Electronic Colloquium on Computational Complexity, 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 $\text{orb}(f) := \{f(A\mathbf{x} + \mathbf{b}) : A \in \text{GL}(n, \mathbb{F}) \text{ and } \mathbf{b} \in \mathbb{F}^n\}$. This paper studies explicit hitting sets for the *orbits* of polynomials computable by certain well-studied circuit classes. This version of the hitting set problem is interesting as $\text{orb}(f)$ is a natural 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 $\mathsf{IMM}_{3,\text{poly}(s)}$. To our knowledge, no efficient hitting set construction was known for even $\text{orb}(\mathsf{IMM}_{3,d})$ before this work.

In this work, we give efficient constructions of hitting sets for the orbits of several interesting 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 ROABP*. This implies quasi-polynomial time hitting sets for the orbits of *multilinear sparse polynomials* and the orbits of the *elementary symmetric polynomials*.
2. Multilinear polynomials computable by *constant-width ROABP*. This implies a quasi-polynomial time hitting set for the orbit of $\mathsf{IMM}_{3,d}$.
3. Polynomials computable by *constant-depth, constant-occur formulas* with low-individual-degree sparse polynomials at the leaves. This implies quasi-polynomial time hitting sets for the orbits of *multilinear depth-4 circuits with constant top fan-in*, and also poly-time hitting sets for the orbits of the *power symmetric polynomials* and the *sum-product polynomials*.
4. Polynomials computable by *occur-once formulas* with low-individual-degree sparse polynomials at the leaves.

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

The first two results are obtained by building upon the rank concentration by translation technique of [ASS13]; the second result also 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 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 case of PIT over a ring [AB03]. Also, multivariate polynomial factorization 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 known for constant depth 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 known as the problem of constructing *hitting sets* for $\mathcal{C}$.

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* (ROABP) 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 polynomials can be computed by linear size ROABP. A polynomial-time PIT algorithm and a quasi-polynomial time hitting set construction for ROABP are known [RS05, FS13b, AGKS15]. Hitting sets for ROABP, 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-commutative algebraic branching programs [FS13b]. The constant-occur formula model is also reasonably natural;

[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] In this case, the input circuit computes a polynomial of degree $\text{poly}(n)$, 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.

it captures other interesting classes like multilinear depth-four circuits with bounded top fan-in [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] gave polynomial-time constructible hitting sets for constant-depth, constant-occur formulas.

Hitting sets for orbits. In this paper, we study hitting sets for *orbits* of ROABP and constant-occur formulas. 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. Orbit of a circuit class is the union of the orbits of the polynomials computable by circuits in the class. The reasons for studying hitting sets for orbits of ROABP 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 $\text{aproj}(f) := \{f(A\mathbf{x}+\mathbf{b}) \ : \ A \in \mathbb{F}^{n\times n} \text{ and } \mathbf{b} \in \mathbb{F}^n\}$; the orbit of f is the set $\text{orb}(f) = \{f(A\mathbf{x}+\mathbf{b}) \ : \ A \in \text{GL}(n,\mathbb{F}) \text{ and } \mathbf{b} \in \mathbb{F}^n\} \subseteq \text{aproj}(f)$.[4] Affine projections of polynomials computable by poly-size ROABP 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 $\text{aproj}(\mathsf{IMM}_{s,s})$. In fact, every polynomial computable by a size-s arithmetic formula is in $\text{aproj}(\mathsf{IMM}_{3,\text{poly}(s)})$ [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 $\text{aproj}(\mathsf{SP}_{s,d})$. As demonstrated by the depth reduction results in [GKKS16, Tav15, Koi12, AV08, VSBR83], depth-3 circuits are incredibly powerful. Orbit of f being an interesting subset of $\text{aproj}(f)$, it is thus natural to ask if we can give efficient hitting set constructions for orbits of the above-mentioned polynomial families. Moreover, $\text{orb}(f)$ is not 'much smaller' than $\text{aproj}(f)$, as the latter is contained in the *orbit closure* of f if $\text{char}(\mathbb{F}) = 0$ (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 $\overline{\text{orb}(f)}$) is defined as the Zariski closure of $\text{orb}(f)$. Polynomials in $\overline{\text{orb}(f)}$, and hence also $\text{aproj}(f)$, can be approximated infinitesimally closely by polynomials in $\text{orb}(f)$ over $\mathbb{C}$.[6]

2. *Geometry of the circuit classes:* Consider an n-variate polynomial $f \in \mathbb{R}[\mathbf{x}]$ that is computable by a poly-size ROABP or constant-occur formula, and 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 transformation[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

[4] Ideally, we should use the notations $\text{aproj}_{\mathbb{F}}$ and $\text{orb}_{\mathbb{F}}$, but we are dropping the subscripts here for simplicity, and as 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, $\overline{\text{orb}(f)}$ can be strictly larger than $\text{aproj}(f)$.

[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 $T(\mathbf{x}) = R\mathbf{x}+\mathbf{b}$.

slightly by replacing $R \in O(n,\mathbb{R})$ with $A \in \mathrm{GL}(n,\mathbb{R})$.[8] A hitting set for ROABP or constant-occur formulas does not imply 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 ROABP and constant-occur formulas are tied to the choice of the coordinate system. In fact, we show in Appendix D that there is an explicit polynomial g in the orbit of a polynomial computable by a poly-size ROABP such that any ROABP computing g has exponential size. Thus, it is 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:* As mentioned above, hitting sets for ROABP and constant-occur formulas do not automatically give hitting sets for their orbits. But, can the techniques used to design these hitting sets 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 orbits of the above-mentioned circuit classes investigates a line of research that, to our knowledge, has remained largely unexplored. 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 (ABP) were defined by Nisan in [Nis91]. As the name suggests, read-once oblivious algebraic branching programs (ROABP) are read-once variant of ABP. While Nisan defined ABP using directed graphs, in this work we use the following conventional definition of 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.

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

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

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 $\mathbb{F}$. *Orbit* of f, denoted by $\text{orb}(f)$, is the set $\{f(A\mathbf{x}) : A \in \text{GL}(n, \mathbb{F})\}$. 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 here continue to hold even if we define orbit of an n-variate polynomial f as $\text{orb}(f) = \{f(A\mathbf{y} + \mathbf{b}) \ : \ |\mathbf{y}| = m \geq n,\ A \in \mathbb{F}^{n\times m} \text{ has rank } n, \text{ and } \mathbf{b} \in \mathbb{F}^n\}$. However, we work with the conventional definition of $\text{orb}(f)$ for simplicity of exposition, and because the proofs in this 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 circuits in the class $\mathcal{C}$. Our main results are efficient constructions of hitting sets (Definition 5) for the orbits of the models mentioned above, namely commutative ROABP, constant-width ROABP, constant-depth constant-occur formulas, and occur-once formulas.

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,

[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.

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.

Theorem 6 (Hitting sets for orbits of commutative ROABP 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 ROABP. 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.*

Remarks.

1. As every s-sparse polynomial can be computed by a commutative ROABP of width s, Theorem 7 follows as a corollary from the above theorem.

2. The elementary symmetric polynomial $\mathsf{ESym}_{n,D} = \sum_{S \in \binom{[n]}{D}} \prod_{i \in S} x_i$ can be computed by a commutative ROABP of width $n+1$. This is due to an interpolation trick (attributed to Ben-Or in [NW97, Shp02]) that gives a formula for $\mathsf{ESym}_{n,D}$ which is a sum of $n+1$ products of univariate affine forms. Such a formula can be expressed as a multilinear commutative ROABP of width $n+1$. So, the theorem implies an $n^{O(\log n)}$-time hitting set for $\mathrm{orb}(\mathsf{ESym}_{n,D})$.

3. The theorem also implies a quasi-polynomial time hitting set for the orbits of *sums of products of low degree univariates*. The sums of products of univariates model has found interesting applications in several other works [Sax08, SSS13, GKKS16].

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

Remarks.

1. It is well known that hitting sets for sparse polynomials can be constructed in polynomial time [KS01, LV03]. The above result gives a quasi-polynomial time hitting set construction for the orbits of *multilinear* sparse polynomials.

2. The algorithm in [KS01] is based on an efficient mechanism to generate *monomial isolating weight assignments* for sparse polynomials. A weight vector $(w_1, \ldots, w_n) \in \mathbb{N}^n$ is monomial isolating for an n-variate, s-sparse polynomial f if the s monomials of f map to different univariate monomials under the substitution $x_i \mapsto x^{w_i}$. The complexity of computing such weight vectors in [KS01] depends polynomially on s. As polynomials in the orbit of even a monomial can have exponential sparsity, it is unclear if the monomial isolation technique can be applied directly to design hitting sets for orbits of sparse polynomials. As stated before, we extend the rank concentration technique of [ASS13] to design such hitting sets, albeit the running time of our construction depends exponentially on the individual degree.

3. Theorem 7 plays a crucial role in the proofs of Theorem 9 and Theorem 10 which apply the algebraic independence based analysis from [ASSS16, BMS13] and the Shpilka-Volkovich generator based argument from [SV15] to reduce to the case of constructing hitting sets for the orbits of sparse polynomials.

Theorem 8 (Hitting sets for orbits of multilinear constant-width ROABP)**.** *Let $\mathcal{C}$ be the set of n-variate multilinear polynomials that are computable by width-w ROABP. Then a hitting set for* $\mathrm{orb}(\mathcal{C})$ *can be computed in $n^{O(w^6 \cdot \log n)}$ time provided that $|\mathbb{F}| > n^{O(w^4)}$.*

Remarks.

1. The theorem gives a quasi-polynomial time hitting set for $\overline{\mathrm{orb}(\mathsf{IMM}_{3,d})}$, 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. Affine projections of $\mathsf{IMM}_{2,d}$ is also quite interesting, 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 orbit closures of arithmetic formulas [BIZ18]. The above theorem implies a quasi-polynomial time hitting set for $\overline{\mathrm{orb}(\mathsf{IMM}_{2,d})}$.

Theorem 9 (Hitting sets for orbits of constant-depth, constant-occur formulas with low individual degree)**.** *Let $\mathcal{C}$ be the set of n-variate, degree-D polynomials that are computable by depth-Δ occur-k formulas whose leaves are labelled by s-sparse polynomials with individual degree at most d. Let $R := (2k)^{2\Delta\cdot 2^{\Delta}}$. If $|\mathbb{F}| > (nR+1)D$ and $\mathrm{char}(\mathbb{F}) = 0$ or $> D^R$, then a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in $(nRD)^{O\left(R^2 d(\log R+\Delta\log k+\Delta\log s)+\Delta R\right)}$ 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 $(nD)^{O(d\log s)}$ 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. As mentioned in Section 1.4, there is a polynomial-time constructible hitting set for multilinear depth-4 circuits with constant top fan-in [SV18, KMSV13]. Theorem 9 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. The proof of the above theorem also implies a quasi-polynomial time hitting set for the orbit of this model, when $f_1,\ldots,f_m$ have low individual degree (in particular, when $f_1,\ldots,f_m$ are multilinear).

2. The theorem also 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.

Theorem 10 (Hitting sets for orbits of occur-once formulas with low individual degree)**.** *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 with individual 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 $(nD)^{O(\log n+d\log s)}$ 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.*

A polynomial time construction of hitting sets for read-once formulas is known [SV15, MV18]. Also, a quasi-polynomial time construction of hitting sets for occur-once formulas (without powering gates) follows from [AvMV15]. But we show in Appendix E 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}.

So, the results in [SV15, AvMV15, MV18] do not directly imply efficient hitting sets for the orbits of occur-once formulas. Nonetheless, we are able to apply the arguments in [SV15] to prove the above theorem.

1.3 Proof techniques

In this section, we briefly discuss the techniques we use to prove the above results.

Commutative ROABP with low individual degree. Theorem 6 is proved by adapting the rank concentration by translation technique of [ASS13][11] to work for orbits of commutative ROABP. 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, when viewed as a polynomial in $\mathbb{F}[\mathbf{x}, \mathbf{z}]$, has a low support monomial. Then, we use the SV generator to construct a hitting set generator for 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 us construct a substitution $t_i \mapsto t_i(\mathbf{z})$ that preserves the nonzeroness of these polynomials.

Multilinear constant-width ROABP. Theorem 8 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 ROABP, 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 ROABP, the $\mathbf{x}$-variables were translated only once, here the translations can be thought of as happening sequentially and

[11] [ASS13] proved their result for products of univariate polynomials over a Hadamard algebra which is a certain kind of commutative ROABP. However, their analysis also works for general commutative ROABP.

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

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 $\mathbf{y}$-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 ROABP, we show that $G(\mathbf{x} + \mathbf{t})$ has low support rank concentration if each polynomial in a certain collection of non-zero polynomials is kept non-zero by the substitution $t_i \mapsto t_i(\mathbf{z})$. However, in the 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 some linear forms in the $\mathbf{t}$-variables.

Constant-depth, constant-occur formulas. We prove Theorem 9 by combining Theorem 7 with the algebraic independence based technique in [ASSS16]. 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 16) 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 7 to construct such a ψ.

Occur-once formulas. We prove Theorem 10 by building upon the arguments in [SV15] and linking it with Theorem 7. At first, we show two structural results (Lemma 38 and 39) 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 12). However, in our case, we want a generator for $\text{orb}(f)$ and not just f. For this reason, we first use the chain rule for derivatives to relate the gradient of a $g \in \text{orb}(f)$ with that of f, and then argue that there exists a x_j such that a generator for $\text{orb}\left(\frac{\partial f}{\partial x_i}\right)$ 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 7.

1.4 Related work

We give a brief account of the known results on PIT and hitting sets for arithmetic circuits. The results on hitting sets for constant-read models are most relevant to our work here. However, for the sake of completeness, we will 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; a PROF is a special case of an occur-once formula. Later, a polynomial-time hitting set construction for the same model was given by [MV18]. Notice that 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), but the arguments in [AvMV15] need the multilinearity assumption.

A polynomial-time PIT for ROABP follows from the PIT algorithm for non-commutative formulas in [RS05]. [FS13b] gave a quasi-polynomial time construction of hitting sets for ROABP, 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 ROABP 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 ROABP (more generally, low individual degree ROABP). Finally, [AGKS15] obtained a quasi-polynomial time constructible hitting set for ROABP 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 to design hitting sets for sums of constantly many ROABPs in quasi-polynomial time [GKST17]; they also gave a polynomial-time PIT algorithm for the same model. A conjunction of the basis isolation and the rank concentration techniques have also been used to give more efficient constructions of hitting sets for ROABP [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 ROABP, 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]. This is the only result on hitting sets for orbits of natural families of polynomials that we are aware of. Observe that PSym is computable by a constant-depth occur-once formula with univariate polynomials at the leaves. So, Theorem 9 subsumes this result. Our hitting-set construction is different from the one in [KS19] which involves second order derivatives (in particular, the Hessian), whereas the proofs here work with first order derivatives. For 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] gave a polynomial-time PIT for sums 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 (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 ROABP. [dOSlV16] gave subexponential-time hitting sets for multilinear depth-3 and depth-4 formulas (more generally, constant-depth multilinear regular formulas) by reducing the problem to constructing hitting sets for ROABP. 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 $i \in [n]$ and $A_0 \in \mathbb{F}^{n\times n}$ is an arbitrary ma-

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

trix, check if $f \equiv 0$ [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 the 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 orbits of polynomials computable by this model imply hitting sets for the orbit of the determinant polynomial and also orbit of the iterated matrix multiplication polynomial via a reduction from ABP to symbolic determinant [Val79].

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

2 Preliminaries

Definition 11 (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_0D. 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 $|\mathbb{F}| > d_0D$.

2.1 The Shpilka-Volkovich generator

Definition 12 (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 in 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 13. *Let $f \in \mathbb{F}[\mathbf{x}]$ be a polynomial that depends on only b of the $\mathbf{x}$ variables, and $g \in \mathrm{orb}(f)$. Then, $g \neq 0$ implies $g(\mathcal{G}_b^{SV}) \neq 0$.*

Proof: Let $g \in \mathrm{orb}(f)$ be non-zero. As f depends on only b variables, there are b variables (say, $x_1, x_2, \ldots, x_b$) such that $g(x_1, x_2, \ldots, x_b, 0, \ldots, 0) \neq 0$. Now observe that $\mathcal{G}_b^{SV}{}_{|(y_1=\alpha_1, y_2=\alpha_2, \ldots, y_b=\alpha_b)} = (z_1, z_2, \ldots, z_b, 0, \ldots, 0)$. Hence, $g(\mathcal{G}_b^{SV}) \neq 0$. □

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.

Observation 14. *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 $\mathrm{Supp}\left(\mathbf{x}^{\boldsymbol{\alpha}}\right) \leq m$, 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 $\left\{C_{\boldsymbol{\beta}} : \mathrm{Supp}\left(\mathbf{x}^{\boldsymbol{\beta}}\right) \leq m\right\}$. Thus, there is a $\boldsymbol{\beta}$ with $\mathrm{Supp}\left(\mathbf{x}^{\boldsymbol{\beta}}\right) \leq m$, for which $\mathbf{1}^T \cdot C_{\boldsymbol{\beta}} \cdot \mathbf{1}$ is non-zero, as $\mathbf{1}^T \cdot C_{\boldsymbol{\alpha}} \cdot \mathbf{1}$ is non-zero. □

2.3 Algebraic rank and faithful homomorphisms

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 (i.e., the algebraic rank) of $\mathbf{f}$ over $\mathbb{F}$ – denoted by $\text{tr-deg}_{\mathbb{F}}(\mathbf{f})$ – to the rank of the Jacobian.

Lemma 15 (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 $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > D^r$, then $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = \mathrm{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f})$.*

Definition 16 (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 17 (Theorem 2.4 in [ASSS16]). *If the 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 18 (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*, 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 19. *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 $\mathbf{g}$. Let $p \in \mathbb{F}[y_1, \ldots, y_r]$ be such that $p(g_1, \ldots, g_r) = 0$. Now, $p(g_1, \ldots, g_r) = p(f_1, \ldots, f_r)(A\mathbf{x})$. As A is invertible this means that $p(f_1, \ldots, f_m) = 0$. 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 $r + 1 \leq j \leq m$ 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 orbits of commutative ROABP

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 $A \in \text{GL}(n, \mathbb{F})$, 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 $G = M_1(y_1) M_2(y_2) \cdots M_n(y_n)$. 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 14, 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 obtain a hitting set generator for g. More precisely, we will prove the following theorem in this section (at the end of Section 3.2).

Theorem 20. *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* $\text{orb}(f)$*.*

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 $M_i(y_i) = \sum_{e_i} u_{i,e_i} y_i^{e_i}$, where $u_{i,e_i} \in \mathbb{A}$ and $M_i(y_i + r_i) = \sum_{e_i} v_{i,b_i} y_i^{b_i}$, 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 that we prove in Appendix A implies that v_{i,e_i} and v_{j,e_j} also commute for $i \neq j$.

Observation 21. *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. $u_{i,e_i} = \sum_{b_i=0}^{d} \cdot \binom{b_i}{e_i} (-r_i)^{b_i - e_i} \cdot v_{i,b_i}$,

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 $\deg_{\mathbf{r}_S}(h_S(\mathbf{r}_S)) \leq m d^{m+1}$, remain non-zero under the substitution $t_i \mapsto t_i(\mathbf{z})$.[14] The following lemma will help us achieve this goal.

Lemma 22. *Let $G, \mathbf{t}, \mathbf{z}, \mathbf{y}$ and $\mathbf{r}_S$ be as defined above. Suppose that the following two conditions are satisfied:*

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 $\mathbf{y}$-variables over $\mathbb{F}(\mathbf{z})$.

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

Proof: Consider a $\mathbf{b} = (b_i : i \in [n]) \in \{0, \ldots, d\}^n$ with $\mathrm{Supp}(\mathbf{b}) \geq m$. Pick a $S \subseteq \binom{[n]}{m}$ such that $\mathrm{Supp}\,(b_i : i \in S) = m$. 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 $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$, $G(x_1 + t_1(\mathbf{z}), \ldots, x_n + t_n(\mathbf{z}))$ has support-$(m-1)$ rank concentration in the $\mathbf{y}$-variables over $\mathbb{F}(\mathbf{z})$. $\square$

3.2 Achieving rank concentration

We will now see how to satisfy conditions 1 and 2 of Lemma 22 such that $\deg_{\mathbf{r}_S}(h_S(\mathbf{r}_S)) \leq md^{m+1}$, $t_i(\mathbf{z})$ 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 23. *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

[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.

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 24. *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 $(d+1)^m \times d^m$ 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 $\mathbf{n_b}$. 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 22.

Corollary 25. *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 $\mathbb{F}[\mathbf{t}]\text{-span}\left\{v_{\mathbf{b}'} : \mathbf{b}' \in \{0,\ldots,d\}^m \text{ and } \mathrm{Supp}(\mathbf{b}') < m\right\}$. □

The following claim about $h(\mathbf{r})$ gives us a way to satisfy condition 2 of Lemma 22.

Claim 26. *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: $h(\mathbf{r}) = h(\ell_1(\mathbf{t}),\ldots,\ell_m(\mathbf{t})) \neq 0$ as $[CMDN]_F$ is an invertible matrix and $\ell_1,\ldots,\ell_m$ are $\mathbb{F}$-linearly independent. Let B be the $m \times n$ matrix whose i-th row is the coefficient vector of ℓ_i. As $\ell_1,\ldots,\ell_m$ are linearly independent, $\mathrm{rank}(B) = m$. Thus, there are m columns $j_1,\ldots,j_m$ of B that are also linearly independent. This means the linear forms $\ell'_1(\mathbf{t}),\ldots,\ell'_m(\mathbf{t})$ obtained from $\ell_1(\mathbf{t}),\ldots,\ell_m(\mathbf{t})$ after setting variables other than $t_{j_1},\ldots,t_{j_m}$ to 0 are also linearly independent. Thus, $h(\ell'_1(\mathbf{t}),\ldots,\ell'_m(\mathbf{t})) \neq 0$ which is only possible if $h(\mathbf{r})$ has a $\mathbf{t}$-monomial of support at most m. □

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. The SV generator requires $|\mathbb{F}| \geq n$. We are now in a position to prove Theorem 20.

Proof of Theorem 20

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 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 14, 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 $2\lceil\log w^2\rceil+1$ proves the theorem. Note that the SV generator needs $|\mathbb{F}| \geq n$.

3.3 Proof of Theorem 6

Let f be a 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 20, $g\left(\mathcal{G}_{(2\lceil\log w^2\rceil(d+1)+1)}^{SV}\right) \neq 0$ whenever $g \neq 0$. Now, $\mathcal{G}_{(2\lceil\log w^2\rceil(d+1)+1)}^{SV}$ has $2(2\lceil\log w^2\rceil(d+1)+1)$ variables, and is of degree n. So $g\left(\mathcal{G}_{(2\lceil\log w^2\rceil(d+1)+1)}^{SV}\right)$ also has $2\left(2\lceil\log w^2\rceil(d+1)+1\right)$ variables. Since the individual degree of f is at most d, the $\deg(f) = \deg(g) = nd$. 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 $\left(n^2d+1\right)^{\left(2\lceil\log w^2\rceil(d+1)+1\right)} = (nd)^{O(d\log w)}$.

3.4 Hitting set generator for 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 with individual degree at most d, where $c_j \in \mathbb{F}$ for $j \in [s]$. Observe that f can be computed by a commutative ROABP as follows: Let $M_1(x_1) := \mathrm{diag}(c_1x_1^{e_{1,1}},\ldots c_sx_1^{e_{s,1}})$ and, for $2 \leq i \leq n$, let $M_i(x_i) := \mathrm{diag}(x_i^{e_{1,i}},\ldots x_i^{e_{s,i}})$. Then, $f = \mathbf{1}^T\cdot M_1(x_1)\cdots M_n(x_n)\cdot\mathbf{1}$; notice that the width of the ROABP is s. The following theorem follows as a corollary of Theorem 20.

Theorem 27. *Let f be an n-variate, s-sparse polynomial with individual degree at most d, and $g \in \mathrm{orb}(f)$. Also, let $|\mathbb{F}| \geq n$. Then, $g \neq 0$ implies $g\left(\mathcal{G}_{(2\lceil\log s\rceil(d+1)+1)}^{SV}\right) \neq 0$. In fact, if g is not a constant, then neither is $g\left(\mathcal{G}_{(2\lceil\log s\rceil(d+1)+1)}^{SV}\right)$.*

[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: From Theorem 20, it is clear that $g \neq 0$ implies $g\left(\mathcal{G}^{SV}_{(2\lceil \log s^2 \rceil (d+1)+1)}\right) \neq 0$. As $M_1(x_1), \ldots, M_n(x_n)$ are diagonal matrices, the space spanned by matrices u_{i,e_i} defined in the paragraph after the statement of Theorem 20 on page 13 is of dimension s. Then, a close examination of the proof of Lemma 24 shows that the parameter m can be made $2\lceil \log s \rceil + 1$ instead of $2\lceil \log w^2 \rceil + 1 = 2\lceil \log s^2 \rceil + 1$ in this case. This implies that $(2\lceil \log s \rceil (d+1) + 1)$ is a hitting set generator for g and so, $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1)}\right) \neq 0$. Now, suppose, for the sake of contradiction, that g is not a constant, but $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1)}\right)$ is a constant, say β. Then the constant terms of both f and g are β. Consider $h = f - \beta$; then h is also an n-variate, s-sparse polynomial with individual degree at most d and $q := g - \beta \in \mathrm{orb}(h)$. As, g is not a constant, $q \neq 0$. However, $q\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1)}\right) = g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1)}\right) - \beta = 0$, a contradiction. Thus, if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1)}\right)$. $\square$

4 Hitting sets for orbits of multilinear constant-width ROABP

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 $A \in \mathrm{GL}(n, \mathbb{F})$, 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 $G = M_1(y_1)M_2(y_2)\cdots M_n(y_n)$. 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 $\mathbf{t}$-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 $r_{p,q,i} := \ell_i(\mathbf{t}_{p,q})$; 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 section and the next section, we will prove the following lemma. Let $\mathbb{A} := \mathbb{F}^{w\times w}$.

Lemma 28. *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, 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 13. For any $w \geq 2$, $2 \leq 2\mu$. As a product of at most 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, 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 over $\mathbb{F}\left(s_{p,q}, z_{p,q} : (p,q) \preccurlyeq (p^*,q^*)\right)$. If $q^* < \mu$, then this means 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 observation implies that the induction hypothesis holds immediately before translation by $\mathbf{t}_{p^*+1,q^*}$.

Observation 29. *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, good $\{\beta(i) : i \in [m]\}$ such that after substituting $t_i \to s \cdot z^{\beta(i)}$, this 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 notation 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 $i \in [m]$, $M_i(y_i) = \sum_{e_i} u_{i,e_i} y_i^{e_i}$, where $u_{i,e_i} \in \mathbb{A}$ and $M_i(y_i + r_i) = \sum_{b_i} v_{i,b_i} y_i^{b_i}$, 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 $\mathbb{A}^{2^m}$. $C := \mathrm{diag}(\mathbf{r}^\mathbf{b} : \mathbf{b} \in \{0,1\}^m)$ and $D := \mathrm{diag}(\mathbf{r}^{-\mathbf{e}} : \mathbf{e} \in \{0,1\}^m)$; both C and D 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 23.

Claim 30. *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 31. *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 some 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 32. *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 25. □

Corollary 33. *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 34. *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 $n^{O\left(w^4\right)}$ and each $\beta(i) \leq n^{O\left(w^4\right)}$.*

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 $\binom{n+2w^2\mu}{2w^2\mu} = n^{O\left(w^4\right)}$ 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 28 stated in Section 4.1.

Proof of Lemma 28. So far we have proved that there exist $\left\{\beta_{p,q}(i) : p \in [\lceil \log n \rceil], q \in [\mu], i \in [n]\right\}$, such that

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

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 8

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 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 28, 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 $\mathbb{F}\left(\left\{s_{p,q}, z_{p,q}\right\}_{p,q}\right)$. Moreover, these polynomials can be computed in time $n^{O\left(w^4\right)}$. Suppose that $g \neq 0$. Then, from Observation 14, $g(x_1 + t_1, \ldots, x_n + t_n)$ has a support-μ, $\mathbf{y}$-monomial when viewed as a polynomial over $\mathbb{F}\left[\left\{s_{p,q}, z_{p,q}\right\}_{p,q}\right]$ in the $\mathbf{y}$-variables. Since f is multilinear, as seen in the proof of Theorem 20, $g(x_1 + t_1, \ldots, x_n + t_n)$ 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 $\mu + 2\mu \cdot \lceil \log n \rceil$ 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 orbits of depth four, constant-occur formulas

In this section, we will show the existence of quasi-polynomial time hitting sets for orbits of depth-4, occur-k formulas whose leaves are labelled by low individual degree sparse polynomials. 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 9 can be generalised to work for arbitrary depths. The general argument can be found in Appendix C.

For some $k \in \mathbb{N}$, let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate, degree-D polynomial computed by a $(4, k, s, d)$ formula, i.e., a depth-4, occur-k formula of size-s whose leaves are labelled by sparse polynomials of individual degree at most d. 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 35 below). We then argue in Claim 36 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 35. *Let f be a $(4, k, s, d)$ formula. Then, for every $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a $(4, 2k^2, 2ks, d)$ 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 $r \leq k$. 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 $q_{i,1}, \ldots q_{i,r_i}$, $\sum_{i \in [r]} r_i \leq k$. Here, $q_{i,j}$ are s-sparse polynomials with individual degree at most d. 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 $\sum_{i \in [r]} r_i \leq k$, its depth is 4, and as the leaves are still $q_{i,j}$ or $\frac{\partial q_{i,j}}{\partial x}$, the individual degrees of the polynomials labelling the leaves are also at most d. 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

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

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

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

□

We now show 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 $\mathrm{orb}(\frac{\partial f}{\partial x_i})$.

Claim 36. *Let $f \in \mathbb{F}[\mathbf{x}]$ be an n-variate polynomial of degree D, and $\mathrm{char}(\mathbb{F}) = 0$ or $> D$. There is an $i \in [n]$ such that if $\mathcal{G}$ is a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then $\widetilde{\mathcal{G}} := \mathcal{G} + \mathcal{G}_1^{SV}$ is a hitting set generator for $\mathrm{orb}(f)$, 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 $\mathrm{orb}(f)$ is trivial. Otherwise, there exists an $i \in [n]$ such that $\frac{\partial f}{\partial x_i} \neq 0$ (because $\mathrm{char}(\mathbb{F}) = 0$ or $> D$). Suppose that a polynomial map $\mathcal{G}$ 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, $\frac{\partial f}{\partial x_i}(A\mathcal{G}) \neq 0 \implies [\nabla f](A\mathcal{G}) \neq 0 \implies [\nabla g](\mathcal{G}) \neq 0 \implies \exists j \in [n]$ such that $\frac{\partial g}{\partial x_j}(\mathcal{G}) \neq 0$. This means that there is a $(\beta_1, ..., \beta_n) \in \mathrm{Img}(\mathcal{G})$ such that

$$\frac{\partial g}{\partial x_j}(\beta_1, ..., \beta_n) \neq 0,$$

because $\deg\left(\frac{\partial g}{\partial x_j}(\mathcal{G})\right) \leq \deg(\mathcal{G}) \cdot D$ and $|\mathbb{F}| > \deg(\mathcal{G}) \cdot D$. Let $r(z_1) := g(\beta_1, ..., \beta_{j-1}, \beta_j + z_1, \beta_{j+1}, ..., \beta_n)$. Then,

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

and so, $g(\beta_1, ..., \beta_{j-1}, \beta_j + z_1, \beta_{j+1}, ..., \beta_n)$ is not a constant. Now, recall that $\mathcal{G}_1^{SV}{}_{|(y_1 = \alpha_j)} = \mathbf{e}_j \cdot z_1$. Let $\mathrm{Img}_{z_1}(\mathcal{G} + \mathcal{G}_1^{SV})$ 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 $(\beta_1, ..., \beta_{j-1}, \beta_j + z_1, \beta_{j+1}, ..., \beta_n) \in \mathrm{Img}_{z_1}(\mathcal{G} + \mathcal{G}_1^{SV})$, and so $\widetilde{\mathcal{G}} := \mathcal{G} + \mathcal{G}_1^{SV}$ is a hitting set generator for $\mathrm{orb}(f)$. □

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

All we need to do now is construct a hitting set generator for 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 $(4, k, s, d)$ formula whose top fan-in is at most k.

5.2 Constructing a faithful homomorphism for orbits

Let $f = \sum_{i \in [m]} f_i$ be a $(4, k, s, d)$ 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 $\text{tr-deg}_{\mathbb{F}}(\mathbf{g}) = \text{tr-deg}_{\mathbb{F}}(\phi(\mathbf{g}))$. Also, from Lemma 17, if ϕ is faithful to $\mathbf{g}$, then for any m-variate 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 orb(f). The following lemma helps us construct such a homomorphism.

Lemma 37. *Let $\mathbf{f} = (f_1, ..., f_m) \in \mathbb{F}[\mathbf{x}]^m$ be a tuple of n-variate polynomials of degree at most D, $A \in \mathrm{GL}(n, \mathbb{F})$, $g_i = f_i(A\mathbf{x})$ for all $i \in [m]$, and $\mathbf{g} = (g_1, \ldots, g_m)$. Further, suppose that $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) \leq r$, 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 $\text{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) = \text{rank}_{\mathbb{F}(\mathbf{z})} \psi(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}))$. 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,

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

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,

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

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

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

Also, from Observation 19, $\text{tr-deg}(\mathbf{g}) = \text{tr-deg}(\mathbf{f}) \leq r$, and $\deg(g_i) = \deg(f_i) \leq D$. So, using Lemma 18, we can construct a homomorphism ϕ faithful to $\mathbf{g}$ from ψ, as stated in the lemma. □

Let us apply Lemma 37 to the $(4, k, s, d)$ formula $f = \sum_{i \in [m]} f_i$, where $m \leq k$. Let $\mathbf{f} = (f_1, ..., f_m)$ and $\text{tr-deg}_{\mathbb{F}}(\mathbf{f}) = r \leq k$. Then, from Lemma 15, $\text{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f}) = r$. As A is invertible, this means that $\text{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x}) = r$. Assume without loss of generality that $f_1, \ldots, f_r$ is a transcendence basis of $\mathbf{f}$. Then, again from Lemma 15, the sub-matrix of $J_{\mathbf{x}}(\mathbf{f})$ consisting of the rows corresponding to $f_1, \ldots, f_r$ 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 $x_1, \ldots, x_r$, 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 $\text{rank}_{\mathbb{F}(\mathbf{z})} \psi\left(J_{\mathbf{x}}(\mathbf{f})(A\mathbf{x})\right)$ 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 with individual degree at most d. For $i \in [r]$, let the number of $q_{i,j}$ containing any of $x_1, \ldots, x_r$ be c_i. As f is an occur-k formula, $\sum_{i \in [m]} c_i \leq kr \leq k^2$. 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 $x_1, \ldots, x_r$. Moreover, even if $q_{i,j}$ contains some variable from $x_1, \ldots, x_r$, 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 individual degree at most $c_i d$. Thus, $\det(M')$ is a polynomial with sparsity at most $r! \cdot \prod_{i \in [r]} c_i s^{c_i} \leq k! \cdot k^k \cdot s^{k^2} \leq k^{2k} \cdot s^{k^2}$ and individual degree at most $\sum_{i \in [r]} c_i d \leq k^2 d$. So, $\det(M)$ is a product of polynomials with sparsity at most $k^{2k} \cdot s^{k^2}$ and individual degree at most $k^2 d$. From Theorem 27, $\psi = \mathcal{G}^{SV}_{\left(2\left\lceil \log\left(k^{2k} \cdot s^{k^2}\right)\right\rceil (k^2 d+1)+1\right)} = \mathcal{G}^{SV}_{O(k^4 d(\log k + \log s))}$ is a hitting set generator for $\mathrm{orb}(\det(M))$, as $|\mathbb{F}| \geq n$.
If the $q_{i,j}$ are b-variate polynomials, then $\det(M')$ is a polynomial in $\sum_{i \in [r]} c_i b \leq k^2 b$ variables. From Observation 13, $\psi = \mathcal{G}^{SV}_{k^2 b}$ is a hitting set generator for $\mathrm{orb}(\det(M))$.

Constructing ϕ. Using ψ and Lemma 37, we get a homomorphism ϕ that is faithful to $\mathbf{g}$. Observe that ϕ is a polynomial map in at most $O\left(k^4 d\left(\log k + \log s\right)\right) + k + 1 = O\left(k^4 d\left(\log k + \log s\right)\right)$ variables and of degree at most $nk+1$ (as degree of the polynomial map ψ is at most n and, in Lemma 37, $\deg\left(\sum_{j=1}^{r} y_j t^{ij}\right) \leq nk+1$).

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 9: the depth-4 case

For $\Delta = 4$, the value of R in the statement of Theorem 9 is $(2k)^{128}$. However, in this special case, one can work with a much smaller value for R. We choose $R = 2k^4$.

Let f be a $(4, k, s, d)$ formula. If f is a constant, then so is every polynomial in $\mathrm{orb}(f)$. In this case, the set containing any non-zero point in $\mathbb{F}^n$ is a hitting set for $\mathrm{orb}(f)$; 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 $\mathrm{char}(\mathbb{F}) = 0$ or $> D$). From Claim 35, $\frac{\partial f}{\partial x_i} \neq 0$ can be computed by a $(4, 2k^2, 2ks, d)$ formula with top fan-in at most k. Moreover, from the proof of Claim 36, if $\mathcal{G}$ is a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then $\widetilde{\mathcal{G}} = \mathcal{G} + \mathcal{G}^{SV}_1$ is a hitting set generator for $\mathrm{orb}(f)$, provided $\mathrm{char}(\mathbb{F}) = 0$ or $> D$ and $|\mathbb{F}| > \deg(\mathcal{G}) \cdot D$. From Section 5.2, there exists a $\mathcal{G}$ that has at most $O\left(\left(2k^2\right)^4 d\left(\log 2k^2 + \log 2ks\right)\right) = O\left(k^8 d\left(\log k + \log s\right)\right)$ many variables and has degree at most $2nk^2 + 1$. Observe that the conditions on $\mathrm{char}(\mathbb{F})$ and $|\mathbb{F}|$ in Claim 36 are satisfied due to the choice of R. As $\mathcal{G}^{SV}_1$ has 2 variables and has degree n, $\widetilde{\mathcal{G}}$ has $O\left(k^8 d\left(\log k + \log s\right)\right)$ variables and has degree at most $2nk^2 + 1$. Thus, for any $g \in \mathrm{orb}(f)$, $g(\widetilde{\mathcal{G}})$ has $O\left(k^8 d\left(\log k + \log s\right)\right)$ variables and has degree at most $(2nk^2 + 1)D$. So, a hitting set for $\mathrm{orb}(f)$ can be computed in time $(nk^2 D)^{O\left(k^8 d(\log k + \log s)\right)} = (nRD)^{O\left(R^2 d(\log k + \log s)\right)}$.

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 $\mathrm{orb}(f)$ can be computed in $(nk^2 D)^{O(k^4 b)}$ time.

6 Hitting sets for orbits of occur-once formulas

In this section, we give a quasi-polynomial time construction of hitting sets for orbits of polynomials that are computable by occur-once formulas whose leaves are labelled by multilinear polynomials (more generally, by polynomials with low individual degree). We will identify an occur-once formula with the polynomial f it computes and define the width of f - denoted by $\mathrm{width}(f)$ - to be the number of non-constant sparse polynomials at the leaves of the formula. As mentioned in Section 1.3, we reduce the problem of finding a hitting set generator for $\mathrm{orb}(f)$ to that of finding a generator for $\mathrm{orb}(\frac{\partial f}{\partial x_i})$, where x_i is such that $\frac{\partial f}{\partial x_i}$ is a product of occur-once formulas of widths at most $\frac{\mathrm{width}(f)}{2}$; this is done in Theorem 40. 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. The lemmas are inspired by similar structural results for read-once formulas given in [SV15], but the arguments need to be strengthened here as occur-once formula is a more powerful model.

6.1 Structural results

We will call an occur-once formula an (s,d) *occur-once formula* if the leaves of the formula are labelled by s-sparse polynomials with individual degree at most d. Without loss of generality, assume that an (s,d) 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 38. *Let f be an (s,d) occur-once formula having $\mathrm{width}(f) \geq 2$. 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,d) occur-once formulas. Further, $\mathrm{width}(f_1) + \mathrm{width}(f_2) = \mathrm{width}(f)$ *in the first two forms, and* $\mathrm{width}(f_1) = \mathrm{width}(f)$ *and* $\mathrm{depth}(f_1) < \mathrm{depth}(f)$ *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 $\mathrm{width}(h) \geq 2$. 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 with $\mathrm{width}(h') \geq 2$ and at layers less than k for some $1 < k \leq \Delta - 2$. Let h be a gate in the k-th layer with $\mathrm{width}(h) \geq 2$.

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-constant, then h is in form 1. On the other hand, if only one child, say h_1, is 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-constant, then h is in form 2. On the other hand, if only one child, say h_1, is 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. $\square$

Lemma 39. *Let f be an (s,d) occur-once formula. Then for any $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a product of (s,d) 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 with individual degree at most d is also an s-sparse polynomial with individual degree at most d. Assume, by the way of induction, that the lemma is true for all gates at layers less than k, for some $1 < k \leq \Delta - 2$ 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,d) 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,d) 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,d) 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,d) occur-once formulas of widths at most $\text{width}(h)$. Thus, the lemma is true for h. $\square$

6.2 Proof of Theorem 10

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

Theorem 40. *Let $f(\mathbf{x})$ be an n-variate, degree-D polynomial that is computable by an (s,d) 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 $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right) \neq 0$. In fact, if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$.*

Proof: Notice that if g is a non-zero constant, then $g\left(\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right) \neq 0$ for all t. So, to prove the theorem, we need to show that if g is not a constant, then neither is $g\left(\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right)$.

Let h be an (s,d) occur-once formula satisfying 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 with individual degree at most d, $e_1,...,e_m \in \mathbb{N}$, $\alpha_1,...,\alpha_m \in \mathbb{F}\setminus\{0\}$ and $\beta_1,...,\beta_m \in \mathbb{F}$. Let $A \in \mathrm{GL}(n,\mathbb{F})$. If $h(A\mathbf{x})$ is a not a constant, then neither is $p(A\mathbf{x})$. Thus, from Theorem 27 and the fact that $\mathrm{Img}(\mathcal{G}^{SV}_k) \subseteq \mathrm{Img}(\mathcal{G}^{SV}_{k+1})$ for any $k \geq 0$, we have that $p\left(A\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right)$ is not a constant for any $t \geq 0$. Hence, $h\left(A\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right)$ 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,d) occur-once formula f' with $1 \leq \mathrm{width}(f') < \ell \leq \mathrm{width}(f)$. Let h be an n-variate, degree-D, (s,d) occur-once formula having width$(h)=\ell \geq 2$, and $A \in \mathrm{GL}(n,\mathbb{F})$. From Lemma 38, there are three cases,

Case 1: $h=\alpha(h_1+h_2)+\beta$. Then, we can assume without loss of generality that width$(h_1) \leq \frac{\mathrm{width}(h)}{2}=\frac{\ell}{2}$, as width$(h_1)$ + width(h_2) = 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 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), $\frac{\partial h}{\partial x_i} \neq 0$. Now, from Lemma 39, $\frac{\partial h_1}{\partial x_i}$ is a product of (s,d) occur-once formulas of width at most $\frac{\ell}{2}$. Then, from the induction hypothesis, $\frac{\partial h}{\partial x_i}\left(A\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right) \neq 0$ 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 $\left(\frac{\partial p}{\partial x_1}\ \frac{\partial p}{\partial x_2}\ \cdots\ \frac{\partial p}{\partial x_n}\right)^T$. 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 $\frac{\partial q}{\partial x_j}\left(\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right) \neq 0$ for any $t \geq \log\ell-1$. This means that there is a $(\beta_1,...,\beta_n) \in \mathrm{Img}\left(\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right)$ such that

$$\frac{\partial q}{\partial x_j}(\beta_1,...,\beta_n) \neq 0,$$

because $\deg\left(\frac{\partial q}{\partial x_j}\left(\mathcal{G}^{SV}_{(2\lceil\log s\rceil(d+1)+1+t)}\right)\right) \leq nD$ and $|\mathbb{F}| > nD$. Now, set $k=2\lceil\log s\rceil(d+1)+1+t$. Let $r(z_{k+1}) := q(\beta_1,...,\beta_{j-1},\beta_j+z_{k+1},\beta_{j+1},...,\beta_n)$. Then,

$$\frac{\partial r}{\partial z_{k+1}}(0)=\frac{\partial q}{\partial x_j}(\beta_1,...,\beta_n) \neq 0,$$

and so, $q(\beta_1,...,\beta_{j-1},\beta_j+z_{k+1},\beta_{j+1},...,\beta_n)$ is not a constant. Now, recall that $\mathcal{G}^{SV}_{k+1|(y_{k+1}=\alpha_j)} = \mathcal{G}^{SV}_k+\mathbf{e}_j\cdot z_{k+1}$. Let $\mathrm{Img}_{z_{k+1}}(\mathcal{G}^{SV}_{k+1})$ be the "partial image" of $\mathcal{G}^{SV}_{k+1}$ obtained by keeping the z_{k+1} variable alive and setting all other variables to field elements. This means that $(\beta_1,...,\beta_{j-1},\beta_j+$

$z_{k+1}, \beta_{j+1}, ..., \beta_n) \in \mathrm{Img}_{z_{k+1}}(\mathcal{G}^{SV}_{k+1})$, and hence, $q(\mathcal{G}^{SV}_{k+1})$ is not a constant; i.e., $h(A\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)})$ 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 $t \geq \log \ell$, $h_1\left(A\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right), h_2\left(A\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$ are not constants and so neither is $h\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$.

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 $h\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$ is not a constant if and only if $h_1\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$ is not a constant, the problem reduces to showing that for any $g_1 \in \mathrm{orb}(h_1)$, $g_1\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+t)}\right)$ 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 38). $\square$

A non-zero polynomial $f \in \mathcal{C}$ is computable by an (s, d) occur-once formula. Observe that $\mathrm{width}(f) \leq n$. Let $g \in \mathrm{orb}(f)$. From Theorem 40, we have that $g\left(\mathcal{G}^{SV}_{(2\lceil \log s \rceil (d+1)+1+\lceil \log n \rceil)}\right)$ is a non-zero polynomial in $2\left(2\lceil \log s \rceil (d+1) + 1 + \lceil \log n \rceil\right)$ variables of degree at most nD. As $|\mathbb{F}| > nD$, a hitting set for $\mathrm{orb}(\mathcal{C})$ can be computed in time $(nD+1)^{2(2\lceil \log s \rceil (d+1)+1+\lceil \log n \rceil)} = (nD)^{O(\log n + d \log s)}$.

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 13 instead of Theorem 27 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 sparse polynomials, commutative ROABP, constant-width ROABP, constant-depth constant-occur formulas, and occur-once formulas (albeit under the low individual degree restriction). 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 some natural questions that, if resolved, will strengthen and complete the set of results presented in this work. We leave these questions 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 ROABP and constant-width ROABP?

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

- **Hitting sets for 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* ROABP is a hitting set for $\mathrm{orb}(\mathsf{IMM})$. 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 IMM. His question, asked during a discussion with CS on equivalence testing for IMM, has spurred us to think about the problems 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 7 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.

[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.

[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 proofs of Observation 21, Claim 23 and Lemma 24.

A.1 Proof of Observation 21

The proof of Observation 21 follows from the following claim.

Claim 41. *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 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 23

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

$$\begin{aligned} u_{\mathbf{e}} &= \prod_{i \in [m]} u_{i,e_i} \\ &= \prod_{i \in [m]} \left(\sum_{b_i=0}^{d} \cdot \binom{b_i}{e_i} (-r_i)^{b_i - e_i} \cdot v_{i,b_i} \right) && \text{(from Observation 21)} \end{aligned}$$

$$= \sum_{\mathbf{b}=(b_1,\ldots,b_n)\in\{0,\ldots,d\}^m} \binom{\mathbf{b}}{\mathbf{e}} \prod_{i\in[m]} (-r_i)^{b_i} \cdot \prod_{i\in[m]} v_{i,b_i} \cdot \prod_{i\in[m]} (-r_i)^{-e_i}$$
$$= \sum_{\mathbf{b}\in\{0,\ldots,d\}^m} \binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{\mathbf{b}} \cdot v_{\mathbf{b}} \cdot \mathbf{r}^{-\mathbf{e}}$$
$$= \sum_{\mathbf{b}\in\{0,\ldots,d\}^m} v_{\mathbf{b}} \cdot \mathbf{r}^{\mathbf{b}} \cdot \binom{\mathbf{b}}{\mathbf{e}} \cdot \mathbf{r}^{-\mathbf{e}}.$$

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

A.3 Proof of Lemma 24

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 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 $\mathcal{B} := \{\mathbf{e} \in \{0,\ldots,d\}^m : u_{\mathbf{e}} \text{ is in the minimal basis of } U \text{ w.r.t. } \prec\}$.

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 $\{u_{\mathbf{e}} : \mathbf{e} \in \mathcal{B} \text{ and } \mathbf{e} \prec \mathbf{b}\}$. 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 $\mathbf{b} \in \{0,\ldots,d\}^m$, 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 42. *There exists a way of numbering all $\mathbf{b} \in F \cap \mathcal{B}$ as $\mathbf{b}_1,\ldots,\mathbf{b}_p$ and there exist non-empty sets $E_1,\ldots,E_p \subseteq [m]$, each of size at most $\log p \leq \log w^2$ such that for all $k \in [p-1]$,*

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

Proof: Suppose that we have already identified $\mathbf{b}_1,\ldots,\mathbf{b}_{k-1}$ for some $k \in [p-1]$ and have constructed $E_1,\ldots,E_{k-1}$ 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 $\mathbf{b}_1, \ldots, \mathbf{b}_{p-1}$, call the last remaining vector $\mathbf{b}_p$ and pick E_p to be any arbitrary singleton set. $\square$

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 43. *For every $k \in [p]$, there exists a vector $\mathbf{b}'_k \in \{0, \ldots, d\}^m \setminus (F \cup \mathcal{B})$ 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 $\mathbf{b}' \in \{0, \ldots, d\}^m$ 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 $\{0, \ldots, d\}^m \setminus F$. As $|\mathcal{B}| \leq w^2$, at least one of these $2w^2 - 1$ vectors is in $\{0, \ldots, d\}^m \setminus (F \cup \mathcal{B})$. Pick any such vector and call it $\mathbf{b}'_k$. $\square$

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 43 as a proxy for $\mathbf{b}_k$. Notice that $u_{\mathbf{b}'_k}$ is dependent on $\{u_{\mathbf{e}} :\ \mathbf{e} \in \mathcal{B} \text{ and } \mathbf{e} \prec \mathbf{b}'_k\}$. 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 44. *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 non-zero entries 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, $R_m[M_m]_F = (R_{m-1} \otimes R_1)([M_{m-1}]_F \otimes [M_1]_F) = (R_{m-1}[M_{m-1}]_F) \otimes (R_1[M_1]_F)$. 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 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 45. *$[N]_A$ 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 46. *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 44, 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 42 and 43, $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k}$ for all $\ell > k$. As every coordinate of $\mathbf{b}_k$ is non-zero, it follows from Claim 44 that the entry in the row indexed by $\mathbf{b}_\ell$ must be 0 for every $\ell > k$. Also, from Claim 43, as $\mathbf{b}_k$ and $\mathbf{b}'_k$ agree at all coordinates $\mathbf{b}'_k$ is non-zero. So, from Claim 44, the entry in the row indexed by $\mathbf{b}_k$ must be non-zero. Also, recall from Claim 44 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 31

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 47. *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 $\{\mathbf{n_b} : \mathbf{b} \in F\}$ 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 $\mathbf{e} \prec_{\text{dlex}} \mathbf{b}\}$. Pick this dependence vector as $\mathbf{n_b}$.

Case 2: $\mathbf{b} \in F \cap \mathcal{B}$. Let there be p such $\mathbf{b}$, $\mathbf{b}_1, \ldots, \mathbf{b}_p$, where $p \le |\mathcal{B}| \le w^2$. 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 48. *There exist sets $E_1, \ldots, E_p \subseteq [m]$, each of size $w^2 - 1$ such that for all $k \in [p]$,*

1. $\text{Supp}\left((\mathbf{b}_k)_{E_k}\right) = w^2 - 1$,
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 $\text{Supp}(\mathbf{b}_k) = |\mathcal{S}(\mathbf{b}_k)| = 2\mu - (q^* - 1) \ge \mu + 2 = w^2 + \lceil \log w^2 \rceil + 2$. For $\ell \neq k$, as $\text{Supp}(\mathbf{b}_k) = \text{Supp}(\mathbf{b}_\ell)$ 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 49. *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 43. □

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

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 49 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 $\det([CMDN]_F)$ is a product of a bunch of non-zero linear forms in $\mathbb{F}[\mathbf{t}]$ and a polynomial in $\mathbb{F}[\mathbf{t}]$ which contains a monomial of degree at most $2w^2\mu$.

$[CMDN]_F$ 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 50. *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 $F_0 := \{\mathbf{b} \in F : \text{Supp}(\mathbf{b}) > 2\mu - (q^* - 1)\}$ and $F_1 := \{\mathbf{b} \in F : \text{Supp}(\mathbf{b}) = 2\mu - (q^* - 1)\}$. 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 $A := (F \setminus \mathcal{B}) \cup \left\{\mathbf{b}'_1, \ldots, \mathbf{b}'_p\right\}$. 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 $k \in [p]$. For any $S \subseteq \{0,1\}^m$ of size $|S| = |F|$, let $[M]_{F,S}$ 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 $S \not\subseteq A \uplus \mathcal{B}$, $\det(N_S) = 0$. 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

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

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

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 51. *$[N]_A$ is an identity matrix.*

Proof: Same as that of Claim 45. □

Claim 52. *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 50, 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 48 and 49, $(\mathbf{b}'_k)_{E_k} = (\mathbf{b}_k)_{E_k} \neq (\mathbf{b}_\ell)_{E_k}$ for all $\ell \neq k$. As every coordinate of $(\mathbf{b}_k)_{E_k}$ is non-zero, it follows from Observation 50 that the entry in the row indexed by $\mathbf{b}_\ell$ must be 0 for every $\ell \neq k$. Also, from Claim 49, as $\mathbf{b}_k$ and $\mathbf{b}'_k$ agree at all coordinates $\mathbf{b}'_k$ is non-zero. So, from Observation 50, the entry in the row indexed by $\mathbf{b}_k$ must be non-zero. Also, recall from Observation 50 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 53. $\det\left([M]_{F,A}\right) \cdot \det(N_A) \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus A} \mathbf{r}^{\mathbf{e}} = \prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}} \neq 0$ *and has $\mathbf{t}$-degree at most $2w^2\mu$.*

Proof: $\det\left([M]_{F,A}\right) \cdot \det(N_A) \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus A} \mathbf{r}^{\mathbf{e}} = \prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}} \neq 0$ follows from Claims 51 and 52 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, $\deg_{\mathbf{t}}\left(\prod_{\mathbf{e} \in \mathcal{B}} \mathbf{r}^{\mathbf{e}}\right) \leq w^2 \cdot (2\mu - (q^* - 1)) \leq 2w^2\mu$, as $|\mathcal{B}| \leq w^2$. □

Claim 54. *For any $S \in A \uplus \mathcal{B}$ such that $\det(N_S)$ 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 $\det\left([M]_{F,S}\right) \cdot \det(N_S) \cdot \prod_{\mathbf{e} \in A \setminus S} \mathbf{r}^{\mathbf{e}} \cdot \prod_{\mathbf{e} \in \mathcal{B} \setminus S} \mathbf{r}^{\mathbf{e}}$ 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 53 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 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, d) formula i.e., a depth-Δ, occur-k formula of size-s whose leaves are sparse polynomials of individual degree at most d. Let us identify f with a (Δ, k, s, d) formula computing it. Just like we did in Section 5, we first upper bound the top fan-in of f in Section C.1 and then use the notion of faithful homomorphisms to construct hitting sets for $\mathrm{orb}(f)$ in C.2.

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

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

Claim 55. *If f is a (Δ, k, s, d) formula, then it can also be computed by a $(\Delta, k, s^{\Delta}, d)$ formula in a canonical form with the following properties:*

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

Proof: As the sum of sparse polynomials is also a sparse polynomial, if there exists a leaf which is a sum gate, then we simply replace it with the sparse polynomial that it computes. Notice that this does not increase the depth, size or occur of f, nor does it increase the individual degree of the leaves of f. 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, occur or individual degree of the leaves. 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, occur or individual degree; 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 $\ell + 1$ to gates at level ℓ be $r_\ell \leq s$, for all $\ell \in [\Delta' - 1]$. Also, let the sum of weights of edges connecting the inputs to gates at level Δ' be $r_{\Delta'}$. As, all edge weights are non-negative, $\sum_{i \in [m]} e_i \leq \prod_{\ell \in [\Delta']} r_\ell \leq s^{\Delta'} \leq s^{\Delta - 2}$. Since, there can be no more than s such sub-trees, the size of f can increase by at most $s^{\Delta - 1}$. Thus, size of f is at most $s + s^{\Delta-1} \leq s^{\Delta}$ (because, $s = 1$ means that f can not contain any $\times\curlywedge$ gate). $\square$

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 $(\Delta - 1, k, s^{\Delta}, D)$ formulas, with $+$ gates at the top or are sparse polynomials. We now make the following claim which will allow us to assume that the top fan-in of f is at most k.

Claim 56. *Let f be a (Δ, k, s, d) formula in the canonical form of Claim 55, with either a $+$ gate at the top or $\Delta = 2$. Then, for any $i \in [n]$, $\frac{\partial f}{\partial x_i}$ is a $(\Delta, (2k)^{\Delta/2}, (2k)^{\Delta/2}s, d)$[22] 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, occur and individual degree 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 $f_1, \dots, f_r$, $r \leq k$. 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 $q_{i,1}, \dots q_{i,r_i}$, $\sum_{i \in [r]} r_i \leq k$. 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. Similarly, the individual degrees of all leaves of $\frac{\partial f}{\partial x}$ is also at most d. 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 $(2k)^{\frac{\Delta-2}{2}}$. So, the occur of $\frac{\partial f}{\partial x}$ is at most $\sum_{i \in [r]} r_i \left((2k)^{\frac{\Delta-2}{2}} + p_i \right)$, 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 $\mathrm{char}(\mathbb{F}) > D^R \geq D$). 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 $\left(\frac{\partial p}{\partial x_1} \; \frac{\partial p}{\partial x_2} \; \dots \; \frac{\partial p}{\partial x_n} \right)^T$. 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 $\frac{\partial g}{\partial x_j}(\mathcal{G}) \neq 0$. This means that there is a $(\beta_1, ..., \beta_n) \in \mathrm{Img}(\mathcal{G})$ such that

[22] Notice 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 at level $\Delta - 2$ are $\times\curlywedge$ gates. So, $\Delta/2$ is an integer.

$$\frac{\partial g}{\partial x_j}(\beta_1, ..., \beta_n) \neq 0,$$

because $\deg\left(\frac{\partial g}{\partial x_i}(\mathcal{G})\right) \leq (nR+1)D$ and $|\mathbb{F}| > (nR+1)D$. Let $r(z) := g(\beta_1, ..., \beta_{j-1}, \beta_j + z, \beta_{j+1}, ..., \beta_n)$. Then,

$$\frac{\partial r}{\partial z}(0) = \frac{\partial g}{\partial x_j}(\beta_1, ..., \beta_n) \neq 0,$$

and so, $g(\beta_1, ..., \beta_{j-1}, \beta_j + z, \beta_{j+1}, ..., \beta_n)$ is not a constant. Notice that, $(\beta_1, ..., \beta_{j-1}, \beta_j + z, \beta_{j+1}, ..., \beta_n) \in \mathrm{Img}(\mathcal{G} + \mathcal{G}_1^{SV})$, and so $\widetilde{\mathcal{G}} := \mathcal{G} + \mathcal{G}_1^{SV}$ is a hitting set generator for $\mathrm{orb}(f)$. So, all we need to do now is construct a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_j}\right)$. Overloading the notation, we refer to $\frac{\partial f}{\partial x_j}$ as f, which is computed by a (Δ, k, s, d) formula in the canonical form and with a $+$ gate at the top whose fan-in is at most k.

C.2 Constructing a faithful homomorphism

Let $f = f_1 + \cdots + f_k$, $g_i = f_i(A\mathbf{x})$ 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 17, 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 $\mathrm{rank}_{\mathbb{F}(\mathbf{x})} J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x}) = \mathrm{rank}_{\mathbb{F}(\mathbf{z})} \phi_{\ell+1}(J_{\mathbf{x}}(\mathbf{h})(A\mathbf{x}))$, then using Lemma 37, 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 57 (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_is 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 $\mathbf{h}$, the above 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 17, $\phi_{\ell+1}\left(p_i^{e_i}(A\mathbf{x})\right) \neq 0$ 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 $r_{\Delta-2}$ may sparse polynomials (because derivatives 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 $\mathbf{h}$. Every entry of any $|R'| \times |R'|$ minor of $J_{\mathbf{x}}(\mathbf{h})$ is a polynomial with sparsity at most s and individual degree at most d. So the determinant of any such minor is a polynomial with sparsity at most $R'! \cdot s^{R'} \leq R! \cdot s^R$ and individual degree dR. Hence, from Theorem 27, $\mathcal{G}^{SV}_{(2\lceil \log R! \cdot s^R \rceil (dR+1)+1)} = \mathcal{G}^{SV}_{O(R^2 d(\log R + \log s))}$ is a hitting set generator for orbits of these determinants. We then repeatedly use Lemma 37 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 $\mathcal{G}^{SV}_{O(R^2 d(\log R + \log s))} + \psi$ is a hitting set generator for $\mathrm{orb}(f)$. We are now ready to prove Theorem 9.

C.3 Proof of Theorem 9

A non-zero polynomial $f \in \mathcal{C}$ is computed by an (Δ, k, s, d) formula. Then, f is also computed by a (Δ, k, s^Δ, d) formula in the canonical form of Claim 55. 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. There exists a x_i such that $\frac{\partial f}{\partial x_i} \neq 0$ and can be computed by a $(\Delta, (2k)^{\Delta/2}, (2k)^{\Delta/2} s^\Delta, d)$ formula. Moreover, if $\mathcal{G}$ is a hitting set generator for $\mathrm{orb}\left(\frac{\partial f}{\partial x_i}\right)$, then $\widetilde{\mathcal{G}} = \mathcal{G} + \mathcal{G}^{SV}_1$ is a hitting set generator for $\mathrm{orb}(f)$. Now, there exists a $\mathcal{G}$ that has at most

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

many variables and of degree $nR+1$. As G^{SV}_1 has 2 variables and is of degree n, the number of variables in $\widetilde{\mathcal{G}}$ is $O\left(R^2 d\left(\log R + \Delta \log k + \Delta \log s\right) + \Delta R\right)$ and its degree is $nR+1$. Thus, for any $A \in \mathrm{GL}_n(\mathbb{F})$, $g(\widetilde{\mathcal{G}})$ is a polynomial in $O\left(R^2 d\left(\log R + \Delta \log k + \log s\right) + \Delta R\right)$ variables and of degree at most $(nR+1)D$. So, a hitting set can be constructed in time $(nRD)^{O\left(R^2 d(\log R + \Delta \log k + \Delta \log s) + \Delta R\right)}$.

Case 2: The top most gate of the formula is a $\times\!\curlywedge$ gate. Then, all inputs to this gate are computed by $(\Delta, k-1, s^\Delta, d)$ formulas in the canonical form and with a $+$ gate at the top. 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 $\mathcal{G}^{SV}_{Rb}$ is a hitting set generator for b-variate polynomials. So, we can use $\mathcal{G} = \mathcal{G}^{SV}_{Rb} + \psi$.

D A lower bound for ROABP

In this section, we show that there is a $3n+2$ variate $O(n)$-sparse polynomial f satisfying the following property: there exists a polynomial $g \in \mathrm{orb}(f)$ such that any ROABP computing g

must have 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 58 (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 59. *The double cover of a k-regular graph is also k-regular.*

Observation 60. *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 59, $\widetilde{G}$ 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 61. *Each x_i, y_i and z_i appears in exactly one of the affine forms associated with vertices in L and in exactly one of the affine forms associated with vertices in R.*

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

Fact 62. *[KNS20] Over any field $\mathbb{F}$, any ROABP computing p must have 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 a ROABP computing g, we get a 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[23] associated with it is $x_i + y_j + z_k$. Before constructing f, we prove the following claim.

Claim 63. *Let the linear forms associated with vertices in L be $\ell_1, \ldots, \ell_n$ and those associated with 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 $2n - 1$.*

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

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 61,

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

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 60, 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 and has $O(n)$ monomials, as desired.

A **and b mapping** *f* **to** *g***.** As $\mathbb{F}\text{-span}\langle \ell_1, \dots \ell_n, r_1, \dots, r_n\rangle$ 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

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

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 every variable in $y_n + \sum_{i\in[n]} z_i$ by the corresponding linear form in A. Then it is easy to see that $g(\mathbf{x}, \mathbf{y}, \mathbf{z}, s_1, s_2) = f(A(\mathbf{x}, \mathbf{y}, \mathbf{z}, s_1, s_2) + \mathbf{b})$.

E 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 64. *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 of formula of the smallest possible depth Δ. From Lemma 38, 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 38 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. $\square$

Claim 65. *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}. $\square$

F Affine projections and orbit closures

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

Definition 66 (Orbit closure)**.** The orbit closure of f over $\mathbb{F}$, denoted by $\overline{\mathrm{orb}_{\mathbb{F}}(f)}$, is the smallest affine variety[24] 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 67. $\mathrm{aproj}_{\mathbb{F}}(f) \subseteq \overline{\mathrm{orb}_{\mathbb{F}}(f)}$.

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

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, $\mathbf{c} \in \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$. $\square$

Claim 69. $\overline{\mathrm{orb}_{\mathbb{F}}(f)} = \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$.

[24]By a 'variety' we mean an 'algebraic set' that is not necessarily an irreducible variety.

Proof: From Observation 68, $\text{orb}_{\mathbb{F}}(f) \subseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$, as $\text{orb}_{\mathbb{F}}(f) \subseteq \text{aproj}_{\mathbb{F}}(f)$. Since $\overline{\text{orb}_{\mathbb{F}}(f)}$ is the smallest variety in $\mathbb{F}^m$ containing $\text{orb}_{\mathbb{F}}(f)$, and intersection of two varieties is again a variety, we have $\overline{\text{orb}_{\mathbb{F}}(f)} \subseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$.

To show the direction, i.e., $\overline{\text{orb}_{\mathbb{F}}(f)} \supseteq \mathbb{V}(\text{ann-}\mathbb{I}(\mathbf{g}))$, it is sufficient to show that the ideal of $\overline{\text{orb}_{\mathbb{F}}(f)}$ (denoted as $\mathbb{I}(\overline{\text{orb}_{\mathbb{F}}(f)})$) is contained in $\text{ann-}\mathbb{I}(\mathbf{g})$. This is because, $\mathbb{V}(\mathbb{I}(\overline{\text{orb}_{\mathbb{F}}(f)})) = \overline{\text{orb}_{\mathbb{F}}(f)}$, as $\overline{\text{orb}_{\mathbb{F}}(f)}$ is a variety. Let $p(\mathbf{z}) \in \mathbb{I}(\overline{\text{orb}_{\mathbb{F}}(f)})$ and $\deg_{\mathbf{z}}(p) = D$. Then, $p(\mathbf{c}) = 0$ for all $\mathbf{c} \in \text{orb}_{\mathbb{F}}(f)$. 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 $\text{char}(\mathbb{F}) = 0$). 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,

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

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 $\mathbf{c} \in \text{orb}_{\mathbb{F}}(f)$,

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

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, $p(\mathbf{g}) = 0$. $\square$

The proof of the theorem now follows from Observation 68 and the above claim. $\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il