth-3 powering circuits to depth-4 is sums of powers of constant degree polynomials; [For15] gave a quasi-polynomial time hitting set for this model. Recently, a sequence of work [PS20b,PS20a,Shp19] led to a polynomial-time hitting set for depth-4 circuits with top fan-in at most 3 and bottom fan-in at most 2 via a resolution of a conjecture of [Gup14, BMS13] on the algebraic rank of the factors appearing in such circuits.

Edmonds' model. An important special case of PIT is the following problem: given $f = \det(A_0 + \sum_{i\in[n]} x_i A_i)$, where $A_i \in \mathbb{F}^{n\times n}$ is a rank-1 matrix for every $i \in [n]$ and $A_0 \in \mathbb{F}^{n\times n}$ is an arbitrary matrix, check if $f = 0$ [Edm67]. This case of PIT, which can be thought of as a generalization of PIT for determinants of read-once symbolic matrices, played an instrumental role in devising fast parallel algorithms for several problems such as perfect matching, linear matroid intersection and maximum rank matrix completion [Lov79,KUW86,MVV87,FGT16,ST17,NSV94,Mur93,GT20]. A polynomial-time PIT for this model is known [Edm79,Lov89,Mur93,Gee99,IKS10]. [GT20] gave a quasi-polynomial time hitting set via a certain derandomization of the Isolation Lemma [MVV87]. It is interesting to note that hitting sets for the orbits of polynomials computable by this model imply hitting sets for the orbit of the determinant polynomial and also the orbit of the iterated matrix multiplication polynomial via a known reduction [Val79] from ABPs to p-projections of the determinant polynomial family.

We refer the reader to the surveys [Sax09,Sax14,SY10] for more details on some of the results and the models mentioned above.

2 Preliminaries

Definition 12 (Hitting set generator)**.** Let $\mathcal{C}$ be a set of n-variate polynomials and $t \in \mathbb{N}$. A polynomial map $\mathcal{G} : \mathbb{F}^t \to \mathbb{F}^n$ is a hitting set generator for $\mathcal{C}$ if for every non-zero $f \in \mathcal{C}$, we have $f \circ \mathcal{G} \neq 0$.

We say the number of variables of $\mathcal{G}$ is t, and the degree of $\mathcal{G}$ – denoted by $\deg(\mathcal{G})$ – is the maximum of the degrees of the n polynomials that define $\mathcal{G}$. We will denote the t-variate polynomial $f \circ \mathcal{G}$ by $f(\mathcal{G})$. By treating a matrix $A \in \mathbb{F}^{n\times n}$ as a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^n$, we will denote the polynomial map $A \circ \mathcal{G}$ by $A\mathcal{G}$ and the t-variate polynomial $f \circ A\mathcal{G}$ by $f(A\mathcal{G})$. If the defining polynomials of $\mathcal{G}$ have degree d_0 and the degree of the polynomials in $\mathcal{C}$ is at most D, then the degree of $f(\mathcal{G})$ is at most $d_0 D$. Thus, if we are given the defining polynomials of $\mathcal{G}$, then we can construct a hitting set for $\mathcal{C}$ in time $\mathrm{poly}(n, (d_0D)^t)$ using the Schwartz-Zippel lemma, provided also that $|\mathbb{F}| > d_0 D$.

2.1 The Shpilka-Volkovich generator

Definition 13 (The Shpilka-Volkovich hitting set generator [SV15])**.** Assume that $|\mathbb{F}| \geq n$ and let $\alpha_1, ..., \alpha_n$ be distinct elements of $\mathbb{F}$. For $i \in [n]$, let

$$L_i(y) := \prod_{j\in[n], j\neq i} \frac{y-\alpha_j}{\alpha_i - \alpha_j}$$

be the i-th Lagrange interpolation polynomial. Then, for $t \in \mathbb{N}$, the Shpilka-Volkovich (SV) generator $\mathcal{G}_t^{SV} : \mathbb{F}^{2t} \to \mathbb{F}^n$ is defined as $\mathcal{G}_t^{SV} := \left(\mathcal{G}_t^{(1)}, ..., \mathcal{G}_t^{(n)}\right)$ where,

$$\mathcal{G}_t^{(i)}(y_1, ..., y_t, z_1, ..., z_t) = \sum_{k=1}^{t} L_i(y_k) \cdot z_k.$$

Notice that $\deg\left(\mathcal{G}_t^{(i)}\right) = n$, and $\mathcal{G}^{SV}_{t+1|(y_{t+1}=\alpha_i)} = \mathcal{G}_t^{SV} + \mathbf{e}_i \cdot z_{t+1}$, where $\mathbf{e}_i$ is the i-th standard basis vector of $\mathbb{F}^n$. Thus, $\mathrm{Img}\left(\mathcal{G}_t^{SV}\right) \subseteq \mathrm{Img}\left(\mathcal{G}_{t+1}^{SV}\right)$ and, continuing in this manner, $\mathrm{Img}\left(\mathcal{G}_t^{SV}\right) \subseteq \mathrm{Img}\left(\mathcal{G}_{t'}^{SV}\right)$ for any $t' \geq t$.

Observation 14. *Let $f \in \mathbb{F}[\mathbf{x}]$ be a non-zero polynomial that depends on only b of the $\mathbf{x}$ variables, and $g \in \mathrm{orb}(f)$. Then, g has a monomial of support at most b and $g(\mathcal{G}_b^{SV}) \neq 0$.*

Proof: Suppose that f depends on only the variables $x_1, \ldots, x_b$. Let $g = f(A\mathbf{x}) \neq 0$, where $A \in \mathrm{GL}(n, \mathbb{F})$. Suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$ for all $i \in [n]$. As A is invertible, $\ell_1, \ldots, \ell_n$ are $\mathbb{F}$-linearly independent. Let B be the $b \times n$ matrix whose i-th row is the coefficient vector of ℓ_i for all $i \in [b]$. Then, $\mathrm{rank}(B) = b$ and there are b columns $j_1, \ldots, j_b$ of B that are also linearly independent. This means the linear forms $\ell'_1, \ldots, \ell'_b$ obtained from $\ell_1, \ldots, \ell_b$ after setting the variables other than $x_{j_1}, \ldots, x_{j_b}$ to 0 are also linearly independent. Thus, $g(\ell'_1, \ldots, \ell'_b) \neq 0$ which is only possible if g has a monomial whose support is contained in $\{x_{j_1}, \ldots, x_{j_b}\}$. Now observe that $g\left(\mathcal{G}^{SV}_{b\;|(y_1=\alpha_{j_1}, y_2=\alpha_{j_2}, \ldots, y_b=\alpha_{j_b})}\right) \neq 0$. □

The following observation, which allows us to construct a hitting set generator for a polynomial f from a hitting set generator for $\frac{\partial f}{\partial x_i}$ will be used crucially in the proofs of Theorems 8, 10 and 11.

Observation 15. *Let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate, degree d polynomial, and for some $m \in \mathbb{N}$, let $\mathcal{G} : \mathbb{F}^m \to \mathbb{F}^n$ be a polynomial map of degree at most d'. If $|\mathbb{F}| > dd'$ and there is an $i \in [n]$ such that $\frac{\partial f}{\partial x_i}(\mathcal{G}) \neq 0$, then $f(\mathcal{G} + \mathcal{G}_1^{SV})$ is not a constant.*

Proof: If $\frac{\partial f}{\partial x_i}(\mathcal{G}) \neq 0$, then there is a $(\beta_1, ..., \beta_n) \in \text{Img}\,(\mathcal{G})$ such that

$$\frac{\partial f}{\partial x_i}(\beta_1, ..., \beta_n) \neq 0,$$

because $\deg\left(\frac{\partial f}{\partial x_i}\,(\mathcal{G})\right) \leq dd'$ and $|\mathbb{F}| > dd'$. Let $r(z_1) := f(\beta_1, ..., \beta_{i-1}, \beta_i + z_1, \beta_{i+1}, ..., \beta_n)$. Then,

$$\frac{\partial r}{\partial z_1}(0) = \frac{\partial f}{\partial x_i}(\beta_1, ..., \beta_n) \neq 0,$$

and so, $f(\beta_1, ..., \beta_{i-1}, \beta_i + z_1, \beta_{i+1}, ..., \beta_n)$ is not a constant. Now, $\mathcal{G} + \mathcal{G}_1^{SV}{}_{|(y_1=\alpha_i)} = \mathcal{G} + \mathbf{e}_i \cdot z_1$. Let $\text{Img}_{z_1}(\mathcal{G} + \mathcal{G}_1^{SV})$ be the "partial image" of $\mathcal{G} + \mathcal{G}_1^{SV}$ obtained by keeping the z_1 variable alive and setting all other variables to field elements. This means that $(\beta_1, ..., \beta_{i-1}, \beta_i + z_1, \beta_{i+1}, ..., \beta_n) \in \text{Img}_{z_1}(\mathcal{G} + \mathcal{G}_1^{SV})$, and hence, $f(\mathcal{G} + \mathcal{G}_1^{SV})$ is not a constant. □

2.2 Low support rank concentration

Let F be a polynomial in $\mathbf{x}$-variables with coefficients from $\mathbb{K}^{w \times w}$, where $\mathbb{K}$ is a field and $w \in \mathbb{N}$. For an $m \in \mathbb{N}$, we say that F has *support-m rank concentration* over $\mathbb{K}$ if the coefficient of every monomial in F is in the $\mathbb{K}$-span of the coefficients of the monomials of support at most m in F. Support of a monomial $\mathbf{x}^{\boldsymbol{\alpha}}$ will be denoted as $\text{Supp}\,(\mathbf{x}^{\boldsymbol{\alpha}})$.

Observation 16. *Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1} \in \mathbb{F}[\mathbf{x}]$ be computable by an ROABP of width w, and $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For an $m \in \mathbb{N}$ and $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z}) \in \mathbb{F}[\mathbf{z}]$, where $\mathbf{z}$ is a set of variables different from $\mathbf{x}$, suppose that $F(\mathbf{x} + \mathbf{t}(\mathbf{z})) := M_1(x_1 + t_1(\mathbf{z}))M_2(x_2 + t_2(\mathbf{z}))\cdots M_n(x_n + t_n(\mathbf{z})) \in \mathbb{F}(\mathbf{z})^{w \times w}[\mathbf{x}]$ has support-m rank concentration over $\mathbb{F}(\mathbf{z})$. Then, $f(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$, when viewed as a polynomial in $\mathbf{x}$-variables with coefficients from $\mathbb{F}[\mathbf{z}]$, has an $\mathbf{x}$-monomial of support at most m, provided $f \neq 0$.*

Proof: Let $F(\mathbf{x} + \mathbf{t}(\mathbf{z})) = \sum_{\boldsymbol{\alpha}} C_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, where $C_{\boldsymbol{\alpha}} \in \mathbb{F}[\mathbf{z}]^{w \times w}$. Then, $f(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z})) = \sum_{\boldsymbol{\alpha}} \left(\mathbf{1}^T \cdot C_{\boldsymbol{\alpha}} \cdot \mathbf{1}\right) \mathbf{x}^{\boldsymbol{\alpha}}$. If $f \neq 0$, then there is an $\boldsymbol{\alpha}$ such that $\mathbf{1}^T \cdot C_{\boldsymbol{\alpha}} \cdot \mathbf{1} \neq 0$. If $\text{Supp}\,(\mathbf{x}^{\boldsymbol{\alpha}}) \leq m$, then there is nothing to prove. Otherwise, as $F(\mathbf{x} + \mathbf{t}(\mathbf{z}))$ has support-m rank concentration over $\mathbb{F}(\mathbf{z})$, $C_{\boldsymbol{\alpha}}$ is in the $\mathbb{F}(\mathbf{z})$-span of $\left\{C_{\boldsymbol{\beta}} : \text{Supp}\,\left(\mathbf{x}^{\boldsymbol{\beta}}\right) \leq m\right\}$. Thus, there is a $\boldsymbol{\beta}$ with $\text{Supp}\,\left(\mathbf{x}^{\boldsymbol{\beta}}\right) \leq m$ such that $\mathbf{1}^T \cdot C_{\boldsymbol{\beta}} \cdot \mathbf{1}$ is non-zero, as $\mathbf{1}^T \cdot C_{\boldsymbol{\alpha}} \cdot \mathbf{1}$ is non-zero. □

2.3 Algebraic rank and faithful homomorphisms

We say that polynomials $f_1, \ldots, f_m \in \mathbb{F}[\mathbf{x}]$ are algebraically independent over $\mathbb{F}$, if they do not satisfy any non-trivial polynomial equation over $\mathbb{F}$, i.e., for any $p \in \mathbb{F}[y_1, \ldots, y_m]$, $p(f_1, \ldots, f_m) = 0$ only if $p = 0$. For $\mathbf{f} = (f_1, \ldots, f_m)$, the transcendence degree (i.e., the algebraic rank) of $\mathbf{f}$ over $\mathbb{F}$ is the cardinality of any maximal algebraically independent subset of $\{f_1, \ldots, f_m\}$ over $\mathbb{F}$. The notion of algebraic rank is well defined as algebraic independence satisfies the matroid properties.

For $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{F}[\mathbf{x}]^m$, let

$$J_{\mathbf{x}}(\mathbf{f}) := \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}_{m \times n}$$

denote the Jacobian matrix of $\mathbf{f}$. The following well-known lemma relates the transcendence degree of $\mathbf{f}$ over $\mathbb{F}$ – denoted by $\text{tr-deg}_{\mathbb{F}}(\mathbf{f})$ – to the rank of the Jacobian.

Lemma 17 (The Jacobian criterion)**.** *Let $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{F}[\mathbf{x}]^m$ be a tuple of polynomials of degree at most D and $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = r$. If $\text{char}(\mathbb{F}) = 0$ or $\text{char}(\mathbb{F}) > D^r$, then $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = \text{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f})$.*

Definition 18 (Faithful homomorphisms)**.** A homomorphism $\phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ is said to be *faithful* to $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{F}[\mathbf{x}]^m$ if $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = \text{tr-deg}_{\mathbb{F}}(\phi(\mathbf{f}))$.

Lemma 19 (Theorem 2.4 in [ASSS16])**.** *If a homomorphism $\phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ is faithful to $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{F}[\mathbf{x}]^m$, then for any $p \in \mathbb{F}[y_1, \ldots, y_m]$, $p(\mathbf{f}) = 0$ if and only if $p(\phi(\mathbf{f})) = 0$.*

The following lemma was proved in [ASSS16, BMS13].

Lemma 20 (Lemma 2.7 of [ASSS16])**.** *Let $\mathbf{f} = (f_1, ..., f_m)$ be a tuple of polynomials of degree at most D, $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) \leq r$, and $\text{char}(\mathbb{F}) = 0$ or $> D^r$. Let $\psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ be a homomorphism such that $\text{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f}) = \text{rank}_{\mathbb{F}(\mathbf{z})} \psi(J_{\mathbf{x}}(\mathbf{f}))$. Then, the map $\phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}, t, y_1, ..., y_r]$ that, for all $i \in [n]$, maps*

$$x_i \to \left(\sum_{j=1}^{r} y_j t^{ij} \right) + \psi(x_i)$$

is faithful to $\mathbf{f}$.

We will also need the following observation in our proofs.

Observation 21. *Let $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{F}[\mathbf{x}]^m$ be a tuple of polynomials with $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = r$. For any $A \in \text{GL}(n, \mathbb{F})$, let $g_i = f_i(A\mathbf{x})$ for all $i \in [m]$ and $\mathbf{g} = (g_1, \ldots, g_m)$. Then, $\text{tr-deg}_{\mathbb{F}}(\mathbf{g}) = r$.*

Proof: Assume without loss of generality that $f_1, \ldots, f_r$ is a transcendence basis of $\mathbf{f}$. We will show that $g_1, \ldots, g_r$ is a transcendence basis of $\mathbf{g}$. For contradiction, let $p \in \mathbb{F}[y_1, \ldots, y_r]$ be such that $p(g_1, \ldots, g_r) = 0$. Then, $p(g_1, \ldots, g_r) = p(f_1, \ldots, f_r)(A\mathbf{x}) = 0$. As A is invertible, $p(f_1, \ldots, f_r) = 0$. Because $f_1, \ldots, f_r$ are algebraically independent, this implies that $p = 0$, and so, $g_1, \ldots, g_r$ are algebraically independent. Also, if there exists a $j \in [r+1, m]$ such that $g_1, \ldots, g_r, g_j$ are algebraically independent, then for all non-zero $p \in \mathbb{F}[y_1, \ldots, y_{r+1}]$, $p(g_1, \ldots, g_r, g_j) \neq 0$. But, as $p(g_1, \ldots, g_r, g_j) = p(f_1, \ldots, f_r, f_j)(A\mathbf{x})$ and A is invertible, for all $p \neq 0$, $p(f_1, \ldots, f_r, f_j) \neq 0$. This means that $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) > r$, which contradicts the hypothesis of the observation. □

3 Hitting sets for the orbits of commutative ROABPs

The strategy. (Recap) Let $f = \mathbf{1}^T \cdot M_1(x_1) M_2(x_2) \cdots M_n(x_n) \cdot \mathbf{1}$ be a width-w commutative ROABP; here $M_i(x_i) \in \mathbb{F}^{w \times w}[x_i]$ for all $i \in [n]$. Also, let $F = M_1(x_1) M_2(x_2) \cdots M_n(x_n)$. For any

$A \in \mathrm{GL}(n, \mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. For $i \in [n]$, suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$, where ℓ_i is a linear form, and let $y_i = \ell_i(\mathbf{x})$ and $\mathbf{y} = \{y_1, \ldots, y_n\}$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. We will show that if $g \neq 0$, then there exist explicit "low" degree polynomials $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z})$, where $\mathbf{z}$ is a "small" set of variables such that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has a "low" support monomial. This will be done by proving that $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has low support rank concentration in the "$\mathbf{y}$-variables". Applying Observation 16, we will get that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has a low support $\mathbf{y}$-monomial. This will then imply that $g(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has a low support $\mathbf{x}$-monomial, provided f has low individual degree. Finally, we will plug in the SV generator to preserve the non-zeroness of g. More precisely, we will prove the following theorem at the end of Section 3.2.

Theorem 22. *Let f be an n-variate polynomial with individual degree at most d that is computable by a width-w commutative ROABP. If $|\mathbb{F}| \geq n$, then $\mathcal{G}^{SV}_{(2\lceil \log w^2 \rceil (d+1)+1)}$ is a hitting set generator for* $\mathrm{orb}(f)$.

Notations and conventions. In the analysis, we will treat $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z})$ as formal variables $\mathbf{t} = (t_1, \ldots, t_n)$ while always keeping in mind the substitution map $t_i \mapsto t_i(\mathbf{z})$. For $i \in [n]$, let $r_i = \ell_i(\mathbf{t})$. For $S \subseteq [n]$, define $\mathbf{r}_S = \{r_i : i \in S\}$. The $\mathbb{F}$-linear independence of $\ell_1, \ldots, \ell_n$ allows us to treat $\mathbf{y}$ and $\mathbf{r}$ as sets of formal variables. Notice that in this notation, $G(\mathbf{x} + \mathbf{t}) = M_1(y_1 + r_1)M_2(y_2 + r_2)\cdots M_n(y_n + r_n)$. Let $\mathbb{A}$ denote the matrix algebra $\mathbb{F}^{w \times w}$. For $i \in [n]$, let $M_i(y_i) = \sum_{e_i=0}^{d} u_{i,e_i} y_i^{e_i}$, where $u_{i,e_i} \in \mathbb{A}$ and $M_i(y_i + r_i) = \sum_{b_i=0}^{d} v_{i,b_i} y_i^{b_i}$, where $v_{i,b_i} \in \mathbb{A}[r_i] \subset \mathbb{A}[\mathbf{t}]$. As f is a commutative ROABP, $M_1(y_1), \ldots, M_n(y_n)$ commute with each other and hence u_{i,e_i} and u_{j,e_j} also commute for $i \neq j$. The following observation, which we prove in Appendix A, implies that v_{i,e_i} and v_{j,e_j} also commute for $i \neq j$.

Observation 23. *For every $i \in [n]$ and $b_i, e_i \in \{0, \ldots, d\}$,*

1. $v_{i,b_i} = \sum_{e_i=0}^{d} \binom{e_i}{b_i} \cdot r_i^{e_i - b_i} \cdot u_{i,e_i}$,

2. $u_{i,e_i} = \sum_{b_i=0}^{d} \binom{b_i}{e_i} \cdot (-r_i)^{b_i - e_i} \cdot v_{i,b_i}$,

where $\binom{a}{b} = 0$ if $a < b$.

For a set $S = \{i_1, i_2, \ldots, i_{|S|}\} \subseteq [n]$, where $i_1 < i_2 < \ldots < i_{|S|}$, the vector $(b_{i_1}, b_{i_2}, \ldots, b_{i_{|S|}})$ will be denoted by $(b_i : i \in S)$. Let $\mathrm{Supp}(\mathbf{b})$ denote the support of the vector $\mathbf{b}$ which is defined as the number of non-zero elements in it. We also define a parameter $m := 2\lceil \log w^2 \rceil + 1$.

3.1 The goal: low support rank concentration

We set ourselves the goal of proving that there exist explicit degree-n polynomials $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z})$, where $|\mathbf{z}| = 2m$, such that $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z})) = M_1(y_1 + r_1)M_2(y_2 + r_2)\cdots M_n(y_n + r_n) \in \mathbb{A}[r_1, \ldots, r_n][\mathbf{y}]$ has support-$(m-1)$ rank concentration over $\mathbb{F}(\mathbf{z})$ in the $\mathbf{y}$-variables. We will show in this and the next section that this happens if all polynomials in a certain collection of non-zero polynomials $\left\{h_S(\mathbf{r}_S) : S \subseteq \binom{[n]}{m}\right\} \subseteq \mathbb{F}[r_1, \ldots, r_n]$, where $\deg_{\mathbf{r}_S}(h_S(\mathbf{r}_S)) \leq md^{m+1}$, remain non-zero under the substitution $t_i \mapsto t_i(\mathbf{z})$.[14] The following lemma will help us achieve this goal.

Lemma 24. *Let $G, \mathbf{t}, \mathbf{z}, \mathbf{y}$ and $\mathbf{r}_S$ be as defined above. Suppose that the following two conditions are satisfied:*

[14] We do not really need the degree bound on $h_S(\mathbf{r}_S)$.

1. *For every $S \subseteq \binom{[n]}{m}$ and $(b_i : i \in S) \in \{0, \ldots, d\}^m$, there is a non-zero polynomial $h_S(\mathbf{r}_S)$ such that*

$$h_S(\mathbf{r}_S) \cdot \prod_{i \in S} v_{i,b_i} \in \mathbb{F}[\mathbf{t}]\text{-span} \left\{ \prod_{i \in S} v_{i,b'_i} \ : \ \mathrm{Supp}\left(b'_i : i \in S\right) < m \right\}.$$

2. *There exists a substitution $t_i \mapsto t_i(\mathbf{z})$ that keeps $h_S(\mathbf{r}_S)$ non-zero for all $S \subseteq \binom{[n]}{m}$.*

Then, for every $\mathbf{b} = (b_i : i \in [n]) \in \{0, \ldots, d\}^n$,

$$\prod_{i \in [n]} v_{i,b_i} \in \mathbb{F}(\mathbf{z})\text{-span} \left\{ \prod_{i \in [n]} v_{i,b'_i} \ : \ \mathrm{Supp}\left(b'_i : i \in [n]\right) < m \right\}, \tag{1}$$

and $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has support-$(m-1)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}(\mathbf{z})$.

Proof: Consider a $\mathbf{b} = (b_i : i \in [n]) \in \{0, \ldots, d\}^n$ with $\mathrm{Supp}(\mathbf{b}) \geq m$. Pick an $S \subseteq \binom{[n]}{m}$ such that $\mathrm{Supp}\left(b_i : i \in S\right) = m$. As $h_S(\mathbf{r}_S)$ is a non-zero polynomial and the substitution $t_i \mapsto t_i(\mathbf{z})$ keeps it non-zero,

$$\prod_{i \in S} v_{i,b_i} \in \mathbb{F}(\mathbf{z})\text{-span} \left\{ \prod_{i \in S} v_{i,b'_i} : \mathrm{Supp}\left(b'_i : i \in S\right) < m \right\}.$$

Also, as v_{i,b_i} and v_{j,b_j} commute when $i \neq j$,

$$\begin{aligned}
\prod_{i \in [n]} v_{i,b_i} &\in \mathbb{F}(\mathbf{z})\text{-span} \left\{ \prod_{i \in S} v_{i,b'_i} \cdot \prod_{j \in [n] \setminus S} v_{j,b_j} : \mathrm{Supp}\left(b'_i : i \in S\right) < m \right\} \\
&= \mathbb{F}(\mathbf{z})\text{-span} \left\{ \prod_{i \in [n]} v_{i,b'_i} : \mathrm{Supp}\left(b'_i : i \in S\right) < m \text{ and } b'_i = b_i \ \forall i \in [n] \setminus S \right\} \\
&\subseteq \mathbb{F}(\mathbf{z})\text{-span} \left\{ \prod_{i \in [n]} v_{i,b'_i} : \mathrm{Supp}\left(b'_i : i \in [n]\right) < \mathrm{Supp}(\mathbf{b}) \right\}.
\end{aligned}$$

Repeat the above argument for every $\mathbf{b}' \in \{0, \ldots, d\}^n$ such that $m \leq \mathrm{Supp}(\mathbf{b}') < \mathrm{Supp}(\mathbf{b})$. Continuing in this manner yields (1) for all $\mathbf{b} \in \{0, \ldots, d\}^n$. Since $\prod_{i \in [n]} v_{i,b_i}$ is the coefficient of the monomial $\mathbf{y}^{\mathbf{b}} := y_1^{b_1} \cdots y_n^{b_n}$ in $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$, the polynomial $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has support-$(m-1)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}(\mathbf{z})$. □

3.2 Achieving rank concentration

We will now see how to satisfy conditions 1 and 2 of Lemma 24 such that $\deg_{\mathbf{r}_S}\left(h_S(\mathbf{r}_S)\right) \leq md^{m+1}$, $t_i(\mathbf{z})$ is an explicit degree-n polynomial, and $|\mathbf{z}| = 2m$. Assume without loss of generality that $S = [m]$. For $\mathbf{b} = (b_1, \ldots, b_m)$ and $\mathbf{e} = (e_1, \ldots, e_m)$ in $\{0, \ldots, d\}^m$, define $\binom{\mathbf{b}}{\mathbf{e}} := \prod_{i \in [m]} \binom{b_i}{e_i}$, where, as before, $\binom{b_i}{e_i} = 0$ if $b_i < e_i$. Also, let $v_{\mathbf{b}} := \prod_{i \in [m]} v_{i,b_i}$ and $u_{\mathbf{e}} := \prod_{i \in [m]} u_{i,e_i}$. Define $\mathbf{r} := (-r_1, \ldots, -r_m)$, $\mathbf{r}^{\mathbf{b}} := \prod_{i \in [m]} (-r_i)^{b_i}$ and $\mathbf{r}^{-\mathbf{e}} := \prod_{i \in [m]} (-r_i)^{-e_i}$. We now define some vectors and matrices by fixing an arbitrary order on the elements of $\{0,\ldots,d\}^m$.

Let $V := \left(v_{\mathbf{b}} : \mathbf{b} \in \{0,\ldots,d\}^m\right)$ and $U := \left(u_{\mathbf{e}} : \mathbf{e} \in \{0,\ldots,d\}^m\right)$; V is a row vector in $\mathbb{A}[\mathbf{r}]^{(d+1)^m}$ whereas U is a row vector in $\mathbb{A}^{(d+1)^m}$. Let $C := \mathrm{diag}(\mathbf{r}^{\mathbf{b}} : \mathbf{b} \in \{0,\ldots,d\}^m)$ and $D := \mathrm{diag}(\mathbf{r}^{-\mathbf{e}} : \mathbf{e} \in \{0,\ldots,d\}^m)$; both C and D are $(d+1)^m \times (d+1)^m$ diagonal matrices. Finally, let M be a $(d+1)^m \times (d+1)^m$ numeric matrix whose rows and columns are indexed by $\mathbf{b} \in \{0,\ldots,d\}^m$ and $\mathbf{e} \in \{0,\ldots,d\}^m$ respectively. The entry of M indexed by $(\mathbf{b},\mathbf{e})$ contains $\binom{\mathbf{b}}{\mathbf{e}}$. We now make the following claim, the proof of which can be found in Appendix A.

Claim 25. *Let U, V, C, M and D be as defined above. Then, $U = VCMD$.*

In [ASS13], a very similar equation was called the *transfer equation* and we will refer to $U = VCMD$ by the same name. Let $F := \left\{\mathbf{b} \in \{0,\ldots,d\}^m : \mathrm{Supp}(\mathbf{b}) = m\right\}$; clearly, $|F| = d^m$. [15] Also, let us call the set of all vectors $\left(n_{\mathbf{e}} : \mathbf{e} \in \{0,\ldots,d\}^m\right) \in \mathbb{F}^{(d+1)^m}$ for which $\sum_{\mathbf{e}\in\{0,\ldots,d\}^m} n_{\mathbf{e}} u_{\mathbf{e}} = 0$ the *null space* of U. Then, we have the following lemma.

Lemma 26. *There are vectors $\{\mathbf{n}_{\mathbf{b}} : \mathbf{b} \in F\}$ in the null space of U such that the following holds: Let N be the $(d+1)^m \times d^m$ matrix whose rows are indexed by $\mathbf{e} \in \{0,\ldots,d\}^m$ and whose columns are indexed by $\mathbf{b} \in F$ and whose column indexed by $\mathbf{b}$ is $\mathbf{n}_{\mathbf{b}}$. Then, the square matrix $[CMDN]_F$ is invertible, where $[CMDN]_F$ is the sub-matrix of $CMDN$ consisting of only those rows of $CMDN$ that are indexed by $\mathbf{b} \in F$.*

We need the value of m in the proof of the lemma which is given in Appendix A. For now, observe that $\det([CMDN]_F) \in \mathbb{F}[\mathbf{r}]$: Every entry of $[CMDN]_F$ is a $\mathbb{F}$-linear combination of some entries of the matrix CMD. The entry of CMD indexed by $(\mathbf{b},\mathbf{e})$ is $\binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}}$, which is non-zero only if $b_i \geq e_i$ for all $i \in [m]$. In this case, $\mathbf{r}^{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}}$ is a monomial in the $\mathbf{r}$-variables. Thus, $\det([CMDN]_F)$ – which is a polynomial in the entries of $[CMDN]_F$ – is a polynomial in the $\mathbf{r}$-variables. This observation leads to the following corollary of the above lemma, which immediately gives a way to satisfy condition 1 of Lemma 24.

Corollary 27. *Let $h(\mathbf{r}) := \det([CMDN]_F)$. Then, $\deg_{\mathbf{r}}(h(\mathbf{r})) \leq md^{m+1}$. Also, for every $\mathbf{b} \in F$,*

$$h(\mathbf{r}) \cdot v_{\mathbf{b}} \in \mathbb{F}[\mathbf{t}]\text{-span}\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,\ldots,d\}^m \text{ and } \mathrm{Supp}\left(\mathbf{b}'\right) < m\right\}.$$

Proof: As mentioned in the previous paragraph, every entry of $[CMDN]_F$ is an $\mathbb{F}$-linear combination of the entries of CMD which themselves are of the form $\binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}}$. As, $\mathbf{b},\mathbf{e} \in \{0,\ldots,d\}^m$ and $\mathbf{r}$ has m variables, the degree of $\mathbf{r}^{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}}$ in the $\mathbf{r}$-variables is at most md. Since $[CMDN]_F$ is a $d^m \times d^m$ matrix, the degree of $\det([CMDN]_F)$ in the $\mathbf{r}$-variables is at most md^{m+1}.

$U = VCMD$ implies that $VCMDN = 0$. Let V_F be the sub-vector of V consisting solely of the entries indexed by $\mathbf{b} \in F$. As $VCMDN = 0$, every entry of $V_F\ [CMDN]_F$ is in

$$\mathbb{F}[\mathbf{t}]\text{-span}\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,\ldots,d\}^m \setminus F\right\} = \mathbb{F}[\mathbf{t}]\text{-span}\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,\ldots,d\}^m \text{ and } \mathrm{Supp}(\mathbf{b}') < m\right\}.$$

So by multiplying $V_F\ [CMDN]_F$ by the adjoint of $[CMDN]_F$, we get that every entry of V_F times $\det([CMDN]_F)$, i.e., $h(\mathbf{r}) \cdot v_{\mathbf{b}}$ where $\mathbf{b} \in F$ is in $\mathbb{F}[\mathbf{t}]$-span $\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,\ldots,d\}^m \text{ and } \mathrm{Supp}(\mathbf{b}') < m\right\}$. □

[15] There is a slight overloading of notation here: We have used F before at the beginning of Section 3 to denote the product $M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. However, since all our arguments involve only $G = F(A\mathbf{x})$ and not F, we would use F in this section to denote the set that is mentioned here.

The following claim about $h(\mathbf{r})$ gives us a way to satisfy condition 2 of Lemma 24.

Claim 28. *The polynomial $h(\mathbf{r})$, when viewed as a polynomial in the $\mathbf{t}$-variables after setting $r_i = \ell_i(\mathbf{t})$, has a $\mathbf{t}$-monomial of support at most m.*

Proof: The polynomial $h(\mathbf{r}) = h(\ell_1(\mathbf{t}), \ldots, \ell_m(\mathbf{t})) \neq 0$ as $[CMDN]_F$ is an invertible matrix and $\ell_1, \ldots, \ell_m$ are $\mathbb{F}$-linearly independent. Then, as there are only m $\mathbf{r}$-variables, the claim follows immediately from Observation 14. □

Thus, by substituting $\mathcal{G}_m^{SV}$ for $\mathbf{t}$, the polynomial $h(\mathbf{r})$ remains non-zero, satisfying condition 2. Note that the number of variables in $\mathcal{G}_m^{SV}$, i.e., $|\mathbf{z}| = 2m$ and its degree is n. We are now in a position to prove Theorem 22.

Proof of Theorem 22

Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$ be a width-w commutative ROABP having individual degree at most d; here $M_i \in \mathbb{F}^{w\times w}[x_i]$ for all $i \in [n]$. Also, let $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For any $A \in \mathrm{GL}(n, \mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. Suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$ and let $y_i = \ell_i(\mathbf{x})$ for all $i \in [n]$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. In Sections 3.1 and 3.2, we have shown that $G\left(\mathbf{x} + \mathcal{G}_m^{SV}\right)$ has support-$(m-1)$ rank concentration (for $m = 2\left\lceil \log w^2 \right\rceil + 1$) over $\mathbb{F}(\mathbf{z})$ in the $\mathbf{y}$-variables; the $\mathbf{z}$-variables are the variables introduced by the $\mathcal{G}_m^{SV}$ generator. From Observation 16, if $g(\mathbf{x}) \neq 0$, then $g\left(\mathbf{x} + \mathcal{G}_m^{SV}\right)$, when viewed as a polynomial over $\mathbb{F}[\mathbf{z}]$ in the $\mathbf{y}$-variables[16], has a $\mathbf{y}$-monomial of support at most $m-1$. Let the $\mathbf{y}$-degree of this monomial be D'. As the individual degree of every $\mathbf{x}$-variable in f is at most d, the individual degree of every $\mathbf{y}$-variable in g is also at most d. Thus, $D' \leq (m-1)d$. As the homogeneous component of $g\left(\mathbf{x} + \mathcal{G}_m^{SV}\right)$ of $\mathbf{y}$-degree D' is non-zero, the homogeneous component of $g\left(\mathbf{x} + \mathcal{G}_m^{SV}\right)$ (now viewed as polynomial over $\mathbb{F}[\mathbf{z}]$ in the $\mathbf{x}$-variables) of $\mathbf{x}$-degree D' must also be non-zero, since $\ell_1, \ldots, \ell_n$ are linearly independent. This means that $g(\mathbf{x} + \mathcal{G}_m^{SV})$, when viewed as a polynomial over $\mathbb{F}[\mathbf{z}]$ in the $\mathbf{x}$-variables, has an $\mathbf{x}$-monomial of support (in fact, degree) at most $D' \leq (m-1)d$. Thus, $g\left(\mathcal{G}_{(m-1)d}^{SV} + \mathcal{G}_m^{SV}\right) \neq 0$. Now, it follows directly from the definition of the SV generator that $\mathcal{G}_{(m-1)d}^{SV} + \mathcal{G}_m^{SV} = \mathcal{G}_{m+(m-1)d}^{SV}$ and so $g\left(\mathcal{G}_{m+(m-1)d}^{SV}\right) \neq 0$. Replacing m by its value $2\left\lceil \log w^2 \right\rceil + 1$ proves the theorem. Note that the SV generator needs $|\mathbb{F}| \geq n$.

3.3 Proofs of Theorems 6 and 7

Proof of Theorem 6: Let f be an n-variate polynomial computed by a width-w commutative ROABP of individual degree at most d, and $g \in \mathrm{orb}\,(f)$. Then, from Theorem 22, $g\left(\mathcal{G}_{(2\lceil \log w^2 \rceil(d+1)+1)}^{SV}\right) \neq 0$ whenever $g \neq 0$. Now, $\mathcal{G}_{(2\lceil \log w^2 \rceil(d+1)+1)}^{SV}$ has $2(2\left\lceil \log w^2 \right\rceil(d+1)+1)$ variables, and is of degree n. So $g\left(\mathcal{G}_{(2\lceil \log w^2 \rceil(d+1)+1)}^{SV}\right)$ also has $2\left(2\left\lceil \log w^2 \right\rceil(d+1)+1\right)$ variables. Since the individual degree of f is at most d, the $\deg(f) = \deg(g) \leq nd$. So the degree of $g\left(\mathcal{G}_{(2\lceil \log w^2 \rceil(d+1)+1)}^{SV}\right)$ is at most n^2d. Thus, as $|\mathbb{F}| > n^2d$, a hitting set for g can be computed in time $\left(n^2d+1\right)^{\left(2\left\lceil \log w^2 \right\rceil(d+1)+1\right)} = (nd)^{O(d\log w)}$. □

[16] This we can do as $g\left(\mathbf{x} + \mathcal{G}_m^{SV}\right) = \mathbf{1}^T \cdot G\left(\mathbf{x} + \mathcal{G}_m^{SV}\right) \cdot \mathbf{1}$, and $G\left(\mathbf{x} + \mathcal{G}_m^{SV}\right)$ can be viewed as a polynomial over $\mathbb{A}[\mathbf{z}]$ in the $\mathbf{y}$-variables.

Proof of Theorem 7: Let f be an n-variate polynomial such that $f = \sum_{i\in[s]} \prod_{j\in[n]} f_{i,j}(x_j)$, where each $f_{i,j}(x_j)$ is a univariate polynomial in x_j. For all $j \in [n]$, define the matrix $M_j = \text{diag}(f_{1,j}, \ldots, f_{s,j})$. Then $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$. Moreover, as the matrices $M_1, \ldots, M_n$ are diagonal, they commute with each other. Hence, f is computed by a width-s commutative ROABP and the theorem follows from Theorem 6. □

Hitting set generator for the orbits of sparse polynomials. Let $f = \sum_{j\in[s]} c_j x_1^{e_{j,1}} \cdots x_n^{e_{j,n}}$ be a sparse polynomial, where $c_j \in \mathbb{F}$ for $j \in [s]$. Observe that f can be computed by a commutative ROABP as follows: Let $M_1(x_1) := \text{diag}(c_1 x_1^{e_{1,1}}, \ldots c_s x_1^{e_{s,1}})$ and, for $2 \le i \le n$, let $M_i(x_i) := \text{diag}(x_i^{e_{1,i}}, \ldots x_i^{e_{s,i}})$. Then, $f = \mathbf{1}^T \cdot M_1(x_1)\cdots M_n(x_n) \cdot \mathbf{1}$. Notice that, as all matrices M_i are diagonal, it is a commutative ROABP and its width is s. Thus, if the individual degree of f is at most d, then Theorem 22 implies that $\mathcal{G}^{SV}_{(2\lceil \log s^2\rceil(d+1)+1)}$ is a hitting set generator for $\text{orb}_{\mathbb{F}}(f)$. However, as mentioned in the introduction, a parallel and independent work [MS21] shows that for the case of sparse polynomials the low individual degree restriction can be removed. They prove the following theorem whose proof we provide in Appendix C.

Theorem 29. *Let f be an n-variate, s-sparse polynomial of degree d and $g \in \text{orb}(f)$. Also, let $|\mathbb{F}| > nd$ and $\text{char}(\mathbb{F}) = 0$ or $> d$. Then, $g \neq 0$ implies $g\left(\mathcal{G}^{SV}_{\lceil \log s\rceil+1}\right) \neq 0$. In fact, if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{\lceil \log s\rceil+1}\right)$.*

We will make use of the above theorem in Sections 5 and 6 to prove Theorems 10 and 11, respectively.

4 Hitting sets for the orbits of multilinear constant-width ROABPs

The strategy. (Recap) Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$ be a multilinear, width-w ROABP; here $M_i(x_i) \in \mathbb{F}^{w\times w}[x_i]$ for all $i \in [n]$. Also, let $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For any $A \in \text{GL}(n, \mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. For $i \in [n]$, suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$, where ℓ_i is a linear form, and let $y_i = \ell_i(\mathbf{x})$ and $\mathbf{y} = \{y_1, \ldots, y_n\}$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. Just like in the previous section, we will show that if $g \neq 0$, then there exist explicit "low" degree polynomials $t_1(\mathbf{z}), \ldots, t_n(\mathbf{z})$, where $\mathbf{z}$ is a "small" set of variables such that $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has "low" support rank concentration in the "$\mathbf{y}$-variables". While in the rank concentration argument in the previous section the $\mathbf{x}$-variables were translated only once, here the translations can be thought of as happening sequentially and in stages. There will be $\lceil \log n\rceil$ stages with each stage also consisting of multiple translations. After the p-th stage, the product of any 2^p consecutive matrices in G will have low support rank concentration in the $\mathbf{y}$-variables. Thus, after $\lceil \log n\rceil$ stages, we will have low support rank concentration in the $\mathbf{y}$-variables for $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$.

Notations and conventions. Much like in the previous section, we will first translate the $\mathbf{x}$-variables by the $\mathbf{t}$-variables and then substitute the $\mathbf{t}$-variables by low degree polynomials in a small set of variables. We will translate the $\mathbf{x}$-variables by $\lceil \log n\rceil$ groups of $\mathbf{t}$-variables, $\mathbf{t}_1, \ldots, \mathbf{t}_{\lceil \log n\rceil}$. For all $p \in \lceil \log n\rceil$, the group $\mathbf{t}_p$ will have $\mu := w^2 + \lceil \log w^2\rceil$ sub-groups of $\mathbf{t}$-variables, $\mathbf{t}_{p,1}, \ldots, \mathbf{t}_{p,\mu}$. For all $p \in \lceil \log n\rceil$ and $q \in [\mu]$, $\mathbf{t}_{p,q} := \{t_{p,q,1}, \ldots, t_{p,q,n}\}$. Thus, finally the translation will look like

$$x_i \to x_i + \sum_{\substack{p\in\lceil\log n\rceil,\\ q\in[\mu]}} t_{p,q,i}$$

for all $i \in [n]$. Finally, we will substitute the $\mathbf{t}$-variables as $t_{p,q,i} \mapsto s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(i)}$, where $\beta_{p,q}(i)$ will be fixed later in the analysis. Let $r_{p,q,i} := \ell_i\left(\mathbf{t}_{p,q}\right)$; notice that for all $i \in [n]$, y_i is translated as

$$y_i \to y_i + \sum_{\substack{p\in\lceil\log n\rceil,\\ q\in[\mu]}} \ell_i\left(\mathbf{t}_{p,q}\right) = y_i + \sum_{\substack{p\in\lceil\log n\rceil,\\ q\in[\mu]}} r_{p,q,i}.$$

For the purpose of analysis, we will think of the translation as happening sequentially in the order $\mathbf{t}_{1,1}, \ldots, \mathbf{t}_{1,\mu}, \mathbf{t}_{2,1}, \ldots, \mathbf{t}_{2,\mu}, \ldots, \mathbf{t}_{n,1}, \ldots \mathbf{t}_{n,\mu}$, i.e., we will first translate by $\mathbf{t}_{1,1}$, then by $\mathbf{t}_{1,2}$, and so on. Let us denote the order thus imposed on the set $\{(p,q) : p \in [\lceil \log n \rceil], q \in [\mu]\}$ by $\prec$.

For a set $S = \{i_1, i_2, \ldots, i_{|S|}\} \subseteq [n]$, where $i_1 < i_2 < \ldots < i_{|S|}$, the vector $(b_{i_1}, b_{i_2}, \ldots, b_{i_{|S|}})$ will be denoted by $(b_i : i \in S)$. Let $\mathrm{Supp}\,(\mathbf{b})$ denote the support of the vector $\mathbf{b}$ which is defined as the number of non-zero elements in it.

The inductive argument given on the next two subsections is inspired by the "merge-and-reduce" idea from [FS13b, FSS14].

4.1 Low support rank concentration: an inductive argument

In this and the next sections, we will prove the following lemma. Let $\mathbb{A} := \mathbb{F}^{w\times w}$.

Lemma 30. *There exist $\{\beta_{p,q}(i) : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]\} \subset \mathbb{Z}_{\geq 0}$, such that*

$$G\left(x_1 + \sum_{\substack{p\in\lceil\log n\rceil,\\ q\in[\mu]}} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{\substack{p\in\lceil\log n\rceil,\\ q\in[\mu]}} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right),$$

when treated as a polynomial in the $\mathbf{y}$-variables over $\mathbb{A}[r_{p,q,i} : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]]$, has support-$\mu$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : p \in [\lceil \log n \rceil], q \in [\mu]\right)$. Moreover, $\{\beta_{p,q}(i) : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]\}$ can be found in time $n^{O(w^4)}$ and each $\beta_{p,q}(i) \leq n^{O(w^4)}$.

We will prove this lemma by induction on (p,q). Let us call $\{\beta_{p,q}(i) : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]\}$ *efficiently computable and good* if they can be found in time $n^{O(w^4)}$ and each $\beta_{p,q}(i) \leq n^{O(w^4)}$. Precisely, the induction hypothesis is as follows.

Induction hypothesis. Just before translating by $\mathbf{t}_{p^*,q^*}$-variables, we will assume that the following is true: there exist efficiently computable and good $\{\beta_{p,q}(i) : (p,q) \prec (p^*,q^*)\}$ such that the product of any 2^{p^*} consecutive matrices in

$$G\left(x_1 + \sum_{(p,q)\prec(p^*,q^*)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{(p,q)\prec(p^*,q^*)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)$$

has support-$(2\mu - (q^* - 1))$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \prec (p^*,q^*)\right)$.

Base case. In the base case, $(p^*, q^*) = (1,1)$. Observe that we can assume that $w \geq 2$; if $w = 1$, then g is a product of univariates and the existence of a polynomial time hitting set follows from Observation 14. For any $w \geq 2$, $2 \leq 2\mu$. As a product of any two consecutive matrices in G has support $2 \leq 2\mu$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}$, the base case is satisfied.

Induction step. We need to show that there exist efficiently computable and good $\left\{\beta_{p^*,q^*}(i) : i \in [n]\right\}$ such that after translating by $\mathbf{t}_{p^*,q^*}$ and substituting $t_{p^*,q^*,i} \to s_{p^*,q^*} \cdot z_{p^*,q^*}^{\beta_{p^*,q^*}(i)}$, the product of any 2^{p^*} consecutive matrices in

$$G\left(x_1 + \sum_{(p,q) \preccurlyeq (p^*,q^*)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{(p,q) \preccurlyeq (p^*,q^*)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)$$

has support-$(2\mu - q^*)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \preccurlyeq (p^*,q^*)\right)$. If $q^* < \mu$, then this would mean that the induction hypothesis holds immediately before translation by $\mathbf{t}_{p^*,q^*+1}$. On the other hand, if $q^* = \mu$, then the following easy-to-verify observation implies that the induction hypothesis holds immediately before translation by $\mathbf{t}_{p^*+1,1}$.

Observation 31. *Suppose that $\left\{\beta_{p,q}(i) : (p,q) \preccurlyeq (p^*,\mu)\right\}$ are such that the product of any 2^{p^*} consecutive matrices in*

$$G\left(x_1 + \sum_{(p,q) \preccurlyeq (p^*,\mu)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{(p,q) \preccurlyeq (p^*,\mu)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)$$

has support-μ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \preccurlyeq (p^*,\mu)\right)$. Then the product of any 2^{p^*+1} consecutive matrices in

$$G\left(x_1 + \sum_{(p,q) \preccurlyeq (p^*,\mu)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{(p,q) \preccurlyeq (p^*,\mu)} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)$$

has support-2μ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \preccurlyeq (p^*,\mu)\right)$.

Simplifying notations for the ease of exposition. By focusing on the induction step, we will henceforth denote $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \prec (p^*,q^*)\right)$ by $\mathbb{F}$, and for all $i \in [n]$,

$$M_i\left(y_j + \sum_{(p,q) \prec (p^*,q^*)} \ell_i\left(s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)\right)$$

by $M_i(y_i)$, $t_{p^*,q^*,i}$ by t_i, $r_{p^*,q^*,i}$ by r_i, s_{p^*,q^*} by s, z_{p^*,q^*} by z and $\beta_{p^*,q^*}(i)$ by $\beta(i)$.

Without loss of generality, we shall consider the product $M_1(y_1 + r_1) \cdots M_m(y_n + r_m)$ of the first $m = 2^{p^*}$ matrices. Our goal is to show that there exist efficiently computable and good $\{\beta(i) : i \in [m]\}$ such that after substituting $t_i \to s \cdot z^{\beta(i)}$, the above product has support-$(2\mu - q^*)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}(s,z)$ *assuming that $M_1(y_1) \cdots M_m(y_m)$ has support-$(2\mu - (q^* - 1))$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}$.*

4.2 Details of the induction step

Recalling some notations. Before we show how to achieve rank concentration, let us recall some notations defined in Section 3. While in Section 3, the individual degree is d, here the individual degree is 1 and so, we modify the definitions accordingly. $\mathbb{A}$ is used to denote the matrix algebra $\mathbb{F}^{w \times w}$. For $i \in [m]$, $M_i(y_i) = \sum_{e_i=0}^{1} u_{i,e_i} y_i^{e_i}$, where $u_{i,e_i} \in \mathbb{A}$ and $M_i(y_i + r_i) = \sum_{b_i=0}^{1} v_{i,b_i} y_i^{b_i}$, where $v_{i,b_i} \in \mathbb{A}[r_i] \subset \mathbb{A}[\mathbf{t}]$. For $\mathbf{b} = (b_1, \ldots, b_m)$ and $\mathbf{e} = (e_1, \ldots, e_m)$ in $\{0,1\}^m$, $\binom{\mathbf{b}}{\mathbf{e}} := \prod_{i \in [m]} \binom{b_i}{e_i}$. Also, $v_{\mathbf{b}} := \prod_{i \in [m]} v_{i,b_i}$ and $u_{\mathbf{e}} := \prod_{i \in [m]} u_{i,e_i}$. Moreover, $\mathbf{r} := (-r_1, \ldots, -r_m)$, $\mathbf{r}^{\mathbf{b}} := \prod_{i \in [m]} (-r_i)^{b_i}$ and $\mathbf{r}^{-\mathbf{e}} := \prod_{i \in [m]} (-r_i)^{-e_i}$. Let $\mathbf{t} := (t_1, \ldots, t_n)$.

The following vectors and matrices are defined by fixing an arbitrary order on the elements of $\{0,1\}^m$. $V := \left(v_{\mathbf{b}} : \mathbf{b} \in \{0,1\}^m\right)$ and $U := \left(u_{\mathbf{e}} : \mathbf{e} \in \{0,1\}^m\right)$; V is a row vector in $\mathbb{A}[\mathbf{r}]^{2^m}$ whereas U is a row vector in $\mathbb{A}^{2^m}$. Both $C := \mathrm{diag}(\mathbf{r}^{\mathbf{b}} : \mathbf{b} \in \{0,1\}^m)$ and $D := \mathrm{diag}(\mathbf{r}^{-\mathbf{e}} : \mathbf{e} \in \{0,1\}^m)$ are $2^m \times 2^m$ diagonal matrices. Finally, M is a $2^m \times 2^m$ numeric matrix whose rows and columns were indexed by $\mathbf{b} \in \{0,1\}^m$ and $\mathbf{e} \in \{0,1\}^m$, respectively. The entry of M indexed by $(\mathbf{b}, \mathbf{e})$ contains $\binom{\mathbf{b}}{\mathbf{e}}$. The proof of the following transfer equation is same as the proof of Claim 25.

Claim 32. *Let U, V, C, M and D be as defined above. Then, $U = VCMD$.*

Let $F := \left\{\mathbf{b} \in \{0,1\}^m : \mathrm{Supp}(\mathbf{b}) > 2\mu - q^*\right\}$.[17] Also, recall that the the *null space* of U is the set of all vectors $\left(n_{\mathbf{e}} : \mathbf{e} \in \{0,1\}^m\right) \in \mathbb{F}^{2^m}$ for which $\sum_{\mathbf{e} \in \{0,1\}^m} n_{\mathbf{e}} u_{\mathbf{e}} = 0$. We have the following lemma.

Lemma 33. *There are vectors $\{\mathbf{n_b} : \mathbf{b} \in F\}$ in the null space of U such that the following holds: Let N be the $2^m \times |F|$ matrix whose rows are indexed by $\mathbf{e} \in \{0,1\}^m$ and whose columns are indexed by $\mathbf{b} \in F$ and whose column indexed by $\mathbf{b}$ is $\mathbf{n_b}$. Then, the square matrix $[CMDN]_F$ is invertible, where $[CMDN]_F$ is the sub-matrix of $CMDN$ consisting of only those rows of $CMDN$ that are indexed by F. Also, $\det\left([CMDN]_F\right) \in \mathbb{F}[\mathbf{r}] \subset \mathbb{F}[\mathbf{t}]$ can be expressed as the ratio of a polynomial in $\mathbb{F}[\mathbf{t}]$ that contains a monomial of degree at most $2w^2\mu$ in the $\mathbf{t}$-variables and a product of linear forms in $\mathbb{F}[\mathbf{t}]$.*

The proof of this lemma, which uses the value of μ, is given in Appendix B. We now complete the induction step using this lemma. As $\det([CMDN]_F)$ is a polynomial in $\mathbb{F}[\mathbf{r}]$ we get the following corollaries.

Corollary 34. *Let $h(\mathbf{r}) := \det([CMDN]_F)$. Then, for every $\mathbf{b} \in F$,*

$$h(\mathbf{r}) \cdot v_{\mathbf{b}} \in \mathbb{F}[\mathbf{t}]\text{-span}\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,1\}^m \text{ and } \mathrm{Supp}\left(\mathbf{b}'\right) \leq 2\mu - q^*\right\}. \tag{2}$$

Proof: Same as the proof of Corollary 27. □

Corollary 35. *Suppose $\{\beta(i) : i \in [n]\}$ are such that the substitution $t_i \mapsto s \cdot z^{\beta(i)}$ keeps all non-zero polynomials in $\mathbb{F}[\mathbf{t}]$ containing a monomial of degree at most $2w^2\mu$ in the $\mathbf{t}$-variables non-zero. Then, the product $M_1(y_1 + r_1) \cdots M_m(y_m + r_m)$ has support-$(2\mu - q^*)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}(s, z)$ after substituting $t_i \to s \cdot z^{\beta(i)}$.*

Proof: Multiply both sides of (2) by $\left(h(\mathbf{r})\right)^{-1}$ after substituting $t_i \mapsto s \cdot z^{\beta(i)}$. □

[17] There is a slight overloading of notation here: We have used F before at the beginning of Section 4 to denote the product $M_1(x_1)M_2(x_2) \cdots M_n(x_n)$. However, since all our arguments involve only $G = F(A\mathbf{x})$ and not F, we would use F in this section to denote the set that is mentioned here.

We now prove that $\{\beta(i) : i \in [n]\}$ as in the above corollary can be computed efficiently.

Claim 36. *There exist $\{\beta(i) : i \in [n]\}$ such that the substitution $t_i \mapsto s \cdot z^{\beta(i)}$ keeps all non-zero polynomials in $\mathbb{F}[\mathbf{t}]$ containing a monomial of degree at most $2w^2\mu$ in the $\mathbf{t}$-variables non-zero. Moreover, we can find all the $\beta(i)$ in time $n^{O(w^4)}$ and each $\beta(i) \leq n^{O(w^4)}$.*

Proof: Because of the presence of s, the substitution $t_i \mapsto s \cdot z^{\beta(i)}$ keeps any two homogeneous polynomials of different degrees distinct (unless it maps both of them to 0). So, we need to find $\{\beta(i) : i \in [n]\}$ such that the substitution $t_i \mapsto z^{\beta(i)}$ maps any two $\mathbf{t}$-monomials of degree at most $2w^2\mu = O(w^4)$ to distinct monomials in z. Now, there are at most $\binom{n+2w^2\mu}{2w^2\mu} = n^{O(w^4)}$ such monomials. So, [KS01] implies that we can find a $\{\beta(i) : i \in [n]\}$ where each $\beta(i) \leq n^{O(w^4)}$ in time $n^{O(w^4)}$. □

This completes the induction step. We now ready to prove Lemma 30 stated in Section 4.1.

Proof of Lemma 30. So far we have proved that there exist $\{\beta_{p,q}(i) : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]\}$, such that

$$G\left(x_1 + \sum_{\substack{p \in \lceil \log n \rceil, \\ q \in [\mu]}} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(1)}, \ldots, x_n + \sum_{\substack{p \in \lceil \log n \rceil, \\ q \in [\mu]}} s_{p,q} \cdot z_{p,q}^{\beta_{p,q}(n)}\right)$$

has support-μ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(s_{p,q}, z_{p,q} : p \in [\lceil \log n \rceil], q \in [\mu]\right)$. Moreover, for each (p,q), we can find all $\beta_{p,q}(i)$ in time $n^{O(w^4)}$ and each $\beta_{p,q}(i) \leq n^{O(w^4)}$. However, since the algorithm that follows from [KS01] is oblivious, the $\beta_{p,q}(i)$ found for some fixed (p,q) can be used for all values of (p,q). This proves the lemma.

4.3 Proof of Theorem 9

Let $f = \mathbf{1}^T \cdot M_1(x_1)M_2(x_2)\cdots M_n(x_n) \cdot \mathbf{1}$ be a multilinear width-w ROABP; here $M_i(x_i) \in \mathbb{F}^{w \times w}[x_i]$ for all $i \in [n]$. Also, let $F = M_1(x_1)M_2(x_2)\cdots M_n(x_n)$. For any $A \in \mathrm{GL}(n, \mathbb{F})$, let $g = f(A\mathbf{x})$ and $G = F(A\mathbf{x})$. For $i \in [n]$, suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$, where ℓ_i is a linear form, and let $y_i = \ell_i(\mathbf{x})$ and $\mathbf{y} = \{y_1, \ldots, y_n\}$. Then, $g = \mathbf{1}^T \cdot M_1(y_1)M_2(y_2)\cdots M_n(y_n) \cdot \mathbf{1}$ and $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. Let $\mu = w^2 + \lceil \log w^2 \rceil$. From Lemma 30, there exist polynomials, say $t_1, \ldots, t_n$, in $\mathbb{F}\left[s_{p,q}, z_{p,q} : p \in [\lceil \log n \rceil], q \in [\mu]\right]$ of degree at most $n^{O(w^4)}$ such that $G(x_1 + t_1, \ldots, x_n + t_n)$ has support-μ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}\left(\{s_{p,q}, z_{p,q}\}_{p,q}\right)$. Moreover, these polynomials can be computed in time $n^{O(w^4)}$. Suppose that $g \neq 0$. Then, from Observation 16, $g(x_1 + t_1, \ldots, x_n + t_n)$ has a support-μ, $\mathbf{y}$-monomial when viewed as a polynomial over $\mathbb{F}\left[\{s_{p,q}, z_{p,q}\}_{p,q}\right]$ in the $\mathbf{y}$-variables. Since f is multilinear, as seen in the proof of Theorem 22, $g(x_1 + t_1, \ldots, x_n + t_n)$ has a support-μ, $\mathbf{x}$-monomial. Thus, $g\left(\mathcal{G}_\mu^{SV} + (t_1, \ldots, t_n)\right) \neq 0$. Now, $g\left(\mathcal{G}_\mu^{SV} + (t_1, \ldots, t_n)\right)$ is a polynomial in $2\mu + \mu \cdot \lceil \log n \rceil$ variables over $\mathbb{F}$. Also, its degree is at most $n^{O(w^4)}$. So, if $|\mathbb{F}| > n^{O(w^4)}$, a hitting set for g can be computed in time

$$n^{O(w^4\cdot\mu\cdot\log n)} = n^{O\left(w^6\cdot\log n\right)}.$$

This, along with the time required to compute $t_1, \ldots, t_n$, still gives a $n^{O\left(w^6\cdot\log n\right)}$-time hitting set for g.

5 Hitting sets for the orbits of depth four, constant-occur formulas

In this section, we will show the existence of quasi-polynomial time hitting sets for the orbits of depth-4, occur-k formulas. Without loss of generality, we will assume that the top-most gate of a formula is a $+$ gate. The argument that we present in this section for the depth $\Delta = 4$ case of Theorem 10 can be generalised to work for arbitrary depths. The general argument can be found in Appendix D.

For some $k \in \mathbb{N}$, let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate, degree-D polynomial computed by a $(4, k, s)$ formula, i.e., a depth-4, occur-k formula of size-s. We will identify f with the formula computing it. As mentioned in Section 1.3, we first upper bound the top fan-in of f in Section 5.1 and then use the notion of faithful homomorphisms to construct hitting sets for $\mathrm{orb}(f)$.

5.1 Upper bounding the top fan-in of f

To upper bound the fan-in of f, we show that for all $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a depth-4, occur-k' formula with top fan-in at most k; here k' is not too large compared to k (see Claim 37 below). We then argue in Claim 38 that there exists an $i \in [n]$ such that a hitting set generator for $\mathrm{orb}(f)$ can be constructed using a hitting set generator for $\mathrm{orb}(\frac{\partial f}{\partial x_i})$. Thus, by overloading the notation and referring to $\frac{\partial f}{\partial x_i}$ as f, we can assume that the top fan-in of f is at most k.

Claim 37. *Let f be a $(4, k, s)$ formula. Then, for every $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a $(4, 2k^2, 2ks)$ formula with top fan-in bounded by k.*

Proof: Let $x = x_i$. Let $f = \sum_{i\in[m]} f_i$, and x be present only in $f_1, \ldots, f_r$, where $r \le k$. Furthermore, for all $i \in [r]$, let $f_i = \prod_{j\in m_i} q_{i,j}^{e_{i,j}}$ and x be present only in $q_{i,1}, \ldots q_{i,r_i}$, $\sum_{i\in[r]} r_i \le k$. Here, $q_{i,j}$ are s-sparse polynomials. Now,

$$\begin{aligned}\frac{\partial f}{\partial x} &= \sum_{i\in[r]} \left(\prod_{j=r_i+1}^{m_i} q_{i,j}^{e_{i,j}}\right) \cdot \left(\sum_{j\in[r_i]} e_{i,j}\frac{\partial q_{i,j}}{\partial x} \cdot q_{i,j}^{e_{i,j}-1} \cdot \prod_{\substack{j'\in[r_i]\\ j'\neq j}} q_{i,j'}^{e_{i,j'}}\right)\\ &= \sum_{i\in[r]} \sum_{j\in[r_i]} \left(e_{i,j}\frac{\partial q_{i,j}}{\partial x} \cdot \prod_{j'\in[m_i]} q_{i,j'}^{e'_{i,j'}}\right),\end{aligned}$$

where $e'_{i,j'} = e_{i,j'}$ for $j' \neq j$ and $e'_{i,j} = e_{i,j} - 1$. First of all, notice that the top fan-in of $\frac{\partial f}{\partial x}$ is at most $\sum_{i\in[r]} r_i \le k$, its depth is 4, and as the leaves are still $q_{i,j}$ or $\frac{\partial q_{i,j}}{\partial x}$, the sparsity of the polynomials labelling the leaves are also at most s. However, the size and the occur may change.

For all $i \in [r]$, let the occur of f_i be $p_i \leq k$; then the occur of $\prod_{j' \in [m_i]} q_{i,j'}^{e'_{i,j'}}$ is at most p_i. Also, as $\frac{\partial q_{i,j}}{\partial x}$ is an s-sparse polynomial, its occur is 1. Then, the occur of $\frac{\partial f}{\partial x}$ is at most

$$\sum_{i \in [r]} r_i (1 + p_i) \leq \sum_{i \in [r]} r_i + \sum_{i \in [r]} r_i k \leq k + k^2 \leq 2k^2.$$

Similarly, suppose that the size of f_i is $s_i \leq s - 1$ [18]; then the size of $\prod_{j' \in [m_i]} q_{i,j'}^{e'_{i,j'}}$ is at most $s_i - 1$ (as $e'_{i,j} = e_{i,j} - 1$). Also, as the size of $q_{i,j}$ is $\leq s$, the size of $\frac{\partial q_{i,j}}{\partial x}$ is at most s. So, the size of $\frac{\partial f}{\partial x}$ is at most

$$\sum_{i \in [r]} r_i (s + s_i + 1) \leq \sum_{i \in [r]} r_i (s + s) \leq 2ks.$$

□

We now show that there exists an $i \in [n]$ such that a hitting set generator for $\operatorname{orb}(f)$ can be constructed using a hitting set generator for $\operatorname{orb}(\frac{\partial f}{\partial x_i})$.

Claim 38. *Let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate polynomial of degree D, and $\operatorname{char}(\mathbb{F}) = 0$ or $> D$. There is an $i \in [n]$ such that $\frac{\partial f}{\partial x_i} \neq 0$, and if $\mathcal{G}$ is a hitting set generator for $\operatorname{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then $\widetilde{\mathcal{G}} := \mathcal{G} + \mathcal{G}_1^{SV}$ is a hitting set generator for $\operatorname{orb}(f)$, provided $|\mathbb{F}| > \deg(\mathcal{G}) \cdot D$.*

Proof: Let $A \in \mathrm{GL}(n, \mathbb{F})$ and $g = f(A\mathbf{x})$. If f is a constant, then constructing a hitting set for $\operatorname{orb}(f)$ is trivial. Otherwise, there exists an $i \in [n]$ such that $\frac{\partial f}{\partial x_i} \neq 0$ (because $\operatorname{char}(\mathbb{F}) = 0$ or $> D$). Suppose that a polynomial map $\mathcal{G}$ is a hitting set generator for $\operatorname{orb}\left(\frac{\partial f}{\partial x_i}\right)$. The gradient of a polynomial $p(\mathbf{x})$, denoted by ∇p, is the column vector $\left(\frac{\partial p}{\partial x_1}\ \frac{\partial p}{\partial x_2}\ \ldots\ \frac{\partial p}{\partial x_n}\right)^T$. By the chain rule of differentiation,

$$\nabla g = A^T \cdot [\nabla f](A\mathbf{x}).$$

As A^T is invertible, $\frac{\partial f}{\partial x_i}(A\mathcal{G}) \neq 0 \implies [\nabla f](A\mathcal{G}) \neq 0 \implies [\nabla g](\mathcal{G}) \neq 0 \implies \exists j \in [n]$ such that $\frac{\partial g}{\partial x_j}(\mathcal{G}) \neq 0$. Since $|\mathbb{F}| > \deg(\mathcal{G}) \cdot D$, by Observation 15, $g(\widetilde{\mathcal{G}})$ is not a constant.

□

All we need to do now is construct a hitting set generator for $\operatorname{orb}\left(\frac{\partial f}{\partial x_i}\right)$. Overloading the notation, we refer to $\frac{\partial f}{\partial x_i}$ as f, which is computed by a $(4, k, s)$ formula whose top fan-in is at most k.

5.2 Constructing a faithful homomorphism for the orbits

Let $f = \sum_{i \in [m]} f_i$ be a $(4, k, s)$ formula. From the discussion in the previous section, we can assume without loss of generality that $m \leq k$. Let $A \in \mathrm{GL}(n, \mathbb{F})$, and $g_i = f_i(A\mathbf{x})$ for all $i \in [m]$. Recall that a homomorphism ϕ is said to be faithful to $\mathbf{g} = (g_1, \ldots, g_m) \in \mathbb{F}[\mathbf{x}]^m$ if $\operatorname{tr-deg}_{\mathbb{F}}(\mathbf{g}) = \operatorname{tr-deg}_{\mathbb{F}}(\phi(\mathbf{g}))$. Also, from Lemma 19, if ϕ is faithful to $\mathbf{g}$, then for any m-variate

[18] 1 less than s, as f_i is connected to the top-most + gate by an edge.

polynomial p, $p(\phi(\mathbf{g})) = 0$ if and only if $p(\mathbf{g}) = 0$. Thus, if we have a homomorphism ϕ that is faithful to $\mathbf{g}$ (irrespective of A), then we can use ϕ as a hitting set generator for $\mathrm{orb}(f)$. The following lemma helps us construct such a homomorphism.

Lemma 39. *Let $\mathbf{f} = (f_1, ..., f_m) \in \mathbb{F}[\mathbf{x}]^m$ be a tuple of n-variate polynomials of degree at most δ, $A \in \mathrm{GL}(n, \mathbb{F})$, $g_i = f_i(A\mathbf{x})$ for all $i \in [m]$, and $\mathbf{g} = (g_1, \dots, g_m)$. Further, suppose that $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) \le r$, and $\mathrm{char}(\mathbb{F}) = 0$ or $> \delta^r$. Let $\psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ be a homomorphism such that $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}))$. Then, the map $\phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}, t, y_1, ..., y_r]$ that, for all $i \in [n]$, maps*

$$x_i \mapsto \left(\sum_{j=1}^{r} y_j t^{ij} \right) + \psi(x_i)$$

is faithful to $\mathbf{g}$.

Proof: Let $J_{\mathbf{x}}(\mathbf{g})$ be the Jacobian matrix of $\mathbf{g}$, and $J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})$ the Jacobian matrix of $\mathbf{f}$ evaluated at $A\mathbf{x}$. From the chain rule of differentiation, $J_{\mathbf{x}}(\mathbf{g}) = J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) \cdot A$. As A in an invertible matrix,

$$\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{g}) = \mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}). \tag{3}$$

Also, for any homomorphism $\psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$, $\psi\left(J_{\mathbf{x}}(\mathbf{g})\right) = \psi\left(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})\right) \cdot A$ and hence,

$$\mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi\left(J_{\mathbf{x}}(\mathbf{g})\right) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi\left(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})\right). \tag{4}$$

So, if we have a homomorphism ψ satisfying $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi\left(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})\right)$, then from (3) and (4),

$$\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{g}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi\left(J_{\mathbf{x}}(\mathbf{g})\right).$$

Also, from Observation 21, $\text{tr-deg}(\mathbf{g}) = \text{tr-deg}(\mathbf{f}) \le r$, and $\deg(g_i) = \deg(f_i) \le \delta$. So, using Lemma 20, we can construct a homomorphism ϕ faithful to $\mathbf{g}$ from ψ, as stated in the lemma. □

Let us apply Lemma 39 to the $(4, k, s)$ formula $f = \sum_{i \in [m]} f_i$, where $m \le k$. Let $\mathbf{f} = (f_1, ..., f_m)$ and $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = r \le k$. Observe that the degree of each f_i is at most $\delta \le s^2$. Then, from Lemma 17, $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{f}) = r$, provided $\mathrm{char}(\mathbb{F}) = 0$ or $> \delta^r$. As A is invertible, this means that $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) = r$. Assume without loss of generality that $\{f_1, \dots, f_r\}$ is a transcendence basis of $\mathbf{f}$. Then, again from Lemma 17, the sub-matrix of $J_{\mathbf{x}}(\mathbf{f})$ consisting of the rows corresponding to $f_1, \dots, f_r$ must be full rank. Thus, we can assume without loss of generality that the minor M of $J_{\mathbf{x}}(\mathbf{f})$ consisting of those rows, and columns corresponding to $x_1, \dots, x_r$, has non-zero determinant. Notice that, as A is invertible, the determinant of M evaluated at $A\mathbf{x}$, i.e., $\det(M(A\mathbf{x})) = [\det(M)](A\mathbf{x})$ is also non-zero. To ensure that the $\mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \psi\left(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})\right)$ is also r, it suffices to construct a homomorphism ψ that is a hitting set generator for $\mathrm{orb}(\det(M))$.

Constructing ψ. Let us look at $\det(M)$ a little more closely. As before, let $f_i = \prod_{j \in m_i} q_{i,j}^{e_{i,j}}$, where $q_{i,j}$ are s-sparse polynomials of degree at most s. For $i \in [r]$, let the number of $q_{i,j}$ containing any of $x_1, \dots, x_r$ be c_i. As f is an occur-k formula, $\sum_{i \in [m]} c_i \le kr \le k^2$. From the i-th row of M, we can factor out $q_{i,j}^{e_{i,j}}$ if $q_{i,j}$ does not contain any of $x_1, \dots, x_r$. Moreover, even if $q_{i,j}$ contains some variable from $x_1, \dots, x_r$, we can still factor out $q_{i,j}^{e_{i,j}-1}$. After we have taken out all these factors, let the residual matrix be M'. Then, each entry of the i-th row of M' is a polynomial with sparsity at most $c_i s^{c_i}$ and degree at most $c_i s$. Thus, $\det(M')$ is a polynomial with sparsity at most $r! \cdot \prod_{i\in[r]} c_i s^{c_i} \le k! \cdot k^k \cdot s^{k^2} \le k^{2k} \cdot s^{k^2}$ and degree at most $\sum_{i\in[r]} c_i s \le k^2 s$. So, $\det(M)$ is a product of polynomials with sparsity at most $k^{2k} \cdot s^{k^2}$ and degree at most $k^2 s$. From Theorem 29, $\psi = \mathcal{G}^{SV}_{(\lceil \log(k^{2k}\cdot s^{k^2})\rceil+1)} = \mathcal{G}^{SV}_{O(k^2(\log k+\log s))}$ is a hitting set generator for $\operatorname{orb}(\det(M))$, if $|\mathbb{F}| > nk^2 s$ and $\operatorname{char}(\mathbb{F}) = 0$ or $> k^2 s$.

If the $q_{i,j}$ are b-variate polynomials, then $\det(M')$ is a polynomial in $\sum_{i\in[r]} c_i b \le k^2 b$ variables. From Observation 14, $\psi = \mathcal{G}^{SV}_{k^2 b}$ is a hitting set generator for $\operatorname{orb}(\det(M))$.

Constructing ϕ. Using ψ and Lemma 39, we get a homomorphism ϕ that is faithful to $\mathbf{g}$. Observe that ϕ is a polynomial map in at most $O\left(k^2\left(\log k + \log s\right)\right) + k + 1 = O\left(k^2\left(\log k + \log s\right)\right)$ variables and of degree at most $nk+1$ (as the degree of the polynomial map ψ is at most n and, in Lemma 39, $\deg\left(\sum_{j=1}^r y_j t^{ij}\right) \le nk+1$).

If the $q_{i,j}$ are b-variate polynomials, then ϕ is a polynomial map in at most $O(k^2 b) + k + 1 = O(k^2 b)$ variables and of degree at most $nk+1$.

5.3 Proof of Theorem 10: the depth-4 case

For $\Delta = 4$, the value of R in the statement of Theorem 10 is $(2k)^{128}$. However, in this special case, one can work with a much smaller value for R. We choose $R = k^4$ so that $\operatorname{char}(\mathbb{F}) = 0$ or $> (2ks)^{6k^2}$. This ensures that the constraints on $\operatorname{char}(\mathbb{F})$ and $|\mathbb{F}|$, coming from Claim 38, Lemma 39 and the application of Theorem 29 in the construction of ψ, are satisfied.

Let f be a $(4,k,s)$ formula. If f is a constant, then so is every polynomial in $\operatorname{orb}(f)$. In this case, the set containing any non-zero point in $\mathbb{F}^n$ is a hitting set for $\operatorname{orb}(f)$; so suppose that f is not a constant. There exists an $i \in [n]$ such that $\frac{\partial f}{\partial x_i} \neq 0$ (as $\operatorname{char}(\mathbb{F}) = 0$ or $> s^2 \ge D$). From Claim 37, $\frac{\partial f}{\partial x_i} \neq 0$ can be computed by a $(4, 2k^2, 2ks)$ formula with top fan-in at most k. Moreover, from the proof of Claim 38, if $\mathcal{G}$ is a hitting set generator for $\operatorname{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then $\widetilde{\mathcal{G}} = \mathcal{G} + \mathcal{G}^{SV}_1$ is a hitting set generator for $\operatorname{orb}(f)$, provided $\operatorname{char}(\mathbb{F}) = 0$ or $> D$ and $|\mathbb{F}| > \deg(\mathcal{G}) \cdot D$. From Section 5.2, there exists a $\mathcal{G}$ that has at most $O\left(\left(2k^2\right)^2\left(\log 2k^2 + \log 2ks\right)\right) = O\left(k^4\left(\log k + \log s\right)\right)$ many variables and has degree at most $2nk^2+1$. As $\mathcal{G}^{SV}_1$ has 2 variables and has degree n, $\widetilde{\mathcal{G}}$ has $O\left(k^4\left(\log k + \log s\right)\right)$ variables and has degree at most $2nk^2+1$. Thus, for any $g \in \operatorname{orb}(f)$, $g(\widetilde{\mathcal{G}})$ has $O\left(k^4\left(\log k + \log s\right)\right)$ variables and has degree at most $(2nk^2+1)D$. So, a hitting set for $\operatorname{orb}(f)$ can be computed in time $(nk^2 D)^{O\left(k^4(\log k+\log s)\right)} = (nRD)^{O(R(\log k+\log s))}$.

The proof for the case where the leaves are labelled by b-variate polynomials is similar. Now, $\mathcal{G}$ has $O(k^4 b)$ variables and has degree at most $2nk^2+1$. Thus, $g(\widetilde{\mathcal{G}})$ has $O(k^4 b)$ variables and is of degree at most $(2nk^2+1)D$, and so, a hitting set for $\operatorname{orb}(f)$ can be computed in $(nk^2 D)^{O(k^4 b)}$ time.

6 Hitting sets for the orbits of occur-once formulas

In this section, we give a quasi-polynomial time construction of hitting sets for the orbits of polynomials computable by occur-once formulas. Assume, without loss of generality, that none of the edge labels of an occur-once formula is zero. We will identify an occur-once formula with the polynomial f it computes and define the width of f - denoted by $\text{width}(f)$ - to be the number of non-constant sparse polynomials at the leaves of the formula. Observe that if $\text{width}(f) \geq 1$, then f is not a constant. As mentioned in Section 1.3, we reduce the problem of finding a hitting set generator for $\text{orb}(f)$ to that of finding a generator for $\text{orb}(\frac{\partial f}{\partial x_i})$, where x_i is such that $\frac{\partial f}{\partial x_i}$ is a product of occur-once formulas of widths at most $\frac{\text{width}(f)}{2}$; this is done in Theorem 42. To prove the theorem, we need a couple of structural results about occur-once formulas and their derivatives, which we prove in the following two lemmas. These lemmas are inspired by similar structural results for read-once formulas given in [SV15], but the arguments need to be appropriately adapted here as occur-once formulas form a more powerful model than read-once formulas.

6.1 Structural results

We will call an occur-once formula an *s-sparse occur-once formula* if the leaves of the formula are labelled by s-sparse polynomials. Without loss of generality, assume that an s-sparse occur-once formula is layered with all the leaves appearing in layer 0. If a gate appears in layer k, then the depth of the occur-once formula rooted at the gate is $k+2$. We will also identify a gate with the occur-once formula rooted at the gate.

Lemma 40. *Let f be an s-sparse occur-once formula having* $\text{width}(f) \geq 2$. *Then, f can be expressed in one of the following three forms:*

1. $f = \alpha(f_1 + f_2) + \beta$,
2. $f = \alpha(f_1 \cdot f_2) + \beta$,
3. $f = \alpha f_1^e + \beta$,

where $\alpha, \beta \in \mathbb{F}$, $\alpha \neq 0$ and f_1, f_2 are non-constant, variable disjoint, s-sparse occur-once formulas. Further, $\text{width}(f_1) + \text{width}(f_2) = \text{width}(f)$ *in the first two forms, and* $\text{width}(f_1) = \text{width}(f)$ *and* $\text{depth}(f_1) < \text{depth}(f)$ *in the third form.*

Proof: Let the depth of f be Δ, which equals the number of layers in f plus 1. Let h be any gate in f in layer 1 (i.e., the layer just above the leaves) and $\text{width}(h) \geq 2$. If h is a $+$ gate, then it can be expressed in form 1. If h is a $\times\curlywedge$ gate, then it can be written in form 2.

Assume, by the way of induction, that the lemma is true for all gates h' in f of $\text{width}(h') \geq 2$ and at layers less than k for some $1 < k \leq \Delta - 2$. Let h be a gate in the k-th layer with $\text{width}(h) \geq 2$. There are two cases:

Case 1: h is a $+$ gate, say $h = \alpha_1 h_1 + \cdots + \alpha_m h_m$. Clearly, if at least two of its children are non-constants, then h is in form 1. On the other hand, if only one child, say h_1, is a non-constant, then $\text{width}(h_1) = \text{width}(h) \geq 2$. As h_1 is in layer $k-1$, from the induction hypothesis, it can be written in one of the three forms with the corresponding constants α and β. Then, by adding

$\alpha_2 h_2 + \cdots + \alpha_m h_m$ (which is a constant) to $\alpha_1\beta$ and multiplying α_1 by α, h can also be written in the same form.

Case 2: h is a $\times\curlywedge$ gate, say $h = {h_1}^{e_1} \cdots {h_m}^{e_m}$. Clearly, if at least two of its children are non-constants, then h is in form 2. On the other hand, if only one child, say h_1, is a non-constant, then $\text{width}(h_1) = \text{width}(h) \geq 2$. In this case, by taking $\alpha = {h_2}^{e_2} \cdots {h_m}^{e_m}$ (which is a constant), and observing that $\text{depth}(h_1) = k - 1 + 2 < k + 2 = \text{depth}(h)$, we see that h is in form 3. □

Lemma 41. *Let f be an s-sparse occur-once formula. Then for any $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a product of s-sparse occur-once formulas of widths at most* $\text{width}(f)$.

Proof: Let the depth of f be Δ. Notice that the lemma is true for all the leaves (i.e., at layer 0) of f as any derivative of an s-sparse polynomial is also an s-sparse polynomial. Assume, by the way of induction, that the lemma is true for all gates at layers less than k, for some $1 \leq k \leq \Delta - 2$ and let h be any gate in the k-th layer of f. There are two cases:

Case 1: h is a $+$ gate, say $h = \alpha_1 h_1 + \cdots + \alpha_m h_m$. As f, and hence h, is an s-sparse occur-once formula, we can assume without loss of generality that x_i appears only in h_1, if it appears at all. Then, $\frac{\partial h}{\partial x_i} = \alpha_1 \frac{\partial h_1}{\partial x_i}$. From the induction hypothesis, $\frac{\partial h_1}{\partial x_i}$ is a product of s-sparse occur-once formulas of widths at most $\text{width}(h_1) \leq \text{width}(h)$, and so, the lemma is true for h.

Case 2: h is a $\times\curlywedge$ gate, say $h = {h_1}^{e_1} \cdots {h_m}^{e_m}$. As, in the previous case, assume that x_i appears only in h_1. Then,

$$\frac{\partial h}{\partial x_i} = e_1 \cdot {h_1}^{e_1 - 1} \cdot {h_2}^{e_2} \cdot \ \cdots \ \cdot {h_m}^{e_m} \cdot \frac{\partial h_1}{\partial x_i}.$$

From the induction hypothesis, $\frac{\partial h_1}{\partial x_i}$ is a product of s-sparse occur-once formulas of widths at most $\text{width}(h_1) \leq \text{width}(h)$. Moreover, ${h_1}^{e_1-1}, {h_2}^{e_2}, ..., {h_m}^{e_m}$ are also s-sparse occur-once formulas of widths at most $\text{width}(h)$. Thus, the lemma is true for h. □

6.2 Proof of Theorem 11

We now show the existence of an efficient hitting set generator for orbits of occur-once formulas.

Theorem 42. *Let f be an n-variate, degree-D polynomial that is computable by an s-sparse occur-once formula, and $g \in \text{orb}(f)$. Also, let $|\mathbb{F}| > nD$ and $\text{char}(\mathbb{F}) = 0$ or $> D$. Then for any $t \geq \log(\text{width}(f))$, $g \neq 0$ implies $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right) \neq 0$. In fact, if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$.*

Proof: Notice that if g is a non-zero constant, then $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right) \neq 0$ for all t. So, to prove the theorem, we need to show that if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$.

Let h be an s-sparse occur-once formula satisfying $\text{width}(h) = 1$. Then, h must be of the form

$$\alpha_m \left(\cdots \left(\alpha_2 \left(\alpha_1 p(\mathbf{x})^{e_1} + \beta_1 \right)^{e_2} + \beta_2 \right) \cdots \right)^{e_m} + \beta_m,$$

where $p(\mathbf{x})$ is an s-sparse polynomial, $e_1, ..., e_m \in \mathbb{N}$, $\alpha_1, ..., \alpha_m \in \mathbb{F} \setminus \{0\}$ and $\beta_1, ..., \beta_m \in \mathbb{F}$. Let $A \in \mathrm{GL}(n, \mathbb{F})$. If $h(A\mathbf{x})$ is not a constant, then neither is $p(A\mathbf{x})$. Thus, from Theorem 29 and the fact that $\mathrm{Img}(\mathcal{G}_k^{SV}) \subseteq \mathrm{Img}(\mathcal{G}_{k+1}^{SV})$ for any $k \geq 0$, we have that $p\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ is not a constant for any $t \geq 0$. Hence, $h\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ is also not a constant for any $t \geq 0$.

Assume, by the way of induction, that the theorem is true for all g' such that $g' \in \mathrm{orb}(f')$ for some n-variate, degree-D, s-sparse occur-once formula f' with $1 \leq \mathrm{width}(f') < \ell \leq \mathrm{width}(f)$. Let h be an n-variate, degree-D, s-sparse occur-once formula having $\mathrm{width}(h) = \ell \geq 2$, and $A \in \mathrm{GL}(n, \mathbb{F})$. From Lemma 40, there are three cases,

Case 1: $h = \alpha(h_1 + h_2) + \beta$. Then, we can assume without loss of generality that $\mathrm{width}(h_1) \leq \frac{\mathrm{width}(h)}{2} = \frac{\ell}{2}$, as $\mathrm{width}(h_1) + \mathrm{width}(h_2) = \mathrm{width}(h)$. Since h_1 is not a constant, there exists an $i \in [n]$ such that $\frac{\partial h_1}{\partial x_i} \neq 0$ (because $\mathrm{char}(\mathbb{F})$ is 0 or $> D$). As $\frac{\partial h}{\partial x_i} = \alpha \cdot \frac{\partial h_1}{\partial x_i}$ (h_1 and h_2 being variable disjoint) and $\alpha \neq 0$, $\frac{\partial h}{\partial x_i} \neq 0$. Now, from Lemma 41, $\frac{\partial h_1}{\partial x_i}$ is a product of s-sparse occur-once formulas of width at most $\frac{\ell}{2}$. Then, from the induction hypothesis, $\frac{\partial h}{\partial x_i}\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right) \neq 0$ for any $t \geq \log \ell - 1$. Let $q = h(A\mathbf{x})$. The gradient of a polynomial $p(\mathbf{x})$, denoted by ∇p, is the column vector $\left(\frac{\partial p}{\partial x_1} \; \frac{\partial p}{\partial x_2} \; \ldots \; \frac{\partial p}{\partial x_n}\right)^T$. By the chain rule of differentiation,

$$\nabla q = A^T \cdot [\nabla h](A\mathbf{x}).$$

As A^T is invertible, there exists a $j \in [n]$ such that $\frac{\partial q}{\partial x_j}\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right) \neq 0$ for any $t \geq \log \ell - 1$. This means, by Observation 15, $q(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)})$ is not a constant for any $t \geq \log \ell$ (as $\deg(q) \leq D$ and $|\mathbb{F}| > nD$). In other words, $h(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)})$ is not a constant for any $t \geq \log \ell$.

Case 2: $h = \alpha(h_1 \cdot h_2) + \beta$. As $\mathrm{width}(h_1), \mathrm{width}(h_2) < \mathrm{width}(h)$, from the induction hypothesis, we have that for any $t \geq \log \ell$, $h_1\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right), h_2\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ are not constants and so neither is $h\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$.

Case 3: $h = \alpha h_1^e + \beta$. In this case, $\mathrm{width}(h_1) = \mathrm{width}(h) = \ell \geq 2$, but $\mathrm{depth}(h_1) < \mathrm{depth}(h)$. As $h\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ is not a constant if and only if $h_1\left(A\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ is not a constant, the problem reduces to showing that for any $g_1 \in \mathrm{orb}(h_1)$, $g_1\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + t)}\right)$ is not a constant for any $t \geq \log \ell$. We now run the argument from the beginning with h replaced by h_1, which has a smaller depth. Eventually, we will land up in Case 1 or 2, as a depth-3 occur-once formula having width ≥ 2 is either in form 1 or 2 (see proof of Lemma 40). □

A non-zero polynomial $f \in \mathcal{C}$ is computable by an s-sparse occur-once formula. Observe that $\mathrm{width}(f) \leq n$. Let $g \in \mathrm{orb}(f)$. From Theorem 42, we have that $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1 + \lceil \log n \rceil)}\right)$ is a non-zero polynomial in $2\left(\lceil \log s \rceil + 1 + \lceil \log n \rceil\right)$ variables of degree at most nD. As $|\mathbb{F}| > nD$, a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in time $(nD+1)^{2(\lceil \log s \rceil + 1 + \lceil \log n \rceil)} = (nD)^{O(\log n + \log s)}$.

The proof is similar if the leaves of the occur-once formulas in $\mathcal{C}$ are labelled by b-variate polynomials. We just need to apply Observation 14 instead of Theorem 29 in the base case.

7 Conclusion

In this paper, we have studied the hitting set problem for the orbits of several important polynomial families and circuit classes that are not closed under affine projections. This line of research is both natural and interesting as affine projections of some of these circuit classes and polynomial families capture much larger circuit classes (in some cases, almost the entire class of VP circuits). The orbit of a polynomial f is a natural and "dense" subset of affine projections of f that, in turn, resides in the orbit closure of f. We have shown efficient hitting set constructions for the orbits of several well-studied circuit classes such as commutative ROABPs, sums of products of univariates and constant-width ROABPs (under the low individual degree restriction), and constant-depth constant-occur formulas and occur-once formulas. In the process, we have obtained efficiently constructible hitting sets for the orbits of the elementary symmetric polynomials, the power symmetric polynomials, the sum-product polynomials, and the iterated matrix multiplication polynomials of width-3, which is a complete family of polynomials for arithmetic formulas under p-projections. Despite the progress made here, there are several natural questions that, if resolved, will strengthen and complete the set of results presented in this work. We leave these for future work:

- **Removing the low individual degree restriction.** The low individual degree restriction is natural as it subsumes the multilinear case. However, it would be ideal if we get rid of this limitation of our results. In particular, can we give an efficient hitting-set construction for the orbits of general commutative ROABPs and constant-width ROABPs?

- **Lower bound and hitting set for the orbits of ROABPs.** We would also like to remove the requirements of commutativity and constant-width from our results on hitting sets for the orbits of ROABPs. It is worth noting that an explicit hitting set for the orbits of ROABPs implies a lower bound for the same model computing some explicit polynomial [Agr05]. To our knowledge, no explicit lower bound is known for the orbits of ROABPs. Can we prove such a lower bound first?

- **Hitting sets for the orbits of Det and IMM.** The determinant (Det) and the iterated matrix multiplication (IMM) polynomial families are complete for the class of algebraic branching programs under p-projections. Can we design efficiently constructible hitting sets for the orbits of Det and IMM? Observe that a hitting set for the orbits of *multilinear* ROABPs is a hitting set for orb(IMM). Also, a hitting set for the orbits of the polynomials computable by the Edmonds' model (see Section 1.4) is a hitting set for the orbits of both Det and IMM.

Acknowledgements

We thank Rohit Gurjar for asking (nearly four years back) whether an explicit hitting set is known for the orbit of the IMM polynomial. His question, asked during a discussion with CS on equivalence testing for the IMM polynomial, has spurred us to think about the problems that we study in this paper. We also thank Ankit Garg, Neeraj Kayal, and Vishwas Bhargava for several stimulating discussions at the onset of this work. Thanks especially to Ankit for pointing out a simplification in the proof of Theorem 6 and Vishwas for asking if Theorem 8 can be applied to orbits of multilinear depth-4 circuits with constant top fan-in.

References

[AB03] Manindra Agrawal and Somenath Biswas. Primality and identity testing via Chinese remaindering. *J. ACM*, 50(4):429–443, 2003. Conference version appeared in the proceedings of FOCS 1999.

[AFS$^+$18] Matthew Anderson, Michael A. Forbes, Ramprasad Saptharishi, Amir Shpilka, and Ben Lee Volk. Identity Testing and Lower Bounds for Read-k Oblivious Algebraic Branching Programs. *ACM Trans. Comput. Theory*, 10(1):3:1–3:30, 2018. Conference version appeared in the proceedings of CCC 2016.

[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.

[Agr05] Manindra Agrawal. Proving lower bounds via pseudo-random generators. In Ramaswamy Ramanujam and Sandeep Sen, editors, *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science, 25th International Conference, Hyderabad, India, December 15-18, 2005, Proceedings*, volume 3821 of *Lecture Notes in Computer Science*, pages 92–105. Springer, 2005.

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of Mathematics*, 160(2):781–793, 2004.

[ASS13] Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 321–330. ACM, 2013.

[ASSS16] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. Jacobian Hits Circuits: Hitting Sets, Lower Bounds for Depth-D Occur-k Formulas and Depth-3 Transcendence Degree-k Circuits. *SIAM J. Comput.*, 45(4):1533–1562, 2016. Conference version appeared in the proceedings of STOC 2012.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic Circuits: A Chasm at Depth Four. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 67–75. IEEE Computer Society, 2008.

[AvMV15] Matthew Anderson, Dieter van Melkebeek, and Ilya Volkovich. Deterministic polynomial identity tests for multilinear bounded-read formulae. *Comput. Complex.*, 24(4):695–776, 2015. Conference version appeared in the proceedings of CCC 2011.

[AW16] Eric Allender and Fengming Wang. On the power of algebraic branching programs of width two. *Comput. Complex.*, 25(1):217–253, 2016. Conference version appeared in the proceedings of ICALP 2011.

[BC92] Michael Ben-Or and Richard Cleve. Computing Algebraic Formulas Using a Constant Number of Registers. *SIAM J. Comput.*, 21(1):54–58, 1992. Conference version appeared in the proceedings of STOC 1988.

[BIZ18] Karl Bringmann, Christian Ikenmeyer, and Jeroen Zuiddam. On algebraic branching programs of small width. *J. ACM*, 65(5):32:1–32:29, 2018. Conference version appeared in the proceedings of CCC 2017.

[BMS13] Malte Beecken, Johannes Mittmann, and Nitin Saxena. Algebraic independence and blackbox identity testing. *Inf. Comput.*, 222:2–19, 2013. Conference version appeared in the proceedings of ICALP 2011.

[CKS18] Chi-Ning Chou, Mrinal Kumar, and Noam Solomon. Hardness vs randomness for bounded depth arithmetic circuits. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*, volume 102 of *LIPIcs*, pages 13:1–13:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[DL78] Richard A. DeMillo and Richard J. Lipton. A Probabilistic Remark on Algebraic Program Testing. *Inf. Process. Lett.*, 7(4):193–195, 1978.

[dOSlV16] Rafael Mendes de Oliveira, Amir Shpilka, and Ben lee Volk. Subexponential Size Hitting Sets for Bounded Depth Multilinear Formulas. *Comput. Complex.*, 25(2):455–505, 2016. Conference version appeared in the proceedings of CCC 2015.

[DS07] Zeev Dvir and Amir Shpilka. Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. *SIAM J. Comput.*, 36(5):1404–1434, 2007. Conference version appeared in the proceedings of STOC 2005.

[DSY09] Zeev Dvir, Amir Shpilka, and Amir Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. Comput.*, 39(4):1279–1293, 2009. Conference version appeared in the proceedings of STOC 2008.

[Edm67] Jack Edmonds. Systems of distinct representatives and linear algebra. *Journal of research of the National Bureau of Standards*, 71:241–245, 1967.

[Edm79] Jack Edmonds. Matroid intersection. In P.L. Hammer, E.L. Johnson, and B.H. Korte, editors, *Discrete Optimization I*, volume 4 of *Annals of Discrete Mathematics*, pages 39–49. Elsevier, 1979.

[FGS18] Michael A. Forbes, Sumanta Ghosh, and Nitin Saxena. Towards blackbox identity testing of log-variate circuits. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPIcs*, pages 54:1–54:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[FGT16] Stephen A. Fenner, Rohit Gurjar, and Thomas Thierauf. Bipartite perfect matching is in quasi-nc. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 754–763. ACM, 2016.

[FK18] Michael A. Forbes and Zander Kelley. Pseudorandom generators for read-once branching programs, in any order. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 946–955. IEEE Computer Society, 2018.

[For15] Michael A. Forbes. Deterministic divisibility testing via shifted partial derivatives. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 451–465. IEEE Computer Society, 2015.

[FS12] Michael A. Forbes and Amir Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 163–172. ACM, 2012.

[FS13a] Michael A. Forbes and Amir Shpilka. Explicit Noether Normalization for Simultaneous Conjugation via Polynomial Identity Testing. In Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX 2013, and 17th International Workshop, RANDOM 2013, Berkeley, CA, USA, August 21-23, 2013. Proceedings*, volume 8096 of *Lecture Notes in Computer Science*, pages 527–542. Springer, 2013.

[FS13b] Michael A. Forbes and Amir Shpilka. Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 243–252. IEEE Computer Society, 2013.

[FS18] Michael A. Forbes and Amir Shpilka. A PSPACE construction of a hitting set for the closure of small algebraic circuits. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1180–1192. ACM, 2018.

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Hitting sets for multilinear read-once algebraic branching programs, in any order. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 867–875. ACM, 2014.

[Gee99] James F. Geelen. Maximum rank matrix completion. *Linear Algebra and its Applications*, 288:211 – 217, 1999.

[GG20] Zeyu Guo and Rohit Gurjar. Improved explicit hitting-sets for roabps. In Jaroslaw Byrka and Raghu Meka, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*, volume 176 of *LIPIcs*, pages 4:1–4:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[GKKS16] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth 3. *SIAM J. Comput.*, 45(3):1064–1079, 2016. Conference version appeared in the proceedings of FOCS 2013.

[GKS17] Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Identity testing for constant-width, and any-order, read-once oblivious arithmetic branching programs. *Theory Comput.*, 13(1):1–21, 2017. Conference version appeared in the proceedings of CCC 2016.

[GKST17] Rohit Gurjar, Arpita Korwar, Nitin Saxena, and Thomas Thierauf. Deterministic Identity Testing for Sum of Read-Once Oblivious Arithmetic Branching Programs. *Comput. Complex.*, 26(4):835–880, 2017. Conference version appeared in the proceedings of CCC 2015.

[GR08] Ariel Gabizon and Ran Raz. Deterministic extractors for affine sources over large fields. *Comb.*, 28(4):415–440, 2008. Conference version appeared in the proceedings of FOCS 2005.

[GSS18] Zeyu Guo, Nitin Saxena, and Amit Sinhababu. Algebraic dependencies and PSPACE algorithms in approximative complexity. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*, volume 102 of *LIPIcs*, pages 10:1–10:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[GT20] Rohit Gurjar and Thomas Thierauf. Linear matroid intersection is in quasi-nc. *Comput. Complex.*, 29(2):9, 2020. Conference version appeared in the proceedings of STOC 2017.

[Gup14] Ankit Gupta. Algebraic geometric techniques for depth-4 PIT & sylvester-gallai conjectures for varieties. *Electron. Colloquium Comput. Complex.*, 21:130, 2014.

[Hal35] P. Hall. On Representatives of Subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 01 1935.

[HS80] Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In Raymond E. Miller, Seymour Ginsburg, Walter A. Burkhard, and Richard J. Lipton, editors, *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 262–272. ACM, 1980.

[IKS10] Gábor Ivanyos, Marek Karpinski, and Nitin Saxena. Deterministic polynomial time algorithms for matrix completion problems. *SIAM J. Comput.*, 39(8):3736–3751, 2010.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In Frank Thomson Leighton and Michael T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 356–364. ACM, 1994.

[JQS10] Maurice J. Jansen, Youming Qiao, and Jayalal Sarma. Deterministic Black-Box Identity Testing pi-Ordered Algebraic Branching Programs. In Kamal Lodaya and Meena Mahajan, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2010, December 15-18, 2010, Chennai, India*, volume 8 of *LIPIcs*, pages 296–307. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2010.

[Kal89] Erich Kaltofen. Factorization of polynomials given by straight-line programs. *Adv. Comput. Res.*, 5:375–412, 1989.

[Kay10] Neeraj Kayal. Algorithms for arithmetic circuits. *Electron. Colloquium Comput. Complex.*, 17:73, 2010.

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complex.*, 13(1-2):1–46, 2004. Conference version appeared in the proceedings of STOC 2003.

[KMSV13] Zohar Shay Karnin, Partha Mukhopadhyay, Amir Shpilka, and Ilya Volkovich. Deterministic Identity Testing of Depth-4 Multilinear Circuits with Bounded Top Fan-in. *SIAM J. Comput.*, 42(6):2114–2131, 2013. Conference version appeared in the proceedings of STOC 2010.

[KNS20] Neeraj Kayal, Vineet Nair, and Chandan Saha. Separation between read-once oblivious algebraic branching programs (roabps) and multilinear depth-three circuits. *ACM Trans. Comput. Theory*, 12(1):2:1–2:27, 2020.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.

[KS01] Adam R. Klivans and Daniel A. Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pages 216–223, 2001.

[KS07] Neeraj Kayal and Nitin Saxena. Polynomial identity testing for depth 3 circuits. *Comput. Complex.*, 16(2):115–138, 2007. Conference version appeared in the proceedings of CCC 2006.

[KS09] Neeraj Kayal and Shubhangi Saraf. Blackbox polynomial identity testing for depth 3 circuits. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 198–207. IEEE Computer Society, 2009.

[KS11] Zohar Shay Karnin and Amir Shpilka. Black box polynomial identity testing of generalized depth-3 arithmetic circuits with bounded top fan-in. *Comb.*, 31(3):333–364, 2011. Conference version appeared in the proceedings of CCC 2008.

[KS19] Pascal Koiran and Mateusz Skomra. Derandomization and absolute reconstruction for sums of powers of linear forms. *CoRR*, abs/1912.02021, 2019.

[KSS15] Swastik Kopparty, Shubhangi Saraf, and Amir Shpilka. Equivalence of polynomial identity testing and polynomial factorization. *Comput. Complex.*, 24(2):295–331, 2015. Conference version appeared in the proceedings of CCC 2014.

[KT90] Erich Kaltofen and Barry M. Trager. Computing with Polynomials Given By Black Boxes for Their Evaluations: Greatest Common Divisors, Factorization, Separation of Numerators and Denominators. *J. Symb. Comput.*, 9(3):301–320, 1990. Conference version appeared in the proceedings of FOCS 1988.

[KUW86] Richard M. Karp, Eli Upfal, and Avi Wigderson. Constructing a perfect matching is in random NC. *Comb.*, 6(1):35–48, 1986. Conference version appeared in the proceedings of STOC 1985.

[Lov79] László Lovász. On determinants, matchings, and random algorithms. In Lothar Budach, editor, *Fundamentals of Computation Theory, FCT 1979, Proceedings of the Conference on Algebraic, Arthmetic, and Categorial Methods in Computation Theory, Berlin/Wendisch-Rietz, Germany, September 17-21, 1979*, pages 565–574. Akademie-Verlag, Berlin, 1979.

[Lov89] László Lovász. Singular spaces of matrices and their application in combinatorics. *Boletim da Sociedade Brasileira de Matemática - Bulletin/Brazilian Mathematical Society*, 20(1):87–99, 1989.

[LV03] Richard J. Lipton and Nisheeth K. Vishnoi. Deterministic identity testing for multivariate polynomials. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 12-14, 2003, Baltimore, Maryland, USA*, pages 756–760. ACM/SIAM, 2003.

[MS21] Dori Medini and Amir Shpilka. Hitting Sets and Reconstruction for Dense Orbits in VP_{e} and $\Sigma\Pi\Sigma$ Circuits. *CoRR*, abs/2102.05632, 2021.

[Mul17] Ketan D. Mulmuley. Geometric complexity theory V: Efficient algorithms for Noether normalization. *J. Amer. Math. Soc.*, 30(1):225–309, 2017. Extended abstract appeared in the proceedings of FOCS 2012.

[Mur93] K. Murota. Mixed matrices: Irreducibility and decomposition. In R. A. Brualdi, S. Friedland, and V. Klee, editors, *Combinatorial and Graph-Theoretical Problems in Linear Algebra. The IMA Volumes in Mathematics and its Applications, vol 50.*, pages 39–71. Springer, New York, NY, 1993.

[MV18] Daniel Minahan and Ilya Volkovich. Complete derandomization of identity testing and reconstruction of read-once formulas. *ACM Trans. Comput. Theory*, 10(3):10:1–10:11, 2018. Conference version appeared in the proceedings of CCC 2017.

[MVV87] Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. Matching is as easy as matrix inversion. *Comb.*, 7(1):105–113, 1987. Conference version appeared in the proceedings of STOC 1987.

[Nis91] Noam Nisan. Lower Bounds for Non-Commutative Computation (Extended Abstract). In Cris Koutsougeras and Jeffrey Scott Vitter, editors, *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 410–418. ACM, 1991.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Comb.*, 12(4):449–461, 1992. Conference version appeared in the proceedings of STOC 1990.

[NSV94] H. Narayanan, Huzur Saran, and Vijay V. Vazirani. Randomized Parallel Algorithms for Matroid Union and Intersection, With Applications to Arboresences and Edge-Disjoint Spanning Trees. *SIAM J. Comput.*, 23(2):387–397, 1994. Conference version appeared in the proceedings of SODA 1992.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994. Conference version appeared in the proceedings of FOCS 1988.

[NW97] Noam Nisan and Avi Wigderson. Lower Bounds on Arithmetic Circuits Via Partial Derivatives. *Computational Complexity*, 6(3):217–234, 1997. Conference version appeared in the proceedings of FOCS 1995.

[PS20a] Shir Peleg and Amir Shpilka. A generalized sylvester-gallai type theorem for quadratic polynomials. In Shubhangi Saraf, editor, *35th Computational Complexity Conference, CCC 2020, July 28-31, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 169 of *LIPIcs*, pages 8:1–8:33. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[PS20b] Shir Peleg and Amir Shpilka. Polynomial time deterministic identity testing algorithm for $\Sigma^{[3]}\Pi\Sigma\Pi^{[2]}$ circuits via Edelstein-Kelly type theorem for quadratic polynomials. *CoRR*, abs/2006.08263, 2020.

[RS05] Ran Raz and Amir Shpilka. Deterministic polynomial identity testing in non-commutative models. *Comput. Complex.*, 14(1):1–19, 2005. Conference version appeared in the proceedings of CCC 2004.

[Sax08] Nitin Saxena. Diagonal circuit identity testing and lower bounds. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part I: Tack A: Algorithms, Automata, Complexity, and Games*, volume 5125 of *Lecture Notes in Computer Science*, pages 60–71. Springer, 2008.

[Sax09] Nitin Saxena. Progress on polynomial identity testing. *Bull. EATCS*, 99:49–79, 2009.

[Sax14] Nitin Saxena. Progress on polynomial identity testing-ii. In M. Agrawal and V. Arvind, editors, *Perspectives in Computational Complexity*, volume 26 of *Progress in Computer Science and Applied Logic*, pages 131–146. Birkhäuser, Cham, 2014.

[Sch80] Jacob T. Schwartz. Fast Probabilistic Algorithms for Verification of Polynomial Identities. *J. ACM*, 27(4):701–717, 1980.

[Shp02] Amir Shpilka. Affine projections of symmetric polynomials. *J. Comput. Syst. Sci.*, 65(4):639–659, 2002. Conference version appeared in the proceedings of CCC 2001.

[Shp19] Amir Shpilka. Sylvester-gallai type theorems for quadratic polynomials. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1203–1214. ACM, 2019.

[SS12] Nitin Saxena and C. Seshadhri. Blackbox Identity Testing for Bounded Top-Fanin Depth-3 Circuits: The Field Doesn't Matter. *SIAM J. Comput.*, 41(5):1285–1298, 2012. Conference version appeared in the proceedings of STOC 2011.

[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. Conference version appeared in the proceedings of FOCS 2010.

[SSS09] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. The power of depth 2 circuits over algebras. In Ravi Kannan and K. Narayan Kumar, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2009, December 15-17, 2009, IIT Kanpur, India*, volume 4 of *LIPIcs*, pages 371–382. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2009.

[SSS13] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. A case of depth-3 identity testing, sparse factorization and duality. *Comput. Complex.*, 22(1):39–69, 2013.

[ST17] Ola Svensson and Jakub Tarnawski. The matching problem in general graphs is in quasi-nc. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 696–707. IEEE Computer Society, 2017.

[ST21] Chandan Saha and Bhargav Thankey. Hitting Sets for Orbits of Circuit Classes and Polynomial Families. *Electron. Colloquium Comput. Complex.*, 28:15, 2021.

[SV15] Amir Shpilka and Ilya Volkovich. Read-once polynomial identity testing. *Comput. Complex.*, 24(3):477–532, 2015. Conference versions appeared in the proceedings of STOC 2008 and APPROX-RANDOM 2009.

[SV18] Shubhangi Saraf and Ilya Volkovich. Black-Box Identity Testing of Depth-4 Multilinear Circuits. *Comb.*, 38(5):1205–1238, 2018. Conference version appeared in the proceedings of STOC 2011.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Inf. Comput.*, 240:2–11, 2015. Conference version appeared in the proceedings of MFCS 2013.

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 249–261, 1979.

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM J. Comput.*, 12(4):641–644, 1983.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, Marseille, France, June 1979, Proceedings*, pages 216–226, 1979.

A Missing proofs from Section 3

In this section, we give the proofs of Observation 23, Claim 25 and Lemma 26.

A.1 Proof of Observation 23

The proof of Observation 23 follows from the following claim.

Claim 43. *Let $p(y) = \sum_{e=0}^{d} w_e y^e$, where $w_e \in \mathbb{A}$ and $p(y+r) = \sum_{b=0}^{d} \widetilde{w}_b y^b$. Then, $\widetilde{w}_b = \sum_{e=0}^{d} \binom{e}{b} r^{e-b} w_e$.*

Proof:

$$\begin{aligned}
p(y+r) &= \sum_{e=0}^{d} w_e (y+r)^e \\
&= \sum_{e=0}^{d} w_e \sum_{b=0}^{d} \binom{e}{b} r^{e-b} y^b \\
&= \sum_{b=0}^{d} \left(\sum_{e=0}^{d} \binom{e}{b} r^{e-b} w_e \right) y^b.
\end{aligned}$$

Thus, $\widetilde{w}_b = \sum_{e=0}^{d} \binom{e}{b} r^{e-b} w_e$. □

For 1, put $\widetilde{w}_b = v_{i,b_i}$, $e = e_i$, $b = b_i, r = r_i$ and $w_e = u_{i,e_i}$. For 2, put $\widetilde{w}_b = u_{i,b_i}$, $e = b_i$, $b = e_i$, $r = -r_i$ and $w_e = v_{i,e_i}$.

A.2 Proof of Claim 25

The entry indexed by $\mathbf{e} \in \{0, \ldots, d\}^m$ of U is $u_{\mathbf{e}}$. Observe that

$$\begin{aligned}
u_{\mathbf{e}} &= \prod_{i \in [m]} u_{i,e_i} \\
&= \prod_{i \in [m]} \left(\sum_{b_i=0}^{d} \binom{b_i}{e_i} \cdot (-r_i)^{b_i - e_i} \cdot v_{i,b_i} \right) && \text{(from Observation 23)} \\
&= \sum_{\mathbf{b}=(b_1,\ldots,b_n) \in \{0,\ldots,d\}^m} \binom{\mathbf{b}}{\mathbf{e}} \prod_{i \in [m]} (-r_i)^{b_i} \cdot \prod_{i \in [m]} v_{i,b_i} \cdot \prod_{i \in [m]} (-r_i)^{-e_i} \\
&= \sum_{\mathbf{b} \in \{0,\ldots,d\}^m} \binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{\mathbf{b}} \cdot v_{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}} \\
&= \sum_{\mathbf{b} \in \{0,\ldots,d\}^m} v_{\mathbf{b}} \cdot \mathbf{r}^{\mathbf{b}} \cdot \binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{-\mathbf{e}}.
\end{aligned}$$

The equation $U = VCMD$ now follows easily from the definitions of these matrices.

A.3 Proof of Lemma 26

The entries of U, the columns of M, the rows and columns of D, and the rows of N are indexed by $\mathbf{e} \in \{0, \ldots, d\}^m$. Impose an order $\prec$, say the lexicographical order, on the indices $\mathbf{e} \in \{0, \ldots, d\}^m$ of U and the other three matrices. Pick the *minimal* basis of the space spanned by the entries of U according to this order, i.e., consider the entries of U in the order dictated by $\prec$ while forming the basis. Let $\mathcal{B} := \{\mathbf{e} \in \{0,\dots,d\}^m : u_{\mathbf{e}}$ is in the minimal basis of U w.r.t. $\prec\}$.

Construction of the matrix N**.** The columns of N are indexed by $\mathbf{b} \in F$. We will now specify a set of column vectors $\{\mathbf{n_b} : \mathbf{b} \in F\}$ in the null space of U such that the column of N indexed by $\mathbf{b} \in F$ is $\mathbf{n_b}$. There are two cases for $\mathbf{b} \in F$:

Case 1: $\mathbf{b} \in F \setminus \mathcal{B}$. In this case, $u_{\mathbf{b}}$ is dependent on $\{u_{\mathbf{e}} : \mathbf{e} \in \mathcal{B}$ and $\mathbf{e} \prec \mathbf{b}\}$. Pick this dependence vector as $\mathbf{n_b}$.

Case 2: $\mathbf{b} \in F \cap \mathcal{B}$. Let there be p such $\mathbf{b}$, where $p \leq |\mathcal{B}| \leq w^2$. For a set $E \subseteq [m]$ and $\mathbf{b} \in \{0,\dots,d\}^m$, let $(\mathbf{b})_E$ denote the vector obtained by projecting $\mathbf{b}$ to the coordinates in E. Roughly speaking, the following claim says that each of these p vectors has a "small signature" that differentiates it from the other $p-1$ vectors.

Claim 44. *There exists a way of numbering all* $\mathbf{b} \in F \cap \mathcal{B}$ *as* $\mathbf{b}_1,\dots,\mathbf{b}_p$ *and there exist non-empty sets* $E_1,\dots,E_p \subseteq [m]$*, each of size at most* $\log p \leq \log w^2$ *such that for all* $k \in [p-1]$*,*

$$(\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k} \ \ \forall \ell \in \{k+1,\dots,p\} \tag{5}$$

Proof: Suppose that we have already identified $\mathbf{b}_1,\dots,\mathbf{b}_{k-1}$ for some $k \in [p-1]$ and have constructed $E_1,\dots,E_{k-1}$ satisfying (5). We will show how to identify $\mathbf{b}_k$ and construct E_k *greedily*.

Initially $E_k = \emptyset$. Let T be the set of the $\mathbf{b}$ vectors that have not been numbered yet; $|T| \leq p$. As each vector in T is unique, there exists an index $i_1 \in [m]$ such that the i_1-th entry is not the same for all $\mathbf{b} \in T$. In fact, there must exist a $j_1 \in [d]$ such that the number of $\mathbf{b}$ whose i_1-th entry is j_1 is at least 1 and at most $|T|/2$. Add i_1 to E_k and remove from T all those $\mathbf{b}$ whose i_1-th entry is not j_1. Again, as each vector in T is unique, there exists an index $i_2 \in [m] \setminus E_k$ and a $j_2 \in [d]$ such that the number of $\mathbf{b} \in T$ whose i_2-th entry is j_2 is at least 1 and at most $|T|/2$. Again, add i_2 to E_k and remove from T all those $\mathbf{b}$ whose i_2-th entry is not j_2. Continuing in this fashion, in $\log p$ or fewer iterations, $|T| = 1$; call the only vector in T, $\mathbf{b}_k$ and stop. It is clear that $|E_k| \leq \log p$ and that $\mathbf{b}_k$ and E_k satisfy (5).

After having identified $\mathbf{b}_1,\dots,\mathbf{b}_{p-1}$, call the last remaining vector $\mathbf{b}_p$ and pick E_p to be any arbitrary singleton set. □

We will call E_k the *signature* of $\mathbf{b}_k$ for $k \in [p]$. The following claim tells us that for each vector $\mathbf{b}_k$, there is a vector that is not in $\mathcal{B}$ and has support at most $m-1$, but agrees with $\mathbf{b}_k$ on its signature and so in some sense can be used as a proxy for $\mathbf{b}_k$.

Claim 45. *For every* $k \in [p]$*, there exists a vector* $\mathbf{b}'_k \in \{0,\dots,d\}^m \setminus (F \cup \mathcal{B})$ *such that* $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k}$ *and also* $\mathbf{b}'_k$ *and* $\mathbf{b}_k$ *agree on all locations where* $\mathbf{b}'_k$ *is non-zero.*

Proof: As $|E_k| \leq \log w^2$ and $m = 2\lceil \log w^2 \rceil + 1$, for any vector $\mathbf{b}' \in \{0,\dots,d\}^m$ satisfying $(\mathbf{b}')_{E_k} = (\mathbf{b}_k)_{E_k}$, there are still at least $\lceil \log w^2 \rceil + 1$ coordinates whose values we are free to choose. For each such free coordinate, we choose its value to be either 0 or the value at the same coordinate in $\mathbf{b}_k$. There are $2^{\lceil \log w^2 \rceil + 1} \geq 2w^2$ such $\mathbf{b}'$, one of which is $\mathbf{b}_k$ and the remaining $2w^2 - 1$ are in $\{0,\dots,d\}^m \setminus F$. As $|\mathcal{B}| \leq w^2$, at least one of these $2w^2 - 1$ vectors is in $\{0,\dots,d\}^m \setminus (F \cup \mathcal{B})$. Pick any such vector and call it $\mathbf{b}'_k$. □

We will now use the above two claims to construct $\mathbf{n}_{\mathbf{b}_k}$ for all $k \in [p]$. We will use $\mathbf{b}'_k$ from Claim 45 as a proxy for $\mathbf{b}_k$. Notice that $u_{\mathbf{b}'_k}$ is dependent on $\left\{u_{\mathbf{e}} : \mathbf{e} \in \mathcal{B} \text{ and } \mathbf{e} \prec \mathbf{b}'_k\right\}$. Let this dependence vector be $\mathbf{n}_{\mathbf{b}_k}$. This completes the construction of N. We will now show that $[CMDN]_F$ is an invertible matrix.

$[CMDN]_F$ **is invertible.** As C is a diagonal matrix with non-zero entries, it is sufficient to show that $[MDN]_F = [M]_F DN$ is an invertible matrix, where $[M]_F$ is the sub-matrix of M consisting of only those rows of M that are indexed by $\mathbf{b} \in F$. The following claim lets us simplify the structure of $[M]_F$ so that it becomes easier to argue that $[M]_F DN$ is invertible.

Claim 46. *There is a row operation matrix $R \in \mathrm{GL}(d^m, \mathbb{F})$ having determinant 1 such that $R[M]_F$ has the following structure: The rows of $R[M]_F$ are indexed by $\mathbf{b} = (b_1, \ldots, b_m) \in F$ and its columns by $\mathbf{e} = (e_1, \ldots, e_m) \in \{0, \ldots, d\}^m$. Its entry indexed by $(\mathbf{b}, \mathbf{e})$ is non-zero if and only if for all $i \in [m]$, $b_i = e_i$ if $e_i \neq 0$. All the non-zero entries of $R[M]_F$ are either 1 or -1.*

Proof: We prove the claim by induction on m. For $m = 1$,

$$[M]_F = \begin{pmatrix} 1 & \binom{d}{d-1} & \binom{d}{d-2} & \cdots & \binom{d}{1} & 1 \\ 0 & 1 & \binom{d-1}{d-2} & \cdots & \binom{d-1}{1} & 1 \\ 0 & 0 & 1 & \cdots & \binom{d-2}{1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}.$$

Let R_1 be the row operation matrix that multiplies the last row of $[M]_F$ by $\binom{2}{1}$ and subtracts it from the second to last row; then it multiplies the last row by $\binom{3}{1}$, the second to last row by $\binom{3}{2}$ and subtracts them from the third to last row, and so on. Then, the first d columns of $R_1[M]_F$ form a $d \times d$ identity matrix. Also, it is not hard to see that the entry in the last column of the row of $R_1[M]_F$ indexed by $e \in [d]$ is $1 - \binom{e}{1} + \binom{e}{2} - \cdots + (-1)^{e-1}\binom{e}{e-1} = (-1)^{e-1}$. Let R_1 be R. Also, ignoring the last column of $R[M]_F$ and $[M]_F$, the remaining sub-matrices of both the matrices are upper triangular with ones on the diagonal. Thus both of them have determinant 1. As R relates them, it also has determinant 1.

Assume that the claim is true for all values of m' up to, but not including $m \geq 2$. Let the matrix M for m' be denoted by $M_{m'}$ and R for m' be denoted by $R_{m'}$. Then, $[M_m]_F = [M_{m-1}]_F \otimes [M_1]_F$. Let $R_m := R_{m-1} \otimes R_1$. Then, $R_m[M_m]_F = (R_{m-1} \otimes R_1)\left([M_{m-1}]_F \otimes [M_1]_F\right) = (R_{m-1}[M_{m-1}]_F) \otimes (R_1[M_1]_F)$. Thus, the claim that $R_m[M_m]_F$ has the desired structure follows from the induction hypothesis. Further, as both R_{m-1} and R_1 have determinant 1, $\det(R_m) = 1$. □

Because of the above claim, showing that $R[M]_F DN$ is invertible would suffice. Just like we did with M, we also impose the order $\prec$ on the columns of $R[M]_F$ that are indexed by $\mathbf{e} \in \{0, \ldots, d\}^m$. Recall that the rows of $R[M]_F$ and the columns of N are indexed by $\mathbf{b} \in F$. We order these indices as follows: we keep the indices $\mathbf{b} \in F \setminus \mathcal{B}$ before $\mathbf{b}_1, \ldots, \mathbf{b}_p$. We will treat $\mathbf{r}^{-\mathbf{e}}$ as a monomial in $(-r_1)^{-1}, \ldots, (-r_m)^{-1}$ "variables" and impose the order $\prec$ on the monomials in these variables. Let $A := \{\mathbf{b} : \mathbf{b} \in F \setminus \mathcal{B}\} \cup \left\{\mathbf{b}'_1, \ldots, \mathbf{b}'_p\right\}$; notice that $|A| = |F|$. Also, the elements of A are ordered as the elements of F but with $\mathbf{b}'_k$ replacing $\mathbf{b}_k$ for $k \in [p]$. Then, from the Cauchy-Binet formula and the construction of the matrix N, $\det(R[M]_F DN)$ equals

$$\det\left([R[M]_F]_{\bullet,A}\right)[N]_A \cdot \prod_{\mathbf{e}\in A} \mathbf{r}^{-\mathbf{e}} + \text{ lower order monomials in the } (-r_1)^{-1}, \ldots, (-r_m)^{-1} \text{ variables.}$$

Here $[R[M]_F]_{\bullet,A}$ denotes the restriction of $R[M]_F$ to the columns indexed by $\mathbf{e} \in A$, and $[N]_A$ denotes the restriction of N to the rows indexed by $\mathbf{e} \in A$. Thus to show that $R[M]_F DN$ (and therefore $[CMDN]_F$) is invertible, it is sufficient to prove the following two claims.

Claim 47. *$[N]_A$ is an identity matrix.*

Proof: This basically follows from the construction of N: Consider a $\mathbf{b} \in F \setminus \mathcal{B}$. As A does not contain any element of $\mathcal{B}$, the column of $[N]_A$ indexed by $\mathbf{b}$ has only one non-zero entry (which is 1) in the row indexed by $\mathbf{b}$. Similarly, the column of $[N]_A$ indexed by $\mathbf{b}_k$ for any $k \in [p]$ has only one non-zero entry (which is 1) in the row indexed by $\mathbf{b}'_k$. The claim then follows from the fact that the elements of A are ordered as the elements of F but with $\mathbf{b}'_k$ replacing $\mathbf{b}_k$ for all $k \in [p]$. □

Claim 48. *The matrix $[R[M]_F]_{\bullet,A}$ is an upper triangular matrix with 1 or -1 entries on the diagonal.*

Proof: Consider the column of $[R[M]_F]_{\bullet,A}$ indexed by some $\mathbf{b} \in F \setminus \mathcal{B}$. From Claim 46, the only non-zero entry in this column is in the row indexed by $\mathbf{b}$ itself. Now consider a column of $[R[M]_F]_{\bullet,A}$ indexed by $\mathbf{b}'_k$ for some $k \in [p]$. From Claims 44 and 45, $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k}$ for all $\ell > k$. As every coordinate of $\mathbf{b}_k$ is non-zero, it follows from Claim 46 that the entry in the row indexed by $\mathbf{b}_\ell$ must be 0 for every $\ell > k$. Also, from Claim 45, $\mathbf{b}_k$ and $\mathbf{b}'_k$ agree at all coordinates $\mathbf{b}'_k$ is non-zero. So, from Claim 46, the entry in the row indexed by $\mathbf{b}_k$ must be non-zero. Also, recall from Claim 46 that the non-zero entries of $R[M]_F$ are either 1 or -1. The claim then follows from the fact that the elements of A are ordered same as elements of F but with $\mathbf{b}'_k$ replacing $\mathbf{b}_k$ for all $k \in [p]$. □

B Missing proof from Section 4

B.1 Proof of Lemma 33

The entries of U, the columns of M, the rows and columns of D, and the rows of N are indexed by $\mathbf{e} \in \{0,1\}^m$. Impose the degree lexicographic order, denoted by $\prec_{\text{dlex}}$, on the indices $\mathbf{e} \in \{0,1\}^m$ of U and the other three matrices[19]. Pick the *minimal* basis of the space spanned by the entries of U according to this order, i.e., consider the entries of U in the order dictated by $\prec_{\text{dlex}}$ while forming the basis. Let $\mathcal{B} := \{\mathbf{e} \in \{0,1\}^m : u_{\mathbf{e}}$ is in the minimal basis of U w.r.t. $\prec_{\text{dlex}}\}$.

Observation 49. *By the induction hypothesis, for every $\mathbf{e} \in F \cap \mathcal{B}$, $\text{Supp}(\mathbf{e}) = 2\mu - (q^* - 1)$.*

Construction of the matrix N. The columns of N are indexed by $\mathbf{b} \in F$. We will now specify a set of column vectors $\{\mathbf{n}_{\mathbf{b}} : \mathbf{b} \in F\}$ in the null space of U such that the column of N indexed by $\mathbf{b} \in F$ is $\mathbf{n}_{\mathbf{b}}$. There are two cases for $\mathbf{b} \in F$:

Case 1: $\mathbf{b} \in F \setminus \mathcal{B}$. In this case, $u_{\mathbf{b}}$ is dependent on $\{u_{\mathbf{e}} : \mathbf{e} \in \mathcal{B}$ and $\mathbf{e} \prec_{\text{dlex}} \mathbf{b}\}$. Pick this dependence vector as $\mathbf{n}_{\mathbf{b}}$.

[19] by identifying $\mathbf{e}$ with an m-variate monomial.

Case 2: $\mathbf{b} \in F \cap \mathcal{B}$. Let there be p such $\mathbf{b}$, $\mathbf{b}_1, \ldots, \mathbf{b}_p$, where $p \leq |\mathcal{B}| \leq w^2$. For a set $E \subseteq [m]$ and $\mathbf{b} \in \{0,1\}^m$, let $(\mathbf{b})_E$ denote the vector obtained by projecting $\mathbf{b}$ to the coordinates in E. Roughly speaking, the following claim says that each of these p vectors has a "small signature" that differentiates it from the other $p-1$ vectors.

Claim 50. *There exist sets $E_1, \ldots, E_p \subseteq [m]$, each of size $w^2 - 1$ such that for all $k \in [p]$,*

1. $\mathrm{Supp}\left((\mathbf{b}_k)_{E_k}\right) = w^2 - 1,$
2. $(\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k}\ \forall \ell \neq k.$

Proof: For $k \in [p]$, let $\mathcal{S}(\mathbf{b}_k)$ be the set of coordinates where $\mathbf{b}_k$ is non-zero. Fix any $k \in [p]$. Notice that $\mathrm{Supp}(\mathbf{b}_k) = |\mathcal{S}(\mathbf{b}_k)| = 2\mu - (q^* - 1) \geq \mu + 2 = w^2 + \lceil \log w^2 \rceil + 2$. For $\ell \neq k$, as $\mathrm{Supp}(\mathbf{b}_k) = \mathrm{Supp}(\mathbf{b}_\ell)$ and $\mathbf{b}_k \neq \mathbf{b}_\ell$, there must exist an $i_\ell \in \mathcal{S}(\mathbf{b}_k)$, such that the i_ℓ-th coordinate of $\mathbf{b}_k$ and $\mathbf{b}_\ell$ are distinct. Put all such i_ℓ for $\ell \neq k$ in E_k. If $|E_k|$ is still less than $w^2 - 1$, then arbitrarily put some more elements in E_k from $\mathcal{S}(\mathbf{b}_k)$ so that $|E_k| = w^2 - 1$. This can be done as $\mathcal{S}(\mathbf{b}_k)$ is sufficiently large. □

As before, we will call E_k the signature of $\mathbf{b}_k$. The following claim tells us that for each vector $\mathbf{b}_k$, there is a vector that is not in $\mathcal{B}$ and has support less than $2\mu - (q^* - 1)$, but agrees with $\mathbf{b}_k$ on its signature and so in some sense can be used as a proxy for $\mathbf{b}_k$.

Claim 51. *For every $k \in [p]$, there exists a vector $\mathbf{b}'_k \in \{0,1\}^m \setminus (F \cup \mathcal{B})$ such that $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k}$ and also $\mathbf{b}'_k$ and $\mathbf{b}_k$ agree on all locations where $\mathbf{b}'_k$ is non-zero.*

Proof: Similar to the proof of Claim 45. □

We will now use the above two claims to construct $\mathbf{n}_{\mathbf{b}_k}$ for all $k \in [p]$. We will use $\mathbf{b}'_k$ from Claim 51 as a proxy for $\mathbf{b}_k$. Notice that $u_{\mathbf{b}'_k}$ is dependent on $\left\{u_{\mathbf{e}} :\ \mathbf{e} \in \mathcal{B} \text{ and } \mathbf{e} \prec_{\text{dlex}} \mathbf{b}'_k\right\}$. Let this dependence vector be $\mathbf{n}_{\mathbf{b}_k}$. This completes the construction of N. We will now show that $[CMDN]_F$ is invertible. In fact, we will show that $\det\left([CMDN]_F\right)$ is the ratio of a polynomial in $\mathbb{F}[\mathbf{t}]$ which contains a monomial of degree at most $2w^2\mu$ and a product of a bunch of non-zero linear forms in $\mathbb{F}[\mathbf{t}]$.

$[CMDN]_F$ **is invertible.** Let $[M]_F$ be the restriction of M to the rows indexed by F, and $[C]_F$ the restriction of C to the rows and columns indexed by F.

Observation 52. *The matrix $[M]_F$ has the following structure: The rows of $[M]_F$ are indexed by $\mathbf{b} = (b_1, \ldots, b_m) \in F$ and its columns by $\mathbf{e} = (e_1, \ldots, e_m) \in \{0,1\}^m$. Its entry indexed by $(\mathbf{b}, \mathbf{e})$ is non-zero if and only if for all $i \in [m]$, $b_i = e_i$ if $e_i \neq 0$. All non-zero entries are* 1.

We order the indices $\mathbf{b} \in F$ as follows: Let $F_0 := \{\mathbf{b} \in F : \mathrm{Supp}(\mathbf{b}) > 2\mu - (q^* - 1)\}$ and $F_1 := \{\mathbf{b} \in F : \mathrm{Supp}(\mathbf{b}) = 2\mu - (q^* - 1)\}$. We first keep the $\mathbf{b} \in F_0$ in (descending) degree lexicographic order[20], followed by $\mathbf{b} \in F_1 \setminus \mathcal{B}$ in (reverse) lexicographic order[21], and then $\mathbf{b}_1, \ldots, \mathbf{b}_p$. Also, let $A := (F \setminus \mathcal{B}) \uplus \left\{\mathbf{b}'_1, \ldots, \mathbf{b}'_p\right\}$. Notice that $|A| = |F|$. Also, the elements of A are ordered as the elements of F but with $\mathbf{b}'_k$ replacing $\mathbf{b}_k$ for $k \in [p]$. For any $S \subseteq \{0,1\}^m$ of size $|S| = |F|$, let $[M]_{F,S}$

[20] i.e., $\mathbf{b}$ comes before $\hat{\mathbf{b}}$ if $\mathrm{Supp}(\mathbf{b}) > \mathrm{Supp}(\hat{\mathbf{b}})$, or if $\mathrm{Supp}(\mathbf{b}) = \mathrm{Supp}(\hat{\mathbf{b}})$ and $\hat{\mathbf{b}} \prec_{lex} \mathbf{b}$.

[21] i.e., $\mathbf{b}$ comes before $\hat{\mathbf{b}}$ if $\hat{\mathbf{b}} \prec_{lex} \mathbf{b}$.

denote the restriction of $[M]_F$ to the columns indexed by $\mathbf{e} \in S$, and $[N]_S$ denote the restriction of N to the rows indexed by $\mathbf{e} \in S$. Now,

$$\begin{aligned}
& \det([CMDN]_F) \\
= & \det([C]_F)\det([M]_F DN) \\
= & \prod_{\mathbf{b}\in F} \mathbf{r}^{\mathbf{b}} \cdot \left(\sum_{\substack{S\subseteq\{0,1\}^m \\ |S|=|F|}} \det\left([M]_{F,S}\right) \cdot \det([N]_S) \cdot \prod_{\mathbf{e}\in S} \mathbf{r}^{-\mathbf{e}} \right) \\
= & \prod_{\mathbf{b}\in F} \mathbf{r}^{\mathbf{b}} \cdot \left(\sum_{\substack{S\subseteq A\uplus\mathcal{B} \\ |S|=|F|}} \det\left([M]_{F,S}\right) \cdot \det([N]_S) \cdot \prod_{\mathbf{e}\in S} \mathbf{r}^{-\mathbf{e}} \right) \\
= & \prod_{\mathbf{b}\in F} \mathbf{r}^{\mathbf{b}} \cdot \left(\sum_{\substack{S\subseteq A\uplus\mathcal{B} \\ |S|=|F|}} \det\left([M]_{F,S}\right) \cdot \det([N]_S) \cdot \prod_{\mathbf{e}\in S\cap A} \mathbf{r}^{-\mathbf{e}} \cdot \prod_{\mathbf{e}\in S\cap\mathcal{B}} \mathbf{r}^{-\mathbf{e}} \right) \\
= & \prod_{\mathbf{b}\in F} \mathbf{r}^{\mathbf{b}} \cdot \prod_{\mathbf{e}\in A\uplus\mathcal{B}} \mathbf{r}^{-\mathbf{e}} \cdot \left(\sum_{\substack{S\subseteq A\uplus\mathcal{B} \\ |S|=|F|}} \det\left([M]_{F,S}\right) \cdot \det([N]_S) \cdot \prod_{\mathbf{e}\in A\setminus S} \mathbf{r}^{\mathbf{e}} \cdot \prod_{\mathbf{e}\in \mathcal{B}\setminus S} \mathbf{r}^{\mathbf{e}} \right),
\end{aligned}$$

where the second equality follows from the Cauchy-Binet formula and the third equality from the fact that for any $S \nsubseteq A \uplus \mathcal{B}$, $\det([N]_S) = 0$. Now, notice that $\prod_{\mathbf{b}\in F} \mathbf{r}^{\mathbf{b}} \cdot \prod_{\mathbf{e}\in A\uplus\mathcal{B}} \mathbf{r}^{-\mathbf{e}}$ is the reciprocal of a product of non-zero linear forms in $\mathbf{t}$-variables, as $F \subseteq A \uplus \mathcal{B}$. We shall now prove that

$$\sum_{\substack{S\subseteq A\uplus\mathcal{B} \\ |S|=|F|}} \det\left([M]_{F,S}\right) \cdot \det([N]_S) \cdot \prod_{\mathbf{e}\in A\setminus S} \mathbf{r}^{\mathbf{e}} \cdot \prod_{\mathbf{e}\in \mathcal{B}\setminus S} \mathbf{r}^{\mathbf{e}} \tag{6}$$

has a $\mathbf{t}$-monomial of degree at most $w^2(2\mu - (q^* - 1))$.

Claim 53. *$[N]_A$ is an identity matrix.*

Proof: Same as that of Claim 47. □

Claim 54. *The matrix $[M]_{F,A}$ is an upper triangular matrix with ones on the diagonal.*

Proof: Consider the column of $[M]_{F,A}$ indexed by some $\mathbf{b} \in F \setminus \mathcal{B}$. Because of the way we have ordered the elements in F and A, it follows from Observation 52, the only non-zero entries in this column are in and above the row indexed by $\mathbf{b}$. Now consider a column of $[M]_{F,A}$ indexed by $\mathbf{b}'_k$ for some $k \in [p]$. From Claims 50 and 51, $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k}$ for all $\ell \neq k$. As every coordinate of $(\mathbf{b}_k)_{E_k}$ is non-zero, it follows from Observation 52 that the entry in the row indexed by $\mathbf{b}_\ell$ must be 0 for every $\ell \neq k$. Also, from Claim 51, as $\mathbf{b}_k$ and $\mathbf{b}'_k$ agree at all coordinates $\mathbf{b}'_k$ is non-zero. So, from Observation 52, the entry in the row indexed by $\mathbf{b}_k$ must be non-zero. Also, recall from Observation 52 that the non-zero entries of $[M]_F$ are ones. The claim then follows from the fact that the elements of A are ordered as that of F but with $\mathbf{b}'_k$ replacing $\mathbf{b}_k$ for $k \in [p]$. □

Claim 55. $\det([M]_{F,A}) \cdot \det([N]_A) \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus A} \mathbf{r}^{\mathbf{e}} = \prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}} \neq 0$ *and has* $\mathbf{t}$*-degree at most* $2w^2\mu$.

Proof: $\det([M]_{F,A}) \cdot \det([N]_A) \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus A} \mathbf{r}^{\mathbf{e}} = \prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}} \neq 0$ follows from Claims 53 and 54 and the fact that $A \cap \mathcal{B}$ is empty. For every $\mathbf{e} \in \mathcal{B}$, $\deg_{\mathbf{t}}(\mathbf{r}^{\mathbf{e}}) \leq 2\mu - (q^* - 1)$. So, $\deg_{\mathbf{t}}(\prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}}) \leq w^2 \cdot (2\mu - (q^* - 1)) \leq 2w^2\mu$, as $|\mathcal{B}| \leq w^2$. □

Claim 56. *For any* $S \subseteq A \uplus \mathcal{B}$ *such that* $|S| = |F|$ *and* $\det([N]_S)$ *is non-zero, there is a one-to-one correspondence between* $A \setminus S$ *and* $S \cap \mathcal{B}$ *such that if* $\mathbf{e} \in A \setminus S$ *corresponds to* $\mathbf{e}' \in S \cap \mathcal{B}$*, then* $\mathbf{e}' \prec_{\text{dlex}} \mathbf{e}$.

Proof: As $\det([N]_S) \neq 0$, there must be a one-to-one correspondence between the rows and columns of $[N]_S$ such that if the column indexed by $\mathbf{b} \in F$ corresponds to a row indexed by $\mathbf{e} \in S$, then the $(\mathbf{e}, \mathbf{b})$-th entry of $[N]_S$ must be non-zero. Obtain a one-to-one correspondence between A and S from the above correspondence by replacing $\mathbf{b}_k$ with $\mathbf{b}'_k$ for all $k \in [p]$. Notice that, if $\mathbf{e} \in A$ corresponds to $\mathbf{e}'$ in S, then either $\mathbf{e}' \prec_{\text{dlex}} \mathbf{e}$ or $\mathbf{e}' = \mathbf{e}$. Now, removing $A \cap S$ from A yields $A \setminus S$, and removing $A \cap S$ from S yields $S \cap \mathcal{B}$. So the correspondence between A and S yields the desired correspondence between $A \setminus S$ and $S \cap \mathcal{B}$. □

The above claim implies that for every $S \in A \uplus \mathcal{B}$ of size $|F|$, either $\det([M]_{F,S}) \cdot \det([N]_S) \cdot \prod_{\mathbf{e} \in A \setminus S} \mathbf{r}^{\mathbf{e}} \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus S} \mathbf{r}^{\mathbf{e}}$ is 0, or $\prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}} \prec_{\text{dlex}} \prod_{\mathbf{e} \in A \setminus S} \mathbf{r}^{\mathbf{e}} \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus S} \mathbf{r}^{\mathbf{e}}$. Hence, $\prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}}$ is the smallest $\mathbf{r}$-monomial in the polynomial given in (6) w.r.t. $\prec_{\text{dlex}}$ order, and so, the homogeneous component of this polynomial that has the same $\mathbf{r}$-degree as that of $\prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}}$ survives. Now, from Claim 55 and the fact that $\ell_1, \ldots, \ell_n$ are linearly independent, the polynomial in (6) has a $\mathbf{t}$-monomial of degree $\leq 2w^2\mu$.

C Hitting sets for the orbits of sparse polynomials

In this section, we provide the proofs of Theorems 29 and 8 due to [MS21].

Proof of Theorem 29: Let $g = f(A\mathbf{x})$, where $A \in \text{GL}(n, \mathbb{F})$, and suppose that A maps $x_i \mapsto \ell_i(\mathbf{x})$ for all $i \in [n]$. If $g(\mathbf{x})$ is a non-zero constant, then $g\left(\mathcal{G}_k^{SV}\right) \neq 0$ for all $k \geq 1$. So, to prove the theorem, we just need to show that if g is not a constant, then neither is $g\left(\mathcal{G}_{(\lceil \log s \rceil + 1)}^{SV}\right)$. We can assume without loss of generality that f is constant-free, i.e., $f(\mathbf{0}) = 0$, as otherwise we can prove the theorem on the constant-free part of f.

The theorem shall be proved by induction on the number of monomials s in f. Assume without loss of generality that the monomials of f do not have a non-trivial GCD, for otherwise we can take the GCD common and then prove the theorem separately for both the GCD (which is a monomial) and the residual polynomial. For the base case of $s = 1$, g is a product of linear forms and hence $g(\mathcal{G}_1^{SV})$ is a non-constant if g is a non-constant. Now, assume that the theorem is true for all n-variate, degree d polynomials with sparsity strictly less than $s > 1$. For f, there are two cases:

Case 1: Every variable x_i appears in at most $s/2$ monomials of f. Assume without loss of generality that x_1 appears in some monomial of f. Then, as $\text{char}(\mathbb{F}) = 0$ or $> d$, $\frac{\partial f}{\partial x_1} \neq 0$. Moreover, the sparsity of $\frac{\partial f}{\partial x_1}$ is at most $s/2$ and so, from the induction hypothesis, $\frac{\partial f}{\partial x_1}\left(A\mathcal{G}_{(\lceil \log \frac{s}{2} \rceil + 1)}^{SV}\right) \neq$

0. Now, recall that the gradient of a polynomial $p(\mathbf{x})$, denoted by ∇p, is the column vector $\left(\frac{\partial p}{\partial x_1}\ \frac{\partial p}{\partial x_2}\ \cdots\ \frac{\partial p}{\partial x_n}\right)^T$. By the chain rule of differentiation,

$$\nabla g = A^T \cdot [\nabla f](A\mathbf{x}).$$

As A^T is invertible, there exists a $j \in [n]$ such that $\frac{\partial g}{\partial x_j}\left(\mathcal{G}^{SV}_{(\lceil \log \frac{s}{2} \rceil + 1)}\right) = \frac{\partial g}{\partial x_j}\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil)}\right) \neq 0$. Then, Observation 15 implies that $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1)}\right)$ is not a constant (as $\deg(g) = d$ and $|\mathbb{F}| > nd$).

Case 2: There is a variable - say x_1 - which appears in more than $s/2$ monomials. As the linear forms $\ell_1, \ldots, \ell_n$ are linearly independent, $\ell_1 \neq 0$. So, there exists an $i \in [n]$ such that the coefficient of x_i in ℓ_1 is non-zero. Let $\mathbf{x}_{-i}$ denote the vector $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ and suppose that $\ell_1(\mathbf{x}) = c(x_i - h(\mathbf{x}_{-i}))$, where $c \in \mathbb{F}\backslash\{0\}$ and h is a linear form. Also, for all $j \in \{2, \ldots, n\}$, define

$$\widetilde{\ell}_j(\mathbf{x}_{-i}) := \ell_j(x_1, \ldots, x_{i-1}, h(\mathbf{x}_{-i}), x_{i+1}, \ldots, x_n)\ = \ell_j \mod \ell_1.$$

As $\ell_1, \ldots, \ell_n$ are $\mathbb{F}$-linearly independent, so are $\widetilde{\ell}_2(\mathbf{x}_{-i}), \ldots, \widetilde{\ell}_n(\mathbf{x}_{-i})$. Now,

$$g(x_1, \ldots, x_{i-1}, h(\mathbf{x}_{-i}), x_{i+1}, \ldots, x_n) = f(0, \widetilde{\ell}_2, \ldots, \widetilde{\ell}_n).$$

As the monomials of f do not have a non-trivial GCD and f is constant-free, $f(0, x_2, \ldots, x_n)$ is not a constant. Moreover, it has at most $s/2$ monomials. Let $t = \lceil \log \frac{s}{2} \rceil + 1$ and recall that $\mathcal{G}^{SV}_t = \left(\mathcal{G}^{(1)}_t, \ldots, \mathcal{G}^{(n)}_t\right)$. Then, from the induction hypothesis,

$$g\left(\mathcal{G}^{(1)}_t, \ldots, \mathcal{G}^{(i-1)}_t, h\left(\mathcal{G}^{(1)}_t, \ldots, \mathcal{G}^{(i-1)}_t, \mathcal{G}^{(i+1)}_t, \ldots \mathcal{G}^{(n)}_t\right), \mathcal{G}^{(i+1)}_t, \ldots, \mathcal{G}^{(n)}_t\right)$$

is a non-constant. This implies, for a fresh variable z_{t+1}, $g\left(\mathcal{G}^{(1)}_t, \ldots, \mathcal{G}^{(i-1)}_t, z_{t+1}, \mathcal{G}^{(i+1)}_t, \ldots, \mathcal{G}^{(n)}_t\right)$ is a non-constant. Hence, $g\left(\mathcal{G}^{(1)}_t, \ldots, \mathcal{G}^{(i-1)}_t, z_{t+1} + \mathcal{G}^{(i)}_t, \mathcal{G}^{(i+1)}_t, \ldots, \mathcal{G}^{(n)}_t\right)$ is a non-constant, and therefore, $g\left(\mathcal{G}^{(1)}_{t+1}, \ldots, \mathcal{G}^{(n)}_{t+1}\right)$ is a non-constant, as $\mathcal{G}^{SV}_{t+1|(y_{t+1}=\alpha_i)} = \mathcal{G}^{SV}_t + \mathbf{e}_i \cdot z_{t+1}$. Putting the value of t, we get that $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1)}\right)$ is a non-constant. □

Theorem 8 follows as a corollary to Theorem 29. Let f be a non-zero n-variate, s-sparse polynomial, and $g \in \mathrm{orb}(f)$. Then, $g(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1)}) \neq 0$. Now, $g\left(\mathcal{G}^{SV}_{(\lceil \log s \rceil + 1)}\right)$ is a $2\left(\lceil \log s \rceil + 1\right)$-variate polynomial of degree nd. As $|\mathbb{F}| > nd$, a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in time $(nd)^{O(\log s)}$.

D Hitting sets for the orbits of constant-depth, constant-occur formulas

Let $f \in \mathbb{F}[\mathbf{x}]$ be a n-variate, degree-D polynomial computed by a (Δ, k, s) formula i.e., a depth-Δ, occur-k formula of size-s. Let us identify f with a (Δ, k, s) formula computing it. In this section, the level of a gate in f will be one plus its distance from the output gate of f. Just like we did in Section 5, we first upper bound the top fan-in of f in Section D.1 and then use the notion of faithful homomorphisms to construct hitting sets for $\mathrm{orb}(f)$ in Section D.2.

D.1 Upper bounding the top fan-in of f

We begin by showing that f can be written in a "canonical" form.

Claim 57. *If f is a (Δ, k, s) formula, then it can also be computed by a $(\Delta, k, (2s)^{\Delta})$ formula in a canonical form with the following properties:*

1. *All gates connected to the leaves of f are $\times\curlywedge$ gates.*
2. *f has alternating levels of $+$ and $\times\curlywedge$ gates.*

Proof: While f contains a $+$ gate connected to the leaves, we merge all the leaves connected to it into a single leaf node computing their sum. Now, if this $+$ gate is not connected to any gate other than this leaf, it can simply be replaced by the leaf after multiplying the sparse polynomial computed by the leaf by the label of the edge between it and the $+$ gate. This does not increase the depth, size or occur of f. Otherwise, we add a $\times\curlywedge$ gate between the $+$ gate and the leaf. While this can increase the size of f by a factor of 2, the occur remains the same. The depth does not increase, because the $+$ gate is also connected to some non-leaf node. Now f has property 1.

If f has a $+$ gate q which is fed another $+$ gate h as input and the edge connecting them is labelled by α, then we can simply remove h, connect all its inputs directly to q and multiply the labels of edges connecting all these inputs to q by α. This modification to f clearly does not increase its depth, size or occur. Also, now each sum gate in f is connected solely to $\times\curlywedge$ gates.

Consider any *maximal* sub-tree of f made up, solely, of $\times\curlywedge$ gates. Let its root be q and its inputs $h_1, \ldots, h_m$. Then, $q = h_1^{e_1} \cdots h_m^{e_m}$, where e_i is the product of the weights of all edges on the path from h_i to q. As the sub-tree is maximal, none of $h_1, \ldots, h_m$ are $\times\curlywedge$ gates and q is also not an input to a $\times\curlywedge$ gate. Thus, if we replace each such sub-tree with a single $\times\curlywedge$ gate computing the same polynomial, f will also satisfy 2. Notice that, doing this does not increase the depth or occur; size on the other hand, may increase. Suppose that the depth of the sub-tree is Δ'. Let the sum of weights of edges connecting gates at level $\ell + 1$ (from q) to gates at level ℓ be $r_\ell \leq 2s$, for all $\ell \in [\Delta' - 1]$. Also, let the sum of weights of edges connecting the leaves be $r_{\Delta'}$. As, all edge weights are non-negative, $\sum_{i \in [m]} e_i \leq \prod_{\ell \in [\Delta']} r_\ell \leq (2s)^{\Delta'} \leq (2s)^{\Delta - 2}$. Since, there can be no more than $(2s)$ such sub-trees, the size of f can increase by at most $(2s)^{\Delta-1}$. Thus, size of f is at most $2s + (2s)^{\Delta - 1} \leq (2s)^{\Delta}$. □

We can also assume that the output gate of f is not a $\times\curlywedge$ gate, for otherwise, we only need to construct a hitting set generator for orbits of all of its factors which themselves are $(\Delta - 1, k, (2s)^{\Delta})$ formulas, with $+$ gates at the top or are sparse polynomials. Thus, we can assume without loss of generality that Δ is an even number: if $\Delta \neq 2$, then the top most gate is a $+$ gate, f has alternating levels of $+$ and $\times\curlywedge$ gates and gates connected to the leaves are $\times\curlywedge$ gates. We now make the following claim which will allow us to assume that the top fan-in of f is at most k.

Claim 58. *Let f be a (Δ, k, s) formula in the canonical form of Claim 57, with either a $+$ gate at the top or $\Delta = 2$. Then, for any $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a $(\Delta, (2k)^{\Delta/2}, (2k)^{\Delta/2} s)$ formula in the canonical form with the top fan-in bounded by k.*

Proof: When $\Delta = 2$, f is a polynomial of sparsity s and $k = 1$. So, the sparsity of $\frac{\partial f}{\partial x_i}$ is at most s and the depth and occur do not increase, making the claim true. Assume, by the way of induction, that the claim is true for all formulas of depth $\Delta - 2$. Let $x = x_i$, $f = \sum_{i \in [m]} f_i$ and x be present only in $f_1, \ldots, f_r$, $r \leq k$. Furthermore, for all $i \in [r]$, let $f_i = \prod_{j \in m_i} q_{i,j}^{e_{i,j}}$ and x be present only in $q_{i,1}, \ldots q_{i,r_i}$, $\sum_{i \in [r]} r_i \leq k$. Then,

$$\begin{aligned}\frac{\partial f}{\partial x} &= \sum_{i \in [r]} \left(\prod_{j=r_i+1}^{m_i} q_{i,j}^{e_{i,j}} \right) \cdot \left(\sum_{j \in [r_i]} e_{i,j} \frac{\partial q_{i,j}}{\partial x} \cdot q_{i,j}^{e_{i,j}-1} \cdot \prod_{\substack{j' \in [r_i] \\ j' \neq j}} q_{i,j'}^{e_{i,j'}} \right) \\ &= \sum_{i \in [r]} \sum_{j \in [r_i]} \left(\frac{\partial q_{i,j}}{\partial x} \cdot \prod_{j' \in [m_i]} q_{i,j'}^{e'_{i,j'}} \right),\end{aligned}$$

where $e'_{i,j'}$ is either $e_{i,j'}$ or $e_{i,j'} - 1$. First of all, notice that, the top fan-in of $\frac{\partial f}{\partial x}$ is at most $\sum_{i \in [r]} r_i \leq k$. As all $q_{i,j}$ are formulas of depth $\Delta - 2$, from the induction hypothesis, $\frac{\partial q_{i,j}}{\partial x}$ is also a depth $\Delta - 2$ formula. Thus, the depth of $\frac{\partial f}{\partial x}$ is at most Δ. However, the size and occur may change.

For all $i \in [r]$, let the occur of f_i be $p_i \leq k$; then the occur of $\prod_{j' \in [m_i]} q_{i,j'}^{e'_{i,j'}}$ is at most p_i. Also, from the induction hypothesis, $\frac{\partial q_{i,j}}{\partial x}$ has occur $(2k)^{(\Delta-2)/2}$. So, the occur of $\frac{\partial f}{\partial x}$ is at most $\sum_{i \in [r]} r_i \left((2k)^{(\Delta-2)/2} + p_i \right)$, which can be bounded from above by $(2k)^{\Delta/2}$. Similarly, suppose that the size of f_i is $s_i \leq s - 1$; then the size of $\prod_{j' \in [m_i]} q_{i,j'}^{e'_{i,j'}}$ is at most s_i. Also, from the induction hypothesis, $\frac{\partial q_{i,j}}{\partial x}$ has size $(2k)^{\frac{\Delta-2}{2}} s$. So, the size of $\frac{\partial f}{\partial x}$ is at most $\sum_{i \in [r]} r_i \left((2k)^{\frac{\Delta-2}{2}} s + s_i + 1 \right) \leq (2k)^{\Delta/2} s$. □

We now upper bound the top fan-in of f using this claim. Let $A \in \mathrm{GL}(n, \mathbb{F})$ and $g(\mathbf{x}) = f(A\mathbf{x})$. If f is a constant, then constructing a hitting set for $\mathrm{orb}(f)$ is trivial. Otherwise, there exists an $i \in [n]$ such that $\frac{\partial f}{\partial x_i} \neq 0$ (because $\mathrm{char}(\mathbb{F}) > (2ks)^{\Delta^3 R} \geq D$). Suppose that a polynomial map, $\mathcal{G} : \mathbb{F}^t \to \mathbb{F}^n$ of degree at most $nR + 1$ is a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$. The gradient of a polynomial $p(\mathbf{x})$, denoted by ∇p, is the column vector $\left(\frac{\partial p}{\partial x_1} \; \frac{\partial p}{\partial x_2} \; \ldots \; \frac{\partial p}{\partial x_n}\right)^T$. By the chain rule of differentiation,

$$\nabla g = A^T \cdot [\nabla f](A\mathbf{x}).$$

As A^T is invertible, $\frac{\partial f}{\partial x_i}(A\mathcal{G}) \neq 0 \implies \nabla f(A\mathcal{G}) \neq 0 \implies \nabla g(\mathcal{G}) \neq 0 \implies \exists \in [n]$, such that $\frac{\partial g}{\partial x_j}(\mathcal{G})$ $\neq 0$. Then, from Observation 15, for $\widetilde{\mathcal{G}} := \mathcal{G} + \mathcal{G}_1^{SV}$, $g(\widetilde{\mathcal{G}}) \neq 0$, i.e., $\widetilde{\mathcal{G}}$ is a hitting set generator for $\mathrm{orb}(f)$. So, all we need to do now is construct a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_j}\right)$ and from Claim 58, $\frac{\partial f}{\partial x_j}$ has top fan-in at most k. Overloading the notation, we refer to $\frac{\partial f}{\partial x_j}$ as f, which is computed by a (Δ, k, s) formula in the canonical form and with a $+$ gate at the top whose fan-in is at most k.

D.2 Constructing a faithful homomorphism

Let $f = f_1 + \cdots + f_k$ and $A \in \mathrm{GL}(n, \mathbb{F})$. Let $g_i = f_i(A\mathbf{x})$ for all $i \in [k]$, $\mathbf{f} = (f_1, \ldots, f_k)$ and $\mathbf{g} = (g_1, \ldots, g_k)$. We now show how to create a homomorphism ϕ that is faithful to $\mathbf{g}$; from Lemma 19, this homomorphism will be a hitting set generator for $\mathrm{orb}(f)$. ϕ will be constructed recursively as follows: each level of recursion corresponds to a level in f, with the recursion starting at level 2 and ending at level $\Delta - 2$. At level ℓ, our goal will be to construct a homomorphism ϕ_ℓ which is faithful to every tuple in a certain set C_ℓ of tuples. Each tuple in C_ℓ consists of at most r_ℓ derivatives, of order at most a_ℓ, of disjoint groups of gates at level ℓ of f, evaluated at $A\mathbf{x}$. Note that, as the derivatives are of disjoint groups of gates in f, $|C_\ell| \leq s$.

For $\ell = 2$, C_2 contains only one tuple, viz. $\mathbf{g}$, $r_2 = k$ and $a_2 = 0$. For any $\ell \geq 2$, let $\mathbf{q} \in C_\ell, \mathbf{q} = (q_1, \ldots, q_{r_\ell})$, where $q_i = h_i(A\mathbf{x})$ for all $i \in [r_\ell]$ and let $\mathbf{h} = (h_1, \ldots, h_{r_\ell})$. If $\phi_{\ell+1}$ is such that $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \phi_{\ell+1}\left(J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x})\right)$, then using Lemma 39, we can construct a ϕ_ℓ faithful to $\mathbf{q}$. The following lemma which was proved in [ASSS16], helps us reduce the problem from level ℓ to level $\ell + 1$.

Lemma 59 (Lemma 4.4 of [ASSS16])**.** *Let* $\mathbf{h}$ *be a tuple of* r_ℓ *derivatives, of order at most* a_ℓ*, of gates* G *at level* ℓ *of* f, $\text{tr-deg}_{\mathbb{F}}(\mathbf{h}) = r'_\ell$ *and* $\mathbf{h}'$ *be a transcendence basis of* $\mathbf{h}$*. Any* $r'_\ell \times r'_\ell$ *minor of* $J_{\mathbf{x}}(\mathbf{h}')$ *is of the form* $\prod_i p_i^{e_i}$*, where* p_i*s are polynomials in at most* $r_{\ell+1} := (a_\ell + 1) \cdot 2^{a_\ell+1} k \cdot r_\ell^2$ *many derivatives of order at most* $a_{\ell+1} := a_\ell + 1$ *of disjoint groups of children of* G*.*

For each $\mathbf{h}$, we will use the above lemma for a non-zero $r'_\ell \times r'_\ell$ minor of $J_{\mathbf{x}}(\mathbf{h}')$. Then, the lemma gives a bunch of tuples $\mathbf{h}_1, \ldots, \mathbf{h}_u$, one for each p_i. Suppose that p_i is a polynomial in $p_{i,1}, \ldots, p_{i,m}$, which are derivatives of gates at level $\ell + 1$ of f. Then, $\mathbf{h}_i = (p_{i,1}(A\mathbf{x}), \ldots, p_{i,m}(A\mathbf{x}))$ and $C_{\ell+1}$ is a set of all $\mathbf{h}_i$, for all $\mathbf{h}$. If $\phi_{\ell+1}$ is faithful to each tuple in $C_{\ell+1}$, then from Lemma 19, $\phi_{\ell+1}\left(p_i^{e_i}(A\mathbf{x})\right) \neq 0$ and hence it preserves the rank of $J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x})$.

The base case of the recursion is when $\ell = \Delta - 2$. Our goal is to create a homomorphism $\phi_{\Delta-2}$ which is faithful to every tuple in the set $C_{\Delta-2}$, $|C_{\Delta-2}| \leq s$ of at most $r_{\Delta-2}$ many sparse polynomials (because any derivative of a sparse polynomial is a sparse polynomial) evaluated at $A\mathbf{x}$. $r_{\Delta-2}$ can be bounded from above by $R := (2k)^{2\Delta \cdot 2^\Delta}$. For all $\mathbf{q} = \mathbf{h}(A\mathbf{x}) = (h_1(A\mathbf{x}), \ldots, h_R(A\mathbf{x})) \in C_{\Delta-2}$, we will create a $\phi_{\Delta-1}$ such that $\mathrm{rank}_{\mathbb{F}(\mathbf{x})}\, J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\, \phi_{\Delta-1}\left(J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x})\right)$. Let $h_1, \ldots, h_{R'}$ be a transcendence basis of $\mathbf{h}$. As the size of f is s, every entry of any $|R'| \times |R'|$ sub-matrix of $J_{\mathbf{x}}(\mathbf{h})$ is a polynomial with sparsity and degree at most s. So, the determinant of any such sub-matrix is a polynomial with sparsity at most $R'! \cdot s^{R'} \leq R! \cdot s^R$ and degree at most sR. Hence, from Theorem 29, $\mathcal{G}^{SV}_{(\lceil \log(R! \cdot s^R) \rceil + 1)} = \mathcal{G}^{SV}_{(O(R(\log R + \log s)))}$ is a hitting set generator for orbits of these determinants. Thus, we can put $\phi_{\Delta-1} = \mathcal{G}^{SV}_{(O(R(\log R + \log s)))}$. We then repeatedly use Lemma 39 to construct ϕ_2. At level ℓ of the recursion, we add at most $r_\ell + 1 \leq R + 1$ many new variables for a total of at most $(\Delta - 2)(R + 1)$ new variables. Also, notice that at level ℓ, the polynomial that we add to $\phi_{\ell+1}$ to create ϕ_ℓ has degree at most $nr_\ell + 1 \leq nR + 1$. Thus, there exists a homomorphism ψ in at most $(\Delta - 2)(R + 1)$ variables and of degree at most $nR + 1$, such that $\mathcal{G}^{SV}_{(O(R(\log R + \log s)))} + \psi$ is a hitting set generator for $\mathrm{orb}(f)$. We are now ready to prove Theorem 10.

D.3 Proof of Theorem 10

A non-zero polynomial $f \in \mathcal{C}$ is computed by a (Δ, k, s) formula. Then, f is also computed by a $(\Delta, k, (2s)^{\Delta})$ formula in the canonical form of Claim 57. There are two cases:

Case 1: The top most gate of the formula is a $+$ gate. If f is constant, then so is every polynomial in $\mathrm{orb}(f)$. In this case, the set containing any point in $\mathbb{F}^n$ is a hitting set for $\mathrm{orb}(f)$; so we will assume that f is not constant. Then, there exists a x_i such that $\frac{\partial f}{\partial x_i} \neq 0$ (as $\mathrm{char}(\mathbb{F}) > (2ks)^{\Delta^3 R} \geq D$) and as argued in Section D.1, $\frac{\partial f}{\partial x_i}$ can be computed by a $(\Delta, (2k)^{\Delta/2}, (2k)^{\Delta/2}(2s)^{\Delta})$ formula with $+$ gate at the top and top fan-in bounded by k. Moreover, if $\mathcal{G}$ is a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then since $\mathrm{char}(\mathbb{F}) > (2ks)^{\Delta^3 R} \geq (nR+1)D$, $\widetilde{\mathcal{G}} = \mathcal{G} + \mathcal{G}_1^{SV}$ is a hitting set generator for $\mathrm{orb}(f)$. As $\mathrm{char}(\mathbb{F}) = 0$ or $> (2ks)^{\Delta^3 R}$, Lemma 39 works, since the degree of polynomials computed by gates in f can be at most $\left((2k)^{\Delta/2}(2s)^{\Delta}\right)^{\Delta} \leq (2ks)^{\Delta^3}$. Thus, as shown in Section D.2, there exists a $\mathcal{G}$ that has at most

$$O\left(R\left(\log R + \log\left((2k)^{\Delta/2}(2s)^{\Delta}\right)\right)\right) + (\Delta - 2)(R+1) = O\left(R\left(\log R + \Delta \log k + \Delta \log s\right) + \Delta R\right)$$

many variables and of degree $nR+1$. As G_1^{SV} has 2 variables and is of degree n, the number of variables in $\widetilde{\mathcal{G}}$ is $O\left(R\left(\log R + \Delta \log k + \Delta \log s\right) + \Delta R\right)$ and its degree is $nR+1$. Thus, for any $A \in \mathrm{GL}(n, \mathbb{F})$, and $g(\mathbf{x}) := f(A\mathbf{x})$, $g(\widetilde{\mathcal{G}})$ is a polynomial in $O\left(R\left(\log R + \Delta \log k + \log s\right) + \Delta R\right)$ variables and of degree at most $(nR+1)D$. So, a hitting set for g can be constructed in time $(nRD)^{O(R(\log R + \Delta \log k + \Delta \log s) + \Delta R)}$.

Case 2: The top most gate of the formula is a ×人 gate. Then, all inputs to this gate are computed by $(\Delta - 1, k, (2s)^{\Delta})$ formulas in the canonical form of Claim 57 and with a $+$ gates at the top. Hence, all inputs of f are in case 1.

The proof for the case where the leaves are labelled by b-variate polynomials is similar; all we need to do is observe that $\mathcal{G}_{Rb}^{SV}$ is a hitting set generator for b-variate polynomials. So, we can use $\mathcal{G} = \mathcal{G}_{Rb}^{SV} + \psi$.

E Lower bounds for ROABPs and occur-once formulas against their orbits

E.1 A lower bound for ROABPs

In this section, we show that there is a $(3n+2)$-variate, $O(n)$-sparse, degree-$(n+1)$ polynomial f satisfying the following property: there exists a polynomial $g \in \mathrm{orb}(f)$ such that any ROABP computing g has width $2^{\Omega(n)}$. The polynomial g is the obtained by suitably modifying a polynomial constructed in [KNS20], so let us first describe their construction.

Definition 60 (Double cover of a graph)**.** For a graph $G = (V, E)$ on n-vertices, the double cover of G is a bipartite graph $\widetilde{G} = (L \uplus R, \widetilde{E})$, where $|L| = |R| = n$ with the following properties:

1. For every $u \in V$, there is a vertex $u^{(L)} \in L$ and a vertex $u^{(R)} \in R$,
2. For every edge $\{u, v\} \in E$, there are edges $\left\{u^{(L)}, v^{(R)}\right\}$ and $\left\{v^{(L)}, u^{(R)}\right\}$ in $\widetilde{E}$.

Observation 61. *The double cover of a k-regular graph is also k-regular.*

Observation 62. *Let $u, v \in V$. If there is a path of odd length between them, then there is a path between $u^{(L)}$ and $v^{(R)}$ in $\widetilde{G}$. If there is a path of even length between them, then there is a path between $u^{(L)}$ and $v^{(L)}$ in $\widetilde{G}$.*

Proof: Let $u \to u_1 \to \cdots \to u_m \to v$ be a path of odd length between u and v. As the length of the path is odd, m is even. Then, $u^{(L)} \to u_1^{(R)} \to u_2^{(L)} \to \cdots \to u_{m-1}^{(R)} \to u_m^{(L)} \to v^{(R)}$ is a path between $u^{(L)}$ and $v^{(R)}$ in $\widetilde{G}$. The proof of the other case is similar. □

Construction of g [KNS20]. Let $G = (V, E)$ be a 3-regular expander graph with n vertices and let $\widetilde{G} = (L \uplus R, \widetilde{E})$ be its double cover. From Observation 61, $\widetilde{G}$ is also a 3-regular graph. So, it follows from Hall's Marriage Theorem [Hal35] that there exist perfect matchings $M_1, M_2, M_3 \subseteq \widetilde{E}$ such that $\widetilde{E} = M_1 \uplus M_2 \uplus M_3$. Label the edges in M_1 by the variables $\mathbf{x} = (x_1, \ldots, x_n)$, the edges in M_2 by the variables $\mathbf{y} = (y_1, \ldots, y_n)$, and the edges in M_3 by the variables $\mathbf{z} = (z_1, \ldots, z_n)$. With every vertex in $u \in L \uplus R$, associate the affine form $1 + x_i + y_j + z_k$ such that the only edges incident on u in $\widetilde{G}$ are labelled by x_i, y_j and z_k.

Observation 63. *Each x_i, y_i and z_i appears in exactly one of the affine forms associated with the vertices in L and in exactly one of the affine forms associated with the vertices in R.*

Let p_1 be the product of all the affine forms associated with the vertices in L and p_2 the product of all the affine forms associated with the vertices in R; define $p := p_1 + p_2$. The following fact was proved in [KNS20].

Fact 64. *[KNS20] Over any field $\mathbb{F}$, any ROABP computing p has width $2^{\Omega(n)}$.*

Using p we construct g as follows $g := s_1 p + s_2 q$, where s_1, s_2 are variables distinct from $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ and q is a polynomial in $\mathbb{F}[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ which we will define later. Notice that any ROABP computing g must also have width $2^{\Omega(n)}$. This is true, since by setting $s_1 \to 1$ and $s_2 \to 0$ in an ROABP computing g, we get an ROABP computing p.

For a vertex $u \in L \uplus R$, with the affine form associated with it being $1 + x_i + y_j + z_k$, we will say that the linear form[22] associated with it is $x_i + y_j + z_k$. Before constructing f, we prove the following claim.

Claim 65. *Let the linear forms associated with the vertices in L be $\ell_1, \ldots, \ell_n$ and those associated with the vertices in R be $r_1, \ldots, r_n$. Then, $\mathbb{F}$-span$\langle \ell_1, \ldots \ell_n, r_1, \ldots, r_n \rangle$ has dimension $2n - 1$.*

Proof: Assume without loss of generality that, for all $i \in [n]$, ℓ_i and r_i are the linear forms containing x_i. Now, from Observation 63,

$$\sum_{i \in [n]} \ell_i = \sum_{i \in [n]} x_i + \sum_{i \in [n]} y_i + \sum_{i \in [n]} z_i = \sum_{i \in [n]} r_i.$$

[22] A linear form is a linear polynomial whose constant term is 0.

So the vector $\mathbf{1} \in \mathbb{F}^{2n}$ whose first n coordinates are 1 and last n coordinates are -1 is a dependence vector of $\ell_1 \dots \ell_n, r_1, \dots, r_n$. We now show that it is the only dependence vector (up to scaling by any field element). This would immediately imply the claim.

Suppose that $\sum_{i\in[n]} c_i\ell_i = \sum_{i\in[n]} d_i r_i$. Then, since x_i appears only in ℓ_i and r_i, $c_i = d_i$ for all $i \in [n]$. Identify the vertices in L and R by the linear forms associated with them. Observe that if there is an edge between ℓ_i and r_j, then they share a variable. Moreover, they are the only linear forms containing that variable. So, $c_i = d_j = c_j$. Fix an $i \neq 1$. As G is an expander, it is connected. So, from Observation 62, there is either a path between ℓ_1 and r_i or a path between ℓ_1 and ℓ_i. Thus, $c_i = c_1$ for all $i \in [n]$, i.e., $\mathbf{1}$ is the only possible dependence vector. □

The polynomial f.

$$f := s_1\left(\prod_{i\in[n]} x_i + \prod_{i\in[n-1]} y_i\left(\sum_{i\in[n]} x_i + \sum_{i\in[n-1]} -y_i\right)\right) + s_2\left(y_n + \sum_{i\in[n]} z_i\right).$$

Notice that f is a polynomial in $3n+2$ variables, it has degree $n+1$, and has $O(n)$ monomials.

A **and b mapping** *f* **to** *g***.** As $\mathbb{F}$-span$\langle \ell_1, \dots \ell_n, r_1, \dots, r_n\rangle$ has dimension $2n-1$, we can assume without loss of generality that $\{\ell_1, \dots, \ell_n, r_1, \dots, r_{n-1}\}$ is its basis. Also, as the space spanned by the linear forms in $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ variables is a vector space of dimension $3n$, there exist linear forms $t_1, \dots t_{n+1}$ such that $\ell_1, \dots, \ell_n, r_1, \dots, r_{n-1}, t_1, \dots, t_{n+1}$ are linearly independent. Let A be the matrix of the linear transformation that maps

$$\begin{aligned}
&x_i \mapsto \ell_i,\ \forall i \in [n],\\
&y_i \mapsto r_i,\ \forall i \in [n-1],\\
&y_n \mapsto t_{n+1},\\
&z_i \mapsto t_i,\ \forall i \in [n],\\
&s_i \mapsto s_i,\ \text{for}\ \ i = 1, 2.
\end{aligned}$$

As $\ell_1, \dots, \ell_n, r_1, \dots, r_{n-1}, t_1, \dots, t_{n+1}$ are linearly independent, and as s_1 and s_2 are variables distinct from $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, $A \in \mathrm{GL}(3n+2, \mathbb{F})$. Define $\mathbf{b}$ as follows: $b_i = 1$ for all $i \in [2n-1]$ (i.e., for coordinates corresponding to $\mathbf{x}$ and $y_1, \dots, y_{n-1}$) and 0 otherwise.

Let q be the polynomial that is obtained after substituting the variables in $y_n + \sum_{i\in[n]} z_i$ by the corresponding linear forms. Then, it is easy to see that $g = f(A \cdot (\mathbf{x}, \mathbf{y}, \mathbf{z}, s_1, s_2)^T + \mathbf{b}) \in \mathrm{orb}(f)$.

E.2 A lower bound for occur-once formulas

Let $f(\mathbf{x}) = x_1x_2\cdots x_n$; clearly, f can be computed by an occur-once formula of size $O(n)$. Let $\ell_1 = x_1$, $\ell_i(\mathbf{x}) = x_1 + x_i$ for $i \in [2, n]$, and $A \in \mathrm{GL}(n, \mathbb{F})$ such that $A\mathbf{x} = (\ell_1\ \ell_2\ \cdots \ell_n)^T$. Let $g := f(A\mathbf{x}) = x_1(x_1+x_2)(x_1+x_3)\cdots(x_1+x_n)$. We will show that any occur-once formula computing g has size at least 2^{n-1}. The proof is divided into the following two claims.

Claim 66. *g cannot be computed by any occur-once formula of width more than* 1.

Proof: For the sake of contradiction, assume that g can be computed by an occur-once formula of width ≥ 2. Consider such a formula of the smallest possible depth Δ. From Lemma 40, there are three cases:

Case 1: $g = \alpha(g_1 + g_2) + \beta$, where g_1 and g_2 are non-constant, variable disjoint, occur-once formulas and $\alpha \neq 0$. As $x_1 \cdots x_n$ is a monomial of g, $x_1, ..., x_n$ must appear in either g_1 or g_2. But then, the other will have to be a constant – a contradiction.

Case 2: $g = \alpha(g_1 \cdot g_2) + \beta$, where g_1 and g_2 are non-constant, variable disjoint, occur-once formulas and $\alpha \neq 0$. Assume without loss of generality that x_1 appears in g_1 and therefore, does not appear in g_2. Then, as every monomial of g contains x_1, the constant term of g_1 must be zero. This means that the constant term of $\alpha(g_1 \cdot g_2)$ is also 0, which forces β to be 0, as g has no constant term. As $\mathbb{F}[\mathbf{x}]$ is a unique factorization domain, $x_1, (x_1 + x_2), ..., (x_1 + x_n)$ are the only irreducible factors of $g = \alpha(g_1 \cdot g_2)$. But then, x_1 is absent in g_2, and so, g_2 must be a constant – a contradiction.

Case 3: $g = \alpha g_1^e + \beta$, where g_1 is a non-constant occur-once formula having $\text{width}(g_1) = \text{width}(g) \geq 2$ and $\text{depth}(g_1) < \text{depth}(g) = \Delta$, and $\alpha \neq 0$. If h is the highest degree homogeneous part of g_1, then αh^e is the highest degree homogeneous part of $\alpha g_1^e + \beta = g$. Since g is homogeneous and square-free, we must have $e = 1$.

Thus, we have shown that $g = \alpha g_1 + \beta$, where g_1 is a non-constant occur-once formula having $\text{width}(g_1) \geq 2$ and $\text{depth}(g_1) \leq \Delta - 1$. If we apply Lemma 40 on g_1, we once again get three cases, out of which, Case 1 and 2 can be refuted as above. Suppose $g_1 = \alpha_1 g_{1,1}^{e_1} + \beta_1$, where $g_{1,1}$ is a non-constant occur-once formula having $\text{width}(g_{1,1}) \geq 2$ and $\text{depth}(g_{1,1}) < \Delta - 1$. Then, $g = \alpha\alpha_1 g_{1,1}^{e_1} + \alpha\beta_1 + \beta$. Arguing as before, we can show that $e_1 = 1$. The expression $\alpha\alpha_1 g_{1,1} + \alpha\beta_1 + \beta$ can be computed by an occur-once formula of width ≥ 2 and depth $\leq \Delta - 1$, as $\text{depth}(g_{1,1}) < \Delta - 1$. This contradicts the minimality of Δ. □

Claim 67. *If g is computable by an occur-once formula of width 1, then the size of the formula is $\geq 2^{n-1}$.*

Proof: If g is computable by an occur-once formula of width 1, then the formula is of the form

$$\alpha_m \left(\cdots \left(\alpha_2 \left(\alpha_1 p(\mathbf{x})^{e_1} + \beta_1 \right)^{e_2} + \beta_2 \right) \cdots \right)^{e_m} + \beta_m, \tag{7}$$

where $p(\mathbf{x})$ is a depth-2 occur-once formula, $e_1, ..., e_m \in \mathbb{N}$, $\alpha_1, ..., \alpha_m \in \mathbb{F} \setminus \{0\}$ and $\beta_1, ..., \beta_m \in \mathbb{F}$. Let h be the highest degree homogeneous part of $p(\mathbf{x})$. Then, $\alpha h^{e_1 e_2 \cdots e_m}$ is the highest degree homogeneous part of g, for some $\alpha \neq 0$. As g is a homogeneous and square-free polynomial, we must have $e_1 = e_2 = \ldots = e_m = 1$. But then, $g = \alpha p(\mathbf{x}) + \beta$ for some $\alpha \in \mathbb{F} \setminus \{0\}$ and $\beta \in \mathbb{F}$. As $p(\mathbf{x})$ is a depth-2 occur-once formula and g has 2^{n-1} monomials, the size of the formula $p(\mathbf{x})$, and therefore also the size of the formula (7) above, is at least 2^{n-1}. □

F Affine projections and orbit closures

Let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate, degree-d polynomial over $\mathbb{F}$, and $\text{char}(\mathbb{F}) = 0$. The set of *affine projections* of f over a field $\mathbb{F}$ is $\text{aproj}_{\mathbb{F}}(f) := \{f(A\mathbf{x} + \mathbf{b}) : A \in \mathbb{F}^{n \times n} \text{ and } \mathbf{b} \in \mathbb{F}^n\}$; the *orbit* of f over $\mathbb{F}$ is the set $\text{orb}_{\mathbb{F}}(f) := \{f(A\mathbf{x} + \mathbf{b}) : A \in \text{GL}(n, \mathbb{F}) \text{ and } \mathbf{b} \in \mathbb{F}^n\} \subseteq \text{aproj}_{\mathbb{F}}(f)$. Let $m := \binom{n+d}{d}$. By identifying a polynomial in $\text{aproj}_{\mathbb{F}}(f)$ with its coefficient vector in $\mathbb{F}^m$, we will view $\text{aproj}_{\mathbb{F}}(f)$ and $\text{orb}_{\mathbb{F}}(f)$ as subsets of $\mathbb{F}^m$.

Definition 68 (Orbit closure)**.** The orbit closure of f over $\mathbb{F}$, denoted by $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$, is the smallest affine variety in $\mathbb{F}^m$ that contains $\mathrm{orb}_{\mathbb{F}}(f)$.

In other words, $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$ is the Zariski closure of the set $\mathrm{orb}_{\mathbb{F}}(f) \subseteq \mathbb{F}^m$ over $\mathbb{F}$. We give a proof of the following well-known theorem, which implies $\mathrm{orb}_{\mathbb{F}}(f) \subseteq \mathrm{aproj}_{\mathbb{F}}(f) \subseteq \overline{\mathrm{orb}_{\mathbb{F}}(f)} \subseteq \mathbb{F}^m$.

Theorem 69. $\mathrm{aproj}_{\mathbb{F}}(f) \subseteq \overline{\mathrm{orb}_{\mathbb{F}}(f)}$.

Proof: Let $M(n,d) := \{\boldsymbol{\alpha} = (\alpha_1,\ \alpha_2, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0} \ : \ \sum_{i\in[n]} \alpha_i \leq d\}$. Let $Y := (y_{i,j})_{i,j\in[n]}$ be a generic $n \times n$ matrix, and $\mathbf{u} := (u_1\ u_2 \ldots u_n)$ be a generic n-dimensional vector. We will treat $y_{i,j}$ and u_i as formal variables and denote these set of variables as $\mathbf{y} := \{y_{i,j} : i, j \in [n]\} \cup \{u_i : i \in [n]\}$. Consider the polynomial $f(Y\mathbf{x} + \mathbf{u}) \in \mathbb{F}[\mathbf{x}, \mathbf{y}]$. By treating $f(Y\mathbf{x} + \mathbf{u})$ as a polynomial in $\mathbf{x}$ variables with coefficients from $\mathbb{F}[\mathbf{y}]$, we write it as,

$$f(Y\mathbf{x} + \mathbf{u}) = \sum_{\boldsymbol{\alpha}\in M(n,d)} g_{\boldsymbol{\alpha}}(\mathbf{y}) \cdot \mathbf{x}^{\boldsymbol{\alpha}},$$

where $g_{\boldsymbol{\alpha}}(\mathbf{y}) \in \mathbb{F}[\mathbf{y}]$ and $\deg_{\mathbf{y}}(g_{\boldsymbol{\alpha}}) \leq d$. Let $\mathbf{g} := \{g_{\boldsymbol{\alpha}}(\mathbf{y}) : \boldsymbol{\alpha} \in M(n,d)\} \subset \mathbb{F}[\mathbf{y}]$. For simplicity, we denote the elements of $\mathbf{g}$ as $g_1, g_2, \ldots, g_m$. Let $\mathbf{z} := \{z_1, z_2, \ldots, z_m\}$ be a set of m variables. The *annihilating ideal* of $\mathbf{g}$ is the set

$$\text{ann-}\mathbb{I}(\mathbf{g}) := \{h(\mathbf{z}) \in \mathbb{F}[\mathbf{z}] : h(\mathbf{g}) = h(g_1, g_2, \ldots, g_m) = 0\} \subset \mathbb{F}[\mathbf{z}].$$

Observe that $\text{ann-}\mathbb{I}(\mathbf{g})$ is an ideal of $\mathbb{F}[\mathbf{z}]$. The affine variety of this ideal over $\mathbb{F}$ will be denoted as $\mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g})) \subseteq \mathbb{F}^m$.

Observation 70. $\mathrm{aproj}_{\mathbb{F}}(f) \subseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$.

Proof: An element $\mathbf{c} \in \mathrm{aproj}_{\mathbb{F}}(f)$ is the coefficient vector of $f(A\mathbf{x} + \mathbf{b})$ for some $A \in \mathbb{F}^{n\times n}$ and $\mathbf{b} \in \mathbb{F}^n$. The matrix A and the vector $\mathbf{b}$ naturally assign a value $\mathbf{a} \in \mathbb{F}^{n^2+n}$ to the $\mathbf{y}$ variables so that

$$f(A\mathbf{x} + \mathbf{b}) = \sum_{\boldsymbol{\alpha}\in M(n,d)} g_{\boldsymbol{\alpha}}(\mathbf{a}) \cdot \mathbf{x}^{\boldsymbol{\alpha}}.$$

Notice that $\mathbf{g}(\mathbf{a}) := (g_1(\mathbf{a}),\ g_2(\mathbf{a}),\ \ldots, g_m(\mathbf{a}))$ is the coefficient vector $\mathbf{c}$ of $f(A\mathbf{x} + \mathbf{b})$. As $h(\mathbf{g}) = 0$ for every $h \in \text{ann-}\mathbb{I}(\mathbf{g})$, we have $h(\mathbf{g}(\mathbf{a})) = 0$ for every $h \in \text{ann-}\mathbb{I}(\mathbf{g})$. Hence, $\mathbf{c} \in \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$. □

Claim 71. $\overline{\mathrm{orb}_{\mathbb{F}}(f)} = \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$.

Proof: From Observation 70, $\mathrm{orb}_{\mathbb{F}}(f) \subseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$, as $\mathrm{orb}_{\mathbb{F}}(f) \subseteq \mathrm{aproj}_{\mathbb{F}}(f)$. Since $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$ is the smallest variety in $\mathbb{F}^m$ containing $\mathrm{orb}_{\mathbb{F}}(f)$, and intersection of two varieties is again a variety, we have $\overline{\mathrm{orb}_{\mathbb{F}}(f)} \subseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$.

To show the other direction, i.e., $\overline{\mathrm{orb}_{\mathbb{F}}(f)} \supseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$, it is sufficient to show that the ideal of $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$ (denoted as $\mathbb{I}(\overline{\mathrm{orb}_{\mathbb{F}}(f)})$) is contained in $\text{ann-}\mathbb{I}(\mathbf{g})$. This is because, $\mathbb{V}(\mathbb{I}(\overline{\mathrm{orb}_{\mathbb{F}}(f)})) = \overline{\mathrm{orb}_{\mathbb{F}}(f)}$, as $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$ is a variety. Let $p(\mathbf{z}) \in \mathbb{I}(\overline{\mathrm{orb}_{\mathbb{F}}(f)})$ and $\deg_{\mathbf{z}}(p) = D$. Then, $p(\mathbf{c}) = 0$ for all $\mathbf{c} \in \mathrm{orb}_{\mathbb{F}}(f)$. Consider the polynomial $p(\mathbf{g}) = p(g_1, g_2, \ldots, g_m) \in \mathbb{F}[\mathbf{y}]$. If $p(\mathbf{g}) = 0$, then $p \in \text{ann-}\mathbb{I}(\mathbf{g})$ and we are done. So, suppose $p(\mathbf{g}) \neq 0$. Note that $\deg_{\mathbf{y}}(p(\mathbf{g})) \leq Dd$, as $\deg_{\mathbf{y}}(g_i) \leq d$.

Pick a set $S \subset \mathbb{F}$ of size $|S| = n + Dd + 1$ (such an S exists as $\mathrm{char}(\mathbb{F}) = 0$). By the Schwartz-Zippel lemma,

$$\Pr_{\mathbf{a} \in_r S^{n^2+n}} \{p(\mathbf{g}(\mathbf{a})) = 0\} \leq \frac{Dd}{|S|}.$$

On the other hand,

$$\Pr_{\mathbf{a} \in_r S^{n^2+n}} \{\mathbf{g}(\mathbf{a}) \in \mathrm{orb}_{\mathbb{F}}(f)\} \geq 1 - \frac{n}{|S|},$$

as a random $A \in S^{n \times n}$ is invertible with probability at least $1 - \frac{n}{|S|}$ (from the Schwartz-Zippel lemma again). Since $p(\mathbf{c}) = 0$ for all $\mathbf{c} \in \mathrm{orb}_{\mathbb{F}}(f)$,

$$\Pr_{\mathbf{a} \in_r S^{n^2+n}} \{\mathbf{g}(\mathbf{a}) \in \mathrm{orb}_{\mathbb{F}}(f)\} \leq \Pr_{\mathbf{a} \in_r S^{n^2+n}} \{p(\mathbf{g}(\mathbf{a})) = 0\}.$$

Hence, $1 - \frac{n}{|S|} \leq \frac{Dd}{|S|}$, implying $|S| \leq n + Dd$. But, this is a contradiction as $|S| = n + Dd + 1$. Therefore, $p(\mathbf{g}) = 0$. □

The proof of the theorem now follows from Observation 70 and the above claim. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il