Electronic Colloquium on Computational Complexity, Report No. 191 (2018)

Reconstruction of non-degenerate homogeneous depth three circuits

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

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

November 10, 2018

Abstract

A homogeneous depth three circuit C computes a polynomial

$$f = T_1 + T_2 + ... + T_s \ ,$$

where each T_i is a product of d linear forms in n variables over some underlying field $\mathbb{F}$. Given black-box access to f, can we efficiently reconstruct (i.e. proper learn) a homogeneous depth three circuit computing f? Learning various subclasses of circuits is natural and interesting from both theoretical and practical standpoints and in particular, properly learning homogeneous depth three circuits efficiently is stated as an open problem in a work by Klivans and Shpilka (COLT 2003) and is well-studied. Unfortunately, there is substantial amount of evidence to show that this is a hard problem *in the worst case*.

We give a (randomized) poly(n, d, s)-time algorithm to reconstruct *non-degenerate* homogeneous depth three circuits for $n = \Omega(d^2)$ (with some additional mild requirements on s and the characteristic of $\mathbb{F}$). We call a circuit C as non-degenerate if the dimension of the partial derivative space of f equals the sum of the dimensions of the partial derivative spaces of the terms $T_1, T_2, \ldots, T_s$. In this sense, the terms are "independent" of each other in a non-degenerate circuit. A random homogeneous depth three circuit (where the coefficients of the linear forms are chosen according to the uniform distribution or any other reasonable distribution) is almost surely non-degenerate. In comparison, previous learning algorithms for this circuit class were either improper (with an exponential dependence on d), or they only worked for $s < n$ (with a doubly exponential dependence of the running time on s).

The main contribution of this work is to formulate the following paradigm for efficiently handling addition gates and to successfully implement it for the class of homogeneous depth three circuits. The problem of finding the children of an addition gate with large fan-in s is first reduced to the problem of decomposing a suitable vector space U into a (direct) sum of *simpler* subspaces $U_1, U_2, \ldots, U_s$. One then constructs a suitable space of operators $\mathcal{S}$ consisting of linear maps acting on U such that analyzing the *simultaneous global structure* of $\mathcal{S}$ enables us to efficiently decompose U. In our case, we exploit the structure of the set of low rank matrices in $\mathcal{S}$ and of the invariant subspaces of U induced by $\mathcal{S}$. We feel that this paradigm is novel and powerful: it should lead to efficient reconstruction of many other subclasses of circuits for which the efficient reconstruction problem had hitherto looked unapproachable because of the presence of large fan-in addition gates.

ISSN 1433-8092

1 Introduction

Reconstruction of arithmetic circuits is the algebraic analogue of exact learning [Ang88] for Boolean circuits. It is the following fundamental learning theoretic problem: Given black-box access (i.e. membership query access) to a multivariate polynomial f that is computable by an arithmetic circuit of size s, construct a small circuit (ideally, a $\text{poly}(s)$-size circuit) computing f. By definition, reconstruction is closely related to approximating the minimum circuit[1] size and it is expected to be an inherently hard problem. Research has, therefore, focused on investigating the computational tractability of reconstruction for restricted (albeit, quite interesting) models of circuits. Depth three circuits is one such model.

Depth three circuit reconstruction. Bounded depth circuits have alternating layers of plus gates (Σ layer) and product gates (Π layer). Reconstruction of depth two circuits is either the problem of sparse polynomial interpolation or the problem of polynomial factorization into linear factors, depending on whether there is a plus gate or a product gate at the top layer. Reconstruction of depth three $\Pi\Sigma\Pi$ circuits is the problem of polynomial factorization into sparse factors. The sparse polynomial interpolation problem can be solved in deterministic polynomial time [BT88,GKS94,KS01], and the polynomial factorization problem in randomized polynomial time [KT90]. Thus, studying the reconstruction problem for depth three $\Sigma\Pi\Sigma$ circuits is the natural next step towards pushing the frontier of efficient reconstruction. However, it turns out that reconstruction of $\Sigma\Pi\Sigma$ circuits is directly linked with reconstruction of general circuits: A polynomial-time reconstruction algorithm for $\Sigma\Pi\Sigma$ circuits implies a sub-exponential time reconstruction algorithm for general circuits. This follows immediately from the depth reduction result of [GKKS16]. In this article, depth three circuit(s) would always mean $\Sigma\Pi\Sigma$ circuit(s).

Restricted depth three circuits. A depth three circuit $\mathtt{C}$ computes a n-variate polynomial

$$f(\mathbf{x}) = T_1 + \ldots + T_s \ ,$$

where each term T_i is a product of d affine forms. The parameter s is the top fan-in of the circuit. We will refer to $\mathtt{C}$ as a (n, d, s) depth three circuit. Circuit $\mathtt{C}$ is a *powering* circuit if every term is the d-th power of a linear form[2], $\mathtt{C}$ is a *set-multilinear* circuit if there is a partition of the variables $\mathbf{x} = \mathbf{x}_1 \uplus \ldots \uplus \mathbf{x}_d$ such that every term is a product of d linear forms with the j-th form depending on only $\mathbf{x}_j$-variables[3], and $\mathtt{C}$ is a *multilinear* circuit if every term computes a multilinear polynomial. If every term is a product of d linear forms then $\mathtt{C}$ is a *homogeneous* depth three circuit – we are primarily interested in this model in this work.

Previous work on reconstruction of restricted depth three circuits. [BBB$^+$00] gave a randomized $\text{poly}(n, d, s)$ time learning algorithm for depth three powering circuits and set-multilinear depth three circuits. It was observed in [KS03] that the algorithm gives $\text{poly}(n, 2^d, s)$ time learning for general depth three circuits. The learning algorithms in [BBB$^+$00, KS03] are not proper as the

[1] In this article, circuit means arithmetic circuit unless specified otherwise.

[2] A linear form is an affine form with constant term zero.

[3] One could alternatively define both depth three powering circuits and set-multilinear depth three circuits using affine forms instead of linear forms. It turns out however that using affine forms instead of linear forms does not significantly change the expressive power nor does it change the computational difficulty of the corresponding reconstruction problem and so for ease of exposition we restrict ourselves to using linear forms in defining these two subclasses.

output hypothesis is a read-once oblivious branching program (ROABP) and not a depth three circuit. Over finite fields $\mathbb{F}$, [Shp07] gave a randomized quasi-polynomial in $(n, d, |\mathbb{F}|)$ time algorithm to reconstruct depth three circuits with top fan-in two; the running time is $\text{poly}(n, |\mathbb{F}|)$ for multilinear depth three circuits with top fan-in two. This learning algorithm[4] was derandomized and generalized in [KS09], who gave a $\text{poly}(n) \cdot |\mathbb{F}|^{(\log d)^{O(s^3)}}$ time algorithm to reconstruct (n, d, s) depth three circuits[5]; the complexity reduces to $(n + |\mathbb{F}|)^{2^{O(s^2)}}$ for multilinear depth three circuits. Over the field of real numbers[6], [Sin16] gave a randomized $\text{poly}(n, d)$ time algorithm[4] to reconstruct depth three circuits with top fan-in two. There are a few reconstruction algorithms for depth three powering circuits that are proper, but these algorithms work efficiently only if s is somewhat small. For instance, the equivalence test for the d-th power symmetric polynomial, in [Kay11], implies a randomized $\text{poly}(n, d, s)$ time reconstruction for depth three powering circuits with $s \leq n$, provided the s linear forms occurring in the circuit are linearly independent. In [Kay12], a randomized $\text{poly}(n, s^{\log_d s})$ time reconstruction algorithm was given for random[7] depth three powering circuits. For $s \leq \binom{n+1}{2}$ and $d \geq 5$, [GKP18] gave a randomized $\text{poly}(n, d)$ time algorithm to reconstruct random depth three powering circuits.

Our contribution. To our knowledge, there is no known efficient reconstruction algorithm for the much stronger homogeneous depth three circuit model. Indeed, this was posed as an open problem in [KS03]. The existing techniques either give an exponential dependence on d [BBB$^+$00,KS03] or they work efficiently for substantially small values of s [Shp07,KS09,Sin16]. Our main contribution is a $\text{poly}(n, d, s)$ time reconstruction algorithm for homogeneous depth three circuits, assuming a *non-degeneracy* condition that holds for almost all circuits from this class. In particular, a random homogeneous depth three circuit is non-degenerate with very high probability (in fact, with probability one in the measure theoretic sense). Qualitatively speaking, a circuit is non-degenerate if the terms of the circuit are "independent" of each other. Removing the non-degeneracy condition from our result would imply sub-exponential time reconstruction for general circuits (this point is discussed further in Section 1.2). We also give $\text{poly}(n, d, s)$ time reconstruction algorithms for depth three powering circuits and set-multilinear depth three circuits under similar non-degeneracy conditions, thereby improving the previous results [Kay12, GKP18] on random depth three powering circuits. Our reconstruction algorithms are proper. We note that there are many other reconstruction algorithms for different circuit classes but almost all[8] of them either consider only circuits in which addition gates have constant fan-in (typically fan-in 2) or their running time degrades very badly in the presence of addition gates with large fan-in. Conceptually, our contribution is to formulate a paradigm to efficiently handle addition gates with large fan-in (when the children are non-degenerate) and implement it in the case of homogeneous depth three circuits. We elaborate on this further in Section 2 after giving an overview of our technique.

[4] which is proper, provided the input circuit satisfies a rank condition

[5] When the input does not satisfy a certain rank condition, the algorithm in [KS09] outputs a "generalized" (a notion that augments the depth three circuit model in a certain way) depth three circuit.

[6] The algorithm of [Sin16] also works over the fields of rational and complex numbers

[7] The coefficients of the linear forms occurring in the circuit are chosen uniformly at random from a sufficiently large finite set.

[8] Except perhaps some of the ones pertaining to set-multilinear depth three and depth three powering circuits.

1.1 The results

Notations. Let $\mathtt{C}$ be a (n,d,s) depth three circuit computing a polynomial $f = T_1 + \ldots + T_s$ in variables $\mathbf{x} = \{x_1, \ldots, x_n\}$. We fix a few notations before stating our results:

$$\begin{aligned} \mathbf{x}^{\boldsymbol{\alpha}} &:= x_1^{\alpha_1} \cdot x_2^{\alpha_2} \ldots x_n^{\alpha_n}, \quad \text{where } \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}, \\ |\boldsymbol{\alpha}| &:= \alpha_1 + \alpha_2 + \ldots + \alpha_n, \end{aligned}$$

Let $k \geq 1$ be an integer parameter.

$$\begin{aligned} \partial^k_{\boldsymbol{\alpha}} f &:= \frac{\partial^k f}{\partial \mathbf{x}^{\boldsymbol{\alpha}}}, \\ U &:= \mathbb{F}\text{-span} \left\{ \partial^k_{\boldsymbol{\alpha}} f \ : \ |\boldsymbol{\alpha}| = k \right\}, \\ U_i &:= \mathbb{F}\text{-span} \left\{ \partial^k_{\boldsymbol{\alpha}} T_i \ : \ |\boldsymbol{\alpha}| = k \right\}, \end{aligned}$$

where $\mathbb{F}$ is the underlying field having characteristic 0 or greater than ds^2. We will assume that univariate polynomial factorization over $\mathbb{F}$ can be done in randomized polynomial time which is indeed the case for finite fields and $\mathbb{Q}$ [Ber70, LLL82]. For a better clarity, we will think of $\mathbb{F} = \mathbb{Q}$ in this article.

Non-degeneracy condition. We choose $k = \left\lceil \frac{\log s}{\log \frac{n}{ed}} \right\rceil$. Let $T_i = \ell_{i1} \cdot \ell_{i2} \ldots \ell_{id}$ be a product of d linear forms. A (n,d,s) homogeneous depth three circuit $\mathtt{C} = T_1 + \ldots + T_s$ is *non-degenerate* if it satisfies the following condition: For every $i \in [s]$, there are $2k+1$ linear forms $\ell_{ir_1}, \ldots, \ell_{ir_{2k+1}}$ such that

$$\dim \left(\sum_{j \in [s] \setminus \{i\}} U_j \mod \langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle \right) = (s-1) \cdot \binom{d}{k},$$

where $\langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle$ is the $\mathbb{F}$-linear space spanned by $\ell_{ir_1}, \ldots, \ell_{ir_{2k+1}}$. It is easy to check that the condition implies $U = U_1 \oplus \ldots \oplus U_s$ and $\dim(U_i) = \binom{d}{k}$ for every $i \in [s]$. Hence, $\dim(U) = s \cdot \binom{d}{k}$.

Theorem 1 (Homogeneous depth three circuit reconstruction)**.** *Let $n, d, s \in \mathbb{N}$, $n \geq (3d)^2$ and $s \leq (\frac{n}{3d})^{\frac{d}{3}}$. There is a randomized* $\text{poly}(n,d,s) = \text{poly}(n,s)$ *time algorithm which takes as input black-box access to a polynomial f that is computable by a non-degenerate (n,d,s) homogeneous depth three circuit and outputs a non-degenerate (n,d,s) homogeneous depth three circuit computing f.*

The algorithm works even if the input is evaluations of f and the k-th order partial derivatives of f at $\text{poly}(n,s)$ number of points chosen uniformly at random from a sufficiently large subset of $\mathbb{F}^n$. We show in Appendix A that a homogeneous depth three circuit is non-degenerate with high probability if the coefficients of the linear forms occurring in the circuit are chosen uniformly and independently at random from an arbitrary set of size $(nds)^4$. In fact, degenerate circuits correspond to a proper algebraic variety and so a random homogeneous depth three circuit is non-degenerate with probability one in the measure theoretic sense.

Reconstruction of subclasses. It is worth mentioning what we get for the two interesting subclasses of homogeneous depth three circuits – set-multilinear depth three circuits and depth three powering circuits. We state the result for set-multilinear depth three circuits, a very similar statement also holds for depth three powering circuits. Suppose the set of variables $\mathbf{x}$ admits a partition into d sets: $\mathbf{x} = \mathbf{x}_1 \uplus \mathbf{x}_2 \uplus \cdots \uplus \mathbf{x}_d$. For simplicity, assume $|\mathbf{x}_1| = |\mathbf{x}_2| = \ldots = |\mathbf{x}_d|$. A set-multilinear depth three circuit is a homogeneous depth three circuit in which every term T_i is of the form

$$T_i = \ell_{i1}(\mathbf{x}_1) \cdot \ell_{i2}(\mathbf{x}_2) \cdot \ldots \cdot \ell_{id}(\mathbf{x}_d), \tag{1}$$

where each ℓ_{ij} is a linear form over the indicated variable set $\mathbf{x}_j$. The polynomial computed by any such term (and therefore also by the circuit) has the property that every monomial in the support contains exactly one variable from each set $\mathbf{x}_j$. Such a polynomial f is referred to as a set-multilinear polynomial and also as a *tensor*, and the minimal number s of terms of the above form required to sum up to f is called the rank of the tensor. Because of its relevance to a wide variety of problems, including the complexity of matrix multiplication, tensor rank and tensor decomposition have been intensely studied. We state our result for this subclass by postponing the statement of the precise non-degeneracy condition to Section 5.

Theorem 2 (Set-multilinear depth three circuit reconstruction)**.** *Let $n, d, s \in \mathbb{N}$, and $s \leq (\frac{n}{d})^{\frac{d}{3}}$. There is a randomized $\mathrm{poly}(n, d, s) = \mathrm{poly}(n, s)$ time algorithm which takes as input black-box access to a polynomial f that is computable by a non-degenerate (n, d, s) set-multilinear depth three circuit and outputs a non-degenerate (n, d, s) set-multilinear depth three circuit computing f.*

We obtain a very similar result for depth three powering circuits where the maximum top fan-in that can be handled is $s \leq \binom{n+\frac{d}{3}-1}{n}$.

1.2 Evidence for the hardness of reconstruction

It is natural to wonder if the non-degeneracy condition can be removed from the above results. In particular, can we get rid of the non-degeneracy condition and the degree restriction ($n \geq (3d)^2$) from Theorem 1? We collect some evidence which indicates that this is likely a hard problem.

1. *NP-hardness of reconstructing subclasses.* Finding the smallest depth three powering circuit computing a given polynomial is NP-hard [Shi16]; it amounts to computing the symmetric-rank. The same is also true for finding the smallest set-multilinear depth three circuit computing a set-multilinear polynomial [Hås90]; it is the problem of computing the tensor-rank. One might now expect (perhaps somewhat naively) that the reconstruction problem for a more expressive circuit class $\mathcal{C}$ should be at least as difficult as the corresponding problem for a less expressive subclass, and this is then some sort of heuristic evidence that the optimal circuit reconstruction version of the problem for homogeneous depth three circuits is likely a hard one in the worst-case. We now give some evidence that even when we are allowed to output an approximately minimal circuit from the class $\mathcal{C}$ of homogeneous depth three circuits, the problem is likely very challenging.

2. *Reconstruction of general circuits.* Suppose we are able to prove Theorem 1 without any kind of non-degeneracy and degree restrictions, and output a $\mathrm{poly}(s)$ size homogeneous depth three circuit. Then, a simple homogenization trick gives a $\mathrm{poly}(n, d, s)$ time reconstruction algorithm for general depth three circuits: If f is computable by a (n, d, s) depth three circuit then $z^d \cdot f(z^{-1}x_1, \ldots, z^{-1}x_n)$ is computable by a $(n + 1, d, s)$ homogeneous depth three

circuit. This, in turn, implies a $n^{O(\sqrt{d})}$ time reconstruction algorithm for poly(n) size general circuits computing n-variate polynomials of degree $d = n^{O(1)}$, owing to the depth reduction to depth three results [GKKS16, Tav13, Koi12, AV08]. In other words, we would get a sub-exponential time reconstruction algorithm for general circuits if Theorem 1 holds without non-degeneracy and degree restrictions.

3. *Reconstruction implies lower bound.* Reconstruction is harder than proving lower bounds. In [FK09], it was shown that a randomized polynomial-time reconstruction algorithm for an arithmetic circuit class $\mathcal{C}$ implies the existence of a function in BPEXP that does not have polynomial size circuits from $\mathcal{C}$ [9]. Also, a deterministic polynomial-time reconstruction algorithm for $\mathcal{C}$ can be used to construct a function in EXP that does not have polynomial size circuits from $\mathcal{C}$ [Vol16]. Thus, dispensing with non-degeneracy and degree restrictions entirely from Theorem 1 would give a super-polynomial lower bound for general depth three circuits (via the homogenization trick mentioned above). Proving such a lower bound is a long standing open problem in algebraic complexity [SW01, Wig07], the current best being only a nearly cubic lower bound [KST16, BLS16, Yau16].

Nevertheless, it may be possible to relax the non-degeneracy condition by a more careful application of existing techniques and the technique introduced here. It may also be possible to weaken the degree restriction substantially[10].

1.3 Do natural lower bound proofs imply reconstruction?

As reconstruction implies lower bound, research on reconstruction has focused on models for which non-trivial lower bounds are known. Do lower bound proofs, particularly natural lower bound proofs, lead to learning? This question has been asked and investigated before for Boolean circuits by multiple prior work. A natural lower bound proof for a circuit class $\mathcal{C}$ has (as a part of it) a "separator" algorithm that distinguishes functions computable by small $\mathcal{C}$-circuits from other functions efficiently. This is the *constructivity* feature of natural proofs [RR97, FSV17, GKSS17]. It is an intriguing possibility that such a separator has enough structure to imply an efficient learning algorithm for $\mathcal{C}$. Indeed, an interesting result [CIKK16] on Boolean circuits showed that the natural lower proof for $\mathsf{AC}^0[p]$ circuits can be used to give a quasi-polynomial time PAC learning algorithm (with membership queries and under the uniform distribution) for the same class. Prior to this work, similar results were known for AC^0 circuits [LMN93], and AC^0 circuits with poly-logarithmic number of majority gates [JKS02]. We note that the learning algorithms in [CIKK16, LMN93, JKS02] are not proper.

It would be really nice to have such 'natural lower bound to learning algorithm' translations for arithmetic circuit classes. However, there are a few aspects of arithmetic circuits that make this task very demanding. First, we are forced to do exact learning (instead of PAC learning) as two distinct polynomial functions differ at a large fraction of points. Second, there are the issues of homogenization and depth reduction: A $(\frac{n}{d})^{\Omega(d)}$ lower bound is known for homogeneous depth

[9] The result in [FK09] is stated for randomized *zero-error* learning algorithms and hard functions in $\mathsf{ZPEXP}^{\mathsf{RP}}$, but the same argument applies to two-sided error randomized learning and hard functions in the bigger class BPEXP.

[10] There is a $2^{\Omega(n)}$ lower bound known for homogeneous depth three circuits for d up to $2^{o(n)}$ [KST16]. Based on the 'lower bound to learning' theme discussed later, it is conceivable that there is an efficient reconstruction algorithm for non-degenerate homogeneous depth three circuits for d up to $2^{o(n)}$.

three circuits for $d \leq n$ [NW97], and a $2^{\Omega(n)}$ lower bound is known for the same model for any $d \geq n$ [KST16]. These lower bound proofs are natural. But, as mentioned before, efficient reconstruction of homogeneous depth three circuits (particularly, for $d = n^{O(\sqrt{n})}$) implies efficient reconstruction of general depth three circuits via homogenization which further implies subexponential time reconstruction for general circuits via depth reduction. Thus, an unconditional translation from natural lower bound to reconstruction for arbitrary arithmetic circuit classes is probably quite difficult. However, it is conceivable that a natural lower bound framework for an arithmetic circuit class can be used to efficiently reconstruct *almost all* circuits from the class; the notion of 'almost all circuits' is formally captured by a non-degeneracy condition that holds with probability one in the measure theoretic sense. Our work here fits in the theme of 'natural lower bound to learning' (under non-degeneracy condition) and shows that this is indeed true for the class of homogeneous depth circuits [11]. Could this also hold for other circuit classes, like homogeneous depth four circuits or constant depth multilinear circuits, for which exponential natural lower bounds are known? We leave these as open problems.

2 Overview of our techniques

How can a lower bound proof (an impossibility result) lead to an efficient learning algorithm? We begin by first sketching a *weak* implication/connection of this nature. We address rank methods, which have long been recognized as encompassing and abstracting almost all known arithmetic circuit lower bounds to-date. Roughly speaking, we first observe that if a rank method yields a lower bound for circuit class $\mathcal{C}$ then we can do reconstruction given the output polynomial of a circuit $\mathtt{C} \in \mathcal{C}$ *together with an additional information pertaining to the space spanned by some of the internal nodes of* $\mathtt{C}$. This weak connection leads almost immediately to efficient learning of the two subclasses – set-multilinear depth three circuits and depth three powering circuits. However, it falls short of directly yielding an efficient reconstruction algorithm for homogeneous depth three circuits. For this general model, we build on the intuition gained from the weak connection and reduce the problem of finding children of the top addition gate with large fan-in to decomposing a certain vector space U into a (direct) sum of *simpler subspaces*. We then show how to do this decomposition efficiently using a carefully chosen operator space acting on U.

Lower bounds via rank methods. For most arithmetic circuit classes $\mathcal{C}$ for which a lower bound is known, the proof is along the following lines. One first shows that a circuit $\mathtt{C}$ from $\mathcal{C}$ computes a polynomial in the following way:

$$f(\mathbf{x}) = T_1 + T_2 + \ldots + T_s\,, \tag{2}$$

where $T_1, T_2, \ldots, T_s$ are *simple* polynomials (typically simpler in the sense that T_i admits a nontrivial factorization wherein the factors are computed by lower depth subclasses of $\mathcal{C}$). Lower

[11] On a related note, the authors of this work and Nair [KNS18] have given an efficient reconstruction algorithm for low-width (in particular, constant-width) homogeneous algebraic branching programs (ABP), under a non-degeneracy condition that holds with high probability. A modest linear width lower bound is known for homogeneous ABP [Kum17]. If we could drop the non-degeneracy condition from [KNS18] then that would imply (via homogenization) efficient reconstruction for constant-width general ABP, which is exactly the class of arithmetic formulas [BC92]. The results in [FK09] would then give a super-polynomial lower bound for arithmetic formulas, thereby solving a longstanding open problem. This then underscores the difficulty of removing non-degeneracy entirely from the result in [KNS18].

bounds for $\mathcal{C}$ are then tantamount to showing that the number of simple summands s required to express some explicit polynomial $h(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is *large*. Rank-based methods achieve it in the following way: one devises *a linear map* $\mu : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}^{r \times c}$ such that for every simple polynomial T, $\operatorname{rank}(\mu(T))$ is relatively *small* as compared to $\operatorname{rank}(\mu(h))$. Since rank is sub-additive, this yields the lower bound $s \geq \frac{\operatorname{rank}(\mu(h))}{\operatorname{rank}(\mu(T))}$.

From lower bounds to reconstruction - recovering a basis of simple polynomials. For reconstruction of $\mathcal{C}$, it suffices to solve the following problem: given a polynomial f which admits an expression of the form (2), can we efficiently recover the simple summands (i.e. the individual T_i)? We observe that if instead of just one polynomial $f \in \mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s)$, we have an entire basis of $\mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s)$ and if the T_i's are also *μ-independent* in the sense that

$$\rho := \operatorname{rank}(\mu(T_1 + T_2 + \ldots + T_s)) = \operatorname{rank}(\mu(T_1)) + \operatorname{rank}(\mu(T_2)) + \ldots + \operatorname{rank}(\mu(T_s)), \tag{3}$$

then this can be done efficiently. This is the weak connection (or the *elementary approach*) we were alluding to before. Let $M = \mathbb{F}\text{-span}(\mu(T_1), \mu(T_2), \ldots, \mu(T_s)) \subseteq \mathbb{F}^{r \times c}$ be the vector space of matrices spanned by the $\mu(T_i)$. Given a basis of $\mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s)$, we can construct M by applying μ to these basis polynomials and taking their $\mathbb{F}$-span. Now viewing M as the ambient space, the set of matrices in M of rank at most $(\rho - 1)$ forms an algebraic variety which has a particular nice structure when the non-degeneracy condition (3) is satisfied: it is simply the union of s hyperplanes[12]. Moreover, these hyperplanes can be computed efficiently (in $\operatorname{poly}(r \cdot c)$-time) and the equations of these hyperplanes enable us to efficiently recover $T_1, T_2, \ldots, T_s$.

The structure of μ and reduction to a vector space decomposition problem. The elementary approach sketched above for reconstruction from lower bounds requires the knowledge of a basis of $\mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s)$ which we do not have *a priori*. But, we observe that in our case this issue can be alleviated to some extent by the linear map μ that gives us access to a basis of the $\mathbb{F}$-span of another set of simple polynomials[13] $\{T'_1, T'_2, \ldots T'_{s'}\}$ for $s' = \operatorname{poly}(s)$. The map μ used to prove lower bounds for homogeneous depth three circuits has some additional structure that we now describe. It turns out that almost all the known lower bound proofs are via the construction of a finite set of linear operators

$$\mathcal{D} = \{\psi_1, \psi_2, \ldots, \psi_c\}, \quad \text{where each } \psi_i : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$$

is a linear map. The matrix $\mu(f)$ referred to above has c columns, the i-th column simply being the coefficient vector of $\psi_i(f)$ so that the rank of $\mu(f)$ is precisely the dimension of the vector space spanned by $\{\psi_1(f), \psi_2(f), \ldots, \psi_c(f)\}$. For homogeneous depth three circuits, Nisan and Wigderson [NW97] employed the differential operators of a suitable order $k \geq 1$

$$\mathcal{D} := \left\{ \partial^k_{\boldsymbol{\alpha}} \ : \ \boldsymbol{\alpha} \in \mathbb{Z}^n_{\geq 0} \quad \text{and } |\boldsymbol{\alpha}| = k \right\}.$$

When we apply this set of linear operators to the identity (2) we obtain

$$U \subseteq U_1 + U_2 + \ldots + U_s, \quad \text{where } U := \mathbb{F}\text{-span}(\mathcal{D} \circ f) \quad \text{and } U_i := \mathbb{F}\text{-span}(\mathcal{D} \circ T_i).$$

[12] This is captured more precisely in its algebraic version given in Claim 5.1.

[13] that are related to $T_1, T_2, \ldots, T_s$

Now it turns out that generically (i.e. when $T_1, T_2, \ldots, T_s$ are independently chosen products of random linear forms), this containment is in fact an equality and the vector space on the right is in fact a direct sum of the U_i, i.e.

$$U = U_1 \oplus \ldots \oplus U_s \quad \text{or equivalently: } \dim(U) = \dim(U_1) + \dim(U_2) + \ldots + \dim(U_s). \tag{4}$$

Furthermore, in this generic/non-degenerate situation, each U_i has a basis consisting of products of some subset of the linear forms in T_i; we denote such a product by T'_j. Given the output polynomial f, a basis of the space $U = \mathbb{F}\text{-span}(\mathcal{D} \circ f) = \mathbb{F}\text{-span}(T'_1, T'_2, \ldots, T'_{s'})$, where $s' = \text{poly}(s)$, can be efficiently computed so that in the non-degenerate case captured by condition (4), it suffices to *efficiently* compute the above decomposition of U. We now sketch how to do this decomposition.

Decomposing U for the two subclasses of homogeneous depth three circuits. Let us focus on set-multilinear depth three circuits, essentially the same approach works for depth three powering circuits as well. We observe that in implementing the strategy of decomposing U for set-multilinear depth three circuits in Section 5 (but with a slightly different choice of the space of linear operators $\mathcal{D}$), each $U_i = \mathbb{F}\text{-span}(\mathcal{D} \circ T_i)$ that we get is in fact a one-dimensional vector space consisting of scalar multiples of a *simple* polynomial T'_i (simple in the sense of being a product of linear forms). Moreover, in the generic case, these simple polynomials $T'_1, \ldots, T'_s$ are μ-independent (as in Equation (3)). This reduces the problem of decomposing $U = \mathbb{F}\text{-span}(T'_1, T'_2, \ldots, T'_s)$ to the problem of finding a basis of simple polynomials of the space U which can then be tackled using the elementary approach mentioned before. This process of decomposing U can also be viewed as recovering low rank matrices in an operator space $\mathcal{S}$ acting on U.

Decomposing U in the general case. For homogeneous depth three circuits, each U_i will generically have a basis of simple polynomials so that U itself has a basis of simple polynomials $\{T'_1, T'_2, \ldots, T'_{s'}\}$. Unfortunately however, $T'_1, T'_2, \ldots, T'_{s'}$ do not satisfy the μ-independence condition given by Equation (3) and hence we cannot directly apply the elementary approach outlined before to decompose U. This is the main difficulty that we show how to handle in Section 4. The new idea behind the decomposition of U is to choose a space $\mathcal{S}$ of linear operators acting on U such that the non-degeneracy condition implies that U_i is an invariant subspace of U induced by $\mathcal{S}$ for every $i \in [s]$. Then, certain 'nice properties' of $\mathcal{S}$ ensure that $U_1, \ldots, U_s$ are in fact the only irreducible invariant subspaces of U induced by $\mathcal{S}$, and bases of these subspaces can be found efficiently from a basis of U by computing $\mathcal{S}$-closures of appropriately chosen vectors in U. Our choice of $\mathcal{S}$ is the *shifted differential operator space* (which is defined in Section 4.2), and the main technical work involves showing that this $\mathcal{S}$ has the required 'nice properties' (as stated in Claim 4.2 and 4.3) that aid in the decomposition of U into irreducible invariant subspaces.

Summary. There are many subclasses of circuits which admit lower bounds via rank methods. But the presence of addition gates of large fan-in in such circuits had hitherto made efficient reconstruction look unapproachable. The main conceptual novelty/contribution of this work is to formulate the following paradigm for efficiently handling addition gates and to successfully implement it for the class of homogeneous depth three circuits. The problem of finding the children of an addition gate with large fan-in s is first reduced to the problem of finding a decomposition of a suitable vector space U into a (direct) sum of *simpler* subspaces $U_1, U_2, \ldots, U_s$. One then constructs a suitable space of operators $\mathcal{S}$ consisting of linear maps acting on U such that analyzing the *simultaneous global structure*[14] of $\mathcal{S}$ enables us to efficiently decompose U. We feel that this paradigm is novel and powerful: by enabling us to handle large addition gates, it should lead to efficient reconstruction of many other such subclasses of circuits. In implementing this paradigm, our conceptual contribution is the construction of the particular $\mathcal{S}$ that works for the class of homogeneous depth three circuits.

3 Basic defintions and facts

We record a few basic definitions and facts which will be used in the arguments later. A n-variate polynomial of degree-d will be called a (n, d) polynomial.

Fact 1 (Computing partial derivatives)**.** *Given black-box access to a (n, d) polynomial f and a monomial $\mathbf{x}^{\boldsymbol{\alpha}}$, a black-box access to $\partial^k_{\boldsymbol{\alpha}} f$ can be computed in deterministic* $\text{poly}(n, d^k)$ *time.*

This follows from the fact that black-box access to a first-order derivative of f can be computed in deterministic polynomial time from black-box access to f.

Fact 2 (Space of linear dependencies)**.** *Given black-box access to (n, d) polynomials $g_1, \ldots, g_r$, a basis of the space $G^{\perp} := \{(c_1, \ldots, c_r) \in \mathbb{F}^r \ : \ \sum_{i \in [r]} c_i g_i = 0\}$ can be computed in randomized* $\text{poly}(n, d, r)$ *time. As a corollary, we have that black-box access to the elements of a basis of* $\mathbb{F}\text{-span}\{g_1, \ldots, g_r\}$ *can be computed in randomized* $\text{poly}(n, d, r)$ *time from black-box access to $g_1, \ldots, g_r$.*

The above can be proved by applying the Schwartz-Zippel lemma [Sch80,Zip79] and reducing the problem to solving a system of linear equations over $\mathbb{F}$.

Definition 3.1 (Invariant subspace)**.** Let U be a vector space and $\mathcal{S}$ a space of linear operators on U. A subspace $V \subseteq U$ is called an *invariant subspace* of U induced by $\mathcal{S}$ if $\mathcal{S}V \subseteq V$. An invariant subspace $V \neq 0$ is *irreducible* if V cannot be expressed as $V = V_1 \oplus V_2$, where V_1 and V_2 are invariant subspaces properly contained in V.

Definition 3.2 (Closure of a vector)**.** Let U be a vector space and $\mathcal{S}$ a space of linear operators on U. The closure of a vector $\mathbf{v} \in U$ with respect to $\mathcal{S}$ is the smallest invariant subspace of U induced by $\mathcal{S}$ that contains $\mathbf{v}$.

The following fact states that the closure of a vector can be computed efficiently. Suppose U is a vector space of dimension m over $\mathbb{F}$. Once we fix a basis of U, it can be identified with $\mathbb{F}^m$ and a linear operator on U with a matrix in $\mathbb{F}^{m \times m}$. Thus, a basis of a space $\mathcal{S}$ of linear operators on U can be given as a list of matrices $\{M_1, \ldots, M_t\}$ in $\mathbb{F}^{m \times m}$.

Fact 3 (Computing closure of a vector)**.** *Given a $\mathbf{v} \in \mathbb{F}^m$ and a list of matrices $\{M_1, \ldots, M_t\}$ in $\mathbb{F}^{m \times m}$, the closure of $\mathbf{v}$ with respect to* $\mathbb{F}\text{-span}\{M_1, \ldots, M_t\}$ *can be computed in deterministic* $\text{poly}(m)$ *time.*

We refer the reader to Algorithm 4 in [KNST17] for a proof of the above fact.

[14] In this case, it is the structure of the set of low rank matrices in $\mathcal{S}$ and of invariant subspaces of U induced by $\mathcal{S}$. There is a high-level similarity between our approach for reconstruction of homogeneous depth three circuits and the approach in [KNST17] for reconstruction of full rank algebraic branching programs. In [KNST17], the ambient space $\mathbb{F}^n$ is decomposed into irreducible invariant subspaces of the Lie algebra (which is a certain space of linear operators on $\mathbb{F}^n$) of the iterated matrix multiplication polynomial, whereas in this work, the partial derivative space U is decomposed into irreducible invariant subspaces of a shifted differential operator space (which is a space of linear operators on U).

4 Homogeneous depth three circuits: Proof of Theorem 1

We follow the outline given in Section 2 for reconstruction of homogeneous depth three circuits and fill in the details here.

The algorithm. We state the algorithm and argue the correctness and complexity of its steps. As in the statement of Theorem 1, assume that $n \geq (3d)^2$ and $s \leq (\frac{n}{3d})^{\frac{d}{3}}$, and let $m := s \cdot \binom{d}{k}$. We record a few relations among the parameters n, d, s and k in the following easy to verify remark.

Remark 1. *If $n \geq (3d)^2$, $s \leq (\frac{n}{3d})^{\frac{d}{3}}$ and $k = \left\lceil \frac{\log s}{\log \frac{n}{ed}} \right\rceil$ then the following relations hold: $k \leq \frac{d}{3} + 1$, $s \leq (\frac{n-2k-1}{d})^{k+1}$, $\binom{n+k-1}{k} \geq s \cdot \binom{d}{k}$, $d^k \leq d \cdot s$ and $n^k \leq n \cdot s^2$.*

Clearly, $m \leq d \cdot s^2$. In this section, we will assume that a basis of a vector space is an <u>ordered</u> set.

Algorithm 1 Reconstruction of non-degenerate homogeneous depth three circuits

```
Input: Black-box access to a $f$ that is computed by a non-degenerate $(n, d, s)$ homogeneous depth three circuit $T_1 + \ldots + T_s$.
Output: A non-degenerate $(n, d, s)$ homogeneous depth three circuit computing $f$.
1. Compute black-box access to a basis of $U$, the $k$-th order partial derivative space of $f$.
2. Decompose $U = U_1 \oplus \ldots \oplus U_s$, i.e., compute black-box access to elements of bases of $U_1, \ldots, U_s$, the $k$-th order partial derivative spaces of $T_1, \ldots, T_s$, using Algorithm 2.
3. Obtain the terms $T_1, \ldots, T_s$ from the bases of $U_1, \ldots, U_s$.
```

4.1 Step 1: Computing a basis of U

Observation 4.1. *From black-box access to f, we can compute black-box access to the elements of a basis $\Gamma = (g_1, g_2, \ldots, g_m)$ of U in randomized* $\text{poly}(n, s)$ *time.*

Proof. The number of k-th order derivatives in n variables is $\binom{n+k-1}{k} \leq n \cdot s^2$, by Remark 1. Using Fact 1, we get black-box access to all the k-th order derivatives of f in $\text{poly}(n, d^k, s) = \text{poly}(n, s)$ time (by Remark 1 again). A basis of the derivatives can then be computed in randomized $\text{poly}(n, s)$ time using Fact 2. □

4.2 Step 2: Decomposing $U = U_1 \oplus \ldots \oplus U_s$

Notations. Let us fix a few notations and terminologies.

- $\text{Norm}(k) := \{\boldsymbol{\alpha} \in \mathbb{Z}^n_{\geq 0} : |\boldsymbol{\alpha}| = k\}$.
- For a polynomial $T \in \mathbb{F}[\mathbf{x}]$, $\text{coeff}_{\boldsymbol{\beta}}(T)$ is the coefficient of the monomial $\mathbf{x}^{\boldsymbol{\beta}}$ in T.
- Recall $T_i = \ell_{i1} \cdot \ell_{i2} \ldots \ell_{id}$ is a product of d linear forms. Let

$$T_{iA} := \prod_{j \in A} \ell_{ij}, \qquad \text{for } A \in \binom{[d]}{k},$$

$$T_{iB} := \prod_{j \in B} \ell_{ij}, \qquad \text{for } B \in \binom{[d]}{d-k}.$$

- The set $\Lambda := \left\{T_{iB} \ : \ i \in [s] \text{ and } B \in \binom{[d]}{d-k}\right\}$ is the *canonical basis* of U. Think of Λ as an ordered set in which the elements $\{T_{iB} \ : \ B \in \binom{[d]}{d-k}\}$, which is a basis of U_i, precede the elements $\{T_{i+1\ B} : B \in \binom{[d]}{d-k}\}$, which is a basis of U_{i+1}, for every $i \in [s-1]$.
- Let $\Gamma = (g_1, g_2, \ldots, g_m)$ be a basis of U and $\mathcal{S}$ a space of linear operators on U. Then, U can be naturally identified with $\mathbb{F}^m$, a subspace $U_i \subseteq U$ can be identified with a subspace $U_{i,\Gamma} \subseteq \mathbb{F}^m$ and an operator $\psi \in \mathcal{S}$ can be identified with a matrix $M_\Gamma(\psi) \in \mathbb{F}^{m\times m}$. Conversely, a subspace $W \subseteq \mathbb{F}^m$ can be identified with a subspace $\Gamma \cdot W \subseteq U$, which consists of all polynomials $(g_1\ g_2\ \ldots g_m)\cdot \mathbf{w}$ for $\mathbf{w} \in W$. Observe that $\Gamma \cdot U_{i,\Gamma} = U_i$.
- For $B, B' \in \binom{[d]}{d-k}$, the distance between them is defined as $\text{dist}(B, B') := (d-k) - |B \cap B'|$.

The shifted differential operator space. The following operator space $\mathcal{SD}_{k,U}$ plays a vital role in our algorithm and its analysis. Let

$$\begin{aligned}\mathcal{SD}_k &:= \mathbb{F}\text{-span}\left\{\mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \ : \ \boldsymbol{\alpha}, \boldsymbol{\beta} \in \text{Norm}(k)\right\}, \text{ and}\\ \mathcal{SD}_{k,U} &:= \{\psi \in \mathcal{SD}_k \ : \ \psi(U) \subseteq U\}.\end{aligned}$$

Observe that $\dim(\mathcal{SD}_k) = \binom{n+k-1}{k}^2 = \text{poly}(n,s)$, by Remark 1, and $\mathcal{SD}_{k,U}$ is a vector space over $\mathbb{F}$. The shifted differential operators in $\mathcal{SD}_k$ act linearly on polynomials and hence $\mathcal{SD}_{k,U}$ is a space of linear operators on U. We will refer to $\mathcal{SD}_{k,U}$ as $\mathcal{S}$ for brevity.

Observation 4.2. *We can compute a basis $(\psi_1, \ldots, \psi_t)$ of $\mathcal{S}$ in randomized* $\text{poly}(n,s)$ *time from black-box access to the elements of a basis $\Gamma = (g_1, g_2, \ldots, g_m)$ of U.* [15]

Proof. An operator $\sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \text{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}$, where $c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathbb{F}$, is in $\mathcal{S}$ if and only if the following holds for every $i \in [m]$.

$$\sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \text{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}\, g_i = \sum_{j\in[m]} d_{ij} \cdot g_j, \quad \text{for some } d_{ij} \in \mathbb{F}.$$

From black-box access to g_i, we get black-box access to $\mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}\, g_i$, for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \text{Norm}(k)$, in $\text{poly}(n,s)$ time using Fact 1 and Remark 1. Then, we compute a basis of the space $G_i^\perp$ of linear dependencies of the polynomials $\{\mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}\, g_i \ : \ \boldsymbol{\alpha}, \boldsymbol{\beta} \in \text{Norm}(k)\} \cup \{g_j \ : \ j \in [m]\}$ in randomized $\text{poly}(n,s)$ time using Fact 2. By treating $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and d_{ij} as formal variables, the coordinates of a vector in $G_i^\perp$ can be naturally indexed by these variables. Restrict the vectors in the basis of $G_i^\perp$ to coordinates indexed by the $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ variables and call the space spanned by these restricted vectors C_i. It is easy to observe that a basis of $C_1 \cap \ldots \cap C_m$ gives a basis of $\mathcal{S}$ directly. □

Henceforth, the correctness of the decomposition step proceeds by proving three important claims – Claim 4.1, 4.2 and 4.3. We state these claims below and prove them in Section 6.

Three claims and two corollaries.

Claim 4.1. *For every $i \in [s]$, U_i is an invariant subspace of U induced by $\mathcal{S}$.*

[15] The exact expression for t, as a function of n and s, is not relevant for our analysis.

The claim is proved using the non-degeneracy condition. The next two claims would help us infer that U_i is in fact an irreducible invariant subspace of U induced by $\mathcal{S}$ and a basis of it can be computed in polynomial time using Algorithm 2. The proofs of these two claims also use the non-degeneracy condition, in particular the fact that $\dim(U) = s \cdot \binom{d}{k}$.

Claim 4.2. *There is an operator $\psi \in \mathcal{S}$ such that*

$$\psi(T_{iB}) = \rho_{iB} \cdot T_{iB}, \quad \textit{where } \rho_{iB} \in \mathbb{F},$$

for every $i \in [s]$ and $B \in \binom{[d]}{d-k}$. Furthermore, the field elements $\{\rho_{iB} : i \in [s] \text{ and } B \in \binom{[d]}{d-k}\}$ are distinct.

The claim implies that there is a $\psi \in \mathcal{S}$ such that $M_\Lambda(\psi)$ is a diagonal matrix with distinct diagonal entries. As $M_\Lambda(\psi)$ and $M_\Gamma(\psi)$ are similar matrices, we have the following corollary.

Corollary 4.1. *There is a $\psi \in \mathcal{S}$ such that $M_\Gamma(\psi)$ has distinct eigenvalues, where Γ is a basis of U.*

Claim 4.3. *For every $i \in [s]$ and $B, B' \in \binom{[d]}{d-k}$ with $\operatorname{dist}(B, B') = 1$, there is a $\psi \in \mathcal{S}$ such that*

$$\psi(T_{iB}) = \rho_{iBB'} \cdot T_{iB'}, \quad \textit{where } \rho_{iBB'} \in \mathbb{F} \setminus \{0\}.$$

The above three claims imply the following corollary, which we prove in Section 6.

Corollary 4.2. *For every $i \in [s]$ and $\mathbf{v} \in U_i$, the closure of $\mathbf{v}$ with respect to $\mathcal{S}$ is U_i. So, $U_1, \ldots, U_s$ are irreducible invariant subspaces of U induced by $\mathcal{S}$. Further, these are the only irreducible invariant subspaces of U induced by $\mathcal{S}$.*

Algorithm 2 Finding bases of $U_1, \ldots, U_s$

```
Input: Black-box access to the elements of a basis $\Gamma = (g_1, g_2, \ldots, g_m)$ of $U$.
Output: Black-box access to the elements of bases of $U_1, \ldots, U_s$.
1. Compute a basis $(\psi_1, \ldots, \psi_t)$ of $\mathcal{S}$ from $\Gamma$ using Observation 4.2. Derive the $m \times m$ matrices $M_\Gamma(\psi_1), \ldots, M_\Gamma(\psi_t)$.
2. Pick a random $M_\Gamma = r_1 \cdot M_\Gamma(\psi_1) + \ldots + r_t \cdot M_\Gamma(\psi_t)$, where $r_1, \ldots, r_t$ are chosen independently and uniformly at random from $[m^3]$.
3. Compute the characteristic polynomial $h(y)$ of $M_\Gamma$. Output 'Fail' if $h$ is not square-free. Otherwise, factorize $h = h_1 \cdot h_2 \ldots h_l$ into irreducible factors over $\mathbb{F}$.
4. Find bases of the null spaces $N_1, \ldots, N_l$ of $h_1(M_\Gamma), \ldots, h_l(M_\Gamma)$ respectively.
5. For every $j \in [l]$, pick a vector $\mathbf{v}$ from the basis of $N_j$ and compute the closure of $\mathbf{v}$ with respect to $\mathbb{F}$-span$\{M_\Gamma(\psi_1), \ldots, M_\Gamma(\psi_t)\}$ (using Fact 3).
6. Let $\{W_1, \ldots, W_p\}$ be the list of the closure spaces after removing repetitions. If $p \neq s$, return 'Fail'. Else, return black-box access to the bases of $\{\Gamma \cdot W_1, \ldots, \Gamma \cdot W_s\}$.
```

Analysis of Algorithm 2.

Steps 1 and 2 of the algorithm are self-explanatory. In step 3, the characteristic polynomial $h(y)$ can be computed in deterministic poly(m) time using interpolation.

Observation 4.3. *Polynomial $h(y)$ is square-free with probability at least $1 - \frac{2}{m}$.*

Proof. As h is a monic polynomial, the resultant $\mathrm{res}_y(h, \frac{\partial h}{\partial y})$ is non-zero if and only if h is square-free. For a moment, think of $r_1, \ldots, r_t$ as formal variables in step 2. Then, h is a monic polynomial whose coefficients are polynomials in $r_1, \ldots, r_t$ of degree at most m, implying that the degree of $\mathrm{res}_y(h, \frac{\partial f}{\partial y})$ as a polynomial in $r_1, \ldots, r_t$ is at most $2m^2$. By Corollary 4.1, $\mathrm{res}_y(h, \frac{\partial f}{\partial y})$ is a non-zero polynomial in $r_1, \ldots, r_t$, and hence an application of the Schwartz-Zippel lemma implies that h is square-free with probability at least $1 - \frac{2}{m}$ over the random choices of $r_1, \ldots, r_t$ in step 2. □

We have assumed that univariate degree-m polynomials over $\mathbb{F}$ can be factored into irreducible factors in $\mathrm{poly}(m)$ time. So, step 3 can be executed in $\mathrm{poly}(n, s)$ time. Computing a basis of the null space of a matrix reduces to solving a system of linear equations and hence, step 4 also takes $\mathrm{poly}(n, s)$ time.

Claim 4.4. *At step 5, the following holds : For every $j \in [l]$, $N_j \subseteq U_{i,\Gamma}$ for some $i \in [s]$. Also, for every $i \in [s]$, there is a $j \in [l]$ such that $N_j \subseteq U_{i,\Gamma}$.*

Assuming the claim, it follows from Corollary 4.2 that the set of spaces $\{W_1, \ldots, W_s\}$ is in fact $\{U_{1,\Gamma}, \ldots, U_{s,\Gamma}\}$ in step 6. The correctness of the algorithm follows by noting that $\Gamma \cdot U_{i,\Gamma} = U_i$.

Proof. Recall that Λ is the canonical basis of U. Let $D \in \mathrm{GL}_m(\mathbb{F})$ be the basis-change matrix from Γ to Λ, i.e. $M_\Gamma(\psi) = D^{-1} \cdot M_\Lambda(\psi) \cdot D$ for every operator $\psi \in \mathcal{S}$.

$$M_\Gamma = \sum_{i \in [t]} r_i \cdot M_\Gamma(\psi_i) = D^{-1} \cdot \left(\sum_{i \in [t]} r_i \cdot M_\Lambda(\psi_i) \right) \cdot D = D^{-1} \cdot M_\Lambda \cdot D,$$

where $M_\Lambda := \sum_{i \in [t]} r_i \cdot M_\Lambda(\psi_i)$, implying that the characteristic polynomial of M_Λ is $h(y)$. By Claim 4.1, M_Λ is a block-diagonal matrix,

$$M_\Lambda = \begin{bmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_s \end{bmatrix},$$

where each R_i is a $\binom{d}{k} \times \binom{d}{k}$ matrix. Let $p_i(y)$ be the characteristic polynomial of R_i. Then, $h = p_1 \cdot p_2 \cdots p_s$, and the polynomial h_j (for j as in step 5) divides p_i for some $i \in [s]$.

Suppose $\mathbf{v} \in N_j$. Then, $h_j(M_\Gamma) \cdot \mathbf{v} = 0$ implying $p_i(M_\Gamma) \cdot \mathbf{v} = 0$ and hence $p_i(M_\Lambda) \cdot D\mathbf{v} = 0$, where

$$p_i(M_\Lambda) = \begin{bmatrix} p_i(R_1) & & & \\ & p_i(R_2) & & \\ & & \ddots & \\ & & & p_i(R_s) \end{bmatrix}.$$

Split the column vector $D\mathbf{v} \in \mathbb{F}^m$ into s continuous chunks of size $\binom{d}{k}$ each, and call the vector defined by the q-th chunk $\mathbf{v}_q \in \mathbb{F}^{\binom{d}{k}}$ for $q \in [s]$. Thus, we have $p_i(R_q) \cdot \mathbf{v}_q = 0$ for every $q \in [s]$.

Observe that p_i and p_q are coprime for $i \neq q$, as h is square-free at step 5. Hence, for $i \neq q$, there exist two polynomials $e_1(y), e_2(y) \in \mathbb{F}[y]$ such that

$$\begin{aligned} e_1(y) \cdot p_i(y) + e_1(y) \cdot p_q(y) &= 1, \\ \Rightarrow e_1(R_q) \cdot p_i(R_q) &= I_m \quad \text{(the } m \times m \text{ identity matrix), as } p_q(R_q) = 0, \\ \Rightarrow e_1(R_q) \cdot p_i(R_q) \cdot \mathbf{v}_q &= \mathbf{v}_q, \\ \Rightarrow \mathbf{v}_q &= 0. \end{aligned}$$

Therefore, $D\mathbf{v} \in U_{i,\Lambda}$ implying $\mathbf{v} \in U_{i,\Gamma}$. The second part of the claim follows easily from the above proof. □

4.3 Step 3: Obtaining the terms $T_1, \ldots, T_s$ from bases of $U_1, \ldots, U_s$

From step 2 of Algorithm 1, we have black-box access to the elements of a basis $(g_{i1}, \ldots g_{i\binom{d}{k}})$ of U_i for every $i \in [s]$. Now, for every $\boldsymbol{\alpha} \in \text{Norm}(k)$, we solve for $c_{\boldsymbol{\alpha}\ i1}, \ldots, c_{\boldsymbol{\alpha}\ i\binom{d}{k}} \in \mathbb{F}$ such that

$$\sum_{i \in [s]} c_{\boldsymbol{\alpha}\ i1} \cdot g_{i1} + \ldots + c_{\boldsymbol{\alpha}\ i\binom{d}{k}} \cdot g_{i\binom{d}{k}} = \partial^k_{\boldsymbol{\alpha}} f.$$

This can be done efficiently using Fact 2. Observe that such a solution satisfies

$$c_{\boldsymbol{\alpha}\ i1} \cdot g_{i1} + \ldots + c_{\boldsymbol{\alpha}\ i\binom{d}{k}} \cdot g_{i\binom{d}{k}} = \partial^k_{\boldsymbol{\alpha}} T_i,$$

for every $i \in [s]$, as $U = U_1 \oplus \ldots \oplus U_s$. A black-box for T_i can be obtained from the following classical, easy-to-verify identity that holds for any degree-$(d-k)$ homogeneous polynomial T_i:

$$T_i = \frac{(d-2k)!}{(d-k)!} \cdot \sum_{\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n) \in \text{Norm}(k)} \binom{k}{\alpha_1, \ldots, \alpha_n} \cdot \mathbf{x}^{\boldsymbol{\alpha}} \cdot \partial^k_{\boldsymbol{\alpha}} T_i.$$

Finally, we obtain T_i by an application of the black-box polynomial factorization algorithm [KT90].

5 Set-multilinear depth three circuits: Proof of Theorem 2

We follow the outline and the notation given in Section 2 and in this section, we fill in some more details for the subclasses of homogeneous depth three circuits corresponding to tensors and depth three powering circuits. Following the discussion in Section 2, we formulate precisely the problem of recovering a simple basis of a given space of polynomials. For a moment, let us not worry about the exact representation of polynomials (which will be black-box in the subsequent applications). Let us also ignore the size of the matrix associated with the linear map μ which need not be poly(n, d, s). We will see how to handle these issues in the two subsequent applications.

Problem 1. Finding a simple basis. Let $\mu : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}^{r \times c}$ be a linear map. Given (a basis of) the vector space of polynomials

$$V = \mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s),$$

where each T_i is a simple polynomial in the sense that rank$(\mu(T_i))$ is *small* can we recover $T_1, T_2, \ldots, T_s$ (up to scalar multiples)?

We show how this can be done efficiently under the following non-degeneracy condition.

Definition 5.1. Let $\mu : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}^{r \times c}$ be a linear map. We say that a set of polynomials $\{T_1, T_2, \ldots, T_s\}$ are *μ-independent* if

$$\rho := \operatorname{rank}(\mu(T_1 + T_2 + \ldots + T_s)) = \operatorname{rank}(\mu(T_1)) + \operatorname{rank}(\mu(T_2)) + \ldots + \operatorname{rank}(\mu(T_s)). \tag{5}$$

Clearly, if $\{T_1, T_2, \ldots, T_s\}$ are *μ-independent* then they are also $\mathbb{F}$-linearly independent.

```
Algorithm 3 Reconstruction of individual terms from their $\mathbb{F}$-span.
  Input: A basis $\mathbf{f} = (f_1, f_2, \ldots, f_s)$ of the vector space $V = \mathbb{F}\text{-span}(T_1, T_2, \ldots, T_s)$.
  Output: The vector $(T_1, T_2, \ldots, T_s)$ (up to permutation and scalar multiples).
1. Compute $M_i := \mu(f_i)$. Form the matrix $M(\mathbf{z}) := z_1 \cdot M_1 + z_2 \cdot M_2 + \ldots + z_s \cdot M_s$.
2. Let $\rho$ be the rank of $M(\mathbf{z})$ for a random choice of $\mathbf{z}$. Compute $g(\mathbf{z}) := \gcd(\text{Minors}(M(\mathbf{z})), \rho)$
   and factor it using the algorithm in [KT90].
3. If $g(\mathbf{z})$ does not factor into a power of products of $s$ linear forms, abort. Else let
       $g(\mathbf{z}) := \prod_{j=1}^{s} \ell_j(\mathbf{z})^{r_j}$
   be the factorization of $g(\mathbf{z})$. Let $A = (a_{ij})_{i,j \in [s]} \in \mathbb{F}^{s \times s}$, where $a_{ij}$ is the coefficient of $z_i$ in $\ell_j$.
   Output $\mathbf{f} \cdot A^{-1}$.
```

The correctness of the algorithm follows almost immediately from the following claim.

Claim 5.1. *Let $\tilde{M}_i = \mu(T_i)$ have rank $r_i \in \mathbb{Z}_{\geq 1}$. Let $\tilde{M}(\mathbf{u}) := u_1 \cdot \tilde{M}_1 + u_2 \cdot \tilde{M}_2 + \ldots + u_s \cdot \tilde{M}_s$ and $\rho = r_1 + r_2 + \ldots + r_s$. If the non-degeneracy condition (5) is satisfied then*

$$\gcd(\text{Minors}(\tilde{M}), \rho) = u_1^{r_1} \cdot u_2^{r_2} \cdot \ldots \cdot u_s^{r_s}.$$

In step 2, we are computing the gcd of all $\rho \times \rho$ minors of $M(\mathbf{z})$. This can be done efficiently by choosing two random matrices $R_1 \in \mathbb{F}^{\rho \times r}$ and $R_2 \in \mathbb{F}^{c \times \rho}$ and computing the determinant of $R_1 \cdot M(\mathbf{z}) \cdot R_2$. Indeed, $\gcd(\text{Minors}(M(\mathbf{z})), \rho) = \det(R_1 \cdot M(\mathbf{z}) \cdot R_2)$ with high probability.

5.1 Application: tensor decomposition

We now specify the choice of parameters mentioned in Section 2 as applicable to this setting. Let $k \leq \frac{d}{3}$ be an integer and denote the union of the first k sets of variables by $\mathbf{y}$, i.e.

$$\mathbf{y} := \mathbf{x}_1 \uplus \mathbf{x}_2 \uplus \ldots \uplus \mathbf{x}_k.$$

We denote the space of operators that we choose as $\mathcal{D}_\mathbf{y}$ and it is the following set of differential operators of order k supported only over the variable set $\mathbf{y}$, i.e.

$$\mathcal{D}_\mathbf{y} := \left\{ \frac{\partial^k}{\prod_{y \in A} \partial_y} \ : \ A \subseteq \mathbf{y}, \ \text{and} \ |A \cap \mathbf{x}_i| = 1 \ \text{for all} \ i \in [k] \right\}.$$

In this case, each $\mathbb{F}$-span$(\mathcal{D}_{\mathbf{y}} \circ T_i)$ has dimension one so that $\mathbb{F}$-span$(\mathcal{D}_{\mathbf{y}} \circ f)$ has dimension at most s. To see this, for a term T_i of the form (1), let us define

$$P_i := \prod_{j \in [k]} \ell_{ij} \quad \text{and } Q_i := \prod_{j \in [d] \setminus [k]} \ell_{ij} \quad \text{so that} \quad T_i = P_i(\mathbf{y}) \cdot Q_i(\mathbf{x} \setminus \mathbf{y}). \tag{6}$$

Each $\mathbb{F}$-span$(\mathcal{D}_{\mathbf{y}} \circ T_i)$ is then simply $\mathbb{F}$-span(Q_i) so that we have:

Observation 5.1. *For $f = T_1 + T_2 + \ldots + T_s$, where T_i is of the form (6), it holds that $\mathbb{F}$-span$(\mathcal{D}_{\mathbf{y}} \circ f) \subseteq$ $\mathbb{F}$-span$(Q_1, Q_2, \ldots, Q_s)$.*

If $k \le \frac{d}{3}$ is the smallest integer such that $(\frac{n}{d})^k \ge s$ then this containment is in fact an equality generically. Specifically, we have:

Proposition 5.1. *Let k be chosen as above. For $i \in [s]$ and $j \in [k]$, if the forms ℓ_{ij} are chosen independently at random then with high probability it holds that $Q_1, Q_2, \ldots, Q_s$ are $\mathbb{F}$-linearly independent and*

$$\mathbb{F}\text{-span}(\mathcal{D}_{\mathbf{y}} \circ f) = \mathbb{F}\text{-span}(Q_1, Q_2, \ldots, Q_s).$$

We can compute a basis $(f_1, f_2, \ldots, f_s)$ of $\mathbb{F}$-span$(\mathcal{D}_{\mathbf{y}} \circ f)$ using Fact 1, and then call Algorithm 3 on this basis. The polynomials $Q_1, Q_2, \ldots, Q_s$, which are μ-independent in the generic case, play the role of $T_1, T_2, \ldots, T_s$ in the algorithm. The linear map μ is as follows: let $\mathbf{y}' := \mathbf{x}_{k+1} \uplus \ldots \uplus \mathbf{x}_{2k}$ and define $\mathcal{D}_{\mathbf{y}'}$ just like $\mathcal{D}_{\mathbf{y}}$. Suppose $\psi_1, \ldots, \psi_c$ be the $c = (\frac{n}{d})^k$ operators in $\mathcal{D}_{\mathbf{y}'}$. Then, $\mu(f)$ is a matrix with c columns, the j-th column being the coefficient vector of $\psi_j(f)$. Observe that $\rho = s$. Instead of working with $\mu(f)$ in Algorithm 3 which is a large matrix, we work with the $\rho \times c$ matrix $\tilde{\mu}(f) := (\psi_j(f)(\mathbf{a}_i))_{i \in [\rho], j \in [c]}$ for randomly chosen points $\mathbf{a}_1, \ldots, \mathbf{a}_\rho \in \mathbb{F}^n$. This also addresses the fact that f is given as a black-box. The algorithm returns $Q_1, \ldots, Q_s$ (up to scaling) which we factor into products of linear forms. The linear forms in P_i can be similarly obtained by interchanging the roles of $\mathcal{D}_{\mathbf{y}}$ and $\mathcal{D}_{\mathbf{y}'}$ and calling Algorithm 3 again. This process gives us $T_1, \ldots, T_s$ up to scaling; the scaling factors can then be retrieved by solving a system of linear equations.

5.2 Application: Depth three powering circuits

In a depth three powering circuit each term $T_i = \ell_i(\mathbf{x})^d$, i.e.,

$$f = \ell_1(\mathbf{x})^d + \ell_2(\mathbf{x})^d + \ldots + \ell_s(\mathbf{x})^d, \tag{7}$$

where each $\ell_i(\mathbf{x})$ is a linear form. These circuits correspond to symmetric tensors and the approach for reconstructing general tensors carry over in a similar fashion. In particular, let $k \le (d/3)$ and $e = (d - k)$. Let $\mathcal{D}$ be the space of differential operators of order k as defined in Section 2.

Observation 5.2. *If f is of the form (7) then $\mathbb{F}$-span$(\mathcal{D} \circ f) \subseteq \mathbb{F}$-span$(\ell_1^e, \ell_2^e, \ldots, \ell_s^e)$.*

If $k \le \frac{d}{3}$ is the smallest integer such that $\binom{n+k-1}{n} \ge s$ then this containment is in fact an equality generically. Specifically,

Proposition 5.2. *Let k be chosen as above. For $i \in [s]$, if the forms ℓ_i are chosen independently at random then with high probability it holds that $\ell_1^e, \ell_2^e, \ldots, \ell_s^e$ are $\mathbb{F}$-linearly independent and*

$$\mathbb{F}\text{-span}(\mathcal{D} \circ f) = \mathbb{F}\text{-span}(\ell_1^e, \ell_2^e, \ldots, \ell_s^e).$$

The remaining argument is very similar to that of tensor decomposition after we replace the set of operators $\mathcal{D}_{\mathbf{y}}$ and $\mathcal{D}_{\mathbf{y}'}$ by $\mathcal{D}$.

6 Proofs of three claims and a corollary from Section 4

6.1 Proof of Claim 4.1

Fix a $i \in [s]$, a $B \in \binom{[d]}{d-k}$ and an operator $\psi \in \mathcal{S}$ arbitrarily. It is sufficient to show that $\psi(T_{iB}) \in U_i$. Let $\psi = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}$, where $c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \in \mathbb{F}$. As $\psi(U) \subseteq U$, we have

$$\psi(T_{iB}) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \, T_{iB} = \sum_{j\in[s],\, C\in\binom{[d]}{d-k}} \gamma_{jC} \cdot T_{jC}, \quad \text{for some } \gamma_{jC} \in \mathbb{F}. \tag{8}$$

The non-degeneracy condition states that there are $2k+1$ linear forms $\ell_{ir_1}, \ldots, \ell_{ir_{2k+1}}$ such that

$$\dim\left(\sum_{j\in[s]\setminus\{i\}} U_j \mod \langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle\right) = (s-1)\cdot\binom{d}{k}.$$

Observe that the expression $\partial^k_{\boldsymbol{\alpha}} \, T_{iB}$ in Equation (8) is zero modulo $\langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle$. Thus, we get

$$\sum_{j\in[s]\setminus\{i\},\, C\in\binom{[d]}{d-k}} \gamma_{jC} \cdot T_{jC} = 0 \mod \langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle.$$

The non-degeneracy condition implies that $\{T_{jC} \mod \langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle : j \in [s] \setminus \{i\}, C \in \binom{[d]}{d-k}\}$ is a basis of the space $\sum_{j\in[s]\setminus\{i\}} U_j \mod \langle \ell_{ir_1}, \ldots, \ell_{ir_{2k+1}} \rangle$. Hence, $\gamma_{jC} = 0$ for all $j \in [s] \setminus \{i\}$ and $C \in \binom{[d]}{d-k}$. Therefore,

$$\psi(T_{iB}) = \sum_{C\in\binom{[d]}{d-k}} \gamma_{iC} \cdot T_{iC} \ \in \ U_i.$$

6.2 Proof of Claim 4.2

Let $\psi = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}$ be an operator in $\mathcal{S}$. Let us treat $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ as formal variables that take values from $\mathbb{F}$. For every $i \in [s]$ and $B \in \binom{[d]}{d-k}$,

$$\begin{aligned}
\psi(T_{iB}) &= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \, T_{iB} \\
&= \sum_{\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{\boldsymbol{\alpha} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \, T_{iB} \\
&= \sum_{\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{\boldsymbol{\alpha} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \sum_{A\in\binom{B}{k}} (\partial^k_{\boldsymbol{\alpha}} \, T_{iA}) \cdot T_{iB\setminus A} \\
&= \sum_{\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A\in\binom{B}{k}} \left(\sum_{\boldsymbol{\alpha} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \, T_{iA}\right) \cdot T_{iB\setminus A} \\
&= \sum_{\boldsymbol{\beta} \,\in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A\in\binom{B}{k}} q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) \cdot T_{iB\setminus A},
\end{aligned} \tag{9}$$

where $q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) := \sum_{\boldsymbol{\alpha} \,\in\, \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}} \, T_{iA}$ is a linear form in the $\mathbf{c}_{\boldsymbol{\beta}} := \{c_{\boldsymbol{\alpha},\boldsymbol{\beta}} : \boldsymbol{\alpha} \in \mathrm{Norm}(k)\}$ variables.

Observation 6.1. *The linear forms* $\left\{q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) \ : \ i \in [s] \text{ and } A \in \binom{[d]}{k}\right\}$ *in the* $\mathbf{c}_{\boldsymbol{\beta}}$ *variables are* $\mathbb{F}$*-linearly independent.*

Proof. Recall $m = s \cdot \binom{d}{k}$. The coefficient matrix of the linear forms $\left\{q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) : i \in [s] \text{ and } A \in \binom{[d]}{k}\right\}$ is a $m \times \binom{n+k-1}{k}$ matrix M with rows indexed by the set $\{(i, A) : i \in [s], A \in \binom{[d]}{k}\}$ and columns indexed by $\{\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathrm{Norm}(k)\}$. The $((i, A), \boldsymbol{\alpha})$-th entry of M is $\partial^k_{\boldsymbol{\alpha}}\, T_{iA}$; note that the entries of M are field elements. We argue that M has rank m.

Let $\boldsymbol{\tau}$ be a row vector whose columns are indexed by $\{(i, A) : i \in [s], A \in \binom{[d]}{k}\}$, the (i, A)-th entry is $T_{i[d]\setminus A}$. Let $\boldsymbol{\delta}$ be a row vector whose columns are indexed by $\{\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathrm{Norm}(k)\}$, the $\boldsymbol{\alpha}$-th entry is $\partial^k_{\boldsymbol{\alpha}}\, f$. The $\boldsymbol{\alpha}$-th entry of the product $\boldsymbol{\tau} \cdot M$ is

$$\sum_{i \in [s],\ A \in \binom{[d]}{k}} T_{i[d]\setminus A} \cdot \partial^k_{\boldsymbol{\alpha}}\, T_{iA} \ = \ \sum_{i \in [s]} \partial^k_{\boldsymbol{\alpha}}\, T_i \ = \ \partial^k_{\boldsymbol{\alpha}}\, f, \quad \text{and so} \quad \boldsymbol{\tau} \cdot M = \boldsymbol{\delta}.$$

If $\mathrm{rank}(M) < m$ then there is a $D \in \mathrm{GL}_m(\mathbb{F})$ such that the last row of $D \cdot M$ has all zero entries. As $\boldsymbol{\tau} D^{-1} \cdot DM = \boldsymbol{\delta}$, every entry of $\boldsymbol{\delta}$ is in the $\mathbb{F}$-span of the first $m - 1$ entries of the row vector $\boldsymbol{\tau} D^{-1}$. But, this contradicts $\dim(U) = m$, since U is the space spanned by the entries of $\boldsymbol{\delta}$. □

The observation implies, we can choose values for the $\mathbf{c}_{\boldsymbol{\beta}}$ variables such that

$$q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) = e_{iA} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T_{iA}), \quad \text{for all } i \in [s] \text{ and } A \in \binom{[d]}{k},$$

where e_{iA} are field elements. The values of e_{iA} will be specified later, so it is best to think of $\{e_{iA} : i \in [s], A \in \binom{[d]}{k}\}$ as distinct variables for the moment. From Equation (9),

$$\begin{aligned}
\psi(T_{iB}) \ &= \ \sum_{\boldsymbol{\beta}\, \in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A \in \binom{B}{k}} e_{iA} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T_{iA}) \cdot T_{iB\setminus A} \\
&= \ \sum_{A \in \binom{B}{k}} e_{iA} \cdot \sum_{\boldsymbol{\beta}\, \in\, \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T_{iA}) \cdot T_{iB\setminus A} \\
&= \ \sum_{A \in \binom{B}{k}} e_{iA} \cdot T_{iA} \cdot T_{iB\setminus A} \\
&= \ \left(\sum_{A \in \binom{B}{k}} e_{iA}\right) \cdot T_{iB}.
\end{aligned}$$

The linear forms $\{\sum_{A \in \binom{B}{k}} e_{iA} : i \in [s], B \in \binom{[d]}{d-k}\}$ are distinct. So, by the Schwartz-Zippel lemma, a random assignment of the e_{iA} variables to field elements makes

$$\sum_{A \in \binom{B}{k}} e_{iA} = \rho_{iB} \ \in \mathbb{F},$$

such that $\{\rho_{iB} : i \in [s], B \in \binom{[d]}{d-k}\}$ are distinct field elements. Therefore, $\psi(T_{iB}) = \rho_{iB} \cdot T_{iB}$ and this completes the proof of the claim.

6.3 Proof of Claim 4.3

Fix a $i \in [s]$ and $B, B' \in \binom{[d]}{d-k}$ with $\mathrm{dist}(B, B') = 1$ arbitrarily. There exist distinct $r, r' \in [d]$ such that $T_{iB} = \ell_{ir} \cdot h$ and $T_{iB'} = \ell_{ir'} \cdot h$, where h is a product of $(d-k-1)$ linear forms and $\gcd(\ell_{ir}, h) = \gcd(\ell_{ir'}, h) = 1$. Let $\psi = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathrm{Norm}(k)} c_{\boldsymbol{\alpha},\boldsymbol{\beta}} \cdot \mathbf{x}^{\boldsymbol{\beta}} \cdot \partial^k_{\boldsymbol{\alpha}}$ be an operator in $\mathcal{S}$. As before, treat $c_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ as variables that take values from $\mathbb{F}$. By Equation (9),

$$\psi(T_{jC}) = \sum_{\boldsymbol{\beta} \in \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A \in \binom{C}{k}} q_{jA}(\mathbf{c}_{\boldsymbol{\beta}}) \cdot T_{jC \setminus A}\,, \quad \text{for every } j \in [s] \text{ and } C \in \binom{[d]}{d-k}.$$

Setting the q_{jA} linear forms. The linear forms $\{q_{jA}(\mathbf{c}_{\boldsymbol{\beta}}) \; : \; j \in [s] \text{ and } A \in \binom{[d]}{k}\}$ are linearly independent, by Observation 6.1. Hence, we can choose values for the $\mathbf{c}_{\boldsymbol{\beta}}$ variables such that the following conditions are satisfied:

1. $q_{jA}(\mathbf{c}_{\boldsymbol{\beta}}) = 0$, for every $j \in [s] \setminus \{i\}$.
2. $q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) = 0$, if $r \notin A$.
3. $q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) = e_{iA} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T'_{iA})$, if $r \in A$, where $T'_{iA} := \frac{T_{iA}}{\ell_{ir}} \cdot \ell_{ir'}$ and e_{iA} are field elements. Note that T'_{iA} need not be square-free. The values of e_{iA} will be chosen appropriately later, so it is best to think of $\{e_{iA} : i \in [s], A \in \binom{[d]}{k}\}$ as distinct variables for the moment.

Hence,

$$\begin{aligned} \psi(T_{jC}) &= 0, \quad \text{for all } j \in [s] \setminus \{i\} \text{ and } C \in \binom{[d]}{d-k}, \text{ and} \\ \psi(T_{iC}) &= 0, \quad \text{if } r \notin C. \end{aligned}$$

Suppose $r \in C$. Then

$$\begin{aligned} \psi(T_{iC}) &= \sum_{\boldsymbol{\beta} \in \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A \in \binom{C}{k}} q_{iA}(\mathbf{c}_{\boldsymbol{\beta}}) \cdot T_{iC \setminus A} \\ &= \sum_{\boldsymbol{\beta} \in \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T'_{iA}) \cdot T_{iC \setminus A} \\ &= \sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \cdot \sum_{\boldsymbol{\beta} \in \mathrm{Norm}(k)} \mathbf{x}^{\boldsymbol{\beta}} \cdot \mathrm{coeff}_{\boldsymbol{\beta}}(T'_{iA}) \cdot T_{iC \setminus A} \\ &= \sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \cdot T'_{iA} \cdot T_{iC \setminus A} \\ &= \sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \cdot \frac{T_{iA}}{\ell_{ir}} \cdot \ell_{ir'} \cdot T_{iC \setminus A} \\ &= \sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \cdot \frac{T_{iC}}{\ell_{ir}} \cdot \ell_{ir'}. \qquad (10) \end{aligned}$$

The notation "$A \in \binom{C}{k} : r \in A$" means the sum runs over all $A \in \binom{C}{k}$ with $r \in A$. We analyze the last equation for two different cases:

Case 1: Suppose $r' \notin C$. By Equation (10),

$$\psi(T_{iC}) = \left(\sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \right) \cdot T_{iC'},$$

where $C' = (C \setminus \{r\}) \uplus \{r'\}$ and $T_{iC'}$ is square-free. For $C = B$, we have $C' = B'$ and

$$\psi(T_{iB}) = \left(\sum_{A \in \binom{B}{k} \,:\, r \in A} e_{iA} \right) \cdot T_{iB'}.$$

This means, we need to assign values to the variables $\{e_{iA} : A \in \binom{[d]}{k}, r \in A\}$ in such a way that the following condition is satisfied.

Condition 1. $\sum_{A \in \binom{B}{k} \,:\, r \in A} e_{iA} = \rho_{iBB'} \quad \in \mathbb{F} \setminus \{0\}$.

Case 2: Suppose $r' \in C$. By Equation (10),

$$\psi(T_{iC}) = \left(\sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} \right) \cdot \frac{T_{iC}}{\ell_{ir}} \cdot \ell_{ir'}.$$

As $\frac{T_{iC}}{\ell_{ir}} \cdot \ell_{ir'}$ is not square-free, we desire to assign values to the variables $\{e_{iA} : A \in \binom{[d]}{k}, r \in A\}$ in such a way that the following condition is satisfied.

Condition 2. *For every $C \in \binom{[d]}{d-k}$ satisfying $r, r' \in C$,* $\sum_{A \in \binom{C}{k} \,:\, r \in A} e_{iA} = 0$.

Observe that if the above two conditions are satisfied then ψ is indeed an operator in $\mathcal{S}$ satisfying the statement of Claim 4.3. The following lemma completes the proof of the claim.

Lemma 6.1. *There exists an assignment of the variables $\left\{ e_{iA} \;:\; A \in \binom{[d]}{k}, r \in A \right\}$ to field elements such that both Condition 1 and 2 are satisfied.*

Proof. Let us focus on Condition 2 first. Consider the $\binom{d-2}{k} \times \binom{d-1}{k-1}$ incidence matrix M whose rows are indexed by $C \in \binom{[d]}{d-k}$ such that $r, r' \in C$, and columns indexed by $A \in \binom{[d]}{k}$ such that $r \in A$. The (C, A)-th entry of M is 1 if $A \subset C$, otherwise the entry is 0. Assume the following ordering among the columns of M: The $\binom{d-2}{k-1}$ columns that are indexed by $A \in \binom{[d]}{k}$ with $r' \notin A$ precede the $\binom{d-2}{k-2}$ columns that are indexed by $A \in \binom{[d]}{k}$ with $r' \in A$.

Let M_1 be the sub-matrix of M defined by the first $\binom{d-2}{k-1}$ columns of M, and M_2 the sub-matrix defined by the last $\binom{d-2}{k-2}$ columns of M.

Observation 6.2. *The $\binom{d-2}{k} \times \binom{d-2}{k-1}$ matrix M_1 has full rank, i.e., it has rank $\binom{d-2}{k-1}$.*

Proof. Matrix M_1 is identical to an incidence matrix J whose rows are indexed by $S \in \binom{[d-2]}{d-k-2}$ and columns are indexed by $T \in \binom{[d-2]}{k-1}$, the (S,T)-th entry of J contains 1 if $T \subset S$, otherwise the entry is 0. It was shown in [Got66] that J has full rank, which is $\binom{d-2}{k-1}$ as $k \leq \frac{d-1}{2}$ by Remark 1. □

Observation 6.3. *Every column of M_2 is in the* $\mathbb{F}$-span *of the columns of M_1.*

Proof. Pick a column from M_2. Suppose that the column is indexed by $\{r, r'\} \uplus A_2$, where $A_2 \in \binom{[d]\setminus\{r,r'\}}{k-2}$. Consider all those columns in M_1 that are indexed by $\{r\} \uplus A_1$, where $A_1 \in \binom{[d]\setminus\{r,r'\}}{k-1}$ and $A_2 \subset A_1$. There are $d-k$ such columns in M_1. It is easy to verify that the sum of these $d-k$ columns in M_1 is $d-2k$ times the column picked from M_2. □

By Observation 6.3, there is a $\binom{d-1}{k-1} \times \binom{d-1}{k-1}$ matrix D whose rows and columns are indexed by $A \in \binom{[d]}{k}$ with $r \in A$ (in the same order as the columns of M) such that D is of the form

$$\begin{aligned}
D &= \left[\begin{array}{c|c} I_{\binom{d-2}{k-1}} & E \\ \hline 0 & I_{\binom{d-2}{k-2}} \end{array}\right], \quad \text{where } E \text{ is a } \binom{d-2}{k-1} \times \binom{d-2}{k-2} \text{ matrix, and} \\
M \cdot D &= \left[\begin{array}{c|c} M_1 & 0 \end{array}\right]. \qquad (11)
\end{aligned}$$

Note that D^{-1} is of the form

$$D^{-1} = \left[\begin{array}{c|c} I_{\binom{d-2}{k-1}} & -E \\ \hline 0 & I_{\binom{d-2}{k-2}} \end{array}\right]. \qquad (12)$$

Let $\mathbf{e}$ be the column vector whose rows are indexed by $A \in \binom{[d]}{k}$ with $r \in A$ (in the same order as the columns of M). The A-th entry of $\mathbf{e}$ is the variable e_{iA}. The solution space of the e_{iA} variables defined by the following linear system exactly captures those assignments of the e_{iA} variables that satisfy Condition 2,

$$\begin{aligned}
M \cdot \mathbf{e} &= 0 \\
\Rightarrow M \cdot D \cdot D^{-1} \cdot \mathbf{e} &= 0 \\
\Rightarrow \left[\begin{array}{c|c} M_1 & 0 \end{array}\right] \cdot D^{-1} \cdot \mathbf{e} &= 0.
\end{aligned}$$

As M_1 has full rank (by Observation 6.2), the solution space is defined by the vanishing of the first $\binom{d-2}{k-1}$ rows of $D^{-1} \cdot \mathbf{e}$. Hence, by the form of D^{-1} in Equation (12), the solution space is defined by

$$\left[\begin{array}{c|c} I_{\binom{d-2}{k-1}} & -E \end{array}\right] \cdot \mathbf{e} = 0. \qquad (13)$$

Clearly, this solution space has dimension $\binom{d-2}{k-2}$, and the variables $\{e_{iA} : A \in \binom{[d]}{k} \text{ and } r, r' \in A\}$ are the free variables. Every solution in this solution space satisfies Condition 2.

Now, let us turn to Condition 1. We wish to show the existence of a solution in the above solution space that satisfies Condition 1. This is argued next.

Consider adding all those rows of M that are indexed by $C \in \binom{[d]}{d-k}$ with $r, r' \in C$ such that $C \setminus \{r'\} \subset B$. There are $d-k-1$ such rows. Call the resulting sum $\mathbf{w}$, a row vector whose columns are indexed by $A \in \binom{[d]}{k}$ with $r \in A$ (in the same order as the columns of M). It is easy to verify that $\mathbf{w}$ has the following structure:

- If $r \in A, r' \notin A$ and $A \subset B$ then the A-th entry of $\mathbf{w}$ is $d-2k$.
- If $r, r' \in A$ and $A \setminus \{r'\} \subset B$ then the A-th entry of $\mathbf{w}$ is $d-2k+1$.
- The remaining entries of $\mathbf{w}$ are zero.

Let $\mathcal{Q} := \{A : A \in \binom{[d]}{k}$ and $r, r' \in A$ and $A \setminus \{r'\} \subset B\}$. Any solution in the space defined by Equation (13) also satisfies

$$\begin{aligned} \mathbf{w} \cdot \mathbf{e} &= 0 \\ \Rightarrow (d-2k) \cdot \sum_{A \in \binom{B}{k} \,:\, r \in A} e_{iA} \quad + \quad (d-2k+1) \cdot \sum_{A \in \mathcal{Q}} e_{iA} &= 0 \\ \Rightarrow \sum_{A \in \binom{B}{k} \,:\, r \in A} e_{iA} &= -\left(\frac{d-2k+1}{d-2k}\right) \cdot \sum_{A \in \mathcal{Q}} e_{iA}. \end{aligned}$$

The e_{iA} variables in the RHS of the above equation belong to the set of free variables of the system defined by Equation (13). Surely, these variables can be chosen in such a way that $\sum_{A \in \mathcal{Q}} e_{iA} \neq 0$, thereby implying $\sum_{A \in \binom{B}{k} \,:\, r \in A} e_{iA} \neq 0$. In other words, there is a solution in the space defined by Equation (13) that satisfies Condition 1.

This finishes the proof of the lemma thereby completing the proof of Claim 4.3. □

6.4 Proof of Corollary 4.2

The statement of the corollary does not depend on the choice of basis of U. We will work with the canonical basis Λ of U. Let $\mathbf{u}_j \in \mathbb{F}^m$ be the unit vector with 1 at the j-th coordinate and 0 elsewhere. It follows from the ordering of the elements in Λ that $\left\{\mathbf{u}_j \ : \ j \in [(i-1) \cdot \binom{d}{k} + 1,\ i \cdot \binom{d}{k}]\right\}$ is a basis of $U_{i,\Lambda} \subset \mathbb{F}^m$ for every $i \in [s]$. Let $\{\psi_1, \ldots, \psi_t\}$ be a basis of $\mathcal{S}$.

Observation 6.4. *Let $\mathbf{v} = \sum_{j \in [m]} \rho_j \cdot \mathbf{u}_j \in \mathbb{F}^m$, where $\rho_j \in \mathbb{F}$. For every $j \in [m]$, if $\rho_j \neq 0$ then $\mathbf{u}_j$ belongs to the closure of $\mathbf{v}$ with respect to $\mathbb{F}$-span$\{M_\Lambda(\psi_1), \ldots, M_\Lambda(\psi_t)\}$.*

Proof. Let us abuse notation slightly and denote $\mathbb{F}$-span$\{M_\Lambda(\psi_1), \ldots, M_\Lambda(\psi_t)$ by $\mathcal{S}$. By Claim 4.2 and Corollary 4.1, there is a diagonal matrix $M_\Lambda = \text{diag}(a_1, \ldots, a_m) \in \mathcal{S}$, where $a_1, \ldots, a_m \in \mathbb{F}$ are distinct. Thus, the closure of $\mathbf{v}$ with respect to $\mathcal{S}$ contains $\mathbf{v}_l := \sum_{j \in [m]} a_j^l \cdot \rho_j \cdot \mathbf{u}_j$ for all $l \in \mathbb{N}$. Suppose $\rho_j \neq 0$. It is easy to verify that we can find $c_1, \ldots, c_m \in \mathbb{F}^m$ such that $\sum_{l \in [m]} c_l \cdot \mathbf{v}_l = \mathbf{u}_j$ by solving a system of linear equations which involves a Vandermonde coefficient matrix. □

The above observation implies that the closure of every $\mathbf{v} \in U_i$ with respect to $\mathcal{S}$ is U_i if and only if for every $j \in [(i-1) \cdot \binom{d}{k} + 1,\ i \cdot \binom{d}{k}]$ the closure of the unit vector $\mathbf{u}_j \in U_{i,\Lambda}$ with respect to $\mathbb{F}$-span$\{M_\Lambda(\psi_1), \ldots, M_\Lambda(\psi_t)\}$ is $U_{i,\Lambda}$. The following observation completes the argument.

Observation 6.5. *For every $i \in [m]$ and $j \in [(i-1) \cdot \binom{d}{k} + 1,\ i \cdot \binom{d}{k}]$, the closure of the unit vector $\mathbf{u}_j \in U_{i,\Lambda}$ with respect to $\mathbb{F}$-span$\{M_\Lambda(\psi_1), \ldots, M_\Lambda(\psi_t)\}$ is $U_{i,\Lambda}$.*

Proof. Fix $i \in [m]$ and $j \in [(i-1) \cdot \binom{d}{k} + 1,\ i \cdot \binom{d}{k}]$ arbitrarily. Let us abuse notation again and denote $\mathbb{F}\text{-span}\{M_\Lambda(\psi_1), \ldots, M_\Lambda(\psi_t)\}$ by $\mathcal{S}$. The unit vector $\mathbf{u}_j$ corresponds to a term T_{iB} for some $B \in \binom{[d]}{d-k}$. Let $\mathbf{u}_l \neq \mathbf{u}_j$ be another unit vector in $U_{i,\Lambda}$, and T_{iC} be the term corresponding to $\mathbf{u}_l$ for some $C \in \binom{[d]}{d-k}$. Suppose $\text{dist}(B, C) = \Delta$. By Claim 4.3, there are Δ matrices $M_1, \ldots, M_\Delta \in \mathcal{S}$ such that $M_1 \cdot M_2 \ldots M_\Delta \cdot \mathbf{u}_j = \rho \cdot \mathbf{u}_l$, where $\rho \in \mathbb{F} \setminus \{0\}$. As $\mathbf{u}_l \neq \mathbf{u}_j$ is chosen arbitrarily, the closure of $\mathbf{u}_j$ with respect to $\mathcal{S}$ is indeed $U_{i,\Lambda}$. □

From the above two observations, it follows readily that $U_1, \ldots, U_s$ are the only irreducible invariant subspaces of U induced by $\mathcal{S}$.

Acknowledgments

We would like to thank Gaurav Sinha for showing us a simpler way to execute Step 3 of Algorithm 1 and allowing us to present his argument in Section 4.3. It has also helped us weaken the non-degeneracy condition slightly. Thanks also to Antonina Kolokolova for asking some interesting questions on this work at the Dagstuhl seminar on "Algebraic Methods in Computational Complexity" 2018. One of her questions has made us observe that Algorithm 1 works even if the input is evaluations of f and its derivatives at sufficiently many random points.

References

[Ang88] Dana Angluin. Queries and concept learning. *Machine Learning.*, 2(4):319–342, 1988.

[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, 2008.

[BBB$^+$00] Amos Beimel, Francesco Bergadano, Nader H. Bshouty, Eyal Kushilevitz, and Stefano Varricchio. Learning functions represented as multiplicity automata. *J. ACM*, 47(3):506–530, 2000.

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

[Ber70] Elwyn R Berlekamp. Factoring polynomials over large finite fields. *Mathematics of Computation*, 24:713–735, 1970.

[BLS16] Nikhil Balaji, Nutan Limaye, and Srikanth Srinivasan. An almost cubic lower bound for $\Sigma\Pi\Sigma$ circuits computing a polynomial in VP. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:143, 2016.

[BT88] Michael Ben-Or and Prasoon Tiwari. A deterministic algorithm for sparse multivariate polynominal interpolation (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 301–309, 1988.

[CIKK16] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Learning Algorithms from Natural Proofs. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 10:1–10:24, 2016.

[FK09] Lance Fortnow and Adam R. Klivans. Efficient learning algorithms yield circuit lower bounds. *J. Comput. Syst. Sci.*, 75(1):27–36, 2009.

[FSV17] Michael A. Forbes, Amir Shpilka, and Ben Lee Volk. Succinct hitting sets and barriers to proving algebraic circuits lower bounds. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 653–664, 2017.

[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.

[GKP18] Ignacio García-Marco, Pascal Koiran, and Timothée Pecatte. Polynomial equivalence problems for sum of affine powers. In *Proceedings of the 2018 ACM on International Symposium on Symbolic and Algebraic Computation, ISSAC 2018, New York, NY, USA, July 16-19, 2018*, pages 303–310, 2018.

[GKS94] Dima Grigoriev, Marek Karpinski, and Michael F. Singer. Computational complexity of sparse rational interpolation. *SIAM J. Comput.*, 23(1):1–11, 1994.

[GKSS17] Joshua A. Grochow, Mrinal Kumar, Michael E. Saks, and Shubhangi Saraf. Towards an algebraic natural proofs barrier via polynomial identity testing. *CoRR*, abs/1701.01717, 2017.

[Got66] D. H. Gottlieb. A Certain Class of Incidence Matrices. *Proceedings of the American Mathematical Society*, 17(6):1233–1237, 1966.

[Hås90] Johan Håstad. Tensor Rank is NP-Complete. *J. Algorithms*, 11(4):644–654, 1990.

[JKS02] Jeffrey C. Jackson, Adam R. Klivans, and Rocco A. Servedio. Learnability beyond AC0. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 776–784, 2002.

[Kay11] Neeraj Kayal. Efficient algorithms for some special cases of the polynomial equivalence problem. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 1409–1421, 2011.

[Kay12] Neeraj Kayal. Affine projections of polynomials: extended abstract. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 643–662, 2012.

[KLSS17] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Formulas. *SIAM J. Comput.*, 46(1):307–335, 2017.

[KNS18] Neeraj Kayal, Vineet Nair, and Chandan Saha. Average-case linear matrix factorization and reconstruction of low width algebraic branching programs. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:29, 2018.

[KNST17] Neeraj Kayal, Vineet Nair, Chandan Saha, and Sébastien Tavenas. Reconstruction of Full Rank Algebraic Branching Programs. In *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, pages 21:1–21:61, 2017.

[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.

[KS03] Adam R. Klivans and Amir Shpilka. Learning arithmetic circuits via partial derivatives. In *Computational Learning Theory and Kernel Machines, 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003, Proceedings*, pages 463–476, 2003.

[KS09] Zohar Shay Karnin and Amir Shpilka. Reconstruction of generalized depth-3 arithmetic circuits with bounded top fan-in. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 274–285, 2009.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 146–153, 2014.

[KST16] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An Almost Cubic Lower Bound for Depth Three Arithmetic Circuits. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 33:1–33:15, 2016.

[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.

[Kum17] Mrinal Kumar. A Quadratic Lower Bound for Homogeneous Algebraic Branching Programs. In *Proceedings of the 32nd Computational Complexity Conference*, CCC '17, pages 19:1–19:16, 2017.

[LLL82] Arjen K Lenstra, Hendrik W Lenstra, and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant Depth Circuits, Fourier Transform, and Learnability. *J. ACM*, 40(3):607–620, 1993.

[NW97] Noam Nisan and Avi Wigderson. Lower Bounds on Arithmetic Circuits Via Partial Derivatives. *Computational Complexity*, 6(3):217–234, 1997.

[RR97] Alexander A. Razborov and Steven Rudich. Natural Proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

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

[Shi16] Yaroslav Shitov. How hard is the tensor rank? *arXiv*, abs/1611.01559, 2016.

[Shp07] Amir Shpilka. Interpolation of depth-3 arithmetic circuits with two multiplication gates. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 284–293, 2007.

[Sin16] Gaurav Sinha. Reconstruction of real depth-3 circuits with top fan-in 2. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 31:1–31:53, 2016.

[SW01] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

[Tav13] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013, Klosterneuburg, Austria, August 26-30, 2013. Proceedings*, pages 813–824, 2013.

[Vol16] Ilya Volkovich. A Guide to Learning Arithmetic Circuits. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, pages 1540–1561, 2016.

[Wig07] Avi Wigderson. P, NP and Mathematics - a computational complexity perspective. In *Proceedings of the International Congress of Mathematicians (ICM), 2006, Madrid, Spain*, pages 665–712, 2007.

[Yau16] Morris Yau. Almost cubic bound for depth three circuits in VP. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:187, 2016.

[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 Non-degeneracy of random circuits

In a random (n, d, s) homogeneous depth three circuit $\mathtt{C} = T_1 + \ldots + T_s$, the coefficients of the linear forms in the terms $T_1, \ldots, T_s$ are chosen independently and uniformly at random from a sufficiently large finite set $S \subseteq \mathbb{F}$. We will call $\mathtt{C}$ a random (n, d, s) homogeneous depth three circuit *over S*. As there are a total of *nds* number of coefficients, every (n, d, s) homogeneous depth three circuit can be identified with a point in $\mathbb{F}^{nds}$. A random homogeneous depth three circuit over S is a random point in S^{nds}. We show that a random homogeneous depth three circuit is non-degenerate with high probability. In fact, it follows from the proof of the next claim that degenerate (n, d, s) homogeneous depth three circuits correspond to points in $\mathbb{F}^{nds}$ that lie in a proper algebraic variety. So, a random homogeneous depth three circuit is non-degenerate with probability one in the measure theoretic sense. This also means that we can define random homogeneous circuits with regard to other reasonable distributions and the implication of 'non-degenerate with high probability' will still hold.

Claim A.1. *Let $S \subseteq \mathbb{F}$ be a finite set. A random (n, d, s) homogeneous depth three circuit over S is non-degenerate with probability at least $1 - \frac{2d^2s^3}{|S|}$.*

Proof. Let $\mathtt{C} = T_1 + \cdots + T_s$ be a random (n, d, s) homogeneous depth three circuit over S, where $T_i = \ell_{i1} \ldots \ell_{id}$ is a product of d linear forms. Fix a $i \in [s]$ arbitrarily. The linear forms $\ell_{i1}, \ldots, \ell_{id}$ are linearly independent with probability at least $1 - \frac{d}{|S|}$ (by the Schwartz-Zippel lemma). As the coefficients of the linear forms in $\{T_j : j \in [s] \setminus \{i\}\}$ are chosen independent of the coefficients of $\ell_{i1}, \ldots, \ell_{id}$ (which we can now assume to be linearly independent), the probability of the event

$$\dim\left(\sum_{j\in[s]\setminus\{i\}} U_j \mod \langle \ell_{i1}, \ldots, \ell_{i\ 2k+1}\rangle\right) = (s-1)\cdot\binom{d}{k}, \tag{14}$$

equals the probability of the event

$$\dim\left(\sum_{j\in[s]\setminus\{i\}} U_j \mod \langle x_1, \ldots, x_{2k+1}\rangle\right) = (s-1)\cdot\binom{d}{k}.$$

Let $\tilde{\ell}_{jr}$ be the linear form derived from ℓ_{jr} by setting the variables $x_1, \ldots, x_{2k+1}$ to zero, $\tilde{T}_j := \tilde{\ell}_{j1} \cdot \tilde{\ell}_{j1} \ldots \tilde{\ell}_{jd}$ and $\tilde{T}_{jB} := \prod_{r\in B} \tilde{\ell}_{jr}$ for $B \in \binom{[d]}{d-k}$. The probability of the above event equals the probability that the polynomials in $P := \{\tilde{T}_{jB} \ : \ j \in [s] \setminus \{i\}, B \in \binom{[d]}{d-k}\}$ are $\mathbb{F}$-linearly independent. Keeping in mind the relations among n, d, s and k from Remark 1, the number of polynomials in P is $(s-1)\cdot\binom{d}{k} \leq ds^2$. The probability that these polynomials are linearly independent is at least $1 - \frac{(ds)^2}{|S|}$ (by an application of the Schwartz-Zippel lemma), if we can show that there exists a setting of the coefficients of the linear forms in $\tilde{T}_{jB}$ such that the polynomials in P (after the setting) are linearly independent. The existence of such a setting of the coefficients follows from the Nisan-Wigderson design polynomial family.

The Nisan-Wigderson design polynomial NW is a multilinear, homogeneous n-variate polynomial of degree d such that any pair of monomials in NW has at most k variables in common. We refer the reader to [KSS14, KLSS17] for an explicit version of the Nisan-Wigderson design polynomial that has $(\frac{n}{d})^{k+1}$ monomials. In our case, we are left with $n - (2k+1)$ variables after setting $x_1, \ldots, x_{2k+1}$ to zero. By Remark 1, we have $(\frac{n-2k-1}{d})^{k+1} \geq s$. So, we can set the coefficients of the linear forms in $\tilde{T}_j$ in such a way that the terms $\{\tilde{T}_j : j \in [s] \setminus \{i\}\}$ map to distinct monomials of NW. Under this setting, the polynomials in P map to distinct monomials (as $k < \frac{d}{2}$ by Remark 1) and hence they are linearly independent.

Thus, for an arbitrarily fixed $i \in [s]$, the condition in Equation (14) is satisfied with probability at least $1 - \frac{(ds)^2+d}{|S|}$. By applying union bound over all $i \in [s]$, it follows that a random circuit $\mathtt{C}$ is non-degenerate with probability at least $1 - \frac{2d^2s^3}{|S|}$. □