THEORY OF COMPUTING, Volume 17 (10), 2021, pp. 1–88
www.theoryofcomputing.org

SPECIAL ISSUE: CCC 2016

Proof Complexity Lower Bounds from Algebraic Circuit Complexity

Michael A. Forbes* Amir Shpilka† Iddo Tzameret‡ Avi Wigderson§

Received June 15, 2016; Revised January 3, 2021; Published November 1, 2021

Abstract. We give upper and lower bounds on the power of subsystems of the *Ideal Proof System (IPS)*, the algebraic proof system recently proposed by Grochow and Pitassi (J. ACM, 2018), where the circuits comprising the proof come from various restricted algebraic circuit classes. This mimics an established research direction in the Boolean setting for subsystems of *Extended Frege* proofs, where proof-lines are circuits from restricted Boolean circuit classes. Except one, all of the subsystems considered in this paper can simulate the well-studied *Nullstellensatz* proof system, and prior to this work there were no known lower

A conference version of this paper appeared in the Proceedings of the 31st Computational Complexity Conference (CCC'16) [29].

*This work was performed when the author was at the Department of Computer Science of Princeton University, supported by the Princeton Center for Theoretical Computer Science.

†The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

‡Partly supported by The National Natural Science Foundation of China Grant (61373002) and has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101002742).

§This research was partially supported by NSF grant CCF-1412958.

ACM Classification: F.1.3, F.1.2

AMS Classification: 68Q17, 68Q15, 03F20

Key words and phrases: Proof complexity, algebraic circuit complexity, lower bounds, algebraic proof systems, IPS, polynomial calculus, hardness versus randomness, functional lower bounds, ABPs

 DOI: 10.4086/toc.2021.v017a010

bounds when measuring proof size by the algebraic complexity of the polynomials (except with respect to degree, or to sparsity).

We give two general methods of converting certain algebraic circuit lower bounds into proof complexity ones. However, we need to strengthen existing lower bounds to hold for either the *functional* model or for *multiplicities* (see below). Our techniques are reminiscent of existing methods for converting Boolean circuit lower bounds into related proof complexity results, such as *feasible interpolation*. We obtain the relevant types of lower bounds for a variety of classes (*sparse polynomials*, *depth-3 powering formulas*, *read-once oblivious algebraic branching programs*, and *multilinear formulas*), and infer the relevant proof complexity results. We complement our lower bounds by giving short refutations of the previously studied *subset-sum* axiom using IPS subsystems, allowing us to conclude strict separations between some of these subsystems.

Our first method is a *functional lower bound*, a notion due to Grigoriev and Razborov (Appl. Algebra Eng. Commun. Comput., 2000), which says that not only does a polynomial f require large algebraic circuits, but that *any* polynomial g agreeing with f on the Boolean cube also requires large algebraic circuits. For our classes of interest, we develop functional lower bounds where $g(\overline{x})$ equals $1/p(\overline{x})$ where p is a constant-degree polynomial, which in turn yield corresponding IPS lower bounds for proving that p is nonzero over the Boolean cube. In particular, we show superpolynomial lower bounds for refuting variants of the subset-sum axiom in various IPS subsystems.

Our second method is to give *lower bounds for multiples*, that is, to give explicit polynomials whose all (nonzero) multiples require large algebraic circuit complexity. By extending known techniques, we are able to obtain such lower bounds for our classes of interest, which we then use to derive corresponding IPS lower bounds. Such lower bounds for multiples are of independent interest, as they have tight connections with the algebraic hardness versus randomness paradigm.

1 Introduction

Propositional proof complexity aims to understand and analyze the computational resources required to prove propositional tautologies, in the same way that circuit complexity studies the resources required to compute Boolean functions. A typical goal would be to establish, for a given proof system, superpolynomial lower bounds on the *size* of any proof of some propositional tautology. The seminal paper of Cook and Reckhow [15] showed that this goal relates quite directly to fundamental hardness questions in computational complexity such as the NP vs. coNP question: establishing superpolynomial lower bounds for *every* propositional proof system would separate NP from coNP (and thus also P from NP). We refer the reader to Krajíček [48] for more on this subject.

Propositional proof systems come in a large variety, as different ones capture different forms of reasoning, either reasoning used to actually prove theorems, or reasoning used by algorithmic techniques for different types of search problems (as failure of the algorithm to find the desired object constitutes a proof of its nonexistence). Much of the research in proof complexity deals with propositional proof systems originating from logic or geometry. Logical proof systems include such systems as *resolution* (whose variants are related to popular algorithms for automated theorem proving and SAT solving), as well as the *Frege* proof system (capturing the most common logic text-book systems) and its many subsystems. Geometric proof systems include *cutting-plane proofs*, capturing reasoning used in algorithms for integer programming, as well as proof systems arising from systematic strategies for rounding linear- or semidefinite-programming such as the *lift-and-project* or *sum-of-squares* hierarchies.

In this paper we focus on algebraic proof systems, in which propositional tautologies (or rather contradictions) are expressed as unsatisfiable systems of polynomial equations and algebraic tools are used to refute them. This study originates with the work of Beame, Impagliazzo, Krajíček, Pitassi and Pudlák [9], who introduced the Nullstellensatz refutation system (based on Hilbert's Nullstellensatz), followed by the Polynomial Calculus system of Clegg, Edmonds, and Impagliazzo [13], which is a "dynamic" version of the Nullstellensatz. In both systems the main measures of proof size that have been studied are the *degree* and *sparsity* of the polynomials appearing in the proof. Substantial work has led to a very good understanding of the power of these systems with respect to these measures (see for example [12, 69, 31, 41, 11, 4] and references therein).

However, the above measures of degree and sparsity are rather rough measures of a complexity of a proof. Accordingly, Grochow and Pitassi [35] have recently advocated measuring the complexity of such proofs by their algebraic circuit size and shown that the resulting proof system can polynomially simulate strong proof systems such as the Frege system (see also prior work on measuring algebraic proofs by their algebraic circuit size [59, 32, 66, 65, 83] as well as the subsequent survey [60]). This naturally leads to the question of establishing lower bounds for this stronger proof system, even for restricted classes of algebraic circuits.

In this article we establish such lower bounds for previously studied restricted classes of algebraic circuits, and show that these lower bounds are interesting by providing non-trivial *upper* bounds in these proof systems for refutations of interesting sets of polynomial equations. This provides what are apparently the first examples of lower bounds on the algebraic circuit size of propositional proofs in the Ideal Proof System (IPS) framework of Grochow and Pitassi [35].

We note that obtaining proof complexity lower bounds from circuit complexity lower bounds is an established tradition that takes many forms. Most prominent are the lower bounds for subsystems of the Frege proof system defined by small-depth Boolean circuits, and lower bounds of Pudlák [61] on Resolution and Cutting Planes system using the so-called feasible interpolation method. We refer the reader again to Krajíček [48] for more details. Our approach here for algebraic systems shares features with both of these approaches.

The rest of this introduction is arranged as follows. In Section 1.1 we give the necessary background in algebraic proof complexity, and explain the IPS system. In Section 1.2 we define the algebraic complexity classes that will underlie the subsystems of IPS we will study. In Section 1.3 we state our results and explain our techniques, for both the algebraic and proof complexity worlds.

1.1 Algebraic proof systems

We now describe the algebraic proof systems that are the subject of this paper. If one has a set of polynomials (called *axioms*) $f_1, \dots, f_m \in \mathbb{F}[x_1, \dots, x_n]$ over some field $\mathbb{F}$, then (the weak version of) Hilbert's Nullstellensatz shows that the system $f_1(\overline{x}) = \cdots = f_m(\overline{x}) = 0$ is unsatisfiable (over the algebraic closure of $\mathbb{F}$) if and only if there are polynomials $g_1,\dots,g_m \in \mathbb{F}[\overline{x}]$ such that $\sum_j g_j(\overline{x})f_j(\overline{x}) = 1$ (as a formal identity), or equivalently, that 1 is in the ideal generated by the $\{f_j\}_j$.

Beame, Impagliazzo, Krajíček, Pitassi, and Pudlák [9] suggested to treat these $\{g_j\}_j$ as a *proof* of the unsatisfiability of this system of equations, called a *Nullstellensatz refutation*. This is in particular relevant for complexity theory as one can restrict attention to *Boolean* solutions to this system by adding the *Boolean axioms*, that is, adding the polynomials $\{x_i^2 - x_i\}_{i=1}^n$ to the system. One can then naturally encode NP-complete problems such as the satisfiability of 3CNF formulas as the satisfiability of a system of constant-degree polynomials, and a Nullstellensatz refutation is then an equation of the form $\sum_{j=1}^m g_j(\overline{x})f_j(\overline{x}) + \sum_{i=1}^n h_i(\overline{x})(x_i^2 - x_i) = 1$ for $g_j, h_i \in \mathbb{F}[\overline{x}]$. This proof system is sound (only refuting unsatisfiable systems over $\{0,1\}^n$) and complete (refuting any unsatisfiable system, by Hilbert's Nullstellensatz).

Given that the above proof system is sound and complete, it is then natural to ask what is its power to refute unsatisfiable systems of polynomial equations over $\{0,1\}^n$. To understand this question one must define the notion of the *size* of the above refutations. Two popular notions are that of the *degree*, and the *sparsity* (number of monomials). One can then show (see for example Pitassi [59]) that for any unsatisfiable system which includes the Boolean axioms, there exists a refutation where the g_j are multilinear and the h_i have degree at most $O(n+d)$, given that each f_j has degree at most d. In particular, this implies that for any unsatisfiable system with $d = O(n)$ there is a refutation of degree $O(n)$ and involving at most $\exp(O(n))$ monomials. This intuitively agrees with the fact that coNP is a subset of non-deterministic exponential time.

Building on the suggestion of Pitassi [59] and various investigations into the power of strong algebraic proof systems ([32, 66, 65]), Grochow and Pitassi [35] have recently considered more *succinct* descriptions of polynomials where one measures the size of a polynomial by the size of an algebraic circuit needed to compute it. This is potentially much more powerful as there are polynomials such as the determinant which are of high degree and involve exponentially many monomials and yet can be computed by small algebraic circuits. They named the resulting system the *Ideal Proof System (*IPS*)* which we now define.

Definition 1.1 (Ideal Proof System (IPS), Grochow-Pitassi [35])**.** Let $f_1(\overline{x}),\dots,f_m(\overline{x}) \in \mathbb{F}[x_1,\dots,x_n]$ be a system of polynomials. An **IPS refutation** for showing that the polynomials $\{f_j\}_j$ have no common solution in $\{0,1\}^n$ is an algebraic circuit $C(\overline{x},\overline{y},\overline{z}) \in \mathbb{F}[\overline{x},y_1,\dots,y_m,z_1,\dots,z_n]$, such that

1. $C(\overline{x},\overline{0},\overline{0}) = 0$.

2. $C(\overline{x}, f_1(\overline{x}),\dots,f_m(\overline{x}), x_1^2 - x_1,\dots,x_n^2 - x_n) = 1$.

The **size** of the IPS refutation is the size of the circuit C. If C is of individual degree ≤ 1 in each y_j and z_i, then this is a **linear** IPS refutation (called *Hilbert* IPS by Grochow-Pitassi [35]), which we will abbreviate as $\mathrm{IPS}_{\mathrm{LIN}}$. If C is of individual degree ≤ 1 only in the y_j then we say this is a $\mathrm{IPS}_{\mathrm{LIN}'}$ refutation. If C comes from a restricted class of algebraic circuits $\mathcal{C}$, then this is a called a $\mathcal{C}$-IPS refutation, and further called a $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation if C is linear in $\overline{y},\overline{z}$, and a $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}'}$ refutation if C is linear in $\overline{y}$.

Notice also that our definition here by default adds the equations $\{x_i^2 - x_i\}_i$ to the system $\{f_j\}_j$. For convenience we will often denote the equations $\{x_i^2 - x_i\}_i$ as $\overline{x}^2 - \overline{x}$. One need not add the equations

$\bar{x}^2 - \bar{x}$ to the system in general, but this is the most interesting regime for proof complexity and thus we adopt it as part of our definition.

The $\mathcal{C}$-IPS system is sound for any $\mathcal{C}$, and Hilbert's Nullstellensatz shows that $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}}$ is complete for any complete class of algebraic circuits $\mathcal{C}$ (that is, classes which can compute any polynomial, possibly requiring exponential complexity). We note that we will also consider non-complete classes such as multilinear-formulas (which can only compute *multilinear* polynomials, but are complete for multilinear polynomials), where we will show that the multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ system is not complete for the language of all unsatisfiable sets of multilinear polynomials (Theorem 4.7), while the stronger multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}'}$ version is complete (Theorem 4.12). However, for the standard conversion of unsatisfiable CNFs into polynomial systems of equations, the multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ system is complete (Theorem 1.2).

The following theorem, due to Grochow and Pitassi [59, 35], shows that the IPS system has surprising power and that lower bounds on this system give rise to *computational* lower bounds.

Theorem 1.2 (Pitassi [59],Grochow-Pitassi [35]). *Let $\varphi = C_1 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF on n-variables, and let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_m]$ be its encoding as a polynomial system of equations. If there is a size-s Frege proof (or Extended Frege proof) that $\{f_j\}_j, \{x_i^2 - x_i\}_i$ is unsatisfiable, then there is a formula-*$\mathrm{IPS}_{\mathrm{LIN}}$ *(circuit-*$\mathrm{IPS}_{\mathrm{LIN}}$*, resp.) refutation of size* $\mathsf{poly}(n,m,s)$ *that is checkable in randomized* $\mathsf{poly}(n,m,s)$ *time.*[1]

Further, $\{f_j\}_j$ has an $\mathrm{IPS}_{\mathrm{LIN}}$ *refutation, where the refutation uses multilinear polynomials in* VNP. *Thus, if every* IPS *refutation of $\{f_j\}_j$ requires formula (or circuit) size $\geq s$, then there is an explicit polynomial (that is, in* VNP*) that requires size $\geq s$ algebraic formulas (circuits, resp.).* □

Remark 1.3. One point to note is that the transformation from Extended Frege to IPS refutations yields circuits of polynomial size but without any guarantee on their degree. In particular, such circuits can compute polynomials of exponential degree. In contrast, the conversion from Frege to IPS refutations yields polynomial sized algebraic formulas and those compute polynomials of polynomially bounded degree. This range of parameters, polynomials of polynomially bounded degree, is the more common setting studied in algebraic complexity.

The fact that $\mathcal{C}$-IPS refutations are efficiently checkable (with randomness) follows from the fact that we need to verify the polynomial identities stipulated by the definition. That is, one needs to solve an instance of the *polynomial identity testing (PIT)* problem for the class $\mathcal{C}$: given a circuit from the class $\mathcal{C}$ decide whether it computes the identically zero polynomial. This problem is solvable in (one-sided) probabilistic polynomial time (coRP) for general algebraic circuits, and there are various restricted classes for which deterministic algorithms are known (see Section 3.1).

Motivated by the fact that PIT of non-commutative formulas [2] can be solved deterministically ([64]) and admit exponential-size lower bounds ([52]), Li, Tzameret and Wang [50] have shown that IPS over *non-commutative* polynomials (along with additional *commutator* axioms) can simulate Frege (they also provided a quasipolynomial simulation of IPS over non-commutative formulas by Frege; see Li, Tzameret and Wang [50] for more details).

[1] We note that Grochow and Pitassi [35] proved this for Extended Frege and circuits, but essentially the same proof relates Frege and formula size.

[2] These are formulas over a set of non-commuting variables.

Theorem 1.4 (Li, Tzameret and Wang [50])**.** *Let $\varphi = C_1 \wedge \cdots \wedge C_m$ be an unsatisfiable CNF on n-variables, and let $f_1, \dots, f_m \in \mathbb{F}[x_1, \dots, x_m]$ be its encoding as a polynomial system of equations. If there is a size-s Frege proof that $\{f_j\}_j, \{x_i^2 - x_i\}_i$ is unsatisfiable, then there is a non-commutative-*IPS *refutation of formula-size* $\mathsf{poly}(n,m,s)$*, where the commutator axioms $x_i x_j - x_j x_i$ are also included in the polynomial system being refuted. Further, this refutation is checkable in deterministic* $\mathsf{poly}(n,m,s)$ *time.* □

The above results naturally motivate studying $\mathcal{C}$-IPS for various restricted classes of algebraic circuits, as lower bounds for such proofs then intuitively correspond to restricted lower bounds for the Extended Frege proof system. In particular, as exponential lower bounds are known for non-commutative formulas ([52]), this possibly suggests that such methods could even attack the full Frege system itself.

1.2 Algebraic circuit classes

Having motivated $\mathcal{C}$-IPS for restricted circuit classes $\mathcal{C}$, we now give formal definitions of the algebraic circuit classes of interest to this paper, all of which were studied independently in algebraic complexity. Some of them capture the state-of-art in our ability to prove lower bounds and provide efficient deterministic identity tests, so it is natural to attempt to fit them into the proof complexity framework. We define each and briefly explain what we know about it. As the list is long, the reader may consider skipping to the results (Section 1.3), and refer to the definitions of these classes as they arise.

Algebraic circuits and formula (over a fixed chosen field) compute polynomials via addition and multiplication gates, starting from the input variables and constants from the field. For background on algebraic circuits in general and their complexity measures we refer the reader to the survey of Shpilka and Yehudayoff [78]. We next define the restricted circuit classes that we will be studying in this paper.

1.2.1 Small-depth classes

We start by defining what are the simplest and most restricted classes of algebraic circuits. The first class simply represents polynomials as a sum of monomials. This is also called the *sparse representation* of the polynomial. Notationally we call this model $\sum\prod$ formulas (to capture the fact that polynomials computed in the class are represented simply as sums of products), but we will more often call these polynomials "sparse".

Definition 1.5. The class $\mathcal{C} = \sum\prod$ computes polynomials in their **sparse** representation, that is, as a sum of monomials. The graph of computation has two layers with an addition gate at the top and multiplication gates at the bottom. The **size** of a $\sum\prod$ circuit of a polynomial f is the product of the number of monomials in f, the number of variables, and the degree.

This class of circuits is what is used in the Nullstellensatz proof system. In our terminology $\sum\prod$-IPS_{LIN} is exactly the Nullstellensatz proof system.

Another restricted class of algebraic circuits is that of *depth*-3 *powering formulas* (sometimes also called "diagonal depth-3 circuits"). We will sometimes abbreviate this name as a "$\sum\bigwedge\sum$ formula", where $\bigwedge$ denotes the powering operation. Specifically, polynomials that are efficiently computed by small formulas from this class can be represented as sum of powers of linear functions. This model appears implicitly in Shpilka [75] and explicitly in the paper of Saxena [71].

Definition 1.6. The class of depth-3 powering formulas, denoted $\sum\bigwedge\sum$, computes polynomials of the following form

$$f(\overline{x}) = \sum_{i=1}^{s} \ell_i(\overline{x})^{d_i},$$

where $\ell_i(\overline{x})$ are linear functions. The degree of this $\sum\bigwedge\sum$ representation of f is $\max_i\{d_i\}$ and its size is $n \cdot \sum_{i=1}^{s}(d_i+1)$.

One reason for considering this class of circuits is that it is a simple, but non-trivial model that is somewhat well-understood. In particular, the partial derivative method of Nisan–Wigderson [54] implies lower bounds for this model and efficient polynomial identity testing algorithms are known ([71, 3, 26, 27, 24], as discussed further in Section 3.1).

We also consider a generalization of this model where we allow powering of low-degree polynomials.

Definition 1.7. The class $\sum\bigwedge\sum\prod^t$ computes polynomials of the following form

$$f(\overline{x}) = \sum_{i=1}^{s} f_i(\overline{x})^{d_i}\ ,$$

where the degree of the $f_i(\overline{x})$ is at most t. The size of this representation is $\binom{n+t}{t} \cdot \sum_{i=1}^{s}(d_i+1)$.

We remark that the reason for defining the size this way is that we think of the f_i as represented as sum of monomials (there are $\binom{n+t}{t}$ n-variate monomials of degree at most t) and the size captures the complexity of writing this as an algebraic formula. This model is the simplest that requires the method of *shifted partial derivatives* of Kayal [46, 36] to establish lower bounds, and this has recently been generalized to obtain polynomial identity testing algorithms ([21], as discussed further in Section 3.1).

1.2.2 Oblivious algebraic branching programs

Algebraic branching programs (ABPs) form a model whose computational power lies between that of algebraic circuits and algebraic formulas, and certain *read-once* and *oblivious* ABPs are a natural setting for studying the *partial derivative matrix* lower bound technique of Nisan [52].

Definition 1.8 (Nisan [52])**.** An **algebraic branching program (ABP) with unrestricted weights** of **depth** D and **width** $\leq r$, on the variables $x_1,\dots,x_n$, is a directed acyclic graph such that:

- The vertices are partitioned in $D+1$ layers $V_0,\dots,V_D$, so that $V_0=\{s\}$ (s is the source node), and $V_D=\{t\}$ (t is the sink node). Further, each edge goes from V_{i-1} to V_i for some $0<i\leq D$.
- $\max|V_i| \leq r$.
- Each edge e is weighted with a polynomial $f_e \in \mathbb{F}[\overline{x}]$.

The **(individual) degree** d of the ABP is the maximum (individual) degree of the edge polynomials f_e. The **size** of the ABP is the product $n\cdot r\cdot d\cdot D$.

Each s-t path is said to compute the polynomial which is the product of the labels of its edges, and the algebraic branching program itself computes the sum over all s-t paths of such polynomials.

There are also the following restricted ABP variants.

- An algebraic branching program is said to be **oblivious** if for every layer ℓ, all the edge labels in that layer are univariate polynomials in a single variable x_{i_ℓ}.
- An oblivious branching program is said to be a **read-once** oblivious ABP (roABP) if each x_i appears in the edge label of exactly one layer, so that $D = n$. That is, each x_i appears in the edge labels in exactly one layer. The layers thus define an **order of the variables**, which will be $x_1 < \cdots < x_n$ if not otherwise specified.
- An oblivious branching program is said to be a **read-k** oblivious ABP if each variable x_i appears in the edge labels of exactly k layers, so that $D = kn$.
- An ABP is **non-commutative** if each f_e is from the ring $\mathbb{F}\langle\overline{x}\rangle$ of non-commuting variables and has $\deg f_e \le 1$, so that the ABP computes a non-commutative polynomial.

Intuitively, roABPs are the algebraic analog of read-once Boolean branching programs, the non-uniform model of the class RL, which are well-studied in Boolean complexity. Algebraically, roABPs are also well-studied. In particular, roABPs are essentially equivalent to non-commutative ABPs ([27]), a model at least as strong as non-commutative formulas. That is, as an roABP reads the variables in a fixed order (hence not using commutativity) it can be almost directly interpreted as a non-commutative ABP. Conversely, as non-commutative multiplication is ordered, one can interpret a non-commutative polynomial in a read-once fashion by (commutatively) exponentiating a variable to its index in a monomial. For example, the non-commutative $xy - yx$ can be interpreted commutatively as $x^1y^2 - y^1x^2 = xy^2 - x^2y$, and one can show that this conversion preserves the relevant ABP complexity ([27]). The study of non-commutative ABPs dates to Nisan [52], who proved lower bounds for non-commutative ABPs (and thus also for roABPs, in any order). In a sequence of more recent papers, polynomial identity testing algorithms were devised for roABPs ([64, 25, 27, 24, 2], see also Section 3.1). In terms of proof complexity, Tzameret [83] studied a proof system with lines given by roABPs, and recently Li, Tzameret and Wang [50] (Theorem 1.4) showed that IPS over non-commutative formulas is essentially equivalent in power to the Frege proof system. Due to the close connections between non-commutative ABPs and roABPs, this last result suggests the importance of proving lower bounds for roABP-IPS as a way of attacking lower bounds for the Frege proof system (although we obtain roABP-IPS_{LIN} lower bounds without obtaining non-commutative-IPS_{LIN} lower bounds).

Finally, we mention that recently Anderson, Forbes, Saptharishi, Shpilka, and Volk [6] obtained exponential lower bounds for read-k oblivious ABPs (when $k = o(\log n / \log\log n)$) as well as a slightly subexponential polynomial identity testing algorithm.

1.2.3 Multilinear formulas

The last model that we consider is that of multilinear formulas.

Definition 1.9 (Multilinear formula)**.** An algebraic formula is a **multilinear formula** if the polynomial computed by *each* gate of the formula is multilinear (as a formal polynomial, that is, as an element of $\mathbb{F}[x_1, \ldots, x_n]$). The **product depth** is the maximum number of multiplication gates on any input-to-output path in the formula.

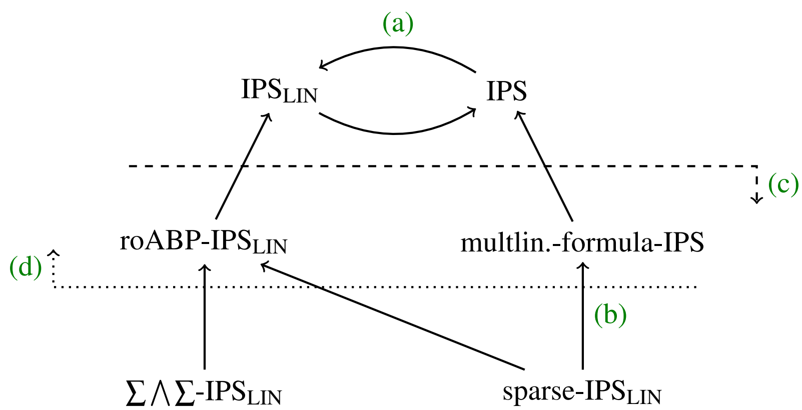

Figure 1: This diagram summarizes the main results of this paper, listing some of the upper bounds given in Section 4, as well as some of the functional lower bounds of Section 5. The lower bounds via multiples (Section 7) are not included here, as those lower bounds are slightly non-standard. Arrows indicate efficient simulation of proof systems, where unlabelled arrows are straightforward. Labelled results are as follows. (a) The simulation of general IPS by linear-IPS is given by Theorem 4.4. (b) The simulation of sparse-IPS$_{\text{LIN}}$ (equivalent to the Nullstellensatz system) by multilinear-formula-IPS is given by Theorem 4.12. (c) The region below the dashed line consists of proof systems where we prove superpolynomial lower bounds (Theorem 5.15). (d) The region above the dotted line can efficiently refute the subset-sum axiom $\sum_i x_i + 1 = 0$ over $\overline{x} \in \{0,1\}^n$ (Corollary 4.14, Theorem 4.15). In contrast, Polynomial Calculus cannot efficiently refute this system of equations ([41]).

Raz [63, 62] proved quasi-polynomial lower bounds for multilinear formulas and separated multilinear formulas from multilinear circuits. Raz and Yehudayoff proved exponential lower bounds for small-depth multilinear formulas [68]. Only slightly sub-exponential polynomial identity testing algorithms are known for small-depth multilinear formulas ([57]).

1.3 Our results and techniques

In this section we summarize our results and techniques, stating some results in less than full generality to more clearly convey the result. Figure 1 presents an extremely abbreviated listing of some highlights of our results. We present the results in the order that we later prove them. We start by giving upper bounds for IPS proofs (Section 1.3.1). We then describe our functional lower bounds and the IPS$_{\text{LIN}}$ lower bounds they imply (Section 1.3.2). Finally, we discuss lower bounds for multiples and state our corresponding lower bounds for IPS (Section 1.3.3).

1.3.1 Upper bounds for proofs within subclasses of IPS

Various previous articles have studied restricted algebraic proof systems and shown non-trivial upper bounds. The general simulation (Theorem 1.2) of Grochow and Pitassi [35] showed that the formula-IPS and circuit-IPS systems can simulate powerful proof systems such as Frege and Extended Frege, respectively. Li, Tzameret and Wang [50] (Theorem 1.4) have shown that even non-commutative-formula-IPS can simulate Frege. Grigoriev and Hirsch [32] have shown that proofs manipulating depth-3 algebraic formulas can refute hard axioms such as the *pigeonhole principle*, the *subset-sum axiom*, and *Tseitin tautologies*. Raz and Tzameret [66, 65] somewhat strengthened their results by restricting the proof to depth-3 *multilinear* proofs (in a *dynamic* system, see Appendix A).

However, these upper bounds are for proof systems (IPS or otherwise) for which no proof lower bounds are known. In this article we also study upper bounds for restricted subsystems of IPS. In particular, we compare linear-IPS versus the full IPS system, and show that even for restricted $\mathcal{C}$, $\mathcal{C}$-IPS can refute interesting unsatisfiable systems of equations arising from NP-complete problems (and we will obtain corresponding proof lower bounds for these $\mathcal{C}$-IPS systems).

Our first upper bound is to show that linear-IPS can simulate the full IPS proof system when the axioms are computationally simple, which essentially resolves a question of Grochow and Pitassi [35, Open Question 1.13[3]].

Theorem 1.10 (Theorem 4.4)**.** *For $|\mathbb{F}| \geq \mathsf{poly}(d)$, if $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ are degree-d polynomials computable by size-s algebraic formulas (or circuits) and they have a size-t formula-*IPS *(or circuit-*IPS*) refutation, then they also have a size-*$\mathsf{poly}(d,s,t)$ *formula-*IPS_{LIN} *(circuit-*IPS_{LIN}*, resp.) refutation.*

This theorem is established by pushing the "non-linear" dependencies on the axioms into the IPS refutation itself, which is possible as the axioms are assumed to themselves be computable by small circuits. We note that Grochow and Pitassi [35] showed such a conversion, but only for IPS refutations computable by sparse polynomials. Also, we remark that this result holds even for circuits of unbounded degree, as opposed to just those of polynomial degree.

We then turn our attention to IPS involving only restricted classes of algebraic circuits, and show that they are complete proof systems. This is clear for complete models of algebraic circuits such as sparse polynomials, depth-3 powering formulas [4] and roABPs. The models of sparse-IPS_{LIN} and roABP-IPS_{LIN} can efficiently simulate the Nullstellensatz proof system measured in terms of number of monomials, as the former is equivalent to this system, and the latter follows as sparse polynomials have small roABPs. Note that depth-3 powering formulas cannot efficiently compute sparse polynomials in general (Theorem 6.9) so cannot efficiently simulate the Nullstellensatz system. For multilinear formulas, showing completeness (much less an efficient simulation of sparse-IPS_{LIN}) is more subtle as not every polynomial is multilinear, however the result can be obtained by a careful multilinearization.

Theorem 1.11 (Theorem 4.7 and Theorem 4.12)**.** *The proof systems of sparse-*IPS_{LIN}*, *$\sum\bigwedge\sum$-IPS_{LIN} *(in large characteristic fields), and roABP-*IPS_{LIN} *are complete proof systems (for systems of polynomials with no Boolean solutions). The multilinear-formula-*IPS_{LIN} *proof system is not complete, but the depth-2 multilinear-formula-*$\text{IPS}_{LIN'}$ *proof system is complete (for multilinear axioms) and can polynomially simulate sparse-*IPS_{LIN} *(for low-degree axioms).*

However, we recall that multilinear-formula-IPS_{LIN} *is* complete when refuting unsatisfiable CNF formulas (Theorem 1.2).

[3]This refers to the following version: `http://arxiv.org/abs/1404.3820v1`.

[4]Showing that depth-3 powering formulas are complete (in large characteristic) can be seen from the fact that any multilinear monomial can be computed in this model, see for example Fischer [19].

We next consider the equation $\sum_{i=1}^{n} \alpha_i x_i - \beta$ along with the Boolean axioms $\{x_i^2 - x_i\}_i$. Deciding whether this system of equations is satisfiable is the NP-complete *subset-sum* problem, and thus we do not expect small refutations in general (unless $\mathsf{NP} = \mathsf{coNP}$). Indeed, Impagliazzo, Pudlák, and Sgall [41] (Theorem A.4) have shown lower bounds for refutations in the *polynomial calculus* system (and thus also the Nullstellensatz system) even when $\overline{\alpha} = \overline{1}$. Specifically, they showed that such refutations require both $\Omega(n)$-degree and $\exp(\Omega(n))$-many monomials, matching the worst-case upper bounds for these complexity measures. In the language of this paper, they gave $\exp(\Omega(n))$-size lower bounds for refuting this system in $\sum\prod$-IPS$_{\text{LIN}}$ (which is equivalent to the Nullstellensatz proof system). In contrast, we establish here $\mathsf{poly}(n)$-size refutations for $\overline{\alpha} = \overline{1}$ in the restricted proof systems of roABP-IPS$_{\text{LIN}}$ and *depth-3* multilinear-formula-IPS$_{\text{LIN}}$ (despite the fact that multilinear-formula-IPS$_{\text{LIN}}$ is not complete).

Theorem 1.12 (Corollary 4.14 and Theorem 4.15)**.** *Let $\mathbb{F}$ be a field of characteristic* $\mathrm{char}(\mathbb{F}) > n$. *Then the system of polynomial equations $\sum_{i=1}^{n} x_i - \beta$, $\{x_i^2 - x_i\}_{i=1}^{n}$ is unsatisfiable for $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$, and there are explicit* $\mathsf{poly}(n)$*-size refutations within roABP*-IPS$_{\text{LIN}}$*, as well as within depth-3 multilinear-formula*-IPS$_{\text{LIN}}$.

This theorem is proven by noting that the polynomial $p(t) := \prod_{k=0}^{n}(t-k)$ vanishes on $\sum_i x_i$ modulo $\{x_i^2 - x_i\}_{i=1}^{n}$, but $p(\beta)$ is a nonzero constant. This implies that $\sum_i x_i - \beta$ divides $p(\sum_i x_i) - p(\beta)$. Denoting the quotient by $f(\overline{x})$, it follows that $\frac{1}{-p(\beta)} \cdot f(\overline{x}) \cdot (\sum_i x_i - \beta) \equiv 1 \mod \{x_i^2 - x_i\}_{i=1}^{n}$, which is nearly a linear-IPS refutation except for the complexity of establishing this relation over the Boolean cube. We show that the quotient f is easily expressed as a depth-3 powering circuit. Unfortunately, proving the above equivalence to 1 modulo the Boolean cube is not possible in the depth-3 powering circuit model. However, by moving to more powerful models (such as roABPs and multilinear formulas) we can give proofs of this multilinearization to 1 and thus give proper IPS refutations.

1.3.2 Linear-IPS lower bounds via functional lower bounds

Having demonstrated the power of various restricted classes of IPS refutations by refuting the subset-sum axiom, we now turn to lower bounds. We give two paradigms for establishing lower bounds, the first of which we discuss here, named a *functional circuit lower bound*. This idea appeared in the work of Grigoriev and Razborov [34] as well as in the recent paper by Forbes, Kumar and Saptharishi [23]. We briefly motivate this type of lower bound as a topic of independent interest in algebraic circuit complexity, and then discuss the lower bounds we obtain and their implications to obtaining proof complexity lower bounds.

In Boolean complexity, the primary object of interest are *functions*. Generalizing slightly, one can even seek to compute functions $f : \{0,1\}^n \to \mathbb{F}$ for some field $\mathbb{F}$. In contrast, in algebraic complexity one seeks to compute *polynomials* as elements of the ring $\mathbb{F}[x_1, \ldots, x_n]$. These two regimes are tied by the fact that every polynomial $f \in \mathbb{F}[\overline{x}]$ induces a function $\hat{f} : \{0,1\}^n \to \mathbb{F}$ via the evaluation $\hat{f} : \overline{x} \mapsto f(\overline{x})$. That is, the polynomial f *functionally computes* the function $\hat{f}$. As an example, $x^2 - x$ functionally computes the zero function despite being a nonzero polynomial.

Traditional algebraic circuit lower bounds for the $n \times n$ permanent are lower bounds for computing perm_n as an element in the ring $\mathbb{F}[\{x_{i,j}\}_{1 \le i,j \le n}]$. This is a strong notion of "computing the permanent", while one can consider the weaker notion of functionally computing the permanent, that is, a polynomial $f \in \mathbb{F}[\{x_{i,j}\}]$ such that $f = \mathrm{perm}_n$ over $\{0,1\}^{n \times n}$, where f is not required to equal perm_n as a polynomial.

As $\mathrm{perm}_n : \{0,1\}^{n\times n} \to \mathbb{F}$ is #P-hard (for fields of large characteristic), assuming plausible conjectures (such as the polynomial hierarchy being infinite) it follows that *any* polynomial f which functionally computes perm_n must require large algebraic circuits. Unconditionally obtaining such a result is what we term a *functional lower bound*.

Goal 1.13 (Functional Circuit Lower Bound ([34, 23]))**.** Obtain an explicit function $\hat{f} : \{0,1\}^n \to \mathbb{F}$ such that for any polynomial $f \in \mathbb{F}[x_1,\ldots,x_n]$ satisfying $f(\overline{x}) = \hat{f}(\overline{x})$ for all $\overline{x} \in \{0,1\}^n$, it must be that f requires large algebraic circuits.

Obtaining such a result is challenging, in part because one must lower bound *all* polynomials agreeing with the function $\hat{f}$ (of which there are infinitely many). Prior work ([33, 34, 49]) has established functional lower bounds for functions when computing with polynomials over constant-sized finite fields, and the recent paper by Forbes, Kumar and Saptharishi [23] has established some lower bounds for any field.

While it is natural to hope that existing methods would yield such lower bounds, many lower-bound techniques inherently use that algebraic computation is *syntactic*. In particular, techniques involving partial derivatives (which include the partial derivative method of Nisan–Wigderson [54] and the shifted partial derivative method of Kayal [46, 36]) cannot as is yield functional lower bounds as knowing a polynomial on $\{0,1\}^n$ is not enough to conclude information about its partial derivatives.

We now explain how functional lower bounds imply lower bounds for linear-IPS refutations in certain cases. Suppose one considers refutations of the unsatisfiable polynomial system $f(\overline{x}), \{x_i^2 - x_i\}_{i=1}^n$. A linear-IPS refutation would yield an equation of the form $g(\overline{x}) \cdot f(\overline{x}) + \sum_i h_i(\overline{x}) \cdot (x_i^2 - x_i) = 1$ for some polynomials $g, h_i \in \mathbb{F}[\overline{x}]$. Viewing this equation modulo the Boolean cube, we have that $g(\overline{x}) \cdot f(\overline{x}) \equiv 1 \mod \{x_i^2 - x_i\}_i$. Equivalently, since $f(\overline{x})$ is unsatisfiable over $\{0,1\}^n$, we see that $g(\overline{x}) = 1/f(\overline{x})$ for $\overline{x} \in \{0,1\}^n$, as $f(\overline{x})$ is never zero so this fraction is well-defined. It follows that *if* the function $\overline{x} \mapsto 1/f(\overline{x})$ induces a functional lower bound then $g(\overline{x})$ must require large complexity, yielding the desired linear-IPS lower bound.

Thus, it remains to instantiate this program. While we are successful, we should note that this approach as is seems to only yield proof complexity lower bounds for systems with one non-Boolean axiom and thus cannot encode polynomial systems where each equation depends on $O(1)$ variables (such as those naturally arising from 3CNFs).

Our starting point is to observe that the subset-sum axiom already induces a weak form of functional lower bound, where the complexity is measured by degree.

Theorem 1.14 (Theorem 5.4)**.** *Let $\mathbb{F}$ be a field of a characteristic at least* $\mathsf{poly}(n)$ *and $\beta \notin \{0,\ldots,n\}$. Then $\sum_i x_i - \beta, \{x_i^2 - x_i\}_i$ is unsatisfiable and any polynomial $f \in \mathbb{F}[x_1,\ldots,x_n]$ with $f(\overline{x}) = \frac{1}{\sum_i x_i - \beta}$ for $\overline{x} \in \{0,1\}^n$, satisfies $\deg f \geq n$.*

A lower bound of $\lceil \frac{n}{2} \rceil + 1$ was previously established by Impagliazzo, Pudlák, and Sgall [41] (Theorem A.4), but the lower bound of n (which is tight) will be crucial for our results.

We then lift this result to obtain lower bounds for stronger models of algebraic complexity. In particular, by replacing "x_i" with "$x_i y_i$" we show that the function $\frac{1}{\sum_i x_i y_i - \beta}$ has maximal *evaluation dimension* between $\overline{x}$ and $\overline{y}$, which is some measure of interaction between the variables in $\overline{x}$ and those in $\overline{y}$ (see Section 3.3). This measure is essentially *functional*, so that one can lower bound this measure by understanding the functional behavior of the polynomial on finite sets such as the Boolean cube. Our lower bound for evaluation dimension follows by examining the above degree bound. Using known relations between this complexity measure and algebraic circuit classes, we can obtain lower bounds for depth-3 powering linear-IPS.

Theorem 1.15 (Theorem 5.10)**.** *Let $\mathbb{F}$ be a field of characteristic $\geq \mathsf{poly}(n)$ and $\beta \notin \{0,\ldots,n\}$. Then $\sum_{i=1}^{n} x_i y_i - \beta, \{x_i^2 - x_i\}_i, \{y_i^2 - y_i\}_i$ is unsatisfiable and any $\sum\bigwedge\sum$-IPS$_{\text{LIN}}$ refutation requires size $\geq$* $\exp(\Omega(n))$.

The above axiom only gets maximal interaction between the variables across a *fixed* partition of the variables. By introducing auxiliary variables we can create such interactions in variables across *any* partition of (some) of the variables. By again invoking results showing such structure implies computational hardness we obtain more linear-IPS lower bounds.

Theorem 1.16 (Theorem 5.15)**.** *Let $\mathbb{F}$ be a field of characteristic $\geq \mathsf{poly}(n)$ and $\beta \notin \{0,\ldots,\binom{2n}{2}\}$. Then $\sum_{i<j} z_{i,j} x_i x_j - \beta, \{x_i^2 - x_i\}_{i=1}^{n}, \{z_{i,j}^2 - z_{i,j}\}_{i<j}$ is unsatisfiable, and any roABP-*IPS$_{\text{LIN}}$ *refutation (in any order of the variables) requires* $\exp(\Omega(n))$ *size. Further, any multilinear-formula-*IPS *refutation requires $n^{\Omega(\log n)}$-size, and any product-depth-d multilinear-formula-*IPS *refutation requires $n^{\Omega((n/\log n)^{1/d}/d^2)}$-size.*

Note that our result for multilinear-formulas is not just for the linear-IPS system, but actually for the full multilinear-formula-IPS system. Thus, we show that even though roABP-IPS$_{\text{LIN}}$ and depth-3 multilinear formula-IPS$_{\text{LIN}'}$ can refute the subset-sum axiom in polynomial size, slight variants of this axiom do not have polynomial-size refutations.

1.3.3 Lower bounds for multiples

While the above paradigm can establish superpolynomial lower bounds for *linear*-IPS, it does not seem able to establish lower bounds for the general IPS proof system over non-multilinear polynomials, even for restricted classes. This is because such systems would induce equations such as $h(\overline{x})f(\overline{x})^2 + g(\overline{x})f(\overline{x}) \equiv 1 \mod \{x_i^2 - x_i\}_{i=1}^{n}$, where we need to design a computationally simple axiom f so that this equation implies at least one of h or g is of large complexity. In a linear-IPS proof it must be that h is zero, so that for any $\overline{x} \in \{0,1\}^n$ we can solve for $g(\overline{x})$, that is, $g(\overline{x}) = 1/f(\overline{x})$. However, in general knowing $f(\overline{x})$ does not uniquely determine $g(\overline{x})$ or $h(\overline{x})$, which makes this approach significantly more complicated. Further, even though we can efficiently simulate IPS by linear-IPS (Theorem 4.4) in general, this simulation increases the complexity of the proof so that even if one started with a $\mathcal{C}$-IPS proof for a restricted circuit class $\mathcal{C}$ the resulting IPS$_{\text{LIN}}$ proof may not be in $\mathcal{C}$-IPS$_{\text{LIN}}$.

For the reasons mentioned above, we introduce a second paradigm, called *lower bounds for multiples*, which can yield $\mathcal{C}$-IPS lower bounds for various restricted classes $\mathcal{C}$. We begin by defining this question.

Goal 1.17 (Lower Bounds for Multiples)**.** Design an explicit polynomial $f(\overline{x})$ such that for any nonzero $g(\overline{x})$ we have that the multiple $g(\overline{x})f(\overline{x})$ is hard to compute.

We now explain how such lower bounds yield IPS lower bounds. Consider the system $f, \{x_i^2 - x_i\}_i$ with a single non-Boolean axiom. An IPS refutation is a circuit $C(\overline{x}, y, \overline{z})$ such that $C(\overline{x}, 0, \overline{0}) = 0$ and $C(\overline{x}, f, \overline{x}^2 - \overline{x}) = 1$, where (as mentioned) $\overline{x}^2 - \overline{x}$ denotes $\{x_i^2 - x_i\}_i$. Expressing $C(\overline{x}, f, \overline{x}^2 - \overline{x})$ as a univariate in f, we obtain that $\sum_{i\geq 1} C_i(\overline{x},\overline{x}^2-\overline{x})f^i = 1-C(\overline{x},0,\overline{x}^2-\overline{x})$ for some polynomials C_i. For most natural measures of circuit complexity $1-C(\overline{x},0,\overline{x}^2-\overline{x})$ has complexity roughly bounded by that of C itself. Thus, we see that a multiple of f has a small circuit, as $\left(\sum_{i\geq 1} C_i(\overline{x},\overline{x}^2-\overline{x})f^{i-1}\right)\cdot f = 1-C(\overline{x},0,\overline{x}^2-\overline{x})$, and one can use the properties of the IPS refutation to show this is in fact a *nonzero* multiple. Thus, if we can show that all nonzero multiples of f require large circuits then we rule out a small IPS refutation.

We now turn to methods for obtaining polynomials with hard multiples. Intuitively if a polynomial f is hard then so should small modifications such as $f^2+x_1 f$, and this intuition is supported by the result of Kaltofen [44] which shows that if a polynomial has a small algebraic circuit then so do all of its factors. As a consequence, if a polynomial requires superpolynomially large algebraic circuits then so do all of its multiples. However, Kaltofen's [44] result is about *general* algebraic circuits, and there are very limited results about the complexity of factors of *restricted* algebraic circuits ([18, 56]) so that obtaining polynomials for hard multiples via factorization results seems difficult.

However, note that lower bound for multiples has a different order of quantifiers than the factoring question. That is, Kaltofen's [44] result speaks about the factors of *any* small circuit, while the lower bound for multiples speaks about the multiples of a *single* polynomial. Thus, it seems plausible that existing methods could yield such explicit polynomials, and indeed we show this is the case.

We begin by noting that obtaining lower bounds for multiples is a natural instantiation of the algebraic *hardness versus randomness* paradigm. In particular, Heintz–Schnorr [39] and Agrawal [1] showed that obtaining deterministic (black-box) polynomial identity testing algorithms implies lower bounds (see Section 3.1 for more on PIT), and we strengthen that connection here to lower bounds for multiples. We can actually instantiate this connection, and we use slight modifications of existing PIT algorithms to show that multiples of the determinant are hard in some models.

Theorem 1.18 (Informal Version of Theorem 6.2 and Theorem 6.7)**.** *Let $\mathcal{C}$ be a restricted class of n-variate algebraic circuits. Full derandomization of PIT algorithms for $\mathcal{C}$ yields a (weakly) explicit polynomial whose nonzero multiples require* $\exp(\Omega(n))$ *size as $\mathcal{C}$-circuits.*

In particular, when $\mathcal{C}$ is the class of sparse polynomials, depth-3 powering formulas, $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas (in characteristic zero), or "every-order" roABPs, then all nonzero multiples of the $n\times n$ determinant are $\exp(\Omega(n))$*-hard in these models.*

The above statement shows that *derandomization* implies *hardness*. We also partly address the converse direction by arguing (Section 6.1) that hardness-to-randomness construction of Kabanets and Impagliazzo [43] only requires lower bounds for multiples to derandomize PIT. Unfortunately, this direction is harder to instantiate for restricted classes as it requires lower bounds for classes with suitable closure properties.[5]

Unfortunately the above result is slightly unsatisfying from a proof complexity standpoint as the (exponential-size) lower bounds for the subclasses of IPS one can derive from the above result would involve the determinant polynomial as an axiom. While the determinant is efficiently computable, it is not computable by the above restricted circuit classes (indeed, the above result proves that). Thus, this would

[5] Although, we note that one can instantiate this connection with depth-3 powering formulas (or even $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas) using the lower bounds for multiples developed in this paper, building on the work of Forbes [21]. However, the resulting PIT algorithms are worse than those developed by Forbes [21].

not fit the real goal of proof complexity which seeks to show that there are statements whose proofs must be *superpolynomially larger* than the length of the statement. Thus, if we measure the size of the IPS proof and the axioms with respect to the same circuit measure, the lower bounds for multiples approach *cannot* establish such superpolynomial lower bounds.

However, we believe that lower bounds for multiples could lead, with further ideas, to proof complexity lower bounds in the conventional sense. That is, it seems plausible that by adding *extension variables* we can convert complicated axioms to simple, local axioms by tracing through the computation of that axiom. That is, consider the axiom $xyzw$. This can be equivalently written as $\{a-xy, b-zw, c-ab, c\}$, where this conversion is done by considering a natural algebraic circuit for $xyzw$, replacing each gate with a new variable, and adding an axiom ensuring the new variables respect the computation of the circuit. While we are unable to understand the role of extension variables in this article we aim to give as simple axioms as possible whose multiples are all hard as this may facilitate future work on extension variables.

We now discuss the lower bounds for multiples we obtain.[6]

Theorem 1.19 (Corollaries 6.9, 6.11, 6.13, 6.21, and 6.23)**.** *We obtain the following lower bounds for multiples.*

- *All nonzero multiples of* $x_1 \cdots x_n$ *require* $\exp(\Omega(n))$ *size as a depth-3 powering formula (over any field), or as a* $\sum \bigwedge \sum \prod^{O(1)}$ *formula (in characteristic zero).*
- *All nonzero multiples of* $(x_1+1)\cdots(x_n+1)$ *require* $\exp(\Omega(n))$*-many monomials.*
- *All nonzero multiples of* $\prod_i (x_i+y_i)$ *require* $\exp(\Omega(n))$ *width as an roABP in any order of the variables where* $\overline{x}$ *precedes* $\overline{y}$*.*
- *All nonzero multiples of* $\prod_{i<j}(x_i+x_j)$ *require* $\exp(\Omega(n))$ *width as an roABP in any order of the variables, as well as* $\exp(\Omega(n))$ *width as a read-twice oblivious ABP.*

We now briefly explain our techniques for obtaining these lower bounds, focusing on the simplest case of depth-3 powering formulas. It follows from the partial derivative method of Nisan and Wigderson [53] (see Kayal [45]) that such formulas require exponential size to compute the monomial $x_1 \dots x_n$ *exactly*. Forbes and Shpilka [26], in giving a PIT algorithm for this class, showed that this lower bound can be *scaled down* and *made robust*. That is, if one has a size-s depth-3 powering formula, it follows that *if* it computes a monomial $x_{i_1} \cdots x_{i_\ell}$ for distinct i_j then $\ell \le O(\log s)$ (so the lower bound is scaled down). One can then show that regardless of what this formula actually computes the *leading* monomial $x_{i_1}^{a_{i_1}} \cdots x_{i_\ell}^{a_{i_\ell}}$ (for distinct i_j and positive a_{i_j}) must have that $\ell \le O(\log s)$. One then notes that leading monomials are *multiplicative*. Thus, for any nonzero g the leading monomial of $g \cdot x_1 \dots x_n$ involves n variables so that if $g \cdot x_1 \dots x_n$ is computed in size-s then $n \le O(\log s)$, giving $s \ge \exp(\Omega(n))$ as desired. One can then obtain the other lower bounds using the same idea, though for roABPs one needs to define a leading *diagonal* (refining an argument of Forbes–Shpilka [25]).

We now conclude our IPS lower bounds.

[6]While we discussed functional lower bounds for multilinear formulas, this class is not interesting for the lower bounds for multiples question. This is because a multiple of a multilinear polynomial may not be multilinear, and thus clearly cannot have a multilinear formula.

Theorem 1.20 (Theorem 7.2, Theorem 7.3)**.** *We obtain the following lower bounds for subclasses of IPS.*

- *In characteristic zero, the system of polynomials* $x_1 \cdots x_n, x_1 + \cdots + x_n - n, \{x_i^2 - x_i\}_{i=1}^n$ *is unsatisfiable, and any* $\sum\bigwedge\sum$-IPS *refutation requires* $\exp(\Omega(n))$ *size.*
- *In characteristic* $> n$*, the system of polynomials,* $\prod_{i<j}(x_i + x_j - 1), x_1 + \cdots + x_n - n, \{x_i^2 - x_i\}_i$ *is unsatisfiable, and any roABP*-IPS *refutation (in any order of the variables), must be of size* $\exp(\Omega(n))$.

Note that the first result is a non-standard encoding of $1 = \mathrm{AND}(x_1, \ldots, x_n) = 0$. Similarly, the second is a non-standard encoding of $\mathrm{AND}(x_1, \ldots, x_n) = 1$ yet $\mathrm{XOR}(x_i, x_j) = 1$ for all i, j.

1.4 Subsequent developments

After the publication of the preliminary version of this article [29], there was follow-up work by Alekseev, Hirsch, Tzameret and Grigoriev [5] who proved a conditional superpolynomial size lower bound against general (unrestricted) IPS refutations over the rational numbers of a subset-sum principle with large coefficients. The hard instance in [5] is the so-called *Binary Value Principle* $\sum_{i=1}^n 2^{i-1} x_i + 1 = 0$, for Boolean variables x_i, and the lower bound is conditioned on the Shub–Smale hypothesis stating that factorial numbers cannot be computed by small (poly-logarithmic size) algebraic expressions consisting of the constants $0, 1, -1$ only (that is, variable-free expressions) [79].

1.5 Organization

The rest of the paper is organized as follows. Section 2 contains the basic notation for the paper. In Section 3 we give background from algebraic complexity, including several important complexity measures such as coefficient dimension and evaluation dimension (see Section 3.2 and Section 3.3). We present our upper bounds for IPS in Section 4. In Section 5 we give our functional lower bounds and from them obtain lower bounds for $\mathrm{IPS}_{\mathrm{LIN}}$. Section 6 contains our lower bounds for multiples of polynomials and in Section 7 we derive lower bounds for IPS using them. In Section 8 we list some problems which were left open by this work.

In Appendix A we describe various other algebraic proof systems and their relations to IPS, such as the dynamic Polynomial Calculus of Clegg, Edmonds, and Impagliazzo [13], the ordered formula proofs of Tzameret [83], and the multilinear proofs of Raz and Tzameret [66]. In Appendix B we give an explicit description of a multilinear polynomial occurring in our IPS upper bounds.

2 Notation

In this section we briefly describe notation used in this paper. We denote $[n] := \{1, \ldots, n\}$. For a vector $\overline{a} \in \mathbb{N}^n$, we denote $\overline{x}^{\overline{a}} := x_1^{a_1} \cdots x_n^{a_n}$ so that in particular $\overline{x}^{\overline{1}} = \prod_{i=1}^n x_i$. The (total) degree of a monomial $\overline{x}^{\overline{a}}$, denoted $\deg \overline{x}^{\overline{a}}$, is equal to $|\overline{a}|_1 := \sum_i a_i$, and the individual degree, denoted $\mathrm{ideg}\,\overline{x}^{\overline{a}}$, is equal to $|\overline{a}|_\infty := \max\{a_i\}_i$. A monomial $\overline{x}^{\overline{a}}$ depends on $|\overline{a}|_0 := |\{i : a_i \neq 0\}|$ many variables. Degree and individual degree can be defined for a polynomial f, denoted $\deg f$ and $\mathrm{ideg}\, f$ respectively, by taking the maximum over all monomials with nonzero coefficients in f. We will sometimes compare vectors $\overline{a}$ and $\overline{b}$ as "$\overline{a} \leq \overline{b}$", which is to be interpreted coordinatewise. We will use $\prec$ to denote a monomial order on $\mathbb{F}[\overline{x}]$, see Section 3.6.

Polynomials will often be written out in their monomial expansion. At various points we will need to extract coefficients from polynomials. When "taking the coefficient of $\overline{y}^{\overline{b}}$ in $f \in \mathbb{F}[\overline{x},\overline{y}]$" we mean that both $\overline{x}$ and $\overline{y}$ are treated as variables and thus the coefficient returned is a scalar in $\mathbb{F}$, and this will be denoted $\mathrm{Coeff}_{\overline{y}^{\overline{b}}}(f)$. However, when "taking the coefficient of $\overline{y}^{\overline{b}}$ in $f \in \mathbb{F}[\overline{x}][\overline{y}]$" we mean that $\overline{x}$ is now part of the ring of scalars, so the coefficient will be a polynomial in $\mathbb{F}[\overline{x}]$, and this coefficient will be denoted $\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)$.

For a vector $\overline{a} \in \mathbb{N}^n$ we denote $\overline{a}_{\leq i} \in \mathbb{N}^i$ to be the restriction of $\overline{a}$ to the first i coordinates. For a set $S \subseteq [n]$ we let $\overline{S}$ denote the complement set. We will denote the size-k subsets of $[n]$ by $\binom{[n]}{k}$. We will use $\mathrm{ml} : \mathbb{F}[\overline{x}] \to \mathbb{F}[\overline{x}]$ to denote the multilinearization operator, defined by Theorem 3.12. We will use $\overline{x}^2 - \overline{x}$ to denote the set of equations $\{x_i^2 - x_i\}_i$.

To present algorithms that are field independent, this paper works in a model of computation where field operations (such as addition, multiplication, inversion and zero-testing) over $\mathbb{F}$ can be computed at unit cost, see for example Forbes [20, Appendix A]. We say that an algebraic circuit is **t-explicit** if it can be constructed in t steps in this unit-cost model.

3 Algebraic complexity theory background

In this section we state some known facts regarding the algebraic circuit classes that we will be studying. We also give some important definitions that will be used later in the paper.

3.1 Polynomial identity testing

In the *polynomial identity testing (PIT)* problem, we are given an algebraic circuit computing some polynomial f, and we have to determine whether "$f \equiv 0$". That is, we are asking whether f is the zero polynomial in $\mathbb{F}[x_1,\ldots,x_n]$. By the Polynomial Identity Lemma[7], if $0 \neq f \in \mathbb{F}[\overline{x}]$ is a polynomial of degree $\leq d$ and $S \subseteq \mathbb{F}$, and $\overline{\alpha} \in S^n$ is chosen uniformly at random, then $f(\overline{\alpha}) = 0$ with probability at most [8] $d/|S|$. Thus, given the circuit, we can perform these evaluations efficiently, giving an efficient randomized procedure for deciding whether "$f \equiv 0$?". It is an important open problem to find a derandomization of this algorithm, that is, to find a *deterministic* procedure for PIT that runs in polynomial time (in the size of circuit) (cf. [78, Chap. 4] or the surveys [72, 73]).

Note that in this randomized algorithm we only use the circuit to compute the evaluation $f(\overline{\alpha})$. Such algorithms are said to run in the *black-box* model. In contrast, an algorithm that can access the internal structure of the circuit runs in the *white-box* model. It is a folklore result that efficient deterministic black-box algorithms are equivalent to constructions of small *hitting sets*. That is, a hitting set is a set of

[7] This is sometimes called the "Schwartz–Zippel–DeMillo–Lipton Lemma" following [86, 74, 17], however, this lemma goes back to Øystein Ore, 1922 [58] and was subsequently rediscovered by other authors; see [8] who unearthed the history of this lemma.

[8] Note that this is non-trivial only if $d < |S| \leq |\mathbb{F}|$, which in particular implies that f is not the zero function.

inputs so that any nonzero circuit from the relevant class evaluates to nonzero on at least one of the inputs in the set. For more on PIT we refer to the survey of Shpilka and Yehudayoff [78].

A related notion to that of a hitting set is that of a *generator*, which is essentially a low-dimensional curve whose image contains a hitting set. The equivalence between hitting sets and generators can be found in the above mentioned survey.

Definition 3.1. Let $\mathcal{C} \subseteq \mathbb{F}[x_1,\ldots,x_n]$ be a set of polynomials. A polynomial $\overline{\mathcal{G}} : \mathbb{F}^\ell \to \mathbb{F}^n$ is a **generator for** $\mathcal{C}$ with **seed length** ℓ if for all $f \in \mathcal{C}$, $f \equiv 0$ iff $f \circ \overline{\mathcal{G}} \equiv 0$. That is, $f(\overline{x}) = 0$ in $\mathbb{F}[\overline{x}]$ iff $f(\overline{\mathcal{G}}(\overline{y})) = 0$ in $\mathbb{F}[\overline{y}]$.

In words, a generator for a circuit class $\mathcal{C}$ is a mapping $\overline{\mathcal{G}} : \mathbb{F}^\ell \to \mathbb{F}^n$, such that for any nonzero polynomial f, computed by a circuit from $\mathcal{C}$, it holds that the composition $f(\overline{\mathcal{G}})$ is nonzero as well. By considering the image of $\overline{\mathcal{G}}$ on S^ℓ, where $S \subseteq \mathbb{F}$ is of polynomial size, we obtain a hitting set for $\mathcal{C}$.

We now list some existing work on derandomizing PIT for some of the classes of polynomials we study in this paper.

Sparse Polynomials: There are many papers giving efficient black-box PIT algorithms for $\sum\prod$ formulas. For example, Klivans and Spielman [47] gave a hitting set of polynomial size.

Depth-3 **Powering Formulas:** Saxena [71] gave a polynomial-time white-box PIT algorithm and Forbes, Shpilka, and Saptharishi [24] gave a $s^{O(\lg\lg s)}$-size hitting set for size-s depth-3 powering formulas.

$\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ **Formulas:** Forbes [21] gave an $s^{O(\lg s)}$-size hitting set for size-s $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas (in large characteristic).

Read-once Oblivious ABPs: Raz and Shpilka [64] gave a polynomial-time white-box PIT algorithm. A long sequence of papers calumniated in the work of Agrawal, Gurjar, Korwar, and Saxena [2], who gave a $s^{O(\lg s)}$-sized hitting set for size-s roABPs.

Read-k **Oblivious ABPs:** Recently, Anderson, Forbes, Saptharishi, Shpilka and Volk [6] obtained a white-box PIT algorithm running in time $2^{\tilde{O}(n^{1-1/2^{k-1}})}$ for n-variate $\mathsf{poly}(n)$-sized read-k oblivious ABPs.

3.2 Coefficient dimension and roABPs

This paper proves various lower bounds on roABPs using a complexity measures known as *coefficient dimension*. In this section, we define this measures and recall basic properties. Full proofs of these claims can be found for example in the thesis of Forbes [20].

We first define the *coefficient matrix* of a polynomial, called the "partial derivative matrix" in the prior work of Nisan [52] and Raz [63]. This matrix is formed from a polynomial $f \in \mathbb{F}[\overline{x},\overline{y}]$ by arranging its coefficients into a matrix. That is, the coefficient matrix has rows indexed by monomials $\overline{x}^{\overline{a}}$ in $\overline{x}$, columns indexed by monomials $\overline{y}^{\overline{b}}$ in $\overline{y}$, and the $(\overline{x}^{\overline{a}},\overline{y}^{\overline{b}})$-entry is the coefficient of $\overline{x}^{\overline{a}}\overline{y}^{\overline{b}}$ in the polynomial f. We now define this matrix, recalling that $\mathrm{Coeff}_{\overline{x}^{\overline{a}}\overline{y}^{\overline{b}}}(f)$ is the coefficient of $\overline{x}^{\overline{a}}\overline{y}^{\overline{b}}$ in f.

Definition 3.2. Consider $f \in \mathbb{F}[\overline{x},\overline{y}]$. Define the **coefficient matrix of** f as the scalar matrix

$$(C_f)_{\overline{a},\overline{b}} := \mathrm{Coeff}_{\overline{x}^{\overline{a}}\overline{y}^{\overline{b}}}(f)\ ,$$

where coefficients are taken in $\mathbb{F}[\overline{x},\overline{y}]$, for $|\overline{a}|_1, |\overline{b}|_1 \le \deg f$.

We now give the related definition of *coefficient dimension*, which looks at the dimension of the row- and column-spaces of the coefficient matrix. Recall that $\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)$ extracts the coefficient of $\overline{y}^{\overline{b}}$ in f, where f is treated as a polynomial in $\mathbb{F}[\overline{x}][\overline{y}]$.

Definition 3.3. Let $\mathbf{Coeff}_{\overline{x}|\overline{y}} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{x}]}$ be the **space of** $\mathbb{F}[\overline{x}][\overline{y}]$ **coefficients**, defined by

$$\mathbf{Coeff}_{\overline{x}|\overline{y}}(f) := \left\{\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)\right\}_{\overline{b}\in\mathbb{N}^n} ,$$

where coefficients of f are taken in $\mathbb{F}[\overline{x}][\overline{y}]$.

Similarly, define $\mathbf{Coeff}_{\overline{y}|\overline{x}} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{y}]}$ by taking coefficients in $\mathbb{F}[\overline{y}][\overline{x}]$.

The following basic lemma shows that the rank of the coefficient matrix equals the coefficient dimension, which follows from simple linear algebra.

Lemma 3.4 (Nisan [52])**.** *Consider $f \in \mathbb{F}[\overline{x},\overline{y}]$. Then the rank of the coefficient matrix C_f obeys*

$$\operatorname{rank} C_f = \dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f) = \dim \mathbf{Coeff}_{\overline{y}|\overline{x}}(f) . \qquad \square$$

Thus, the ordering of the partition ($(\overline{x},\overline{y})$ versus $(\overline{y},\overline{x})$) does not matter in terms of the resulting dimension. The above matrix-rank formulation of coefficient dimension can be rephrased in terms of low-rank decompositions.

Lemma 3.5. *Let $f \in \mathbb{F}[\overline{x},\overline{y}]$. Then $\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f)$ equals the minimum r such that there are $\overline{g} \in \mathbb{F}[\overline{x}]^r$ and $\overline{h} \in \mathbb{F}[\overline{y}]^r$ such that f can be written as $f(\overline{x},\overline{y}) = \sum_{i=1}^r g_i(\overline{x})h_i(\overline{y})$.* $\square$

We now state a convenient normal form for roABPs (see for example Forbes [20, Corollary 4.4.2]).

Lemma 3.6. *A polynomial $f \in \mathbb{F}[x_1,\dots,x_n]$ is computed by width-r roABP iff there exist matrices $A_i(x_i) \in \mathbb{F}[x_i]^{r\times r}$ of (individual) degree $\le \deg f$ such that $f = (\prod_{i=1}^n A_i(x_i))_{1,1}$. Further, this equivalence preserves explicitness of the roABPs up to $\mathsf{poly}(n,r,\deg f)$-factors.* $\square$

By splitting an roABP into such variable-disjoint inner-products one can obtain a lower bound for roABP width via coefficient dimension. In fact, this complexity measure *characterizes* roABP width.

Lemma 3.7. *Let $f \in \mathbb{F}[x_1,\dots,x_n]$ be a polynomial. If f is computed by a width-r roABP then $r \ge \max_i \dim \mathbf{Coeff}_{\overline{x}_{\le i}|\overline{x}_{>i}}(f)$. Further, f is computable by a width-$\big(\max_i \dim \mathbf{Coeff}_{\overline{x}_{\le i}|\overline{x}_{>i}}(f)\big)$ roABP.* $\square$

Using this complexity measure it is rather straightforward to prove the following closure properties of roABPs.

Fact 3.8. *If $f,g \in \mathbb{F}[\overline{x}]$ are computable by width-r and width-s roABPs respectively, then*

- *$f+g$ is computable by a width-$(r+s)$ roABP.*
- *$f\cdot g$ is computable by a width-(rs) roABP.*

Further, roABPs are also closed under the following operations.

- *If $f(\overline{x},\overline{y}) \in \mathbb{F}[\overline{x},\overline{y}]$ is computable by a width-r roABP in some order of the variables then the partial substitution $f(\overline{x},\overline{\alpha})$, for $\overline{\alpha} \in \mathbb{F}^{|\overline{y}|}$, is computable by a width-r roABP in the induced order on $\overline{x}$, where the degree of this roABP is bounded by the degree of the roABP for f.*

- *If $f(z_1,\ldots,z_n)$ is computable by a width-r roABP in order of the variables $z_1 < \cdots < z_n$, then $f(x_1y_1,\ldots,x_ny_n)$ is computable by a $\mathsf{poly}(r, \operatorname{ideg} f)$-width roABP in order of the variables $x_1 < y_1 < \cdots < x_n < y_n$.*

Further, these operations preserve the explicitness of the roABPs up to polynomial factors in all relevant parameters.

We now state the extension of these techniques which yield lower bounds for read-k oblivious ABPs, as recently obtained by Anderson, Forbes, Saptharishi, Shpilka and Volk [6].

Theorem 3.9 ([6]). *Let $f \in \mathbb{F}[x_1,\ldots,x_n]$ be a polynomial computed by a width-w read-k oblivious ABP. Then there exists a partition $\overline{x} = (\overline{u},\overline{v},\overline{w})$ such that*

1. $|\overline{u}|, |\overline{v}| \geq n/k^{O(k)}$.
2. $|\overline{w}| \leq n/10$.
3. $\dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}(f_{\overline{w}}) \leq w^{2k}$, *where $f_{\overline{w}}$ is f as a polynomial in $\mathbb{F}(\overline{w})[\overline{u},\overline{v}]$.* □

3.3 Evaluation dimension

While coefficient dimension measures the size of a polynomial $f(\overline{x},\overline{y})$ by taking all coefficients in $\overline{y}$, *evaluation dimension* is a complexity measure due to Saptharishi [70] that measures the size by taking all possible evaluations in $\overline{y}$ over the field. This measure will be important for our applications as one can restrict such evaluations to the Boolean cube and obtain circuit lower bounds for computing $f(\overline{x},\overline{y})$ as a *polynomial* via its induced *function* on the Boolean cube. We begin with the definition.

Definition 3.10 (Saptharishi [70]). Let $S \subseteq \mathbb{F}$. Let $\mathbf{Eval}_{\overline{x}|\overline{y},S} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{x}]}$ be the **space of** $\mathbb{F}[\overline{x}][\overline{y}]$ **evaluations over** S, defined by

$$\mathbf{Eval}_{\overline{x}|\overline{y},S}(f(\overline{x},\overline{y})) := \left\{ f(\overline{x},\overline{\beta}) \right\}_{\overline{\beta}\in S^{|\overline{y}|}} .$$

Define $\mathbf{Eval}_{\overline{x}|\overline{y}} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{x}]}$ to be $\mathbf{Eval}_{\overline{x}|\overline{y},S}$ when $S = \mathbb{F}$.

Similarly, define $\mathbf{Eval}_{\overline{y}|\overline{x},S} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{y}]}$ by replacing $\overline{x}$ with all possible evaluations $\overline{\alpha} \in S^{|\overline{x}|}$, and likewise define $\mathbf{Eval}_{\overline{y}|\overline{x}} : \mathbb{F}[\overline{x},\overline{y}] \to 2^{\mathbb{F}[\overline{y}]}$.

The equivalence between evaluation dimension and coefficient dimension was shown by Forbes–Shpilka [27] by appealing to interpolation.

Lemma 3.11 (Forbes–Shpilka [27]). *Let $f \in \mathbb{F}[\overline{x},\overline{y}]$. For any $S \subseteq \mathbb{F}$ we have that $\mathbf{Eval}_{\overline{x}|\overline{y},S}(f) \subseteq \operatorname{span} \mathbf{Coeff}_{\overline{x}|\overline{y}}(f)$ so that $\dim \mathbf{Eval}_{\overline{x}|\overline{y},S}(f) \leq \dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f)$. In particular, if $|S| > \operatorname{ideg} f$ then $\dim \mathbf{Eval}_{\overline{x}|\overline{y},S}(f) = \dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f)$.* □

While evaluation dimension and coefficient dimension are equivalent when the field is large enough, when restricting our attention to inputs from the Boolean cube this equivalence no longer holds (in particular, we have to consider all polynomials that obtain the same values on the Boolean cube and not just one polynomial), but evaluation dimension will be still helpful as it always lower bounds coefficient dimension.

3.4 Multilinear polynomials and multilinear formulas

We now turn to multilinear polynomials and classes that respect multilinearity such as multilinear formulas. We first state some well-known facts about multilinear polynomials.

Fact 3.12. *For any two multilinear polynomials $f,g \in \mathbb{F}[x_1,\dots,x_n]$, $f = g$ as polynomials iff they agree on the Boolean cube $\{0,1\}^n$. That is, $f = g$ iff $f|_{\{0,1\}^n} = g|_{\{0,1\}^n}$.*

*Further, there is a **multilinearization** map $\mathrm{ml} : \mathbb{F}[\overline{x}] \to \mathbb{F}[\overline{x}]$ such that for any $f,g \in \mathbb{F}[\overline{x}]$,*

1. *$\mathrm{ml}(f)$ is multilinear.*
2. *f and $\mathrm{ml}(f)$ agree on the Boolean cube, that is, $f|_{\{0,1\}^n} = \mathrm{ml}(f)|_{\{0,1\}^n}$.*
3. *$\deg \mathrm{ml}(f) \le \deg f$.*
4. *$\mathrm{ml}(fg) = \mathrm{ml}(\mathrm{ml}(f)\,\mathrm{ml}(g))$, and if f and g are defined on disjoint sets of variables then $\mathrm{ml}(fg) = \mathrm{ml}(f)\,\mathrm{ml}(g)$.*
5. *ml is linear, so that for any $\alpha,\beta \in \mathbb{F}$, $\mathrm{ml}(\alpha f + \beta g) = \alpha\,\mathrm{ml}(f) + \beta\,\mathrm{ml}(g)$.*
6. *$\mathrm{ml}(x_1^{a_1}\cdots x_n^{a_n}) = \prod_i x_i^{\min\{a_i,1\}}$.*
7. *If f is the sum of at most s monomials (s-sparse) then so is $\mathrm{ml}(f)$.*

Also, if $\hat{f}$ is a function $\{0,1\}^n \to \mathbb{F}$ that only depends on the coordinates in $S \subseteq [n]$, then the unique multilinear polynomial f agreeing with $\hat{f}$ on $\{0,1\}^n$ is a polynomial only in $\{x_i\}_{i\in S}$.

One can also extend the multilinearization map $\mathrm{ml} : \mathbb{F}[\overline{x}] \to \mathbb{F}[\overline{x}]$ to matrices $\mathrm{ml} : \mathbb{F}[\overline{x}]^{r\times r} \to \mathbb{F}[\overline{x}]^{r\times r}$ by applying the map entrywise and the above properties still hold.

Throughout the rest of this paper 'ml' will denote the multilinearization operator. Raz [63, 62] gave lower bounds for multilinear formulas using the above notion of coefficient dimension, and Raz-Yehudayoff [67, 68] gave simplifications and extensions to constant-depth multilinear formulas.

Theorem 3.13 (Raz-Yehudayoff [63, 68])**.** *Let $f \in \mathbb{F}[x_1,\dots,x_{2n},\overline{z}]$ be a multilinear polynomial in the set of variables $\overline{x}$ and auxiliary variables $\overline{z}$. Let $f_{\overline{z}}$ denote the polynomial f in the ring $\mathbb{F}[\overline{z}][\overline{x}]$. Suppose that for any partition $\overline{x} = (\overline{u},\overline{v})$ with $|\overline{u}| = |\overline{v}| = n$ that*

$$\dim_{\mathbb{F}(\overline{z})} \mathbf{Coeff}_{\overline{u}|\overline{v}} f_{\overline{z}} \ge 2^n \ .$$

Then f requires $\ge n^{\Omega(\log n)}$-size to be computed as a multilinear formula, and for $d = o(\log n/\log\log n)$, f requires $n^{\Omega((n/\log n)^{1/d}/d^2)}$-size to be computed as a multilinear formula of product-depth-d. ☐

3.5 Depth-3 powering formulas

In this section we review facts about depth-3 powering formulas. We begin with the *duality trick* of Saxena [71], which shows that one can convert a power of a linear form to a sum of products of univariate polynomials.

Theorem 3.14 (Saxena's Duality Trick [77, 71, 22])**.** *Let $n \geq 1$, and $d \geq 0$. If $|\mathbb{F}| \geq nd+1$, then there are $\mathsf{poly}(n,d)$-explicit univariates $f_{i,j} \in \mathbb{F}[x_i]$ such that*

$$(x_1 + \cdots + x_n)^d = \sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n) \, ,$$

where $\deg f_{i,j} \leq d$ *and* $s = (nd+1)(d+1)$. □

We note that the parameters we use in Theorem 3.14 are from the proof of Forbes, Gupta, and Shpilka [22], which works over any large enough field (irrespective of its characteristic). Saxena [71] proved a stronger statement with $s = nd+1$, however, his proof only worked over fields of large enough characteristic. A similar version of this trick also appeared in Shpilka–Wigderson [77].

Noting that the product $f_{i,1}(x_1) \cdots f_{i,n}(x_n)$ trivially has a width-1 roABP (in any order of the variables), it follows that $(x_1 + \cdots + x_n)^d$ has a $\mathsf{poly}(n,d)$-width roABP over a large enough field. Thus, size-s $\sum \bigwedge \sum$ formulas have $\mathsf{poly}(s)$-size roABPs over large enough fields by appealing to closure properties of roABPs (Theorem 3.8). As it turns out, this result also holds over any field as Forbes–Shpilka [27] adapted Saxena's [71] duality to work over any field. Their version works over any field, but loses the above clean form (sum of product of univariates).

Theorem 3.15 (Forbes–Shpilka [27])**.** *Let $f \in \mathbb{F}[\overline{x}]$ be expressed as $f(\overline{x}) = \sum_{i=1}^{s} (\alpha_{i,0} + \alpha_{i,1}x_1 + \cdots + \alpha_{i,n}x_n)^{d_i}$. Then f is computable by a $\mathsf{poly}(r,n)$-explicit width-r roABP of degree $\max_i\{d_i\}$, in any order of the variables, where $r = \sum_i (d_i + 1)$.* □

One way to see this claim is to observe that for any variable partition, a linear function can be expressed as the sum of two variable-disjoint linear functions $\ell(\overline{x}_1, \overline{x}_2) = \ell_1(\overline{x}_1) + \ell_2(\overline{x}_2)$. By the binomial theorem, the d-th power of this expression is a summation of $d+1$ variable-disjoint products, which implies a coefficient dimension upper bound of $d+1$ (Lemma 3.5) and thus also an roABP-width upper bound (Lemma 3.7). One can then sum over the linear forms.

While this simulation suffices for obtaining roABP upper bounds, we will also want the clean form obtained via duality for application to multilinear-formula IPS proofs of the subset-sum axiom (Theorem 4.15).

3.6 Monomial orders

We recall here the definition and properties of a *monomial order*, following Cox, Little and O'Shea [16]. We first fix the definition of a *monomial* in our context.

Definition 3.16. A **monomial** in $\mathbb{F}[x_1, \ldots, x_n]$ is a polynomial of the form $\overline{x}^{\overline{a}} = x_1^{a_1} \cdots x_n^{a_n}$ for $\overline{a} \in \mathbb{N}^n$.

We will sometimes abuse notation and associate a monomial $\overline{x}^{\overline{a}}$ with its exponent vector $\overline{a}$, so that we can extend this order to the exponent vectors. Note that in this definition "1" is a monomial, and that scalar multiples of monomials such as $2x$ are not considered monomials. We now define a monomial order, which will be total order on monomials with certain natural properties.

Definition 3.17. A **monomial ordering** is a total order $\prec$ on the monomials in $\mathbb{F}[\overline{x}]$ such that

- For all $\overline{a} \in \mathbb{N}^n \setminus \{\overline{0}\}$, $1 \prec \overline{x}^{\overline{a}}$.

- For all $\overline{a},\overline{b},\overline{c} \in \mathbb{N}^n$, $\overline{x}^{\overline{a}} \prec \overline{x}^{\overline{b}}$ implies $\overline{x}^{\overline{a}+\overline{c}} \prec \overline{x}^{\overline{b}+\overline{c}}$.

For nonzero $f \in \mathbb{F}[\overline{x}]$, the **leading monomial of f (with respect to a monomial order $\prec$)**, denoted $\mathrm{LM}(f)$, is the largest monomial in $\mathrm{Supp}(f) := \{\overline{x}^{\overline{a}} : \mathrm{Coeff}_{\overline{x}^{\overline{a}}}(f) \neq 0\}$ with respect to the monomial order $\prec$. The **trailing monomial of** f, denoted $\mathrm{TM}(f)$, is defined analogously to be the smallest monomial in $\mathrm{Supp}(f)$. The zero polynomial has neither leading nor trailing monomial.

For nonzero $f \in \mathbb{F}[\overline{x}]$, the **leading (or trailing) coefficient of** f, denoted $\mathrm{LC}(f)$ ($\mathrm{TC}(f)$, resp.), is $\mathrm{Coeff}_{\overline{x}^{\overline{a}}}(f)$ where $\overline{x}^{\overline{a}} = \mathrm{LM}(f)$ ($\overline{x}^{\overline{a}} = \mathrm{TM}(f)$, resp.).

Henceforth in this paper we will assume $\mathbb{F}[\overline{x}]$ is equipped with some monomial order $\prec$. The results in this paper will hold for *any* monomial order. However, for concreteness, one can consider the lexicographic ordering on monomials, which is easily seen to be a monomial ordering (see also Cox, Little and O'Shea [16]).

We begin with a simple lemma about how taking leading or trailing monomials (or coefficients) is homomorphic with respect to multiplication.

Lemma 3.18. *Let $f,g \in \mathbb{F}[\overline{x}]$ be nonzero polynomials. Then the leading monomial and trailing monomials and coefficients are homomorphic with respect to multiplication, that is,* $\mathrm{LM}(fg) = \mathrm{LM}(f)\,\mathrm{LM}(g)$ *and* $\mathrm{TM}(fg) = \mathrm{TM}(f)\,\mathrm{TM}(g)$*, as well as* $\mathrm{LC}(fg) = \mathrm{LC}(f)\,\mathrm{LC}(g)$ *and* $\mathrm{TC}(fg) = \mathrm{TC}(f)\,\mathrm{TC}(g)$.

Proof: We do the proof for leading monomials and coefficients, the claim for trailing monomials and coefficients is symmetric.

Let $f(\overline{x}) = \sum_{\overline{a}} \alpha_{\overline{a}} \overline{x}^{\overline{a}}$ and $g(\overline{x}) = \sum_{\overline{b}} \beta_{\overline{b}} \overline{x}^{\overline{b}}$. Isolating the leading monomials,

$$f(\overline{x}) = \mathrm{LC}(f)\cdot \mathrm{LM}(f) + \sum_{\overline{x}^{\overline{a}} \prec \mathrm{LM}(f)} \alpha_{\overline{a}} \overline{x}^{\overline{a}}, \qquad g(\overline{x}) = \mathrm{LC}(g)\cdot \mathrm{LM}(g) + \sum_{\overline{x}^{\overline{b}} \prec \mathrm{LM}(g)} \beta_{\overline{b}} \overline{x}^{\overline{b}},$$

with $\mathrm{LC}(f) = \alpha_{\mathrm{LM}(f)}$ and $\mathrm{LC}(g) = \beta_{\mathrm{LM}(g)}$ being nonzero. Thus,

$$\begin{aligned} f(\overline{x})g(\overline{x}) = \mathrm{LC}(f)\,\mathrm{LC}(g)\cdot \mathrm{LM}(f)\,\mathrm{LM}(g) + \mathrm{LC}(f)\,\mathrm{LM}(f)\left(\sum_{\overline{x}^{\overline{b}} \prec \mathrm{LM}(g)} \beta_{\overline{b}} \overline{x}^{\overline{b}}\right) \\ + \mathrm{LC}(g)\,\mathrm{LM}(g)\left(\sum_{\overline{x}^{\overline{a}} \prec \mathrm{LM}(f)} \alpha_{\overline{a}} \overline{x}^{\overline{a}}\right) + \left(\sum_{\overline{x}^{\overline{a}} \prec \mathrm{LM}(f)} \alpha_{\overline{a}} \overline{x}^{\overline{a}}\right)\left(\sum_{\overline{x}^{\overline{b}} \prec \mathrm{LM}(g)} \beta_{\overline{b}} \overline{x}^{\overline{b}}\right). \end{aligned}$$

Using that $\overline{x}^{\overline{a}}\overline{x}^{\overline{b}} \prec \mathrm{LM}(f)\,\mathrm{LM}(g)$ whenever $\overline{x}^{\overline{a}} \prec \mathrm{LM}(f)$ or $\overline{x}^{\overline{b}} \prec \mathrm{LM}(g)$ due to the definition of a monomial order, we have that $\mathrm{LM}(f)\,\mathrm{LM}(g)$ is indeed the maximal monomial in the above expression with nonzero coefficient, and as its coefficient is $\mathrm{LC}(f)\,\mathrm{LC}(g)$. □

We now recall the well-known fact that for any set of polynomials the dimension of their span in $\mathbb{F}[\overline{x}]$ is equal to the number of distinct leading or trailing monomials in their span.

Lemma 3.19. *Let $S \subseteq \mathbb{F}[\overline{x}]$ be a set of polynomials. Then* $\dim \operatorname{span} S = |\mathrm{LM}(\operatorname{span} S)| = |\mathrm{TM}(\operatorname{span} S)|$. *In particular,* $\dim \operatorname{span} S \geq |\mathrm{LM}(S)|, |\mathrm{TM}(S)|$. □

4 Upper bounds for linear-IPS

While the primary focus of this article is on *lower bounds* for restricted classes of the IPS proof system, we begin by discussing *upper bounds* to demonstrate that these restricted classes can prove the unsatisfiability of non-trivial systems of polynomials equations. In particular we go beyond existing work on upper bounds ([32, 66, 65, 35, 50]) and place interesting refutations in IPS subsystems where we will also prove lower bounds, as such upper bounds demonstrate the non-triviality of our lower bounds.

We begin by discussing the power of the linear-IPS proof system. While one of the most novel features of IPS proofs is their consideration of non-linear certificates, we show that in powerful enough models of algebraic computation, linear-IPS proofs can efficiently simulate general IPS proofs, essentially answering an open question of Grochow and Pitassi [35]. A special case of this result was obtained by Grochow and Pitassi [35], where they showed that $\mathrm{IPS}_{\mathrm{LIN}}$ can simulate $\sum\prod$-IPS. We then consider the *subset-sum* axioms, previously considered by Impagliazzo, Pudlák, and Sgall [41], and show that they can be refuted in polynomial size by the $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}}$ proof system where $\mathcal{C}$ is either the class of roABPs, or the class of multilinear formulas.

4.1 Simulating IPS proofs with linear-IPS

We show here that general IPS proofs can be efficiently simulated by linear-IPS, assuming that the axioms to be refuted are described by small algebraic circuits. Grochow and Pitassi [35] showed that whenever the IPS proof computes *sparse* polynomials, one can simulate it by linear-IPS using (possibly non-sparse) algebraic circuits. We give here a simulation of IPS when the proofs use general algebraic circuits.

To give our simulation, we will need to show that if a small circuit $f(\overline{x},y)$ is divisible by y, then the quotient $f(\overline{x},y)/y$ also has a small circuit. Such a result clearly follows from Strassen's [81] elimination of divisions in general, but we give two constructions for the quotient which tailor Strassen's [81] technique to optimize certain parameters.

The first construction assumes that f has degree bounded by d, and produces a circuit for the quotient whose size depends polynomially on d. This construction is efficient when f is computed by a formula or branching program (so that d is bounded by the size of f). In particular, this construction will preserve the depth of f in computing the quotient, and for that reason we only present it for formulas. The construction proceeds via interpolation to decompose $f(\overline{x},y)=\sum_i f_i(\overline{x})y^i$ into its constituent parts $\{f_i(\overline{x})\}_i$ and then directly constructs $f(\overline{x},y)/y=\sum_i f_i(\overline{x})y^{i-1}$.

Lemma 4.1. *Let $\mathbb{F}$ be a field with $|\mathbb{F}|\ge d+1$. Let $f(\overline{x},y)\in\mathbb{F}[x_1,\ldots,x_n,y]$ be a degree $\le d$ polynomial expressible as $f(\overline{x},y)=\sum_{0\le i\le d} f_i(\overline{x})y^i$ for $f_i\in\mathbb{F}[\overline{x}]$. Assume f is computable by a size-s depth-D formula. Then for $a\ge 1$ one can compute*

$$\sum_{i=a}^{d} f_i(\overline{x})y^{i-a}\ ,$$

by a $\mathsf{poly}(s,a,d)$-size depth-$(D+2)$ formula. Further, given d and the formula for f, the resulting formula is $\mathsf{poly}(s,a,d)$-explicit. In particular, if $y^a|f(\overline{x},y)$ then the quotient $f(\overline{x},y)/y^a$ has a $\mathsf{poly}(s,a,d)$-explicit formula of size $\mathsf{poly}(s,a,d)$ and depth $(D+2)$.

Proof: Express $f(\bar{x},y) \in \mathbb{F}[\bar{x}][y]$ by $f(\bar{x},y) = \sum_{0\le i\le d} f_i(\bar{x})y^i$. As $|\mathbb{F}| \ge 1+\deg_y f$, by interpolation there are $\mathsf{poly}(d)$-explicit constants $\alpha_{i,j}, \beta_j \in \mathbb{F}$ such that

$$f_i(\bar{x}) = \sum_{j=0}^{d} \alpha_{i,j} f(\bar{x},\beta_j) .$$

It then follows that

$$\sum_{i=a}^{d} f_i(\bar{x})y^{i-a} = \sum_{i=a}^{d}\left(\sum_{j=0}^{d} \alpha_{i,j} f(\bar{x},\beta_j)\right)y^{i-a} = \sum_{i=a}^{d}\sum_{j=0}^{d} \alpha_{i,j} f(\bar{x},\beta_j)y^{i-a} ,$$

which is clearly a formula of the appropriate size, depth, and explicitness. The claim about the quotient $f(\bar{x},y)/y^a$ follows from seeing that if the quotient is a polynomial then $f(\bar{x},y)/y^a = \sum_{i=a}^{d} f_i(\bar{x})y^{i-a}$. □

The above construction suffices in the typical regime of algebraic complexity where the circuits compute polynomials whose degree is polynomially related to their circuit size. However, the simulation of Extended Frege by general IPS proved by Grochow-Pitassi [35] (Theorem 1.2) yields IPS refutations with circuits of possibly exponential degree (see also Theorem 1.3). This motivates the search for an efficient division lemma in this regime. We now provide such a lemma, which is a variant of Strassen's [81] homogenization technique for efficiently computing the low-degree homogeneous components of an unbounded degree circuit. As weaker models of computation (such as formulas and branching programs) cannot compute polynomials of degree exponential in their size, we only present this lemma for circuits.

Lemma 4.2. *Let $f(\bar{x},y) \in \mathbb{F}[x_1,\ldots,x_n,y]$ be a polynomial expressible as $f(\bar{x},y) = \sum_i f_i(\bar{x})y^i$ for $f_i \in \mathbb{F}[\bar{x}]$, and assume f is computable by a size-s circuit. Then for $a \ge 1$ there is an $O(a^2 s)$-size circuit with output gates computing*

$$f_0(\bar{x}),\ldots,f_{a-1}(\bar{x}), \sum_{i\ge a} f_i(\bar{x})y^{i-a} .$$

Further, given a and the circuit for f, the resulting circuit is $\mathsf{poly}(s,a)$-explicit. In particular, if $y^a | f(\bar{x},y)$ then the quotient $f(\bar{x},y)/y^a$ has a $\mathsf{poly}(s,a)$-explicit circuit of size $O(a^2 s)$.

Proof: The proof proceeds by viewing the computation in the ring $\mathbb{F}[\bar{x}][y]$, and splitting each gate in the circuit for f into its coefficients in terms of y. However, to avoid a dependence on the degree, we only split out the coefficients of $y^0, y^1, \ldots, y^{a-1}$, and then group the rest of the coefficients together. That is, every node v computing the polynomial $f_v(\bar{x},y) = \sum_{i\ge 0} f_{(v,i)}(\bar{x})y^i$ in the circuit Φ for f, is split into $a+1$ nodes computing $f_{(v,0)}(\bar{x}),\ldots,f_{(v,a-1)}(\bar{x})$ and $\sum_{i\ge a} f_i(\bar{x})y^{i-a}$, respectively. This is similar to homogenizing algebraic circuits, see [78]. We can then locally update this split by appropriately keeping track of how addition and multiplication affects this grouping of coefficients. We note that we can assume without loss of generality that the circuit for f has fan-in 2, as this only increases the size of the circuit by a constant factor (measuring the size of the circuit in number of edges) and simplifies the construction.

<u>construction:</u> Let Φ denote the circuit for f. For a gate v in Φ, denote Φ_v to be the sub-circuit rooted at v in Φ and let f_v be the polynomial computed by the gate v. We will define the new circuit Ψ, which will be defined by the gates $\{(v,i) : v \in \Phi, 0 \le i \le a\}$ and the wiring between them, as follows.

- $\Phi_v \in \mathbb{F}$: $\Psi_{(v,0)} := \Phi_v$, $\Psi_{(v,i)} := 0$ for $i \ge 1$.

- $\Phi_v = x_j$: $\Psi_{(v,0)} := x_j$, $\Psi_{(v,i)} := 0$ for $i \geq 1$.
- $\Phi_v = y$: $\Psi_{(v,1)} := 1$, $\Psi_{(v,i)} := 0$ for $i \neq 1$.
- $\Phi_v = \Phi_u + \Phi_w$: $\Psi_{(v,i)} := \Psi_{(u,i)} + \Psi_{(w,i)}$, all i.
- $\Phi_v = \Phi_u \times \Phi_w$, $0 \leq i < a$:
$$\Psi_{(v,i)} := \sum_{0 \leq j \leq i} \Psi_{(u,j)} \times \Psi_{(w,i-j)} \ .$$
- $\Phi_v = \Phi_u \times \Phi_w$, $i = a$:
$$\begin{aligned}\Psi_{(v,a)} := \sum_{\ell=a}^{2(a-1)} y^{\ell-a} \sum_{\substack{i+j=\ell \\ 0 \leq i,j<a}} \Psi_{(u,i)} \times \Psi_{(w,j)} + \sum_{0 \leq i<a} \Psi_{(u,i)} \times \Psi_{(w,a)} \times y^i \\ + \sum_{0 \leq j<a} \Psi_{(u,a)} \times \Psi_{(w,j)} \times y^j + \Psi_{(u,a)} \times \Psi_{(w,a)} \times y^a \ .\end{aligned}$$

complexity: Split the gates in Ψ into two types, those gates (v,i) where $i = a$ and v is a multiplication gate in Φ, and then the rest. For the former type, $\Psi_{(v,a)}$ is computable by a size-$O(a^2)$ circuit in its children, and there are at most s such gates. For the latter type, $\Psi_{(v,i)}$ is computable by a size-$O(a)$ circuit in its children, and there are at most $O(as)$ such gates. Therefore, the total size is $O(a^2 s)$.

correctness: We now establish correctness as a subclaim. For a gate (v,i) in Ψ, let $g_{(v,i)}$ denote the polynomial that it computes.

Subclaim 4.3. *For each gate v in Φ, for $0 \leq i < a$ we have that $g_{(v,i)} = \mathrm{Coeff}_{\overline{x}|y^i}(f_v)$ and for $i = a$ we have that $g_{(v,a)} = \sum_{i \geq a} \mathrm{Coeff}_{\overline{x}|y^i}(f_v) y^{i-a}$. In particular, $f_v = \sum_{i=0}^{a} g_{(v,i)} y^i$.*

Sub-Proof: Note that the second part of the claim follows from the first. We now establish the first part by induction on the gates of the circuit.

- $\Phi_v \in \mathbb{F}$: $f_v = \gamma$ for $\gamma \in \mathbb{F}$. By construction, $g_{(v,0)} = \gamma = \mathrm{Coeff}_{\overline{x}|y^0}(f_v)$, and for $i \geq 1$, $g_{(v,i)} = 0 = \mathrm{Coeff}_{\overline{x}|y^i}(f_v)$.
- $\Phi_v = x_j$: $f_v = x_j$. By construction, $g_{(v,0)} = x_j = \mathrm{Coeff}_{\overline{x}|y^0}(f_v)$, and for $i \geq 1$, $g_{(v,i)} = 0 = \mathrm{Coeff}_{\overline{x}|y^i}(f_v)$.
- $\Phi_v = y$: By construction, $g_{(v,1)} = 1 = \mathrm{Coeff}_{\overline{x}|y^1}(f_v)$, and for $i \neq 1$, $g_{(v,i)} = 0 = \mathrm{Coeff}_{\overline{x}|y^i}(f_v)$.
- $\Phi_v = \Phi_u + \Phi_w$:
$$\begin{aligned} g_{(v,i)} &= g_{(u,i)} + g_{(w,i)} \\ &= \mathrm{Coeff}_{\overline{x}|y^i}(f_u) + \mathrm{Coeff}_{\overline{x}|y^i}(f_w) \\ &= \mathrm{Coeff}_{\overline{x}|y^i}(f_u + f_w) = \mathrm{Coeff}_{\overline{x}|y^i}(f_v) \ . \end{aligned}$$

- $\Phi_v = \Phi_u \times \Phi_w$, $0 \le i < a$:

$$\begin{aligned} g_{(v,i)} &= \sum_{0 \le j \le i} g_{(u,j)} \cdot g_{(w,i-j)} \\ &= \sum_{0 \le j \le i} \mathrm{Coeff}_{\overline{x}|y^j}(f_u) \cdot \mathrm{Coeff}_{\overline{x}|y^{i-j}}(f_w) \\ &= \mathrm{Coeff}_{\overline{x}|y^i}(f_u \cdot f_w) = \mathrm{Coeff}_{\overline{x}|y^i}(f_v)\,. \end{aligned}$$

- $\Phi_v = \Phi_u \times \Phi_w$, $i = a$:

$$\begin{aligned} g_{(v,a)} = \sum_{\ell=a}^{2(a-1)} y^{\ell-a} \sum_{\substack{i+j=\ell \\ 0 \le i,j<a}} g_{(u,i)} \cdot g_{(w,j)} &+ \sum_{0 \le i < a} g_{(u,i)} \cdot g_{(w,a)} \cdot y^i \\ &+ \sum_{0 \le j < a} g_{(u,a)} \cdot g_{(w,j)} \cdot y^j + g_{(u,a)} \cdot g_{(w,a)} \cdot y^a \\ = \sum_{\ell=a}^{2(a-1)} y^{-a} \sum_{\substack{i+j=\ell \\ 0 \le i,j<a}} & \mathrm{Coeff}_{\overline{x}|y^i}(f_u) y^i \cdot \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^j \\ &+ \sum_{0 \le i < a} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \cdot \left(\sum_{j \ge a} \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{j-a} \right) \cdot y^i \\ &+ \sum_{0 \le j < a} \left(\sum_{i \ge a} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) y^{i-a} \right) \cdot \mathrm{Coeff}_{\overline{x}|y^j}(f_w) \cdot y^j \\ &+ \left(\sum_{i \ge a} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) y^{i-a} \right) \cdot \left(\sum_{j \ge a} \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{j-a} \right) \cdot y^a \\ = \sum_{\substack{i+j \ge a \\ 0 \le i,j < a}} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{i+j-a} &+ \sum_{\substack{0 \le i < a \\ j \ge a}} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{i+j-a} \\ + \sum_{\substack{i \ge a \\ 0 \le j < a}} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{i+j-a} &+ \sum_{i,j \ge a} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \mathrm{Coeff}_{\overline{x}|y^j}(f_w) y^{i+j-a} \\ = \sum_{i+j \ge a} \mathrm{Coeff}_{\overline{x}|y^i}(f_u) \mathrm{Coeff}_{\overline{x}|y^j}(f_w) \cdot y^{i+j-a} & \\ = \sum_{\ell \ge a} \mathrm{Coeff}_{\overline{x}|y^\ell}(f_u \cdot f_w) \cdot y^{\ell-a} & \\ = \sum_{\ell \ge a} \mathrm{Coeff}_{\overline{x}|y^\ell}(f_v) \cdot y^{\ell-a}\,. & \end{aligned}$$

□

The correctness then follows by examining v_{out}, the output gate of Φ, so that $f_{v_{\mathrm{out}}} = f$. The gates $(v_{\mathrm{out}}, 0), \ldots, (v_{\mathrm{out}}, a)$ are then outputs of Ψ and by the above subclaim have the desired functionality.

quotient: The claim about the quotient $f(\overline{x},y)/y^a$ follows from seeing that if the quotient is a polynomial then $f(\overline{x},y)/y^a = \sum_{i \ge a} f_i(\overline{x}) y^{i-a}$ which is one of the outputs of the constructed circuit. □

We now give our simulation of general IPS by linear-IPS. In the set of axioms below, we do not separate out the Boolean axioms from the rest, as this simplifies notation.

Proposition 4.4. *Let $f_1,\dots,f_m \in \mathbb{F}[x_1,\dots,x_n]$ be unsatisfiable polynomials with an IPS refutation $C \in \mathbb{F}[\overline{x},y_1,\dots,y_m]$. Then $f_1,\dots,f_m$ have a linear-IPS refutation $C' \in \mathbb{F}[\overline{x},\overline{y}]$ under the following conditions.*

1. *Suppose $f_1,\dots,f_m,C$ are computed by size-s formulas, have degree at most d, and $|\mathbb{F}| \ge d+1$. Then C' is computable by a $\mathsf{poly}(s,d,m)$-size formula of depth-$O(D)$, and C' is $\mathsf{poly}(s,d,m)$-explicit given d and the formulas for $f_1,\dots,f_m,C$.*

2. *Suppose $f_1,\dots,f_m,C$ are computed by size-s circuits. Then C' is computable by a $\mathsf{poly}(s,m)$-size circuit, and C' is $\mathsf{poly}(s,m)$-explicit given the circuits for $f_1,\dots,f_m,C$.*

Proof: Express $C(\overline{x},\overline{y})$ as a polynomial in $\mathbb{F}[\overline{x}][\overline{y}]$, so that $C(\overline{x},\overline{y}) = \sum_{\overline{a}>\overline{0}} C_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}}$, where we use that $C(\overline{x},\overline{0}) = 0$ to see that we can restrict $\overline{a}$ to $\overline{a} > \overline{0}$. Partitioning the $\overline{a} \in \mathbb{N}^n$ based on the index of their first nonzero value, and denoting $\overline{a}_{<i}$ for the first $i-1$ coordinates of $\overline{a}$, we obtain

$$\begin{aligned} C(\overline{x},\overline{y}) &= \sum_{\overline{a}>\overline{0}} C_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}} \\ &= \sum_{i=1}^{n} \sum_{\substack{\overline{a}:\overline{a}_{<i}=\overline{0},\\ a_i>0}} C_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}} . \end{aligned}$$

Now define $C_i(\overline{x},\overline{y}) := \sum_{\substack{\overline{a}:\overline{a}_{<i}=\overline{0},\\ a_i>0}} C_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}-\overline{e}_i}$, where $\overline{e}_i$ is the i-th standard basis vector. Note that this is a valid polynomial as in this summation we assume $a_i > 0$ so that $\overline{a}-\overline{e}_i \ge \overline{0}$. Thus,

$$C(\overline{x},\overline{y}) = \sum_{i=1}^{n} C_i(\overline{x},\overline{y})y_i .$$

We now define $C'(\overline{x},\overline{y}) := \sum_{i=1}^{n} C_i(\overline{x},\overline{f}(\overline{x}))y_i$ and claim it is the desired linear-IPS refutation, where note that we have only *partially* substituted in the f_i for the y_i. First, observe that it is a valid refutation, as $C'(\overline{x},\overline{0}) = \sum_{i=1}^{n} C_i(\overline{x},\overline{f}(\overline{x}))\cdot 0 = 0$, and $C'(\overline{x},\overline{f}(\overline{x})) = \sum_{i=1}^{n} C_i(\overline{x},\overline{f}(\overline{x}))f_i(\overline{x}) = C(\overline{x},\overline{f}(\overline{x})) = 1$ via the above equation and using that C is a valid IPS refutation.

We now argue that C' can be efficiently computed in the two above regimes.

(1): Up to constant loss in the depth and polynomial loss in the size, for bounding the complexity of C' it suffices to bound the complexity of each $C_i(\overline{x},\overline{f}(\overline{x}))$. First, note that

$$C_i(\overline{x},\overline{y})y_i = \sum_{\substack{\overline{a}:\overline{a}_{<i}=\overline{0},\\ a_i>0}} C_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}} = C(\overline{x},\overline{0},y_i,\overline{y}_{>i}) - C(\overline{x},\overline{0},0,\overline{y}_{>i}) ,$$

where each "$\overline{0}$" here is a vector of $i-1$ zeros. Clearly each of $C(\overline{x},\overline{0},y_i,\overline{y}_{>i})$ and $C(\overline{x},\overline{0},0,\overline{y}_{>i})$ have formula size and depth bounded by that of C. From our division lemma for formulas (Theorem 4.1) it follows that $C_i(\overline{x},\overline{y}) = \frac{1}{y_i}(C(\overline{x},\overline{0},y_i,\overline{y}_{>i}) - C(\overline{x},\overline{0},0,\overline{y}_{>i}))$ has a $\mathsf{poly}(s,d)$-size depth-$O(D)$ formula of the desired explicitness (as $\deg_y\left(C(\overline{x},\overline{0},y_i,\overline{y}_{>i})-C(\overline{x},\overline{0},0,\overline{y}_{>i})\right)\le \deg_y C\le d\le|\mathbb{F}|-1$, so that $\mathbb{F}$ is large enough). We replace $\overline{y}\leftarrow\overline{f}(\overline{x})$ to obtain $C_i(\overline{x},\overline{f}(\overline{x}))$ of the desired size and explicitness, using that the f_j themselves have small-depth formulas.

(2): This follows as in (1), using now the division lemma for circuits (Theorem 4.2). □

Grochow and Pitassi [35, Open Question 1.13[9]] asked whether one can relate the complexity of IPS and linear-IPS, as they only established such relations for simulating $\sum\prod$-IPS by (general) linear-IPS. Our above result essentially answers this question for general formulas over sufficiently large fields, and for circuits, over any field, under the assumption that the unsatisfiable polynomial system $f_1=\cdots=f_m=0$ can be expressed using small algebraic formulas or circuits, respectively. This is a reasonable assumption as it is the most common regime for proof complexity. However, the above result does not fully close the question of Grochow-Pitassi [35] with respect to simulating $\mathcal{C}$-IPS by $\mathcal{D}$-IPS$_{\text{LIN}}$ for various restricted subclasses $\mathcal{C},\mathcal{D}$ of algebraic computation. That is, for such a simulation our result requires $\mathcal{D}$ to at the very least contain $\mathcal{C}$ composed with the axioms $f_1,\ldots,f_m$, and when applying this to the models considered in this paper (sparse polynomials, depth-3 powering formulas, roABPs, multilinear formulas) this seems to non-negligibly increase the complexity of the algebraic reasoning.

4.2 Multilinearizing roABP-IPS$_{\text{LIN}}$

We now exhibit instances where one can efficiently *prove* that a polynomial equals its multilinearization modulo the Boolean axioms. That is, for a polynomial f computed by a small circuit we wish to prove that $f\equiv\mathrm{ml}(f)\mod\overline{x}^2-\overline{x}$ by expressing $f-\mathrm{ml}(f)=\sum_i h_i\cdot(x_i^2-x_i)$ so that the h_i also have small circuits. Such a result will simplify the search for linear-IPS refutations by allowing us to focus on the non-Boolean axioms. That is, if we are able to find a refutation of $\overline{f},\overline{x}^2-\overline{x}$ given by

$$\sum_j g_jf_j\equiv 1\mod\overline{x}^2-\overline{x}\,,$$

where the g_j have small circuits, multilinearization results of the above form guarantee that there are h_i so that

$$\sum_j g_jf_j+\sum_i h_i\cdot(x_i^2-x_i)=1\,,$$

which is a proper linear-IPS refutation.

We establish such a multilinearization result when f is an roABP in this section, and consider the cases where f is the product of a low-degree multilinear polynomial and a multilinear formula in the next section. We will use these multilinearization results in our construction of IPS refutations of the subset-sum axiom (Section 4.4).

We begin by noting that multilinearization for these two circuit classes is rather special, as these classes straddle the conflicting requirements of neither being *too weak* nor *too strong*. That is, some circuit classes are simply too weak to compute their multilinearizations. An example is the class of depth-3 powering formulas, where $(x_1+\cdots+x_n)^n$ has a small $\sum\bigwedge\sum$ formula, but its multilinearization has the leading term $n!x_1\cdots x_n$ and thus requires exponential size as a $\sum\bigwedge\sum$ formula (by appealing to

[9]This refers to the following version: http://arxiv.org/abs/1404.3820v1

Theorem 6.8). On the other hand, some circuit classes are too strong to admit efficient multilinearization (under plausible complexity assumptions). That is, consider an $n \times n$ symbolic matrix X where $(X)_{i,j} = x_{i,j}$ and the polynomial $f(X,\overline{y}) := (x_{1,1}y_1 + \cdots + x_{1,n}y_n)\cdots(x_{n,1}y_1 + \cdots + x_{n,n}y_n)$, which is clearly a simple depth-3 ($\prod\sum\prod$) circuit. Viewing this polynomial in $\mathbb{F}[X][\overline{y}]$, one sees that $\mathrm{Coeff}_{X|y_1\cdots y_n} f = \mathrm{perm}(X)$, where $\mathrm{perm}(X)$ is the $n \times n$ permanent. Viewing $\mathrm{ml}(f)$, the multilinearization of f, in $\mathbb{F}[X][\overline{y}]$ one sees that $\mathrm{ml}(f)$ is of degree n and its degree n component is the coefficient of $y_1 \cdots y_n$ in $\mathrm{ml}(f)$, which is still $\mathrm{perm}(X)$. Hence, by interpolation, one can extract this degree n part and thus can compute a circuit for $\mathrm{perm}(X)$ given a circuit for $\mathrm{ml}(f)$. Since we believe $\mathrm{perm}(X)$ does not have small algebraic circuits it follows that the multilinearization of f does not have small circuits. These examples show that efficient multilinearization is a somewhat special phenomenon.

We now give our result for multilinearizing roABPs, where we multilinearize variable by variable via telescoping.

Proposition 4.5. *Let $f \in \mathbb{F}[x_1,\ldots,x_n]$ be computable by a width-r roABP, in order of the variables $x_1 < \cdots < x_n$, and with individual degrees at most d. Then $\mathrm{ml}(f)$ has a $\mathsf{poly}(r,n,d)$-explicit width-r roABP in order of the variables $x_1 < \cdots < x_n$, and there are $\mathsf{poly}(r,n,d)$-explicit width-r roABPs $h_1,\ldots,h_n \in \mathbb{F}[\overline{x}]$ in order of the variables $x_1 < \cdots < x_n$ such that*

$$f(\overline{x}) = \mathrm{ml}(f) + \sum_{j=1}^{n} h_j \cdot (x_j^2 - x_j)\ .$$

Further, $\mathrm{ideg}\, h_j \le \mathrm{ideg}\, f$ *and the individual degree of the roABP for* $\mathrm{ml}(f)$ *is* ≤ 1.

Proof: Let $f(\overline{x}) = \left(\prod_{i=1}^{n} A_i(x_i)\right)_{1,1}$ where $A_i \in \mathbb{F}[x_i]^{r\times r}$ have $\deg A_i \le d$. We apply the multilinearization map $\mathrm{ml} : \mathbb{F}[\overline{x}] \to \mathbb{F}[\overline{x}]$ to matrices $\mathrm{ml} : \mathbb{F}[\overline{x}]^{r\times r} \to \mathbb{F}[\overline{x}]^{r\times r}$ by applying the map entrywise (Theorem 3.12). It follows then that $A_i(x_i) - \mathrm{ml}(A_i(x_i)) \equiv 0 \mod x_i^2 - x_i$, so that $A_i(x_i) - \mathrm{ml}(A_i(x_i)) = B_i(x_i)\cdot(x_i^2 - x_i)$ for some $B_i(x_i) \in \mathbb{F}[x_i]^{r\times r}$ where $\mathrm{ideg}\, B_i(x_i) \le \mathrm{ideg}\, A_i(x_i)$.

The idea of the proof is to construct the new roABP in stages. We first multilinearize the matrix corresponding to x_1 entry wise. This clearly does not change the dimension of the matrix. We can thus write $A_1(x_1) = \mathrm{ml}(A_1(x_1)) + B_1(x_1)\cdot(x_1^2 - x_1)$. We can therefore write our original roABP as the sum of to roABPs, the first one having $\mathrm{ml}(A_1(x_1))$ as the first matrix and the other one having $B_1(x_1)\cdot(x_1^2 - x_1)$, where the matrices in the other layers remain unchanged. The second roABP computes a polynomial of the form $h_1 \cdot (x_1^2 - x_1)$. We now continue multilinearizing the second layer in the first roABP etc. Continuing in this way we get our desired representation of f.

Formally, define $\mathrm{ml}_{\le i}$ to be the map which multilinearizes the first i variables and leaves the others intact, so that $\mathrm{ml}_{\le 0}$ is the identity map and $\mathrm{ml}_{\le n} = \mathrm{ml}$. Telescoping,

$$\begin{aligned}\prod_{i=1}^{n} A_i(x_i) &= \mathrm{ml}_{<1}\left(\prod_{i=1}^{n} A_i(x_i)\right)\\ &= \mathrm{ml}_{\le n}\left(\prod_{i=1}^{n} A_i(x_i)\right) + \sum_{j=1}^{n}\left[\mathrm{ml}_{<j}\left(\prod_{i=1}^{n} A_i(x_i)\right) - \mathrm{ml}_{\le j}\left(\prod_{i=1}^{n} A_i(x_i)\right)\right]\end{aligned}$$

using that the identity $\mathrm{ml}(gh) = \mathrm{ml}(g)\,\mathrm{ml}(h)$, for polynomials on disjoint sets of variables (Theorem 3.12), naturally extends from scalars to matrices, and also to partial-multilinearization (by viewing $\mathrm{ml}_{\le i}$ as multilinearization in $\mathbb{F}[\overline{x}_{>i}][\overline{x}_{\le i}]$),

$$\begin{aligned}
&= \prod_{i=1}^{n} \mathrm{ml}(A_i(x_i)) + \sum_{j=1}^{n} \left[\prod_{i<j} \mathrm{ml}(A_i(x_i)) \prod_{i\ge j} A_i(x_i) - \prod_{i\le j} \mathrm{ml}(A_i(x_i)) \prod_{i>j} A_i(x_i) \right] \\
&= \prod_{i=1}^{n} \mathrm{ml}(A_i(x_i)) + \sum_{j=1}^{n} \left[\prod_{i<j} \mathrm{ml}(A_i(x_i)) \cdot \Big(A_j(x_j) - \mathrm{ml}(A_j(x_j)) \Big) \cdot \prod_{i>j} A_i(x_i) \right] \\
&= \prod_{i=1}^{n} \mathrm{ml}(A_i(x_i)) + \sum_{j=1}^{n} \left[\prod_{i<j} \mathrm{ml}(A_i(x_i)) \cdot B_j(x_j) \cdot \prod_{i>j} A_i(x_i) \cdot (x_j^2 - x_j) \right] ,
\end{aligned}$$

where we have used the definition of $B_i(x_i)$ from above. Taking the $(1,1)$-entry in the above yields that

$$\begin{aligned}
f(\overline{x}) &= \left(\prod_{i=1}^{n} A_i(x_i) \right)_{1,1} \\
&= \left(\prod_{i=1}^{n} \mathrm{ml}(A_i(x_i)) \right)_{1,1} + \sum_{j=1}^{n} \left(\prod_{i<j} \mathrm{ml}(A_i(x_i)) \cdot B_j(x_j) \cdot \prod_{i>j} A_i(x_i) \right)_{1,1} \cdot (x_j^2 - x_j) \,.
\end{aligned}$$

Thus, define

$$\hat{f} := \left(\prod_{i=1}^{n} \mathrm{ml}(A_i(x_i)) \right)_{1,1} , \qquad h_j := \left(\prod_{i<j} \mathrm{ml}(A_i(x_i)) \cdot B_j(x_j) \cdot \prod_{i>j} A_i(x_i) \right)_{1,1} .$$

It follows by construction that $\hat{f}$ and each h_j are computable by width-r roABPs of the desired explicitness in the correct order of the variables. Further, $\operatorname{ideg} h_j \le \operatorname{ideg} f$ and $\hat{f}$ has individual degree ≤ 1. Thus, the above yields that $f = \hat{f} + \sum_j h_j \cdot (x_j^2 - x_j)$, from which it follows that $\mathrm{ml}(f) = \hat{f}$, as desired. □

We now conclude that in designing an roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation $\sum_j g_j \cdot f_j + \sum_i h_i \cdot (x_i^2 - x_i)$ of $f_1(\overline{x}), \ldots, f_m(\overline{x}), \overline{x}^2 - \overline{x}$, it suffices to bound the complexity of the g_j's.

Proposition 4.6. *Let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ be a system of polynomials unsatisfiable over $\overline{x} \in \{0,1\}^n$ computable by width-s roABPs in order of the variables $x_1 < \cdots < x_n$. Suppose that there are $g_j \in \mathbb{F}[\overline{x}]$ such that*

$$\sum_{j=1}^{m} g_j(\overline{x}) f_j(\overline{x}) \equiv 1 \mod \overline{x}^2 - \overline{x} \,,$$

where the g_j have width-r roABPs in the order of the variables $x_1 < \cdots < x_n$. Then there is an roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation $C(\overline{x}, \overline{y}, \overline{z})$ of individual degree at most $1 + \operatorname{ideg} \overline{f}$ and computable in width-$\mathrm{poly}(s, r, n, m)$ in any order of the variables where $x_1 < \cdots < x_n$. Furthermore, this refutation is $\mathrm{poly}(s, r, \operatorname{ideg} \overline{g}, \operatorname{ideg} \overline{f}, n, m)$-explicit given the roABPs for f_j and g_j.

Proof: We begin by noting that we can multilinearize the g_j, so that $\sum_{j=1}^m \mathrm{ml}(g_j(\overline{x}))f_j(\overline{x}) \equiv 1 \mod \overline{x}^2-\overline{x}$, and that $\mathrm{ml}(g_j)$ are $\mathsf{poly}(r, \mathrm{ideg}\,\overline{g}, n, m)$-explicit multilinear roABPs of width-r by Theorem 4.5. Thus, we assume going forward that the g_j are multilinear.

As g_j and f_j are computable by roABPs, their product $g_j f_j$ is computable by a width-rs roABP in the order of the variables $x_1 < \cdots < x_n$ (Theorem 3.8) with individual degree at most $1+\mathrm{ideg}\, f_j$ (Lemma 3.6). Thus, by the above multilinearization (Theorem 4.5), there are $h_{j,i} \in \mathbb{F}[\overline{x}]$ such that

$$g_j(\overline{x})f_j(\overline{x}) = \mathrm{ml}(g_j f_j) + \sum_{i=1}^n h_{j,i}(\overline{x}) \cdot (x_i^2 - x_i) \; .$$

where the $h_{j,i}$ are computable by width-rs roABPs of individual degree at most $1+\mathrm{ideg}\, f_j$. We now define

$$C(\overline{x},\overline{y},\overline{z}) := \sum_{j=1}^m g_j(\overline{x}) y_j - \sum_{i=1}^n \left(\sum_{j=1}^m h_{j,i}(\overline{x}) \right) z_i \; .$$

By the closure operations of roABPs (Theorem 3.8) it follows that C is computable by the appropriately sized roABPs in the desired orders of the variables moreover that C has the desired individual degree, and finally that C has the desired explicitness.

We now show that this is a valid IPS refutation. Observe that $C(\overline{x},\overline{0},\overline{0}) = 0$ and that

$$\begin{aligned} C(\overline{x},\overline{f},\overline{x}^2-\overline{x}) &= \sum_{j=1}^m g_j(\overline{x})f_j(\overline{x}) - \sum_{i=1}^n \left(\sum_{j=1}^m h_{j,i}(\overline{x}) \right)(x_i^2 - x_i) \\ &= \sum_{j=1}^m \left(g_j(\overline{x})f_j(\overline{x}) - \sum_{i=1}^n h_{j,i}(\overline{x})(x_i^2-x_i) \right) \\ &= \sum_{j=1}^m \mathrm{ml}(g_j f_j) \end{aligned}$$

as $\sum_{j=1}^m g_j(\overline{x})f_j(\overline{x}) \equiv 1 \mod \overline{x}^2 - \overline{x}$ we have that

$$\sum_{j=1}^m \mathrm{ml}(g_j f_j) = \mathrm{ml}\left(\sum_{j=1}^m g_j(\overline{x})f_j(\overline{x}) \right) = 1 \; ,$$

where we appealed to linearity, and uniqueness, of multilinearization (Theorem 3.12), so that

$$C(\overline{x},\overline{f},\overline{x}^2-\overline{x}) = \sum_{j=1}^m \mathrm{ml}(g_j f_j) = 1 \; ,$$

as desired. □

4.3 Multilinear-formula-IPS$_{\mathbf{LIN'}}$

We now turn to proving that $g \cdot f \equiv \mathrm{ml}(g \cdot f) \mod \overline{x}^2 - \overline{x}$ when f is low-degree and g is a multilinear formula. This multilinearization can be used to complete our construction of multilinear-formula-IPS refutations of subset-sum axiom (Section 4.4), though our actual construction will multilinearize more directly (Theorem 4.15).

More importantly, the multilinearization we establish here shows that multilinear-formula-IPS can efficiently simulate sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ (when the axioms are low-degree and multilinear). Such a simulation holds intuitively, as multilinear formulas can efficiently compute any sparse (multilinear) polynomial, and as we work over the Boolean cube we are morally working with multilinear polynomials. While this intuition suggests that such a simulation follows immediately, this intuition is false. Specifically, while sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ is a complete refutation system for any system of unsatisfiable polynomials over the Boolean cube, multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ is *incomplete* as seen by the following example (though, by Theorem 1.2, multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ *is* complete for refuting unsatisfiable CNFs).

Example 4.7. Consider the unsatisfiable system of equations $xy+1, x^2-x, y^2-y$. A multilinear linear-IPS proof is a tuple of multilinear polynomials $(f,g,h) \in \mathbb{F}[x,y]$ such that $f\cdot(xy+1)+g\cdot(x^2-x)+h\cdot(y^2-y)=1$. In particular, $f(x,y)=\frac{1}{xy+1}$ for $x,y\in\{0,1\}$, which implies by interpolation over the Boolean cube that $f(x,y)=1\cdot(1-x)(1-y)+1\cdot(1-x)y+1\cdot x(1-y)+\frac{1}{2}\cdot xy = 1-\frac{1}{2}\cdot xy$. Thus $f\cdot(xy+1)$ contains the monomial x^2y^2. However, as g,h are multilinear we see that x^2y^2 cannot appear in $g\cdot(x^2-x)+h\cdot(y^2-y)-1$, so that the equality $f\cdot(xy+1)+g\cdot(x^2-x)+h\cdot(y^2-y)=1$ cannot hold. Thus, $xy+1, x^2-x, y^2-y$ has no linear-IPS refutation only using multilinear polynomials.

Put another way, the above example shows that in a linear-IPS refutation $\sum_j g_j f_j + \sum_i h_i\cdot(x_i^2-x_i)=1$, while one can multilinearize the g_j (with a possible increase in complexity) and still retain a refutation, one cannot multilinearize the h_i in general.

Thus, to simulate sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ (a complete proof system) we must resort to using the more general $\mathrm{IPS}_{\mathrm{LIN}'}$ over multilinear formulas, where recall (Theorem 1.1) that the $\mathrm{IPS}_{\mathrm{LIN}'}$ refutation system refutes $\overline{f}, \overline{x}^2-\overline{x}$ with a polynomial $C(\overline{x},\overline{y},\overline{z})$ where $C(\overline{x},\overline{0},\overline{0})=0$, $C(\overline{x},\overline{f},\overline{x}^2-\overline{x})=1$, with the added restriction that when viewing C as in $\mathbb{F}[\overline{x},\overline{z}][\overline{y}]$ the degree of C with respect to the $\overline{y}$-variables is at most 1, that is, $\deg_{\overline{y}} C \le 1$. In fact, we establish such a simulation using the subclass of refutations of the form $C(\overline{x},\overline{y},\overline{z})=\sum_j g_j y_j + C'(\overline{x},\overline{z})$ where $C'(\overline{x},\overline{0})=0$. Note that such refutations are only linear in the *non*-Boolean axioms, which allows us to circumvent Theorem 4.7.

We now show how to prove $g\cdot f \equiv \mathrm{ml}(g\cdot f) \mod \overline{x}^2-\overline{x}$ when f is low-degree and g is a multilinear formula, where the proof is supplied by the equality $g\cdot f - \mathrm{ml}(g\cdot f) = C(\overline{x},\overline{x}^2-\overline{x})$ where $C(\overline{x},\overline{0})=0$, and where we seek to show that C has a small multilinear formula. We begin with the special case of f and g being the same monomial.

Lemma 4.8. *Let $\overline{x}^{\overline{a}} = \prod_{i=1}^n x_i^{a_i}$. Then, for $\overline{a}\le\overline{1}$,*

$$(\overline{x}^{\overline{a}})^2 - \overline{x}^{\overline{a}} = C(\overline{x},\overline{x}^2-\overline{x}) ,$$

where $C(\overline{x},\overline{z}) \in \mathbb{F}[\overline{x},z_1,\ldots,z_n]$ is defined by

$$C(\overline{x},\overline{z}) := (\overline{z}+\overline{x})^{\overline{a}} - \overline{x}^{\overline{a}} = \sum_{\overline{0}<\overline{b}\le\overline{a}} \overline{z}^{\overline{b}}\overline{x}^{\overline{a}-\overline{b}} ,$$

so that $C(\overline{x},\overline{0})=0$. □

Note that the first expression for C is a $\mathsf{poly}(n)$-sized depth-3 expression, while the second is an $\exp(n)$-sized depth-2 expression. This difference will, going forward, show that we can multilinearize efficiently in depth-3, but can only efficiently multilinearize *low-degree* monomials in depth-2. We now give an IPS proof for showing how a monomial times a multilinear formula equals its multilinearization.

Lemma 4.9. *Let $g \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_d]$ be a multilinear polynomial. Then there is a $C \in \mathbb{F}[\overline{x},\overline{y},z_1,\ldots,z_d]$ such that*

$$g(\overline{x},\overline{y})\cdot\overline{y}^{\overline{1}} - \mathrm{ml}(g(\overline{x},\overline{y})\cdot\overline{y}^{\overline{1}}) = C(\overline{x},\overline{y},\overline{y}^2-\overline{y}) ,$$

and $C(\overline{x},\overline{y},\overline{0}) = 0$.

If g is t-sparse, then C is computable as a $\mathsf{poly}(t,n,2^d)$-size depth-2 multilinear formula (which is $\mathsf{poly}(t,n,2^d)$-explicit given the computation for g), as well as being computable by a $\mathsf{poly}(t,n,d)$-size depth-3 multilinear formula with a $+$-output-gate (which is $\mathsf{poly}(t,n,d)$-explicit given the computation for g). If g is computable by a size-t depth-D multilinear formula, then C is computable by a $\mathsf{poly}(t,2^d)$-size depth-$(D+2)$ multilinear formula with a $+$-output-gate (which is $\mathsf{poly}(t,2^d)$-explicit given the computation for g).

Proof: defining C: Express g as $g = \sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}}$ in the ring $\mathbb{F}[\overline{x}][\overline{y}]$, so that each $g_{\overline{a}}$ is multilinear. Then

$$\begin{aligned} g(\overline{x},\overline{y})\cdot\overline{y}^{\overline{1}} - \mathrm{ml}(g(\overline{x},\overline{y})\cdot\overline{y}^{\overline{1}}) &= \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}}\cdot\overline{y}^{\overline{1}} - \mathrm{ml}\left(\sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}}\cdot\overline{y}^{\overline{1}}\right) \\ &= \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\left(\overline{y}^{\overline{a}+\overline{1}} - \overline{y}^{\overline{1}}\right) = \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{1}-\overline{a}}\left((\overline{y}^{\overline{a}})^2 - \overline{y}^{\overline{a}}\right) , \end{aligned}$$

and appealing to Lemma 4.8,

$$\begin{aligned} &= \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{1}-\overline{a}}\left(((\overline{y}^2-\overline{y})+\overline{y})^{\overline{a}} - \overline{y}^{\overline{a}}\right) \\ &= C(\overline{x},\overline{y},\overline{y}^2-\overline{y}) , \end{aligned}$$

where we define

$$C(\overline{x},\overline{y},\overline{z}) := \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{1}-\overline{a}}\left((\overline{z}+\overline{y})^{\overline{a}} - \overline{y}^{\overline{a}}\right) .$$

As $(\overline{z}+\overline{y})^{\overline{a}} = \overline{y}^{\overline{a}}$ under $\overline{z}\leftarrow\overline{0}$ we have that $C(\overline{x},\overline{y},\overline{0}) = 0$.

g is t-sparse: As g is t-sparse and multilinear, so are each $g_{\overline{a}}$, so that $g_{\overline{a}}(\overline{x}) = \sum_{i=1}^{t}\alpha_{\overline{a},i}\overline{x}^{\overline{c}_{\overline{a},i}}$, and thus

$$\begin{aligned} C(\overline{x},\overline{y},\overline{z}) &= \sum_{\overline{a}}\sum_{i=1}^{t}\alpha_{\overline{a},i}\overline{x}^{\overline{c}_{\overline{a},i}}\overline{y}^{\overline{1}-\overline{a}}\left((\overline{z}+\overline{y})^{\overline{a}} - \overline{y}^{\overline{a}}\right) \\ &= \sum_{\overline{a}}\sum_{i=1}^{t}\alpha_{\overline{a},i}\overline{x}^{\overline{c}_{\overline{a},i}}\overline{y}^{\overline{1}-\overline{a}}(\overline{z}+\overline{y})^{\overline{a}} - \sum_{\overline{a}}\sum_{i=1}^{t}\alpha_{\overline{a},i}\overline{x}^{\overline{c}_{\overline{a},i}}\overline{y}^{\overline{1}} , \end{aligned}$$

where this is clearly an explicit depth-3 ($\sum\prod\sum$) multilinear formula (as $\overline{y}^{\overline{1}-\overline{a}}$ is variable-disjoint from $(\overline{z}+\overline{y})^{\overline{a}}$), and the size of this formula is $\mathsf{poly}(n,t,d)$ as there are at most t such $\overline{a}$ where $g_{\overline{a}} \neq 0$ as g is t-sparse. Continuing the expansion, appealing to Lemma 4.8,

$$
\begin{aligned}
&= \sum_{\overline{a}} \sum_{i=1}^{t} \alpha_{\overline{a},i} \overline{x}^{\overline{c}_{\overline{a},i}} \overline{y}^{\overline{1}-\overline{a}} \sum_{0 \leq \overline{b} \leq \overline{a}} \overline{z}^{\overline{b}} \overline{y}^{\overline{a}-\overline{b}} - \sum_{\overline{a}} \sum_{i=1}^{t} \alpha_{\overline{a},i} \overline{x}^{\overline{c}_{\overline{a},i}} \overline{y}^{\overline{1}} \\
&= \sum_{\overline{a}} \sum_{i=1}^{t} \alpha_{\overline{a},i} \overline{x}^{\overline{c}_{\overline{a},i}} \sum_{\overline{0} < \overline{b} \leq \overline{a}} \overline{z}^{\overline{b}} \overline{y}^{\overline{a}-\overline{b}} \\
&= \sum_{\overline{a}} \sum_{i=1}^{t} \sum_{\overline{0} < \overline{b} \leq \overline{a}} \alpha_{\overline{a},i} \overline{x}^{\overline{c}_{\overline{a},i}} \overline{z}^{\overline{b}} \overline{y}^{\overline{a}-\overline{b}} ,
\end{aligned}
$$

which is then easily seen to be explicit and $\mathsf{poly}(n,t,2^d)$-sparse appealing to the above logic.

g a multilinear formula: By interpolation, it follows that for each $\overline{a}$ there are $\mathsf{poly}(2^d)$-explicit constants $\overline{\alpha}_{\overline{a},\overline{\beta}}$ such that $g_{\overline{a}}(\overline{x}) = \sum_{\overline{\beta} \in \{0,1\}^d} \alpha_{\overline{a},\overline{\beta}} g(\overline{x},\overline{\beta})$. From this it follows that $g_{\overline{a}}$ is computable by a depth $D+1$ multilinear formula of size $\mathsf{poly}(t,2^d)$. Expanding the definition of C we get that

$$
\begin{aligned}
C(\overline{x},\overline{y},\overline{z}) &= \sum_{\overline{a}} g_{\overline{a}}(\overline{x}) \overline{y}^{\overline{1}-\overline{a}} \left((\overline{z}+\overline{y})^{\overline{a}} - \overline{y}^{\overline{a}} \right) \\
&= \sum_{\overline{a}} g_{\overline{a}}(\overline{x}) \overline{y}^{\overline{1}-\overline{a}} (\overline{z}+\overline{y})^{\overline{a}} - \sum_{\overline{a}} g_{\overline{a}}(\overline{x}) \overline{y}^{\overline{1}} \\
&= \sum_{\overline{a}} \sum_{\overline{\beta} \in \{0,1\}^d} \alpha_{\overline{a},\overline{\beta}} g(\overline{x},\overline{\beta}) \overline{y}^{\overline{1}-\overline{a}} (\overline{z}+\overline{y})^{\overline{a}} - \sum_{\overline{a}} \sum_{\overline{\beta} \in \{0,1\}^d} \alpha_{\overline{a},\overline{\beta}} g(\overline{x},\overline{\beta}) \overline{y}^{\overline{1}} ,
\end{aligned}
$$

which is clearly an explicit depth-$(D+2)$ multilinear formula of size $\mathsf{poly}(t,2^d)$, as $D \geq 1$ so that the computation of the $z_i + y_i$ is parallelized with computing $g(\overline{x},\overline{\beta})$, and we absorb the subtraction into the overall top-gate of addition. □

It is then straightforward to extend this to multilinearizing the product of a low-degree sparse multilinear polynomial and a multilinear formula, as we can use that multilinearization is linear.

Corollary 4.10. *Let $g \in \mathbb{F}[x_1,\ldots,x_n]$ be a multilinear polynomial and $f \in \mathbb{F}[\overline{x}]$ a s-sparse multilinear polynomial of degree $\leq d$. Then there is a $C \in \mathbb{F}[\overline{x},z_1,\ldots,z_n]$ such that*

$$
g \cdot f - \mathrm{ml}(g \cdot f) = C(\overline{x},\overline{x}^2 - \overline{x}) ,
$$

and $C(\overline{x},\overline{0}) = 0$.

If g is t-sparse, then C is computable as a $\mathsf{poly}(s,t,n,2^d)$-size depth-2 multilinear formula (which is $\mathsf{poly}(s,t,n,2^d)$-explicit given the computations for f,g), as well as being computable by a $\mathsf{poly}(s,t,n,d)$-size depth-3 multilinear formula with a $+$-output-gate (which is $\mathsf{poly}(s,t,n,d)$-explicit given the computations for f,g). If g is computable by a size-t depth-D multilinear formula, then C is computable by a $\mathsf{poly}(s,t,2^d)$-size depth-$(D+2)$ multilinear formula with a $+$-output-gate (which is $\mathsf{poly}(s,t,2^d)$-explicit given the computations for f,g). □

We now show how to derive multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}'}$ refutations for $\overline{f}, \overline{x}^2 - \overline{x}$ from equations of the form $\sum_j g_j f_j \equiv 1 \mod \overline{x}^2 - \overline{x}$.

Corollary 4.11. *Let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ be degree at most d multilinear s-sparse polynomials that are unsatisfiable over $\overline{x} \in \{0,1\}^n$. Suppose that there are multilinear $g_j \in \mathbb{F}[\overline{x}]$ such that*

$$\sum_{j=1}^{m} g_j(\overline{x}) f_j(\overline{x}) \equiv 1 \mod \overline{x}^2 - \overline{x} \, .$$

Then there is a multilinear-formula-$\mathrm{IPS}_{LIN'}$ refutation $C(\overline{x}, \overline{y}, \overline{z})$ such that

- *If the g_j are t-sparse, then C is computable by a $\mathsf{poly}(s,t,n,m,2^d)$-size depth-2 multilinear formula (which is $\mathsf{poly}(s,t,n,m,2^d)$-explicit given the computations for the f_j, g_j), as well as being computable by a $\mathsf{poly}(s,t,n,m,d)$-size depth-3 multilinear formula (which is $\mathsf{poly}(s,t,n,m,d)$-explicit given the computations for f_j, g_j).*
- *If the g_j are computable by size-t depth-D multilinear formula, then C is computable by a $\mathsf{poly}(s,t,m,2^d)$-size depth-$(D+2)$ multilinear formula (which is $\mathsf{poly}(s,t,m,2^d)$-explicit given the computations for f_j, g_j).*

Proof: construction: By the above multilinearization (Corollary 4.10), there are $C_j \in \mathbb{F}[\overline{x}, \overline{z}]$ such that

$$g_j(\overline{x}) f_j(\overline{x}) = \mathrm{ml}(g_j f_j) + C_j(\overline{x}, \overline{x}^2 - \overline{x}) \, .$$

where $C_j(\overline{x}, \overline{0}) = 0$. We now define

$$C(\overline{x}, \overline{y}, \overline{z}) := \sum_{j=1}^{m} \left(g_j(\overline{x}) y_j - C_j(\overline{x}, \overline{z}) \right) \, .$$

We now show that this is a valid IPS refutation. Observe that $C(\overline{x}, \overline{0}, \overline{0}) = 0$ and that

$$\begin{aligned} C(\overline{x}, \overline{f}, \overline{x}^2 - \overline{x}) &= \sum_{j=1}^{m} \big(g_j(\overline{x}) f_j(\overline{x}) - C_j(\overline{x}, \overline{x}^2 - \overline{x}) \big) \\ &= \sum_{j=1}^{m} \mathrm{ml}(g_j f_j) \, . \end{aligned}$$

As $\sum_{i=1}^{m} g_j(\overline{x}) f_j(\overline{x}) \equiv 1 \mod \overline{x}^2 - \overline{x}$ we have that

$$C(\overline{x}, \overline{f}, \overline{x}^2 - \overline{x}) = \sum_{j=1}^{m} \mathrm{ml}(g_j f_j) = \mathrm{ml}\left(\sum_{i=1}^{m} g_j(\overline{x}) f_j(\overline{x}) \right) = 1 \, ,$$

as desired.

complexity: The claim now follows from appealing to Corollary 4.10 for bounding the complexity of the C_j. That is, if the g_j are t-sparse then $\sum_{j=1}^{m} g_j(\overline{x}) y_j$ is tm-sparse and thus computable by a $\mathsf{poly}(t,m)$-size depth-2 multilinear formula with $+$-output-gate. As each C_j is computable by a $\mathsf{poly}(s,t,n,2^d)$-size depth-2 or $\mathsf{poly}(s,t,n,d)$-size depth-3 multilinear formula (each having a $+$-output-gate), it follows that $C(\overline{x},\overline{y},\overline{z}) := \sum_{j=1}^{m}(g_j(\overline{x})y_j - C_j(\overline{x},\overline{z}))$ can be explicitly computed by a $\mathsf{poly}(s,t,n,m,2^d)$-size depth-2 or $\mathsf{poly}(s,t,n,m,d)$-size depth-3 multilinear formula (collapsing addition gates into a single level).

If the g_j are computable by size-t depth-D multilinear formula then $\sum_{j=1}^{m} g_j(\overline{x})y_j$ is computable by size $\mathsf{poly}(m,t)$-size depth-$(D+2)$ multilinear formula (with a $+$-output-gate), and each C_j is computable by a $\mathsf{poly}(s,t,2^d)$-size depth-$(D+2)$ multilinear formula with a $+$-output-gate, from which it follows that $C(\overline{x},\overline{y},\overline{z}) := \sum_{j=1}^{m}(g_j(\overline{x})y_j - C_j(\overline{x},\overline{z}))$ can be explicitly computed by a $\mathsf{poly}(s,t,m,2^d)$-size depth-$(D+2)$ multilinear formula by collapsing addition gates. □

We now conclude by showing that multilinear-formula-IPS$_{\text{LIN}'}$ can efficiently simulate sparse-IPS$_{\text{LIN}}$ when the axioms are low-degree. As this latter system is complete, this shows the former is as well. That is, we allow the refutation to depend non-linearly on the Boolean axioms, but only linearly on the other axioms.

Corollary 4.12. *Let $f_1,\ldots,f_m \in \mathbb{F}[x_1,\ldots,x_n]$ be degree at most d s-sparse polynomials unsatisfiable over $\overline{x} \in \{0,1\}^n$. Suppose they have an $\sum\prod$-IPS$_{\text{LIN}}$ refutation, that is, that there are t-sparse polynomials $g_1,\ldots,g_m,h_1,\ldots,h_n \in \mathbb{F}[\overline{x}]$ such that $\sum_{j=1}^{m} g_j f_j + \sum_{i=1}^{n} h_i \cdot (x_i^2 - x_i) = 1$. Then $\overline{f}, \overline{x}^2 - \overline{x}$ can be refuted by a depth-2 multilinear-formula-IPS$_{LIN'}$ proof of $\mathsf{poly}(s,t,n,m,2^d)$-size, or by a depth-3 multilinear-formula-IPS$_{LIN'}$ proof of $\mathsf{poly}(s,t,n,m,d)$-size.*

Proof: This follows from Theorem 4.11 by noting that $\sum_j \mathrm{ml}(g_j) f_j \equiv 1 \mod \overline{x}^2 - \overline{x}$, and that the $\mathrm{ml}(g_j)$ are multilinear and t-sparse. □

4.4 Refutations of the subset-sum axiom

We now give efficient IPS refutations of the subset-sum axiom, where these IPS refutations can be even placed in the restricted roABP-IPS$_{\text{LIN}}$ or multilinear-formula-IPS$_{\text{LIN}}$ subclasses. That is, we give such refutations for whenever the polynomial $\sum_i \alpha_i x_i - \beta$ is unsatisfiable over the Boolean cube $\{0,1\}^n$, where the size of the refutation is polynomial in the size of the set $A := \{\sum_i \alpha_i x_i : \overline{x} \in \{0,1\}^n\}$. A motivating example is when $\overline{\alpha} = \overline{1}$ so that $A = \{0,\ldots,n\}$.

To construct our refutations, we first show that there is an efficiently computable polynomial f such that $f(\overline{x}) \cdot (\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}$. This will be done by considering the univariate polynomial $p(t) := \prod_{\alpha \in A}(t-\alpha)$. Using that for any univariate $p(x)$ that $x-y$ divides $p(x)-p(y)$, we see that $p(\sum_i \alpha_i x_i) - p(\beta)$ is a multiple of $\sum_i \alpha_i x_i - \beta$. As $\sum_i \alpha_i x_i - \beta$ is unsatisfiable it must be that $\beta \notin A$. This implies that $p(\sum_i \alpha_i x_i) \equiv 0 \mod \overline{x}^2 - \overline{x}$ while $p(\beta) \neq 0$. Consequently, $p(\sum_i \alpha_i x_i) - p(\beta)$ is equivalent to a nonzero constant modulo $\overline{x}^2 - \overline{x}$, yielding the Nullstellensatz refutation

$$\frac{1}{-p(\beta)} \cdot \frac{p(\sum_i \alpha_i x_i) - p(\beta)}{\sum_i \alpha_i x_i - \beta} \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}\,.$$

By understanding the quotient $\frac{p(\sum_i \alpha_i x_i) - p(\beta)}{\sum_i \alpha_i x_i - \beta}$ we see that it can be efficiently computed by a small $\sum\bigwedge\sum$ formula and thus an roABP, so that using our multilinearization result for roABPs (Theorem 4.5) this yields the desired roABP-IPS$_{\text{LIN}}$ refutation. However, this does not yield the desired multilinear-formula-IPS$_{\text{LIN}}$ refutation. For this, we need to (over a large field) convert the above quotient to a sum of products of univariates using duality (Theorem 3.14). We can then multilinearize this to a sum of products of *linear* univariates, which is a depth-3 multilinear formula. By appealing to our proof-of-multilinearization result for multilinear formula (Theorem 4.11) one obtains a multilinear-formula-$\text{IPS}_{\text{LIN}'}$ refutation, and we give a direct proof which actually yields the desired multilinear-formula-IPS_{LIN} refutation.

We briefly remark that for the special case of $\overline{\alpha} = \overline{1}$, one can explicitly describe the unique multilinear polynomial f such that $f(\overline{x})(\sum_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}$. This description (Theorem B.1) shows that f is a linear combination of elementary symmetric polynomials, which implies the desired complexity upper bounds for this case via known bounds on the complexity of elementary symmetric polynomials ([54]). However, this proof strategy is more technical and thus we pursue the more conceptual outline given above to bound the complexity of f for general A.

Proposition 4.13. *Let $\overline{\alpha} \in \mathbb{F}^n$, $\beta \in \mathbb{F}$ and $A := \{\sum_{i=1}^n \alpha_i x_i : \overline{x} \in \{0,1\}^n\}$ be so that $\beta \notin A$. Then there is a multilinear polynomial $f \in \mathbb{F}[\overline{x}]$ such that*

$$f(\overline{x}) \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x} \,.$$

For any $|\mathbb{F}|$, f is computable by a $\mathsf{poly}(|A|,n)$-explicit $\mathsf{poly}(|A|,n)$-width roABP of individual degree ≤ 1.

If $|\mathbb{F}| \ge \mathsf{poly}(|A|,n)$, then f is computable as a sum of product of linear univariates (and hence a depth-3 multilinear formula)

$$f(\overline{x}) = \sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n) \,,$$

where each $f_{i,j} \in \mathbb{F}[x_i]$ has $\deg f_{i,j} \le 1$, $s \le \mathsf{poly}(|A|,n)$, and this expression is $\mathsf{poly}(|A|,n)$-explicit.

Proof: computing A: We first note that A can be computed from $\overline{\alpha}$ in $\mathsf{poly}(|A|,n)$-steps (as opposed to the naive 2^n steps). That is, define $A_j := \{\sum_{i=1}^j \alpha_i x_i : x_1,\ldots,x_j \in \{0,1\}\}$, so that $A_0 = \emptyset$ and $A_n = A$. It follows that for all j, we have that $A_j \subseteq A$ and thus $|A_j| \le |A|$, and that $A_{j+1} \subseteq A_j \cup (A_j + \alpha_j)$. It follows that we can compute A_{j+1} from A_j in $\mathsf{poly}(|A|)$ time, so that $A = A_n$ can be computed in $\mathsf{poly}(|A|,n)$-time.

defining f: Define $p(t) \in \mathbb{F}[t]$ by $p(t) := \prod_{\alpha \in A}(t - \alpha)$, so that $p(A) = 0$ and $p(\beta) \neq 0$. Express $p(t)$ in its monomial representation as $p(t) = \sum_{k=0}^{|A|} \gamma_k t^k$, where the γ_k can be computed in $\mathsf{poly}(|A|)$ time from $\overline{\alpha}$ by computing A as above. Then observe that

$$\begin{aligned} p(t) - p(\beta) &= \left(\sum_{k=0}^{|A|} \gamma_k \frac{t^k - \beta^k}{t - \beta}\right)(t - \beta) \\ &= \left(\sum_{k=0}^{|A|} \gamma_k \sum_{j=0}^{k-1} t^j \beta^{(k-1)-j}\right)(t - \beta) \\ &= \left(\sum_{j=0}^{|A|-1} \left(\sum_{k=j+1}^{|A|} \gamma_k \beta^{(k-1)-j}\right) t^j\right)(t - \beta) \,. \end{aligned}$$

Thus, plugging in $t \leftarrow \sum_i \alpha_i x_i$, we can define the polynomial $g(\overline{x}) \in \mathbb{F}[\overline{x}]$ by

$$\begin{aligned} g(\overline{x}) &:= \frac{p(\sum_i \alpha_i x_i) - p(\beta)}{\sum_i \alpha_i x_i - \beta} \\ &= \sum_{j=0}^{|A|-1} \left(\sum_{k=j+1}^{|A|} \gamma_k \beta^{(k-1)-j} \right) \left(\sum_i \alpha_i x_i \right)^j . \end{aligned} \tag{4.1}$$

Hence,

$$g(\overline{x})(\textstyle\sum_i \alpha_i x_i - \beta) = p(\sum_i \alpha_i x_i) - p(\beta) .$$

For any $\overline{x} \in \{0,1\}^n$ we have that $\sum_i \alpha_i x_i \in A$. As $p(A) = 0$ it follows that $p(\sum_i \alpha_i x_i) = 0$ for all $\overline{x} \in \{0,1\}^n$. This implies that $p(\sum_i \alpha_i x_i) \equiv 0 \mod \overline{x}^2 - \overline{x}$, yielding

$$g(\overline{x})(\textstyle\sum_i \alpha_i x_i - \beta) \equiv -p(\beta) \mod \overline{x}^2 - \overline{x} .$$

As $-p(\beta) \in \mathbb{F} \setminus \{0\}$, we have that

$$\frac{1}{-p(\beta)} \cdot g(\overline{x}) \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x} .$$

We now simply multilinearize, and thus define the multilinear polynomial

$$f(\overline{x}) := \mathrm{ml}\left(\frac{1}{-p(\beta)} \cdot g(\overline{x}) \right) .$$

First, we see that this has the desired form, using the interaction of multilinearization and multiplication (Theorem 3.12).

$$\begin{aligned} 1 &= \mathrm{ml}\left(\frac{1}{-p(\beta)} g(\overline{x}) \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \right) \\ &= \mathrm{ml}\left(\mathrm{ml}\left(\frac{1}{-p(\beta)} \cdot g(\overline{x}) \right) \mathrm{ml}(\textstyle\sum_i \alpha_i x_i - \beta) \right) \\ &= \mathrm{ml}\left(f \cdot \mathrm{ml}(\textstyle\sum_i \alpha_i x_i - \beta) \right) \\ &= \mathrm{ml}\left(f \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \right) . \end{aligned}$$

Thus, $f \cdot (\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}$ as desired.

computing f as an roABP: By Equation 4.1 we see that $g(\overline{x})$ is computable by a $\mathrm{poly}(|A|, n)$-size $\sum\bigwedge\sum$-formula, and by the efficient simulation of $\sum\bigwedge\sum$-formula by roABPs (Theorem 3.15) $g(\overline{x})$ and thus $\frac{1}{-p(\beta)} \cdot g(\overline{x})$ are computable by $\mathrm{poly}(|A|, n)$-width roABPs of $\mathrm{poly}(|A|, n)$-degree. Noting that roABPs can be efficiently multilinearized (Theorem 4.5) we see that $f = \mathrm{ml}(\frac{1}{-p(\beta)} \cdot g(\overline{x}))$ can thus be computed by such an roABP also, where the individual degree of this roABP is at most 1. Finally, note that each of these steps has the required explicitness.

computing f via duality: We apply duality (Theorem 3.14) to see that over large enough fields there are univariates $g_{j,\ell,i}$ of degree at most $|A|$, where

$$\begin{aligned} g(\bar{x}) &= \sum_{j=0}^{|A|-1} \left(\sum_{k=j+1}^{|A|} \gamma_k \beta^{(k-1)-j} \right) \left(\sum_i \alpha_i x_i \right)^j \\ &= \sum_{j=0}^{|A|-1} \left(\sum_{k=j+1}^{|A|} \gamma_k \beta^{(k-1)-j} \right) \sum_{\ell=1}^{(nj+1)(j+1)} g_{j,\ell,1}(x_1) \cdots g_{j,\ell,n}(x_n) \end{aligned}$$

Absorbing the constant $\left(\sum_{k=j+1}^{|A|} \gamma_k \beta^{(k-1)-j}\right)$ into these univariates and re-indexing,

$$= \sum_{i=1}^{s} g_{i,1}(x_1) \cdots g_{i,n}(x_n)$$

for some univariates $g_{i,j}$, where $s \le |A|(n|A|+1)(|A|+1) = \mathsf{poly}(|A|,n)$.

We now obtain f by multilinearizing the above expression.

$$\begin{aligned} f &= \mathrm{ml}\left(\frac{1}{-p(\beta)} g(\bar{x}) \right) \\ &= \mathrm{ml}\left(\frac{1}{-p(\beta)} \sum_{i=1}^{s} g_{i,1}(x_1) \cdots g_{i,n}(x_n) \right) \end{aligned}$$

absorbing the constant $1/_{-p(\beta)}$ and renaming,

$$\begin{aligned} &= \mathrm{ml}\left(\sum_{i=1}^{s} g'_{i,1}(x_1) \cdots g'_{i,n}(x_n) \right) \\ &= \mathrm{ml}\left(\sum_{i=1}^{s} \mathrm{ml}(g'_{i,1}(x_1)) \cdots \mathrm{ml}(g'_{i,n}(x_n)) \right) \end{aligned}$$

defining $f_{i,j}(x_j) := \mathrm{ml}(g_{i,j}(x_j))$, so that $\deg f_{i,j} \le 1$,

$$= \mathrm{ml}\left(\sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n) \right)$$

and we can drop the outside ml as this expression is now multilinear,

$$= \sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n) \ ,$$

showing that f is computable as desired, noting that this expression has the desired explicitness. □

Note that computing f via duality also implies an roABP for f, but only in large enough fields $|\mathbb{F}| \geq \mathsf{poly}(|A|,n)$. Of course, the field must at least have $|\mathbb{F}| \geq |A|$, but by using the field-independent conversion of $\sum\bigwedge\sum$ to roABP (Theorem 3.15) this shows that $\mathbb{F}$ need not be any larger than A for the refutation to be efficient.

The above shows that one can give an "IPS proof" $g(\overline{x})(\sum_i \alpha_i x_i - \beta) + \sum_i h_i(\overline{x})(x_i^2 - x_i) = 1$, where g is efficiently computable. However, this may be an inefficient IPS proof as it does not bound the complexity of the h_i. We now extend this to an efficient IPS proof by using the above multilinearization results for roABPs (Theorem 4.6), leveraging that $\sum_i \alpha_i x_i - \beta$ is computable by an roABP in any order of the variables (and that the above result works in any order of the variables).

Corollary 4.14. *Let $\overline{\alpha} \in \mathbb{F}^n$, $\beta \in \mathbb{F}$ and $A := \{\sum_{i=1}^n \alpha_i x_i : \overline{x} \in \{0,1\}^n\}$ be so that $\beta \notin A$. Then $\sum_i \alpha_i x_i - \beta, \overline{x}^2 - \overline{x}$ has a $\mathsf{poly}(|A|,n)$-explicit roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation of individual degree 2 computable in width-$\mathsf{poly}(|A|,n)$ in any order of the variables.* □

Note that while the above results give a small $\sum\bigwedge\sum$ formula g such that $g \cdot (\sum_i \alpha_i x_i - \beta) \equiv -p(\beta) \mod \overline{x}^2 - \overline{x}$ for nonzero scalar $-p(\beta)$, this does not translate to a $\sum\bigwedge\sum$-IPS refutation as $\sum\bigwedge\sum$ formulas cannot be multilinearized efficiently (see the discussion in Section 4.2).

We now turn to refuting the subset-sum axioms by multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ (which is not itself a complete proof system, but will suffice here). While one can use the multilinearization techniques for multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN'}}$ of Theorem 4.11 it gives slightly worse results due to its generality, so we directly multilinearize the refutations we built above using that the subset-sum axiom is linear.

Proposition 4.15. *Let $\overline{\alpha} \in \mathbb{F}^n$, $\beta \in \mathbb{F}$ and $A := \{\sum_{i=1}^n \alpha_i x_i : \overline{x} \in \{0,1\}^n\}$ be so that $\beta \notin A$. If $|\mathbb{F}| \geq \mathsf{poly}(|A|,n)$, then $\sum_i \alpha_i x_i - \beta, \overline{x}^2 - \overline{x}$ has a $\mathsf{poly}(|A|,n)$-explicit $\mathsf{poly}(|A|,n)$-size depth-3 multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation.*

Proof: By Theorem 4.13, there is a multilinear polynomial $f \in \mathbb{F}[\overline{x}]$ such that $f(\overline{x}) \cdot (\sum_i \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}$, and f is explicitly given as

$$f(\overline{x}) = \sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n) ,$$

where each $f_{i,j} \in \mathbb{F}[x_i]$ has $\deg f_{i,j} \leq 1$ and $s \leq \mathsf{poly}(|A|,n)$.

We now efficiently prove that $f(\overline{x}) \cdot (\sum_{i=1}^n \alpha_i x_i - \beta)$ is equal to its multilinearization (which is 1) modulo the Boolean cube. The key step is that for a linear function $p(x) = \gamma x + \delta$ we have that $(\gamma x + \delta)x = (\gamma + \delta)x + \gamma(x^2 - x) = p(1)x + (p(1) - p(0))(x^2 - x)$.

Thus,

$$
\begin{aligned}
&f(\overline{x}) \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \\
&= \left(\sum_{i=1}^{s} f_{i,1}(x_1) \cdots f_{i,n}(x_n)\right) \cdot (\textstyle\sum_i \alpha_i x_i - \beta) \\
&= \sum_{i=1}^{s} -\beta f_{i,1}(x_1) \cdots f_{i,n}(x_n) \\
&\qquad + \sum_{i=1}^{s} \sum_{j=1}^{n} \alpha_j \prod_{k \neq j} f_{i,k}(x_k) \cdot \Big(f_{i,j}(1) x_j + (f_{i,j}(1) - f_{i,j}(0))(x_j^2 - x_j) \Big) \\
&= \sum_{i=1}^{s} -\beta f_{i,1}(x_1) \cdots f_{i,n}(x_n) + \sum_{i=1}^{s} \sum_{j=1}^{n} \alpha_j \prod_{k \neq j} f_{i,k}(x_k) \cdot f_{i,j}(1) x_j \\
&\qquad + \sum_{i=1}^{s} \sum_{j=1}^{n} \alpha_j \prod_{k \neq j} f_{i,k}(x_k) \cdot (f_{i,j}(1) - f_{i,j}(0)) \cdot (x_j^2 - x_j)
\end{aligned}
$$

absorbing constants and renaming, using $j = 0$ to incorporate the above term involving β,

$$
= \sum_{i=1}^{s} \sum_{j=0}^{n} \prod_{k=1}^{n} f_{i,j,k}(x_k) + \sum_{j=1}^{n} \left(\sum_{i=1}^{s} \prod_{k=1}^{n} h_{i,j,k}(x_k) \right) (x_j^2 - x_j)
$$

where each $f_{i,j,k}$ and $h_{i,j,k}$ are linear univariates. As $f(\overline{x}) \cdot (\sum_{i=1}^{n} \alpha_i x_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}$ it follows that $\sum_i \sum_j \prod_k f_{i,j,k}(x_k) \equiv 1 \mod \overline{x}^2 - \overline{x}$, but as each $f_{i,j,k}$ is linear it must actually be that $\sum_i \sum_j \prod_k f_{i,j,k}(x_k) = 1$, so that,

$$
= 1 + \sum_{j=1}^{n} \left(\sum_{i=1}^{s} \prod_{k=1}^{n} h_{i,j,k}(x_k) \right) (x_j^2 - x_j) \; .
$$

Define $C(\overline{x}, y, \overline{z}) := f(\overline{x}) y - \sum_{j=1}^{n} h_j(\overline{x}) z_j$, where $h_j(\overline{x}) := \sum_{i=1}^{s} \prod_{k=1}^{n} h_{i,j,k}(x_k)$. It follows that $C(\overline{x}, 0, \overline{0}) = 0$ and that $C(\overline{x}, \sum_i \alpha_i x_i - \beta, \overline{x}^2 - \overline{x}) = 1$, so that C is a linear-IPS refutation. Further, as each f, h_j is computable as a sum of products of linear univariates, these are depth-3 multilinear formulas. By distributing the multiplication of the variables $y, z_1, \ldots, z_n$ to the bottom multiplication gates, we see that C itself has a depth-3 multilinear formula of the desired complexity. □

5 Lower bounds for linear-IPS via functional lower bounds

In this section we give *functional* circuit lower bounds for various measures of algebraic complexity, such as degree, sparsity, roABPs and multilinear formulas. That is, while algebraic complexity typically treats a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ as a *syntactic* object, one can also see that it defines a function on the Boolean cube $\hat{f} : \{0,1\}^n \to \mathbb{F}$. If this function is particularly complicated then one would expect that this implies that the polynomial f requires large algebraic circuits. In this section we obtain such lower bounds, showing that for *any* polynomial f (not necessarily multilinear) that agrees with a certain function on the Boolean cube must in fact have large algebraic complexity.

Our lower bounds will proceed by first showing that the complexity of f is an upper bound for the complexity of its multilinearization $\mathrm{ml}(f)$. While such a statement is known to be false for general circuits (under plausible assumptions, see Section 4.2), such efficient multilinearization can be shown for the particular restricted models of computation we consider. In particular, this multilinearization is easy for degree and sparsity, for multilinear formulas (as f is already multilinear), and for roABPs this is seen in Section 4.2. As then $\mathrm{ml}(f)$ is uniquely defined by the function $\hat{f}$ (Theorem 3.12), we then only need to lower bound the complexity of $\mathrm{ml}(f)$ using standard techniques. We remark that the actual presentation will not follow the above exactly, as for roABPs and multilinear formulas it is just as easy to work directly with the underlying lower bound technique.

We then observe that by deriving such lower bounds for carefully crafted functions $\hat{f}:\{0,1\}^n \to \mathbb{F}$ one can easily obtain linear-IPS lower bounds for the above circuit classes. That is, consider the system of equations $f(\overline{x}), \overline{x}^2-\overline{x}$, where $f(\overline{x})$ is chosen so this system is unsatisfiable, hence $f(\overline{x}) \neq 0$ for all $\overline{x} \in \{0,1\}^n$. Any linear-IPS refutation yields an equation $g(\overline{x}) \cdot f(\overline{x}) + \sum_i h_i(\overline{x})(x_i^2 - x_i) = 1$, which implies that $g(\overline{x}) = 1/f(\overline{x})$ for all $\overline{x} \in \{0,1\}^n$ (that this system is unsatisfiable allows us to avoid division by zero). It follows that the polynomial $g(\overline{x})$ agrees with the function $\hat{g}(\overline{x}) := 1/f(\overline{x})$ on the Boolean cube. If the function $\hat{g}$ has a functional lower bound then this implies g must have large complexity, giving the desired lower bound for the linear-IPS refutation.

The section proceeds as follows. We begin by detailing the above strategy for converting functional lower bounds into lower bounds for linear-IPS. We then derive a tight functional lower bound of n for the degree of $1/(\sum_i x_i + 1)$. We then extend this via random restrictions to a functional lower bound of $\exp(\Omega(n))$ on the sparsity of $1/(\sum_i x_i + 1)$. We can then lift this degree bound to a functional lower bound of 2^n on the evaluation dimension of $1/(\sum_i x_i y_i + 1)$ in the $\overline{x}|\overline{y}$ partition, which we then symmetrize to obtain a functional lower bound on the evaluation dimension in any partition of the related function $1/\left(\sum_{i<j} z_{i,j} x_i x_j + 1\right)$. In each case, the resulting linear-IPS lower bounds are immediate via the known relations of these measures to circuit complexity classes (Section 3).

5.1 The strategy

We give here the key lemma detailing the general reduction from linear-IPS lower bounds to functional lower bounds.

Lemma 5.1. *Let $\mathcal{C} \subseteq \mathbb{F}[x_1, \ldots, x_n]$ be a set of polynomials closed under partial substitutions. Let $f \in \mathbb{F}[\overline{x}]$, where the system $f(\overline{x}), \overline{x}^2 - \overline{x}$ is unsatisfiable. Suppose that for all $g \in \mathbb{F}[\overline{x}]$ with*

$$g(\overline{x}) = \frac{1}{f(\overline{x})}, \qquad \forall \overline{x} \in \{0,1\}^n ,$$

that $g \notin \mathcal{C}$. Then $f(\overline{x}), \overline{x}^2 - \overline{x}$ does not have $\mathcal{C}$-$\mathrm{IPS_{LIN}}$ refutations (nor $\mathcal{C}$-$\mathrm{IPS}_{LIN'}$ refutations).

Proof: Suppose for contradiction that $f(\overline{x}), \overline{x}^2 - \overline{x}$ has the $\mathcal{C}$-$\mathrm{IPS_{LIN}}$ refutation $C(\overline{x}, y, \overline{z}) = g(\overline{x}) \cdot y + \sum_i h_i(\overline{x}) \cdot z_i$ where $C(\overline{x}, f, \overline{x}^2 - \overline{x}) = 1$ (and clearly $C(\overline{x}, 0, \overline{0}) = 0$). As $g = C(\overline{x}, 1, \overline{0})$, it follows that $g \in \mathcal{C}$

from the closure properties we assumed of $\mathcal{C}$. Thus,

$$\begin{aligned} 1 &= C(\overline{x}, f, \overline{x}^2 - \overline{x}) \\ &= g(\overline{x}) \cdot f(\overline{x}) + \sum_i h_i(\overline{x})(x_i^2 - x_i) \\ &\equiv g(\overline{x}) \cdot f(\overline{x}) \mod \overline{x}^2 - \overline{x} \,. \end{aligned}$$

Thus, for any $\overline{x} \in \{0,1\}^n$, as $f(\overline{x}) \neq 0$,

$$g(\overline{x}) = {}^1\!/_{f(\overline{x})} \,.$$

However, this yields the desired contradiction, as this contradicts the assumed functional lower bound for ${}^1\!/_f$. □

We now note that the lower bound strategy of using functional lower bounds actually produces lower bounds for $\mathrm{IPS}_{\mathrm{LIN}'}$ (and even for the full IPS system if we have multilinear polynomials), and not just $\mathrm{IPS}_{\mathrm{LIN}}$. This is because we work modulo the Boolean axioms, so that any non-linear dependence on these axioms vanishes, only leaving a linear dependence on the remaining axioms. This slightly stronger lower bound is most interesting for multilinear-formulas, where the $\mathrm{IPS}_{\mathrm{LIN}}$ version is not complete in general (Theorem 4.7) (though it is still interesting due to its short refutations of the subset-sum axiom (Theorem 4.15)), while the $\mathrm{IPS}_{\mathrm{LIN}'}$ version is complete (Theorem 4.12).

Lemma 5.2. *Let $\mathcal{C} \subseteq \mathbb{F}[x_1, \ldots, x_n]$ be a set of polynomials closed under partial substitutions, and let $\mathcal{D}$ be the set of differences of $\mathcal{C}$, that is, $\mathcal{D} := \{p(\overline{x}) - q(\overline{x}) : p, q \in \mathcal{C}\}$. Let $f \in \mathbb{F}[\overline{x}]$, where the system $f(\overline{x}), \overline{x}^2 - \overline{x}$ is unsatisfiable. Suppose that for all $g \in \mathbb{F}[\overline{x}]$ with*

$$g(\overline{x}) = \frac{1}{f(\overline{x})}, \qquad \forall \overline{x} \in \{0,1\}^n \,,$$

that $g \notin \mathcal{D}$. Then $f(\overline{x}), \overline{x}^2 - \overline{x}$ does not have $\mathcal{C}$-$\mathrm{IPS}_{LIN'}$ refutations.

Furthermore, if $\mathcal{C}$ (and thus $\mathcal{D}$) are a set of multilinear polynomials, then $f(\overline{x}), \overline{x}^2 - \overline{x}$ does not have $\mathcal{C}$-IPS refutations.

Proof: Suppose for contradiction that $f(\overline{x}), \overline{x}^2 - \overline{x}$ has the $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}'}$ refutation $C(\overline{x}, y, \overline{z})$. That $\deg_y C(\overline{x}, y, \overline{z}) \leq 1$ implies there is the decomposition $C(\overline{x}, y, \overline{z}) = C_1(\overline{x}, \overline{z})y + C_0(\overline{x}, \overline{z})$. As $C_1(\overline{x}, \overline{0}) = C(\overline{x}, 1, \overline{0}) - C(\overline{x}, 0, \overline{0})$, the assumed closure properties imply that $C_1(\overline{x}, \overline{0}) \in \mathcal{D}$. By the definition of an IPS refutation, we have that $0 = C(\overline{x}, 0, \overline{0}) = C_1(\overline{x}, \overline{0}) \cdot 0 + C_0(\overline{x}, \overline{0})$, so that $C_0(\overline{x}, \overline{0}) = 0$. By using the definition of an IPS refutation again, we have that $1 = C(\overline{x}, f, \overline{x}^2 - \overline{x}) = C_1(\overline{x}, \overline{x}^2 - \overline{x}) \cdot f + C_0(\overline{x}, \overline{x}^2 - \overline{x})$, so that modulo the Boolean axioms,

$$\begin{aligned} 1 &= C_1(\overline{x}, \overline{x}^2 - \overline{x}) \cdot f + C_0(\overline{x}, \overline{x}^2 - \overline{x}) \\ &\equiv C_1(\overline{x}, \overline{0}) \cdot f + C_0(\overline{x}, \overline{0}) \mod \overline{x}^2 - \overline{x} \end{aligned}$$

using that $C_0(\overline{x}, \overline{0}) = 0$,

$$\equiv C_1(\overline{x}, \overline{0}) \cdot f \mod \overline{x}^2 - \overline{x} \,.$$

Thus, for every $\overline{x} \in \{0,1\}^n$ we have that $C_1(\overline{x},\overline{0}) = 1/f(\overline{x})$ so that by the assumed functional lower bound $C_1(\overline{x},\overline{0}) \notin \mathcal{D}$, yielding the desired contradiction to the above $C_1(\overline{x},\overline{0}) \in \mathcal{D}$.

Now suppose that $\mathcal{C}$ is a set of multilinear polynomials. Any $\mathcal{C}$-IPS refutation $C(\overline{x},y,\overline{z})$ of $f(\overline{x}),\overline{x}^2-\overline{x}$ thus must have $\deg_y C \le 1$ as C is multilinear, so that C is actually a $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}'}$ refutation, thus the above lower bound applies. □

5.2 Degree of a polynomial

We now turn to obtaining functional lower bounds, and deriving the corresponding linear-IPS lower bounds. We begin with a particularly weak form of algebraic complexity, the degree of a polynomial. While it is trivial to obtain such bounds in some cases (as any polynomial that agrees with the AND function on the Boolean cube $\{0,1\}^n$ must have degree $\ge n$), for our applications to proof complexity we will need such degree bounds for functions defined by $\hat{f}(\overline{x}) = 1/f(\overline{x})$ for simple polynomials $f(\overline{x})$.

We show that any multilinear polynomial agreeing with $1/f(\overline{x})$, where $f(\overline{x})$ is the subset-sum axiom $\sum_i x_i - \beta$, must have the maximal degree n. We note that a degree lower bound of $\lceil n/2 \rceil + 1$ was established by Impagliazzo, Pudlák, and Sgall [41] (Theorem A.4). They actually established this degree bound [10] when $f(\overline{x}) = \sum_i \alpha_i x_i - \beta$ for *any* $\overline{\alpha}$, while we only consider $\overline{\alpha} = \overline{1}$ here. However, we need the tight bound of n here as it will be used crucially in the proof of Theorem 5.8.

Proposition 5.3. *Let $n \ge 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0,\ldots,n\}$. Let $f \in \mathbb{F}[x_1,\ldots,x_n]$ be a multilinear polynomial such that*

$$f(\overline{x})\left(\sum_i x_i - \beta\right) = 1 \mod \overline{x}^2 - \overline{x}\,.$$

Then $\deg f = n$.

Proof: $\underline{\le n}$: This is clear as f is multilinear.

$\underline{\ge n}$: Begin by observing that as $\beta \notin \{0,\ldots,n\}$, this implies that $\sum_i x_i - \beta$ is never zero on the Boolean cube, so that the above functional equation implies that for $\overline{x} \in \{0,1\}^n$ the expression

$$f(\overline{x}) = \frac{1}{\sum_i x_i - \beta}\,,$$

is well defined.

Now observe that this implies that f is a symmetric polynomial. That is, define the multilinear polynomial g by symmetrizing f,

$$g(x_1,\ldots,x_n) := \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} f(x_{\sigma(1)},\ldots,x_{\sigma(n)})\,,$$

[10] The degree lower bound of Impagliazzo, Pudlák, and Sgall [41] (Theorem A.4) actually holds for the (dynamic) polynomial calculus proof system (see Appendix A), while we only consider the (static) Nullstellensatz proof system here. Note that for polynomial calculus Impagliazzo, Pudlák, and Sgall [41] also showed a matching upper bound of $\lceil n/2 \rceil + 1$ for $\overline{\alpha} = \overline{1}$.

where $\mathfrak{S}_n$ is the symmetric group on n symbols. Then we see that f and g agree on $\bar{x} \in \{0,1\}^n$, as

$$\begin{aligned} g(\bar{x}) &= \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} f(x_{\sigma(1)},\ldots,x_{\sigma(n)}) \\ &= \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} \frac{1}{\sum_i x_{\sigma(i)}-\beta} = \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\frac{1}{\sum_i x_i-\beta} \\ &= \frac{1}{n!}\cdot n!\cdot\frac{1}{\sum_i x_i-\beta} = \frac{1}{\sum_i x_i-\beta} = f(\bar{x})\,. \end{aligned}$$

It follows then that $g = f$ as polynomials, since they are multilinear and agree on the Boolean cube (Theorem 3.12). As g is clearly symmetric, so is f. Thus f can be expressed as $f = \sum_{k=0}^{d}\gamma_k S_{n,k}(\bar{x})$, where $d := \deg f$, $S_{n,k} := \sum_{S\in\binom{[n]}{k}}\prod_{i\in S} x_i$ is the k-th elementary symmetric polynomial, and $\gamma_k \in \mathbb{F}$ are scalars with $\gamma_d \neq 0$.

Now observe that for $k < n$, we can understand the action of multiplying $S_{n,k}$ by $\sum_i x_i - \beta$.

$$\begin{aligned} \left(\sum_i x_i-\beta\right)S_{n,k}(\bar{x}) &= \sum_{S\in\binom{[n]}{k}}\left(\sum_i x_i-\beta\right)\prod_{j\in S}x_j \\ &= \sum_{S\in\binom{[n]}{k}}\left(\sum_{i\notin S}x_i\prod_{j\in S}x_j + \sum_{i\in S}x_i\prod_{j\in S}x_j - \beta\prod_{j\in S}x_j\right) \\ &= \sum_{S\in\binom{[n]}{k}}\left(\sum_{\substack{|T|=k+1\\ T\supseteq S}}\prod_{j\in T}x_j + (k-\beta)\prod_{j\in S}x_j\right) \mod \bar{x}^2-\bar{x} \\ &= (k+1)S_{n,k+1} + (k-\beta)S_{n,k} \mod \bar{x}^2-\bar{x}\,. \end{aligned}$$

Note that we used that each subset of $[n]$ of size $k+1$ contains exactly $k+1$ subsets of size k.

Putting the above together, suppose for contradiction that $d < n$. Then,

$$\begin{aligned} 1 &= f(\bar{x})\left(\sum_i x_i-\beta\right) \mod \bar{x}^2-\bar{x} \\ &= \left(\sum_{k=0}^{d}\gamma_k S_{n,k}\right)\left(\sum_i x_i-\beta\right) \mod \bar{x}^2-\bar{x} \\ &= \left(\sum_{k=0}^{d}\gamma_k\Big((k+1)S_{n,k+1}+(k-\beta)S_{n,k}\Big)\right) \mod \bar{x}^2-\bar{x} \\ &= \gamma_d(d+1)S_{n,d+1} + (\text{degree} \le d) \mod \bar{x}^2-\bar{x}\,. \end{aligned}$$

However, as $\gamma_d \neq 0$, $d+1 \le n$ (so that $d+1 \neq 0$ in $\mathbb{F}$ and $S_{n,d+1}$ is defined) this shows that 1 (a multilinear degree 0 polynomial) equals $\gamma_d(d+1)S_{n,d+1} + (\text{degree} \le d)$ (a multilinear degree $d+1$ polynomial) modulo $\bar{x}^2-\bar{x}$, which is a contradiction to the uniqueness of representation of multilinear polynomials modulo $\bar{x}^2-\bar{x}$. Thus, we must have $d = n$. □

To paraphrase the above argument, it shows that for multilinear f of $\deg f < n$ with $\mathrm{ml}(f(\overline{x}) \cdot (\sum_i x_i - \beta)) = 1$ it holds that $\deg \mathrm{ml}(f(\overline{x}) \cdot (\sum_i x_i - \beta)) = \deg f + 1$. This contradicts the fact that $\deg 1 = 0$, so that $\deg f = n$. It is tempting to attempt to argue this claim without using that $\mathrm{ml}(f(\overline{x}) \cdot (\sum_i x_i - \beta)) = 1$ in some way. That is, one could hope to argue that $\deg(\mathrm{ml}(f(\overline{x}) \cdot (\sum_i x_i - \beta))) = \deg f + 1$ directly. Unfortunately this is false, as seen by the example $\mathrm{ml}((x+y)(x-y)) = \mathrm{ml}(x^2 - y^2) = x - y$. However, one can make this approach work to obtain a degree lower bound of $\lceil n/2 \rceil + 1$, as shown by Impagliazzo, Pudlák, and Sgall [41].

Putting the above together with the fact that multilinearization is degree non-increasing we obtain that any polynomial agreeing with $\frac{1}{\sum_i x_i - \beta}$ on the Boolean cube must be of degree $\geq n$.

Corollary 5.4. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial such that*

$$f(\overline{x}) \left(\sum_i x_i - \beta \right) = 1 \mod \overline{x}^2 - \overline{x} \,.$$

Then $\deg f \geq n$.

Proof: By simple properties of multilinearization (Theorem 3.12) we see that $1 = \mathrm{ml}\big(f(\overline{x}) \cdot (\sum_i x_i - \beta)\big) = \mathrm{ml}\big(\mathrm{ml}(f) \cdot (\sum_i x_i - \beta)\big)$, so that $\mathrm{ml}(f) \cdot (\sum_i x_i - \beta) = 1 \mod \overline{x}^2 - \overline{x}$. Thus $\deg f \geq \deg \mathrm{ml}(f)$ and $\deg \mathrm{ml}(f) = n$ by the above Theorem 5.3, yielding the claim. □

The above proof shows that the unique multilinear polynomial f agreeing with $1/(\sum_i x_i - \beta)$ on the hypercube has degree n, but does so without actually specifying the coefficients of f. In Theorem B.1 we compute the coefficients of this polynomial, giving an alternate proof that it has degree n (Corollary B.3). In particular, this computation yields a small algebraic circuit for f, expressing it as an explicit linear combination of elementary symmetric polynomials (which have small algebraic circuits).

5.3 Sparse polynomials

We now use the above functional lower bounds for degree, along with random restrictions, to obtain functional lower bounds for sparsity. We then apply this to obtain exponential lower bounds for sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ refutations of the subset-sum axiom. Recall that sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ is equivalent to the Nullstellensatz proof system when we measure the size of the proof in terms of the number of monomials. While we provide the proof here for completeness, we note that this result has already been obtained by Impagliazzo-Pudlák-Sgall [41], who also gave such a lower bound for the stronger polynomial calculus proof system.

We first recall the random restrictions lemma. This lemma shows that by randomly setting half of the variables to zero, sparse polynomials become sums of monomials involving few variables, which after multilinearization is a low-degree polynomial.

Lemma 5.5. *Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be an s-sparse polynomial. Let $\rho : \mathbb{F}[\overline{x}] \to \mathbb{F}[\overline{x}]$ be the homomorphism induced by randomly and independently setting each variable x_i to 0 with probability $1/2$ and leaving x_i intact with probability $1/2$. Then with probability $\geq 1/2$, each monomial in $\rho(f(\overline{x}))$ involves $\leq \lg s + 1$ variables. Thus, with probability $\geq 1/2$, $\deg \mathrm{ml}(\rho(f)) \leq \lg s + 1$.*

Proof: Consider a monomial $\overline{x}^{\overline{a}}$ involving $\geq t$ variables, $t \in \mathbb{R}$. Then the probability that $\rho(\overline{x}^{\overline{a}})$ is nonzero is at most 2^{-t}. Now consider $f(\overline{x}) = \sum_{j=1}^{s} \alpha_j \overline{x}^{\overline{a}_j}$. By a union bound, the probability that any monomial $\overline{x}^{\overline{a}_j}$ involving at least t variables survives the random restriction is at most $s2^{-t}$. For $t = \lg s + 1$ this is at most $1/2$. The claim about the multilinearization of $\rho(f(\overline{x}))$ follows by observing that for a monomial $\overline{x}^{\overline{a}}$ involving $\leq \lg s + 1$ variables it must be that $\deg \mathrm{ml}(\rho(\overline{x}^{\overline{a}})) \leq \lg s + 1$ (Theorem 3.12). $\square$

We now give our functional lower bound for sparsity. This follows from taking any refutation of the subset-sum axiom and applying a random restriction. The subset-sum axiom will be relatively unchanged, but any sparse polynomial will become (after multilinearization) low-degree, to which our degree lower bounds (Section 5.2) can then be applied.

Proposition 5.6. *Let $n \geq 8$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial such that*

$$f(\overline{x}) = \frac{1}{\sum_i x_i - \beta} \ ,$$

for $\overline{x} \in \{0,1\}^n$. Then f requires $\geq 2^{n/4-1}$ monomials.

Proof: Suppose that f is s-sparse so that $f(\overline{x}) = \sum_{j=1}^{s} \alpha_j \overline{x}^{\overline{a}_j}$. Take a random restriction ρ as in Theorem 5.5, so that with probability at least $1/2$ we have that $\deg \mathrm{ml}(\rho(f)) \leq \lg s + 1$. By the Chernoff bound,[11] we see that ρ keeps alive at least $n/4$ variables with probability at least $1 - \mathrm{e}^{-2\cdot(1/4)^2 \cdot n}$, which is $\geq 1 - \mathrm{e}^{-1}$ for $n \geq 8$. Thus, by a union bound the probability that ρ fails to have either that $\deg \mathrm{ml}(\rho(f)) \leq \lg s + 1$ or that it keeps at least $n/4$ variables alive is at most $1/2 + \mathrm{e}^{-1} < 1$. Thus a ρ exists obeying both properties.

Thus, the functional equation for f implies that

$$f(\overline{x}) \left(\sum_i x_i - \beta \right) = 1 + \sum_i h_i(\overline{x})(x_i^2 - x_i) \ ,$$

for some $h_i \in \mathbb{F}[\overline{x}]$. Applying the random restriction and multilinearization to both sizes of this equation, we obtain that

$$\mathrm{ml}(\rho(f)) \cdot \left(\sum_{\rho(x_i) \neq 0} x_i - \beta \right) \equiv 1 \mod \{x_i^2 - x_i\}_{\rho(x_i) \neq 0} \ .$$

Thus, by appealing to the degree lower bound for this functional equation (Theorem 5.3) we obtain that $\lg s + 1 \geq \deg \mathrm{ml}(\rho(f))$ is at least the number of variables which is $\geq n/4$, so that $s \geq 2^{n/4-1}$ as desired. $\square$

We remark that one can actually improve the sparsity lower bound to the optimal "$\geq 2^n$" by computing the sparsity of the unique multilinear polynomial satisfying the above functional equation (Corollary B.3). We now apply these functional lower bounds to obtain lower bounds for sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ refutations of $\sum_i x_i - \beta, \overline{x}^2 - \overline{x}$ via our reduction (Theorem 5.1).

Corollary 5.7. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Then $\sum_{i=1}^{n} x_i - \beta, \overline{x}^2 - \overline{x}$ is unsatisfiable and any sparse-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation requires size $\exp(\Omega(n))$.* $\square$

[11] For independent $[0,1]$-valued random variables $\mathsf{X}_1, \ldots, \mathsf{X}_n$, $\Pr\left[\sum_i \mathsf{X}_i - \sum_i \mathbb{E}[\mathsf{X}_i] \leq -\varepsilon n\right] \leq \mathrm{e}^{-2\varepsilon^2 n}$.

5.4 Coefficient dimension in a fixed partition

We now seek to prove functional circuit lower bounds for more powerful models of computation such as roABPs and multilinear formulas. As recalled in Section 3, the coefficient dimension complexity measure can give lower bounds for such models. However, by definition it is a *syntactic* measure as it speaks about the coefficients of a polynomial. Unfortunately, knowing that a polynomial $f \in \mathbb{F}[\overline{x}]$ agrees with a function $\hat{f} : \{0,1\}^n \to \mathbb{F}$ on the Boolean cube $\{0,1\}^n$ does not in general give enough information to determine its coefficients. In contrast, the *evaluation* dimension measure is concerned with evaluations of a polynomial (which is functional). Obtaining lower bounds for evaluation dimension, and leveraging the fact that the evaluation dimension lower bounds coefficient dimension (Lemma 3.11) we can obtain the desired lower bounds for this complexity measure.

We now proceed to the lower bound. It will follow from the degree lower bound for the subset-sum axiom (Theorem 5.4). That is, this degree bound shows that if $f(\overline{z}) \cdot (\sum_i z_i - \beta) \equiv 1 \mod \overline{z}^2 - \overline{z}$ then f must have degree $\geq n$. We can then "lift" this lower bound by the use of a gadget, in particular by replacing $\overline{z} \leftarrow \overline{x} \circ \overline{y}$, where '$\circ$' is the Hadamard (entrywise) product. Because the degree of f is maximal, this gadget forces $\overline{x}$ and $\overline{y}$ to maximally "interact", and hence the evaluation dimension is large in the $\overline{x}$ versus $\overline{y}$ partition.

Proposition 5.8. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0,\ldots,n\}$. Let $f \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$ be a polynomial such that*

$$f(\overline{x},\overline{y}) = \frac{1}{\sum_i x_i y_i - \beta} ,$$

for $\overline{x},\overline{y} \in \{0,1\}^n$. Then $\dim \mathbf{Coeff}_{\overline{x}|\overline{y}} f \geq 2^n$.

Proof: By showing that the coefficient dimension is not less than the evaluation dimension over the Boolean cube (Lemma 3.11),

$$\begin{aligned}\dim \mathbf{Coeff}_{\overline{x}|\overline{y}} f &\geq \dim \mathbf{Eval}_{\overline{x}|\overline{y},\{0,1\}} f \\ &= \dim\{f(\overline{x}, \mathbb{1}_S) : S \subseteq [n]\} \\ &\geq \dim\{\mathrm{ml}(f(\overline{x}, \mathbb{1}_S)) : S \subseteq [n]\} ,\end{aligned}$$

where $\mathbb{1}_S \in \{0,1\}^n$ is the indicator vector for a set S. Note that we used that dimension is non-increasing under linear maps. Now note that for $\overline{x} \in \{0,1\}^n$,

$$f(\overline{x}, \mathbb{1}_S) = \frac{1}{\sum_{i \in S} x_i - \beta} .$$

It follows that $\mathrm{ml}(f(\overline{x}, \mathbb{1}_S))$ is a multilinear polynomial only depending on $\overline{x}|_S$ (Theorem 3.12), and by its functional behavior it follows from Theorem 5.3 that $\deg \mathrm{ml}(f(\overline{x}, \mathbb{1}_S)) = |S|$. As $\mathrm{ml}(f(\overline{x}, \mathbb{1}_S))$ is multilinear it thus follows that the leading monomial of $\mathrm{ml}(f(\overline{x}, \mathbb{1}_S))$ is $\prod_{i \in S} x_i$, which is distinct for each distinct S. This is also readily seen from the explicit description of $\mathrm{ml}(f(\overline{x}, \mathbb{1}_S))$ given by Theorem B.1.

Thus, we can lower bound the dimension of this space by the number of leading monomials (Lemma 3.19),

$$\begin{aligned}
\dim \mathbf{Coeff}_{\bar{x}|\bar{y}} f &\geq \dim\{\mathrm{ml}(f(\bar{x}, \mathbb{1}_S)) : S \subseteq [n]\} \\
&\geq \left|\mathrm{LM}\left(\{\mathrm{ml}(f(\bar{x}, \mathbb{1}_S)) : S \subseteq [n]\}\right)\right| \\
&= \left|\left\{\prod_{i \in S} x_i : S \subseteq [n]\right\}\right| \\
&= 2^n \,. \qquad \square
\end{aligned}$$

Note that in the above proof we crucially leveraged that the degree bound of Theorem 5.3 is *exactly* n, not just $\Omega(n)$. This exact bound allows us to determine the leading monomials of these polynomials, which seems not to follow from degree lower bounds of $\Omega(n)$.

As coefficient dimension lower bounds roABP-width (Lemma 3.7) and depth-3 powering formulas can be computed by roABPs in any order of the variables (Theorem 3.15), we obtain as a corollary our functional lower bound for these models.

Corollary 5.9. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_n, y_1, \ldots, y_n]$ be a polynomial such that*

$$f(\bar{x}, \bar{y}) = \frac{1}{\sum_i x_i y_i - \beta} \,,$$

for $\bar{x}, \bar{y} \in \{0,1\}^n$. Then f requires width $\geq 2^n$ to be computed as an roABP in any order of the variables where $\bar{x}$ precedes $\bar{y}$. In particular, f requires $\exp(\Omega(n))$ size as a depth-3 powering formula. $\square$

We now conclude with a lower bound for linear-IPS over roABPs in certain orders of the variables, and thus also for depth-3 powering formulas, by appealing to our reduction to functional lower bounds (Theorem 5.1).

Corollary 5.10. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Then $\sum_{i=1}^n x_i y_i - \beta, \bar{x}^2 - \bar{x}, \bar{y}^2 - \bar{y}$ is unsatisfiable and any roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation, where the roABP reads $\bar{x}$ before $\bar{y}$, requires width $\geq \exp(\Omega(n))$. In particular, any $\sum\bigwedge\sum$-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation requires size $\geq \exp(\Omega(n))$.*

Proof: That this system is unsatisfiable is clear from construction. The proof then follows from applying our functional lower bound (Corollary 5.9) to our reduction strategy (Theorem 5.1), where we use that partial evaluations of small roABPs yield small roABPs in the induced order of the variables (Theorem 3.8), and that depth-3 powering formulas are a subclass of roABPs (in any order) (Theorem 3.15). $\square$

The above result shows an roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ lower bound for orders of the variables where $\bar{x}$ precedes $\bar{y}$, and we complement this by giving an upper bound showing there *are* small roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ upper bounds for orders of the variables where $\bar{x}$ and $\bar{y}$ are tightly interleaved. This is achieved by taking the roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ upper bound of Corollary 4.14 for $\sum_i z_i - \beta, \bar{z}^2 - \bar{z}$ under the substitution $z_i \leftarrow x_i y_i$, and observing that such substitutions preserve roABP width in the $x_1 < y_1 < \cdots < x_n < y_n$ order (Theorem 3.8). In particular, as $\sum\bigwedge\sum$ formulas are small roABPs in *every* order of the variables, this allows us to achieve an exponential separation between $\sum\bigwedge\sum$-$\mathrm{IPS}_{\mathrm{LIN}}$ and roABP-$\mathrm{IPS}_{\mathrm{LIN}}$.

Corollary 5.11. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\mathrm{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Then $\sum_{i=1}^n x_i y_i - \beta, \overline{x}^2 - \overline{x}, \overline{y}^2 - \overline{y}$ is unsatisfiable, has a $\mathsf{poly}(n)$-explicit $\mathsf{poly}(n)$-size roABP-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation in the order of the variables $x_1 < y_1 < \cdots < x_n < y_n$, and every $\sum\bigwedge\sum$-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation requires size $\geq \exp(\Omega(n))$.*

Proof: Theorem 5.10 showed that this system is unsatisfiable and has the desired $\sum\bigwedge\sum$-$\mathrm{IPS}_{\mathrm{LIN}}$ lower bound, so that it remains to prove the roABP upper bound.

By Theorem 4.13 the unique multilinear polynomial $f \in \mathbb{F}[\overline{z}]$ such that $f(\overline{z}) \cdot (\sum_{i=1}^n z_i - \beta) \equiv 1 \mod \overline{z}^2 - \overline{z}$ has a multilinear $\mathsf{poly}(n)$-size roABP in the order of the variables $z_1 < \cdots < z_n$. Applying the variable substitution $z_i \leftarrow x_i y_i$, it follows that $f'(\overline{x}, \overline{y}) := f(x_1 y_1, \ldots, x_n y_n)$ obeys $f' \cdot (\sum_{i=1}^n x_i y_i - \beta) \equiv 1 \mod \overline{x}^2 - \overline{x}, \overline{y}^2 - \overline{y}$ (as $z_i^2 - z_i \equiv 0 \mod \overline{x}^2 - \overline{x}, \overline{y}^2 - \overline{y}$ under the substitution $z_i \leftarrow x_i y_i$) and that f' is computable by a $\mathsf{poly}(n)$-size roABP in the order of the variables $x_1 < y_1 < \cdots < x_n < y_n$ (Theorem 3.8, using that f has individual degree 1). Appealing to the efficient multilinearization of roABPs (Theorem 4.6) completes the claim as $\sum_i x_i y_i - \beta$ is computable by a $\mathsf{poly}(n)$-size roABP (in any order). ☐

5.5 Coefficient dimension in any variable partition

The previous section gave functional lower bounds for coefficient dimension, and thus roABP width, in the $\overline{x}|\overline{y}$ variable partition. However, this lower bound fails for other order of the variablesings where $\overline{x}$ and $\overline{y}$ are interleaved because of corresponding upper bounds (Theorem 5.11). In this section we extend the lower bound to *any* order of the variablesing by using suitable auxiliary variables to plant the previous lower bound into any partition we desire by suitably evaluating the auxiliary variables.

We begin by developing some preliminaries for how coefficient dimension works in the presence of auxiliary indicator variables. That is, consider a polynomial $f(\overline{x}, \overline{y}, \overline{z})$ where we wish to study the coefficient dimension of f in the $\overline{x}|\overline{y}$ partition. We can view this polynomial as lying in $\mathbb{F}[\overline{z}][\overline{x}, \overline{y}]$ so that its coefficients are polynomials in $\overline{z}$ and one studies the dimension of the coefficient space in the field of rational functions $\mathbb{F}(\overline{z})$. Alternatively one can evaluate $\overline{z}$ at some point $\overline{z} \leftarrow \overline{\alpha}$ so that $f(\overline{x}, \overline{y}, \overline{\alpha}) \in \mathbb{F}[\overline{x}, \overline{y}]$ and study its coefficient dimension over $\mathbb{F}$. The following straightforward lemma shows the first dimension over $\mathbb{F}(\overline{z})$ is lower-bounded by the second dimension over $\mathbb{F}$.

Lemma 5.12. *Let $f \in \mathbb{F}[\overline{x}, \overline{y}, \overline{z}]$. Let $f_{\overline{z}}$ denote f as a polynomial in $\mathbb{F}[\overline{z}][\overline{x}, \overline{y}]$, so that for any $\overline{\alpha} \in \mathbb{F}^{|\overline{z}|}$ we have that $f_{\overline{\alpha}}(\overline{x}, \overline{y}) = f(\overline{x}, \overline{y}, \overline{\alpha}) \in \mathbb{F}[\overline{x}, \overline{y}]$. Then for any such $\overline{\alpha}$,*

$$\dim_{\mathbb{F}(\overline{z})} \mathbf{Coeff}_{\overline{x}|\overline{y}} f_{\overline{z}}(\overline{x}, \overline{y}) \geq \dim_{\mathbb{F}} \mathbf{Coeff}_{\overline{x}|\overline{y}} f_{\overline{\alpha}}(\overline{x}, \overline{y}) .$$

Proof: Let $f(\overline{x}, \overline{y}, \overline{z})$ be written in $\mathbb{F}[\overline{x}, \overline{y}, \overline{z}]$ as $f = \sum_{\overline{a}, \overline{b}} f_{\overline{a}, \overline{b}}(\overline{z}) \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}$. By Lemma 3.4 we see that $\dim_{\mathbb{F}(\overline{z})} \mathbf{Coeff}_{\overline{x}|\overline{y}} f_{\overline{z}}(\overline{x}, \overline{y})$ is equal to the rank (over $\mathbb{F}(\overline{z})$) of the coefficient matrix $C_{f_{\overline{z}}}$, so that its entries $(C_{f_{\overline{z}}})_{\overline{a}, \overline{b}} = f_{\overline{a}, \overline{b}}(\overline{z})$ are in $\mathbb{F}[\overline{z}]$. Similarly, $\dim_{\mathbb{F}} \mathbf{Coeff}_{\overline{x}|\overline{y}} f_{\overline{\alpha}}(\overline{x}, \overline{y})$ is equal to the rank (over $\mathbb{F}$) of the coefficient matrix $C_{f_{\overline{\alpha}}}$, so that as $f(\overline{x}, \overline{y}, \overline{\alpha}) = \sum_{\overline{a}, \overline{b}} f_{\overline{a}, \overline{b}}(\overline{\alpha}) \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}$ we have that $(C_{f_{\overline{\alpha}}})_{\overline{a}, \overline{b}} = f_{\overline{a}, \overline{b}}(\overline{\alpha})$, which is in $\mathbb{F}$. Thus, it follows that $C_{f_{\overline{z}}}|_{\overline{z} \leftarrow \overline{\alpha}} = C_{f_{\overline{\alpha}}}$.

The claim then follows by noting that for a matrix $M(\overline{w}) \in \mathbb{F}[\overline{w}]^{r \times r}$ it holds that $\mathrm{rank}_{\mathbb{F}(\overline{w})} M(\overline{w}) \geq \mathrm{rank}_{\mathbb{F}} M(\overline{\beta})$ for any $\overline{\beta} \in \mathbb{F}^{|\overline{w}|}$. This follows as the rank of $M(\overline{w})$ is equal to the maximum size of a minor with a non-vanishing determinant. Thus, determinants are polynomials in $\overline{w}$, and they can only further vanish when $\overline{w} \leftarrow \overline{\beta}$. ☐

We now use auxiliary variables to embed the coefficient dimension lower bound from Theorem 5.8 into any order of the variables. We do this by viewing the polynomial $\sum_i u_i v_i - \beta$ as using a matching between variables in $\overline{u}$ and $\overline{v}$. We then wish to embed this matching graph-theoretically into a complete graph, where nodes are labelled with the variables $\overline{x}$. Any equipartition of this graph will induce many edges across this cut, and we can drop edges to find a large matching between the $\overline{x}$ variables which we then identify as instance of $\sum_i u_i v_i - \beta$. We introduce one new auxiliary variable $z_{i,j}$ per edge which, upon setting it to 0 or 1, allows us to have this edge (respectively) dropped from or kept in the desired matching. This leads to the new (symmetrized) equation $\sum_{i<j} z_{i,j} x_i x_j - \beta$, for which we now give the desired lower bound.

Proposition 5.13. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\operatorname{char}(\mathbb{F}) > \binom{2n}{2}$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \dots, \binom{2n}{2}\}$. Let $f \in \mathbb{F}[x_1, \dots, x_{2n}, z_1, \dots, z_{\binom{2n}{2}}]$ be a polynomial such that*

$$f(\overline{x}, \overline{z}) = \frac{1}{\sum_{i<j} z_{i,j} x_i x_j - \beta} ,$$

for $\overline{x} \in \{0,1\}^{2n}$, $\overline{z} \in \{0,1\}^{\binom{2n}{2}}$. Let $f_{\overline{z}}$ denote f as a polynomial in $\mathbb{F}[\overline{z}][\overline{x}]$. Then for any partition $\overline{x} = (\overline{u}, \overline{v})$ with $|\overline{u}| = |\overline{v}| = n$,

$$\dim_{\mathbb{F}(\overline{z})} \mathbf{Coeff}_{\overline{u}|\overline{v}} f_{\overline{z}} \geq 2^n .$$

Proof: We wish to embed $\sum_i u_i v_i - \beta$ in this instance via a restriction of $\overline{z}$. Define the $\overline{z}$-evaluation $\overline{\alpha} \in \{0,1\}^{\binom{2n}{2}}$ to restrict f to sum over those $x_i x_j$ in the natural matching between $\overline{u}$ an $\overline{v}$, so that

$$\alpha_{i.j} = \begin{cases} 1 & x_i = u_k, x_j = v_k \\ 0 & \text{else} \end{cases} .$$

It follows then that $f(\overline{u}, \overline{v}, \overline{\alpha}) = \frac{1}{\sum_{k=1}^{n} u_k v_k - \beta}$ for $\overline{u}, \overline{v} \in \{0,1\}^n$. Thus, by appealing to our lower bound for a fixed partition (Theorem 5.8) and the relation between the coefficient dimension in $f_{\overline{z}}$ versus $f_{\overline{\alpha}}$ (Theorem 5.12),

$$\begin{aligned} \dim_{\mathbb{F}(\overline{z})} \mathbf{Coeff}_{\overline{u}|\overline{v}} f_{\overline{z}}(\overline{u}, \overline{v}) &\geq \dim_{\mathbb{F}} \mathbf{Coeff}_{\overline{u}|\overline{v}} f_{\overline{\alpha}}(\overline{u}, \overline{v}) \\ &\geq 2^n . \end{aligned}$$

□

We remark that this lower bound is only $\exp(\Omega(\sqrt{m}))$ where $m = 2n + \binom{2n}{2}$ is the number of total variables, while one could hope for an $\exp(\Omega(m))$ lower bound as 2^m is the trivial upper bound for multilinear polynomials. One can achieve such a lower bound by replacing the above auxiliary variable scheme (which corresponds to a complete graph) with one derived from a constant-degree expander graph. That is because in such graphs any large partition of the vertices induces a large matching across that partition, where one can then embed the fixed-partition lower bounds of the previous section (Section 5.4). However, we omit the details as this would not qualitatively change the results.

We now obtain our desired functional lower bounds for roABPs and multilinear formulas.

Corollary 5.14. *Let $n \geq 1$ and $\mathbb{F}$ be a field with* $\operatorname{char}(\mathbb{F}) > \binom{2n}{2}$. *Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, \binom{2n}{2}\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_{2n}, z_1, \ldots, z_{\binom{2n}{2}}]$ be a polynomial such that*

$$f(\overline{x}, \overline{z}) = \frac{1}{\sum_{i<j} z_{i,j} x_i x_j - \beta} ,$$

for $\overline{x} \in \{0,1\}^{2n}$, $\overline{z} \in \{0,1\}^{\binom{2n}{2}}$. Then f requires width $\geq 2^n$ to be computed by an roABP in any order of the variables. Also, f requires $n^{\Omega(\log n)}$-size to be computed as a multilinear formula. For $d = o(\log n / \log\log n)$, f requires $n^{\Omega((n/\log n)^{1/d}/d^2)}$-size multilinear formulas of product-depth-d.

Proof: roABPs: Suppose that $f(\overline{x}, \overline{z})$ is computable by a width-r roABP in some order of the variables. By pushing the $\overline{z}$ variables into the fraction field, it follows that $f_{\overline{z}}$ (f as a polynomial in $\mathbb{F}[\overline{z}][\overline{x}]$) is also computable by a width-r roABP over $\mathbb{F}(\overline{z})$ in the induced order of the variables on $\overline{x}$ (Theorem 3.8). By splitting $\overline{x}$ in half along its order of the variables one obtains the lower bound by combining the coefficient dimension lower bound of Theorem 5.13 with its relation to roABPs (Lemma 3.7).

multilinear formulas: This follows immediately from our coefficient dimension lower bound (Theorem 5.13) and the Raz [63] and Raz-Yehudayoff [68] results (Theorem 3.13). □

As before, this immediately yields the desired roABP-$\mathrm{IPS_{LIN}}$ and multilinear-formula-IPS lower bounds.

Corollary 5.15. *Let $n \geq 1$ and $\mathbb{F}$ be a field with* $\operatorname{char}(\mathbb{F}) > \binom{2n}{2}$. *Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, \binom{2n}{2}\}$. Then $\sum_{i<j} z_{i,j} x_i x_j - \beta, \overline{x}^2 - \overline{x}, \overline{z}^2 - \overline{z} \in \mathbb{F}[x_1, \ldots, x_{2n}, z_1, \ldots, z_{\binom{2n}{2}}]$ is unsatisfiable, and any roABP-*$\mathrm{IPS_{LIN}}$ *refutation (in any order of the variables) requires* $\exp(\Omega(n))$ *size. Further, any multilinear-formula-*IPS *refutation requires $n^{\Omega(\log n)}$-size, and any product-depth-d multilinear-formula-*IPS *refutation requires $n^{\Omega((n/\log n)^{1/d}/d^2)}$-size.*

Proof: The system is unsatisfiable as any setting of $\overline{x} \in \{0,1\}^n$ yields a sum over at most $\binom{2n}{2}$ z-variables, which must be in $\{0, \ldots, \binom{2n}{2}\}$ which by hypothesis does not contain β.

The roABP-$\mathrm{IPS_{LIN}}$ lower bound follows immediately from the above functional lower bound (Theorem 5.14) along with our reduction (Theorem 5.1), just as in Theorem 5.10.

The multilinear-formula-IPS lower bound also follows immediately from the above functional lower bound (Theorem 5.14) along with our reduction from IPS lower bounds to functional lower bounds for multilinear polynomials (Theorem 5.2). In particular, this application uses that multilinear formulas are closed under partial evaluations, and that taking the difference of two formulas will only double its size and does not change the product depth. □

6 Lower bounds for multiples of polynomials

In this section we consider the problem of finding explicit polynomials whose nonzero multiples are all hard. Such polynomials are natural to search for, as intuitively if f is hard to compute then so should small modifications such as $x_1 f^2 + 4f^3$. This intuition is buttressed by Kaltofen's [44] result that if a polynomial has a small algebraic circuit then so do all of its factors (up to some pathologies in small characteristic). Taken in a contrapositive, this says that if a polynomial f requires superpolynomial size algebraic circuits, then so must all of its nonzero multiples. Thus, for general circuits the question of lower bounds for multiples reduces to the standard lower bounds question.

Unfortunately, for many restricted classes of circuits where lower bounds are known (depth-3 powering formulas, sparse polynomials, roABPs) Kaltofen's [44] result produces circuits for the factors which do not fall into (possibly stronger) restricted classes of circuits where lower bounds are still known.[12] Therefore, developing lower bounds for multiples against these restricted classes seems to require further work beyond the standard lower bound question.

We will begin by discussing the applications of this problem to the hardness versus randomness paradigm in algebraic complexity. We then use existing derandomization results to show that multiples of the determinant are hard for certain restricted classes. However, this method is very rigidly tied to the determinant. Thus, we also directly study existing lower bound techniques for restricted models of computation (depth-3 powering formulas, sparse polynomials, and roABPs) and extend these results to also apply to multiples. We will show the applications of such polynomials to proof complexity in Section 7.

6.1 Connections to hardness versus randomness and factoring circuits

To motivate the problem of finding polynomials with hard multiples, we begin by discussing the hardness versus randomness approach to derandomizing polynomial identity testing. That is, Kabanets and Impagliazzo [43] extended the hardness versus randomness paradigm of Nisan and Wigderson [53] to the algebraic setting, showing that sufficiently good algebraic circuit lower bounds for an explicit polynomial would qualitatively derandomize PIT. While much of the construction is similar (using combinatorial designs, hybrid arguments, etc.) to the approach of Nisan and Wigderson [53] for Boolean derandomization, there is a key difference. In the Boolean setting, obtaining a hardness versus randomness connection requires converting *worst-case* hardness (no small computation can compute the function everywhere) to *average-case* hardness (no small computation can compute the function on most inputs). Such a reduction (obtained by Impagliazzo and Wigderson [42]) can in fact be obtained using certain error-correcting codes based on multivariate polynomials (as shown by Sudan, Trevisan and Vadhan [82]).

Such a worst-case to average-case reduction is also needed in the algebraic setting, but as multivariate polynomials are one source of this reduction in the Boolean regime, it is natural to expect it to be easier in the algebraic setting. Specifically, the notion of average-case hardness for a polynomial $f(\overline{x})$ used in Kabanets–Impagliazzo [43] is that for any $g(\overline{x},y)$ satisfying $g(\overline{x},f(\overline{x}))=0$, it must be that g then requires large algebraic circuits (by taking $g(\overline{x},y):=y-f(\overline{x})$ this implies f itself requires large circuits). This can be interpreted as average-case hardness because if such a g existed with a small circuit, then for any value $\overline{\alpha}$ we have that $g(\overline{\alpha},y)$ is a univariate polynomial that vanishes on $f(\overline{\alpha})$. By factoring this univariate (which can be done efficiently), we see that such g give a small list (of size at most $\deg g$) of possible values for $f(\overline{\alpha})$. By picking a random element from this list, one can correctly compute $f(\overline{x})$ with noticeable probability, which by an averaging argument one can convert to a (non-uniform) deterministic procedure to compute $f(\overline{x})$ on most inputs (over any fixed finite set). While this procedure (involving univariate

[12] While some results ([18, 56]) can bound the depth of the factors in terms of the depth of the input circuit, there are only very weak lower bounds known for constant-depth algebraic circuits.

factorization) is not an algebraic circuit, the above argument shows that the Kabanets–Impagliazzo [43] notion is a natural form of average case hardness.

To obtain this form of average-case hardness from worst-case hardness, Kabanets and Impagliazzo [43] used a result of Kaltofen [44], who showed that (up to pathologies in low-characteristic fields), factors of small (general) circuits have small circuits. As $g(\overline{x}, f(\overline{x})) = 0$ iff $y - f(\overline{x})$ divides $g(\overline{x}, y)$, it follows that if $g(\overline{x}, y)$ has a small circuit then so does $y - f(\overline{x})$, and thus so does $f(\overline{x})$. Taking the contrapositive, if f requires large circuits (worst-case hardness) then any such $g(\overline{x}, y)$ with $g(\overline{x}, f(\overline{x})) = 0$ also requires large circuits (average-case hardness). Note that this says that *any* worst-case hard polynomial is *also* average-case hard. In contrast, this is provably false for Boolean functions, where such worst-case to average-case reductions thus necessarily modify the original function.

As mentioned above, Kaltofen's [44] factoring algorithm does not preserve structural restrictions (such as multilinearity, homogeneousness, small-depth, read-once-ness, etc.) of the original circuit, so that obtaining average-case hardness for restricted classes of circuits requires worst-case hardness for much stronger classes. While follow-up work has reduced the complexity of the circuits resulting from Kaltofen's [44] algorithm (Dvir–Shpilka–Yehudayoff [18] and Oliveira [56] extended Kaltofen's [44] to roughly preserve the depth of the original computation) this work is limited to factoring polynomials of small individual degree and does not seem applicable to other types of computations such as roABPs. Indeed, it even remains an open question to show any non-trivial upper bounds on the complexity of the factors of sparse polynomials. In fact, we actually have non-trivial *lower* bounds. Specifically, von zur Gathen and Kaltofen [30] gave an explicit s-sparse polynomial (over any field) which has a factor with $s^{\Omega(\log s)}$ monomials, and Volkovich [85] gave, for a prime p, an explicit n-variate n-sparse polynomial of degree-p which in characteristic p has a factor with $\binom{n+p-2}{n-1}$ monomials (an exponential separation for $p \geq \mathsf{poly}(n)$). We refer the reader to the survey of Forbes and Shpilka [28] for more on the challenges in factoring small algebraic circuits.

While showing the equivalence of worst-case and average-case hardness for restricted circuit classes seems difficult, to derandomize PIT via the method of Kabanets–Impagliazzo [43] only requires a *single* polynomial that is average case hard. To facilitate obtaining such hard polynomials, we now record an easy lemma showing that polynomials with only hard multiples are average-case hard.

Lemma 6.1. *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ and $g(\overline{x}, y) \in \mathbb{F}[\overline{x}, y]$ both be nonzero, where $g(\overline{x}, 0) \neq 0$ also. If $g(\overline{x}, f(\overline{x})) = 0$ then $g(\overline{x}, 0)$ is a nonzero multiple of $f(\overline{x})$.*

Proof: Let $g(\overline{x}, y) = \sum_i g_i(\overline{x}) y^i$ and $g_0(\overline{x}) := g(\overline{x}, 0)$. That $g(\overline{x}, f(\overline{x})) = 0$ implies that

$$0 = g(\overline{x}, f(\overline{x})) = \sum_i g_i(\overline{x})(f(\overline{x}))^i = g_0(\overline{x}) + \sum_{i \geq 1} g_i(\overline{x})(f(\overline{x}))^i$$

so that $g_0(\overline{x}) = f(\overline{x}) \cdot \left(-\sum_{i \geq 1} g_i(\overline{x})(f(\overline{x}))^{i-1}\right)$ as desired. □

That is, saying that $f(\overline{x})$ is *not* average-case hard means that $g(\overline{x}, f(\overline{x})) = 0$ for a nonzero $g(\overline{x}, y)$. One can assume that $g(\overline{x}, 0) \neq 0$, as otherwise one can replace g by g/y^i for some $i \leq \deg g$, as this only mildly increases the size for most measures of circuit size (see for example Section 4.1). As then the complexity of $g(\overline{x}, 0)$ is bounded by that of $g(\overline{x}, y)$ (for natural measures), the lemma shows then that f has a nonzero multiple of low-complexity. Taken contrapositively, if f only has hard nonzero multiples then it is average-case hard in the sense needed for applying the technique of Kabanets–Impagliazzo [43]. This shows that lower bounds for multiples is essentially the lower bound needed for algebraic hardness versus randomness.[13]

While in the sections below we are able to give explicit polynomials with hard multiples for various restricted classes of algebraic circuits, some of these classes (such as sparse polynomials and roABPs) still do not have the required closure properties to use Kabanets–Impagliazzo [43] to obtain deterministic PIT algorithms. Even for classes with the needed closure properties (such as $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas, where the hard polynomial is the monomial), the resulting PIT algorithms are only worse than existing results (which for $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas is the result of Forbes [21]). However, it seems likely that future results establishing polynomials with hard multiples would imply new PIT algorithms.

6.2 Lower bounds for multiples via PIT

This above discussion shows that obtaining lower bounds for multiples is sufficient for instantiating the hardness versus randomness paradigm. We now observe the converse, showing that one can obtain such polynomials with hard multiples via derandomizing (black-box) PIT, or equivalently, producing generators with small seed-length. That is, Heintz–Schnorr [39] and Agrawal [1] showed that one can use explicit generators for small circuits to obtain hard polynomials, and we observe here that the resulting polynomials also have only hard multiples.

Note that we give the construction based on a non-trivial *generator* for a class of circuits. While one can analogously prove the *hitting-set* version of this claim, it is weaker. That is, it is possible to consider classes $\mathcal{C}$ of unbounded degree and still have generators with small seed-length (see for example Theorem 6.5 below), but for such classes one must have hitting sets with infinite size (as hitting univariate polynomials of unbounded degree requires an infinite number of points).

Lemma 6.2. *Let $\mathcal{C} \subseteq \mathbb{F}[\overline{x}]$ be a class of polynomials and let $\overline{\mathcal{G}} : \mathbb{F}^{\ell} \to \mathbb{F}^{|\overline{x}|}$ be a generator for $\mathcal{C}$. Suppose $0 \neq h \in \mathbb{F}[\overline{x}]$ has $h \circ \overline{\mathcal{G}} = 0$. Then for any nonzero $g \in \mathbb{F}[\overline{x}]$ we have that $g \cdot h \notin \mathcal{C}$.*

Proof: By definition of $\overline{\mathcal{G}}$, for any $f \in \mathcal{C}$, $f = 0$ iff $f \circ \overline{\mathcal{G}} = 0$. Then for any nonzero g, $g \cdot h \neq 0$ and $(g \cdot h) \circ \overline{\mathcal{G}} = (g \circ \overline{\mathcal{G}}) \cdot (h \circ \overline{\mathcal{G}}) = (g \circ \overline{\mathcal{G}}) \cdot 0 = 0$. Thus, we must have that $g \cdot h \notin \mathcal{C}$. □

That is, if $\ell < n$ then such an h exists (as the coordinates of $\overline{\mathcal{G}}$ are algebraically dependent) and such an h can be found in exponential time by solving an exponentially large linear system. Thus, h is a weakly explicit polynomial with only hard multiples, which is sufficient for instantiating hardness versus randomness.

While there are now a variety of restricted circuit classes with non-trivial (black-box) PIT results, it seems challenging to find for any given generator $\overline{\mathcal{G}}$ an *explicit* nonzero polynomial f with $f \circ \overline{\mathcal{G}} = 0$. Indeed, to the best of our knowledge no such examples have ever been furnished for interesting generators. Aside from the quest for polynomials with hard multiples, this question is independently interesting as it demonstrates the limits of the generator in question, especially for generators that are commonly used. There is not even a consensus as to whether the generators currently constructed could suffice

[13] However, it is not an exact equivalence between lower bounds for multiples and average case hardness, as the converse to Theorem 6.1 is false, as seen by considering $g(x,y) := y - x(x+1)$, so that $x | g(x,0)$ but $g(x,x) \neq 0$.

to derandomize PIT for general circuits. Agrawal [1] has even conjectured that a certain generator for depth-2 circuits (sparse polynomials) would actually suffice for PIT of constant-depth circuits.

We consider here the generator of Shpilka–Volkovich [76]. This generator has a parameter ℓ, and intuitively can be seen as an algebraic analogue of the Boolean pseudorandomness notion of a (randomness efficient) ℓ-wise independent hash function. Just as ℓ-wise independent hash functions are ubiquitous in Boolean pseudorandomness, the Shpilka–Volkovich [76] generator has likewise been used in a number of papers on black-box PIT (for example [76, 7, 26, 24, 85, 21] is a partial list). Therefore, we believe it is important to understand the limits of this generator.

The Shpilka–Volkovich [76] generator is really a family of generators that all share a certain property. Specifically, the map $\overline{\mathcal{G}}^{\mathrm{SV}}_{\ell,n} : \mathbb{F}^r \to \mathbb{F}^n$ has the property [14] that the image $\overline{\mathcal{G}}^{\mathrm{SV}}_{\ell,n}(\mathbb{F}^r)$ contains all ℓ-sparse vectors in $\mathbb{F}^n$. The most straightforward construction of a randomness efficient generator with this property (via Lagrange interpolation, given by Shpilka–Volkovich [76]) has $r = 2\ell$. Even this construction is actually a family of possible constructions, as there is significant freedom to choose the finite set of points where Lagrange interpolation will be performed. Therefore, instead of studying a specific generator we seek to understand the power of the above *property*, and thus we are free to construct another generator $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n}$ with this property for which we can find an explicit nonzero f where $f \circ \overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n} = 0$ for small ℓ. We choose a variant of the original construction so that we can take f as the determinant.

In the original Shpilka–Volkovich [76] generator, one first obtains the $\ell = 1$ construction by using two variables, the control variable y and another variable z. By using Lagrange polynomials to simulate indicator functions, the value of y can be set to choose between the outputs $z\overline{e}_1, \ldots, z\overline{e}_n \in \mathbb{F}[z]^n$, where $\overline{e}_i \in \mathbb{F}^n$ is the i-th standard basis vector. By varying z one obtains all 1-sparse vectors in $\mathbb{F}^n$. To obtain $\overline{\mathcal{G}}^{\mathrm{SV}}_{\ell,n}$ one can sum ℓ independent copies of $\overline{\mathcal{G}}^{\mathrm{SV}}_{1,n}$. In contrast, our construction will simply use a different control mechanism, where instead of using univariate polynomials we use bivariates.

Construction 6.3. *Let $n, \ell \geq 1$. Let $\mathbb{F}$ be a field of size $\geq n$. Let $\Omega := \{\omega_1, \ldots, \omega_n\}$ be distinct elements in $\mathbb{F}$. Define $\overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2} : \mathbb{F}^3 \to \mathbb{F}^{n \times n}$ by*

$$\left(\overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2}(x,y,z)\right)_{i,j} = z \cdot \mathbb{1}_{\omega_i,\Omega}(x) \cdot \mathbb{1}_{\omega_j,\Omega}(y) \; .$$

where $\mathbb{1}_{\omega_i,\Omega}(x) \in \mathbb{F}[x]$ is the unique univariate polynomial of degree $< n$ such that

$$\mathbb{1}_{\omega_i,\Omega}(\omega_j) = \begin{cases} 1 & j = i \\ 0 & else \end{cases} \; .$$

Define $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2} : \mathbb{F}^{3\ell} \to \mathbb{F}^{n \times n}$ by the polynomial map

$$\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}(x_1,y_1,z_1,\ldots,x_\ell,y_\ell,z_\ell) := \overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2}(x_1,y_1,z_1) + \cdots + \overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2}(x_\ell,y_\ell,z_\ell) \; ,$$

working in the ring $\mathbb{F}[\overline{x},\overline{y},\overline{z}]$. ◊

[14] The most obvious algebraic analogue of an ℓ-wise independent hash function would require that for a generator $\overline{\mathcal{G}} : \mathbb{F}^r \to \mathbb{F}^n$ that any subset $S \subseteq [n]$ with $|S| \leq \ell$ the output of $\overline{\mathcal{G}}$ restricted to S is all of $\mathbb{F}^S$. This property is implied by the Shpilka–Volkovich [76] property, but is strictly weaker, and is in fact too weak to be useful for PIT. That is, consider the map $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, x_1 + \cdots + x_n)$. This map has this naive "algebraic n-wise independence" property, but does not even fool linear polynomials (which the Shpilka–Volkovich [76] generator does).

Note that this map has n^2 outputs. Now observe that it is straightforward to see that this map has the desired property.

Lemma 6.4. *Assume the setup of Construction 6.3. Then the image of the generator, $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}(\mathbb{F}^{3\ell})$, contains all ℓ-sparse vectors in $\mathbb{F}^{n\times n}$.* $\square$

To the best of the authors knowledge, existing work using the Shpilka–Volkovich [76] generator [15] only use the above property (and occasionally also the fact that a coordinatewise sum of constantly many such generators is a generator of the original form with similar parameters ([3, 24, 38, 21]), which our alternate construction also satisfies). Thus, we can replace the standard construction with our variant in known black-box PIT results (such as [76, 3, 26, 24, 38, 21]), some of which we now state.

Corollary 6.5. *Assume the setup of Construction 6.3. Then $\overline{\mathcal{G}}^{\mathrm{SV}'}_{O(\log s),n^2}$ is a generator for the following classes of n-variate polynomials, of arbitrary degree.*

- *Polynomials of sparsity s ([76, 38, 21]).*
- *Polynomials computable as a depth-3 powering formula of top-fan-in s ([3, 26]).*
- *Polynomials computable as a $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formula of top-fan-in s ([21]), in characteristic zero.*
- *Polynomials computable by width-s roABPs in every order of the variables, also known as commutative roABPs ([3, 24]).*

The above result shows the power of the $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}$ generator to hit restricted classes of circuits. We now observe that it is also limited by its inability to hit the determinant.

Proposition 6.6. *Assume the setup of Construction 6.3. The output of $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}$ is an $n\times n$ matrix of rank $\le \ell$, when viewed as a matrix over the ring $\mathbb{F}(\overline{x},\overline{y},\overline{z})$. Thus, taking $\det_n \in \mathbb{F}[X]$ to be the $n\times n$ determinant, we have that $\det_n \circ \overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2} = 0$ for $\ell < n$.*

Proof: $\underline{\ell = 1:}$ For a field $\mathbb{K}$, a (nonzero) matrix $M \in \mathbb{K}^{n\times n}$ is rank-1 if it can be expressed as an outer-product, so that $(M)_{i,j} = \alpha_i\beta_j$ for $\overline{\alpha},\overline{\beta} \in \mathbb{K}^n$. Taking $\overline{\alpha},\overline{\beta} \in \mathbb{F}(\overline{x},\overline{y},\overline{z})^n$ defined by $\alpha_i := z\mathbb{1}_{\omega_i,\Omega}(x)$ and $\beta_j := \mathbb{1}_{\omega_j,\Omega}(y)$ we immediately see that $\overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2}(x,y,z)$ is rank-1.

$\underline{\ell > 1:}$ This follows as rank is subadditive, and $\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}$ is the sum of ℓ copies of $\overline{\mathcal{G}}^{\mathrm{SV}'}_{1,n^2}$.

$\underline{\det_n \text{ vanishes:}}$ This follows as the $n\times n$ determinant vanishes on matrices of rank $< n$. $\square$

Note that in the above we could hope to find an f such that $f \circ \overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2} = 0$ for all $\ell < n^2$, but we are only able to handle $\ell < n$. Given the above result, along with Theorem 6.2, we obtain that the multiples of the determinant are hard.

Corollary 6.7. *Let $\det_n \in \mathbb{F}[X]$ denote the $n\times n$ determinant. Then any nonzero multiple $f\cdot\det_n$ of $\det_n$, for $0 \ne f \in \mathbb{F}[X]$, has the following lower bounds.*

[15] Note that for black-box PIT it is important that we use a *generator* that contains all sparse vectors, instead of just the *set* of sparse vectors. As an example, the monomial $x_1\cdots x_n$ is zero on all k-sparse vectors for $k < n$, but is nonzero when evaluated on the Shpilka–Volkovich [76] generator for any $\ell \ge 1$.

- $f \cdot \det_n$ *involves* $\exp(\Omega(n))$ *monomials.*
- $f \cdot \det_n$ *requires size* $\exp(\Omega(n))$ *to be expressed as a depth-*3 *powering formula.*
- $f \cdot \det_n$ *requires size* $\exp(\Omega(n))$ *to be expressed as a* $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ *formula, in characteristic zero.*
- $f \cdot \det_n$ *requires width-*$\exp(\Omega(n))$ *to be computed as an roABP in some order of the variables.*

Proof: By Theorem 6.5, $\overline{\mathcal{G}}^{\mathrm{SV}'}_{O(\log s),n^2}$ is a generator for the above size-s computations in the above classes. However, following Theorem 6.2, $(f \cdot \det_n) \circ \left(\overline{\mathcal{G}}^{\mathrm{SV}'}_{\ell,n^2}\right) = 0$ for $\ell < n$. Thus, if $f \cdot \det_n$ (which is nonzero) is computable in size-s it must be that $O(\log s) \geq n$, so that $s \geq \exp(\Omega(n))$. □

Note that we crucially leveraged that the determinant vanishes on low-rank matrices. Therefore, the above results do not seem to imply similar results for the permanent, despite the fact that the permanent is a harder polynomial. That is, recall that because of VNP-completeness of the permanent the determinant $\det_n(X)$ can be written as a projection of the permanent, so that $\det_n(X) = \mathrm{perm}_m(A(X))$ for an affine matrix $A(X) \in \mathbb{F}[X]^{m \times m}$ with $m \leq \mathsf{poly}(n)$. Then, given a multiple $g(Y) \cdot \mathrm{perm}_m(Y)$ one would like to use this projection to obtain $g(A(X))\,\mathrm{perm}_m(A(X)) = g(A(X))\det_n X$, which is a multiple of $\det_n$. Unfortunately this multiple may not be a *nonzero multiple*: it could be that $g(A(X)) = 0$, from which no lower bounds for $g(A(X))\det_n(X)$ (and hence $g(Y)\,\mathrm{perm}_m(Y)$) can be derived.

6.3 Lower bounds for multiples via leading/trailing monomials

We now use the theory of leading (and trailing) monomials to obtain explicit polynomials with hard multiples. We aim at finding as simple polynomials as possible so they will give rise to simple "axioms" with no small refutations for our proof complexity applications in Section 7. These results will essentially be immediate corollaries of previous work.

6.3.1 Depth-3 powering formulas

Kayal [45] observed that using the partial derivative method of Nisan and Wigderson [54] one can show that these formulas require $\exp(\Omega(n))$ size to compute the monomial $x_1 \cdots x_n$. Forbes and Shpilka [26] later observed that this result can be made *robust* by modifying the *hardness of representation* technique of Shpilka and Volkovich [76], in that similar lower bounds apply when the leading monomial involves many variables, as we now quote.

Theorem 6.8 (Forbes–Shpilka [26])**.** *Let* $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ *be computed a* $\sum\bigwedge\sum$ *formula of size* $\leq s$*. Then the leading monomial* $\overline{x}^{\overline{a}} = \mathrm{LM}(f)$ *involves* $|\overline{a}|_0 \leq \lg s$ *variables.* □

We now note that as the leading monomial is multiplicative (Theorem 3.18) this lower bound automatically extends to multiples of the monomial.

Corollary 6.9. *Any nonzero multiple of* $x_1 \cdots x_n$ *requires size* $\geq 2^n$ *to be computed as a* $\sum\bigwedge\sum$ *formula.*

Proof: Consider any $0 \neq g(\overline{x}) \in \mathbb{F}[x_1, \ldots, x_n]$. Then as the leading monomial is multiplicative (Theorem 3.18) we have that $\mathrm{LM}(g \cdot x_1 \cdots x_n) = \mathrm{LM}(g) \cdot x_1 \cdots x_n$, so that $\mathrm{LM}(g \cdot x_1 \cdots x_n)$ involves n variables. By the robust lower bound (Theorem 6.8) this implies $g(\overline{x}) \cdot x_1 \cdots x_n$ requires size $\geq 2^n$ as a $\sum\bigwedge\sum$ formula. □

6.3.2 $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas

Kayal [46] introduced the method of shifted partial derivatives, and Gupta–Kamath–Kayal–Saptharishi [36] refined it to give exponential lower bounds for various sub-models of depth-4 formulas. In particular, it was shown that the monomial $x_1 \cdots x_n$ requires $\exp(\Omega(n))$ size to be computed as a $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formula. Applying the hardness of representation approach of Shpilka and Volkovich [76], Mahajan-Rao-Sreenivasaiah [51] showed how to develop a deterministic black-box PIT algorithm for *multilinear* polynomials computed by $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas. Independently, Forbes [21] (following Forbes–Shpilka [26]) showed that this lower bound can again be made to apply to leading monomials [16] (which implies a deterministic black-box PIT algorithm for *all* $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas, with the same complexity as Mahajan-Rao-Sreenivasaiah [51]). This leading monomial lower bound, which we now state, is important for its applications to polynomials with hard multiples.

Theorem 6.10 (Forbes [21])**.** *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ be computed as a $\sum\bigwedge\sum\prod^t$ formula of size $\leq s$. If $\mathrm{char}(\mathbb{F}) \geq \mathrm{ideg}(\overline{x}^{\overline{a}})$, then the leading monomial $\overline{x}^{\overline{a}} = \mathrm{LM}(f)$ involves $|\overline{a}|_0 \leq O(t \lg s)$ variables.* □

As for depth-3 powering formulas (Theorem 6.9), this immediately yields that all multiples (of degree below the characteristic) of the monomial are hard.

Corollary 6.11. *All nonzero multiples of $x_1 \cdots x_n$ of degree $< \mathrm{char}(\mathbb{F})$ require size $\geq \exp(\Omega(n/t))$ to be computed as $\sum\bigwedge\sum\prod^t$ formula.* □

6.3.3 Sparse polynomials

While the above approaches for $\sum\bigwedge\sum$ and $\sum\bigwedge\sum\prod^{\mathcal{O}(1)}$ formulas focus on leading monomials, one cannot show that the leading monomials of sparse polynomials involve few variables as sparse polynomials can easily compute the monomial $x_1 \cdots x_n$. However, following the *translation* idea of Agrawal-Saha-Saxena [3], Gurjar-Korwar-Saxena-Thierauf [38] showed that sparse polynomials under full-support translations have *some* monomial involving few variables, and Forbes [21] (using different techniques) showed that in fact the *trailing* monomial involves few variables (translations do not affect the leading monomial, so the switch to trailing monomials is necessary here).

Theorem 6.12 (Forbes [21])**.** *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ be $(\leq s)$-sparse, and let $\overline{\alpha} \in (\mathbb{F} \setminus \{0\})^n$ so that $\overline{\alpha}$ has full-support. Then the trailing monomial $\overline{x}^{\overline{a}} = \mathrm{TM}(f(\overline{x}+\overline{\alpha}))$ involves $|\overline{a}|_0 \leq \lg s$ variables.* □

This result thus allows one to construct polynomials whose multiples are all non-sparse.

Corollary 6.13. *All nonzero multiples of $(x_1+1)\cdots(x_n+1) \in \mathbb{F}[\overline{x}]$ require sparsity $\geq 2^n$. Similarly, all nonzero multiples of $(x_1+y_1)\cdots(x_n+y_n) \in \mathbb{F}[\overline{x},\overline{y}]$ require sparsity $\geq 2^n$.*

Proof: Define $f(\overline{x}) = \prod_{i=1}^n (x_i+1)$. For any $0 \neq g(\overline{x}) \in \mathbb{F}[\overline{x}]$ the multiple $g(\overline{x})f(\overline{x})$ under the translation $\overline{x} \mapsto \overline{x} - \overline{1}$ is equal to $g(\overline{x}-\overline{1})\prod_i x_i$. Then all monomials (in particular the trailing monomial) involve n variables (as $g(\overline{x}) \neq 0$ implies $g(\overline{x}-\overline{1}) \neq 0$). Thus, by Theorem 6.12 it must be that $g(\overline{x})f(\overline{x})$ requires $\geq 2^n$ monomials.

The second part of the claim follows as the first, where we now work over the fraction field $\mathbb{F}(\overline{y})[\overline{x}]$, noting that this can only decrease sparsity. Thus, using the translation $\overline{x} \mapsto \overline{x} - \overline{y}$ the above trailing

[16] The result there is stated for trailing monomials, but the argument equally applies to leading monomials.

monomial argument implies that the sparsity of nonzero multiples $\prod_i(x_i+y_i)$ are $\geq 2^n$ over $\mathbb{F}(\overline{y})[\overline{x}]$ and hence also over $\mathbb{F}[\overline{x},\overline{y}]$. $\square$

Note that it is tempting to derive the second part of the above corollary from the first, using that the substitution $\overline{y} \leftarrow \overline{1}$ does not increase sparsity. However, this substitution can convert nonzero multiples of $\prod_i(x_i+y_i)$ to zero multiples of $\prod_i(x_i+1)$, which ruins such a reduction, as argued in the discussion after Theorem 6.7.

6.4 Lower bounds for multiples of sparse multilinear polynomials

While the previous section established that all multiples of $(x_1+1)\cdots(x_n+1)$ are non-sparse, the argument was somewhat specific to that polynomial and fails to obtain an analogous result for $(x_1+1)\cdots(x_n+1)+1$. Further, while that argument can show for example that all multiples of the $n\times n$ determinant or permanent require sparsity $\geq \exp(\Omega(n))$, this is the best sparsity lower bound obtainable for these polynomials with this method.[17] In particular, one cannot establish a sparsity lower bound of "$n!$" for the determinant or permanent (which would be tight) via this method.

We now give a different argument, due to Oliveira [55] that establishes a much more general result showing that multiples of *any* multilinear polynomial have at least the sparsity of the original polynomial. While Oliveira [55] gave a proof using Newton polytopes, we give a more compact proof here using induction on variables (loosely inspired by a similar result of Volkovich [84] on the sparsity of factors of multi-quadratic polynomials).

Proposition 6.14 (Oliveira [55])**.** *Let $f(\overline{x}) \in \mathbb{F}[x_1,\ldots,x_n]$ be a nonzero multilinear polynomial with sparsity exactly s. Then any nonzero multiple of f has sparsity $\geq s$.*

Proof: By induction on variables.

<u>$n=0$:</u> Then f is a constant, so that $s=1$ as $f\neq 0$. All nonzero multiples are nonzero polynomials so have sparsity ≥ 1.

<u>$n\geq 1$:</u> Partition the variables $\overline{x}=(\overline{y},z)$, so that $f(\overline{y},z)=f_1(\overline{y})z+f_0(\overline{y})$, where $f_i(\overline{y})$ has sparsity s_i and $s=s_1+s_0$. Consider any nonzero $g(\overline{y},z)=\sum_{i=d_0}^{d_1} g_i(\overline{y})z^i$ with $g_{d_0}(\overline{y}), g_{d_1}(\overline{y})\neq 0$ (possibly with $d_0=d_1$). Then

$$\begin{aligned}g(\overline{y},z)f(\overline{y},z) &= \Big(f_1(\overline{y})z+f_0(\overline{y})\Big)\cdot\left(\sum_{i=d_0}^{d_1} g_i(\overline{y})z^i\right)\\ &= f_1(\overline{y})g_{d_1}(\overline{y})z^{d_1+1}+\left[\sum_{d_0<i\leq d_1}\Big(f_1(\overline{y})g_{i-1}(\overline{y})+f_0(\overline{y})g_i(\overline{y})\Big)z^i\right]+f_0(\overline{y})g_{d_0}(\overline{y})z^{d_0}\ .\end{aligned}$$

By partitioning this sum by powers of z so that there is no cancellation, and then discarding the middle terms,

$$\left|\mathrm{Supp}\Big(g(\overline{y},z)f(\overline{y},z)\Big)\right| \geq \left|\mathrm{Supp}\Big(f_1(\overline{y})g_{d_1}(\overline{y})\Big)\right|+\left|\mathrm{Supp}\Big(f_0(\overline{y})g_{d_0}(\overline{y})\Big)\right|$$

[17] Specifically, as the determinant and permanent are degree n multilinear polynomials, and thus so are their translations, their monomials always involve $\leq n$ variables so no sparsity bound better than 2^n can be obtained by using Theorem 6.12.

so that appealing to the induction hypothesis, as f_0 and f_1 are multilinear polynomials of sparsity s_0 and s_1 respectively,

$$\geq s_1 + s_0 = s\,. \qquad \square$$

We note that multilinearity is essential in the above lemma, even for univariates. This is seen by noting that the 2-sparse polynomial $x^n - 1$ is a multiple of $x^{n-1} + \cdots + x + 1$.

Thus, the above not only gives a different proof of the non-sparsity of multiples of $\prod_i(x_i+1)$ (Theorem 6.13), but also establishes that nonzero multiples of $\prod_i(x_i+1)+1$ are $\geq 2^n$ sparse, and nonzero multiples of the determinant or permanent are $n!$ sparse, which is tight. Note further that this lower bound proof is "monotone" in that it applies to any polynomial with the same support, whereas the proof of Theorem 6.13 is seemingly not monotone as seen by contrasting $\prod_i(x_i+1)$ and $\prod_i(x_i+1)+1$.

6.5 Lower bounds for multiples by leading/trailing diagonals

In the previous sections we obtained polynomials with hard multiples for various circuit classes by appealing to the fact that lower bounds for these classes can be reduced to studying the number of variables in leading or trailing monomials. Unfortunately this approach is restricted to circuit classes where monomials (or translations of monomials) are hard to compute, which in particular rules out this approach for roABPs. Thus, to develop polynomials with hard multiples for roABPs we need to develop a different notion of a "leading part" of a polynomial. In this section, we define such a notion called a *leading diagonal*, establish its basic properties, and obtain the desired polynomials with hard multiples. The ideas of this section are a cleaner version of the techniques used in the PIT algorithm of Forbes and Shpilka [25] for commutative roABPs.

6.5.1 Leading and trailing diagonals

We begin with the definition of a leading diagonal, which is a generalization of a leading monomial.

Definition 6.15. Let $f \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$ be nonzero. The **leading diagonal of** f, denoted $\mathrm{LD}(f)$, is the leading coefficient of $f(\overline{x}\circ\overline{z},\overline{y}\circ\overline{z})$ when this polynomial is considered in the ring $\mathbb{F}[\overline{x},\overline{y}][z_1,\ldots,z_n]$, and where $\overline{x}\circ\overline{z}$ denotes the Hadamard product $(x_1z_1,\ldots,x_nz_n)$. The **trailing diagonal of** f is defined analogously. The zero polynomial has no leading or trailing diagonal.

As this notion has not explicitly appeared prior in the literature, we now establish several straightforward properties. The first is that extremal diagonals are homomorphic with respect to multiplication.

Lemma 6.16. *Let $f,g \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$ be nonzero. Then $\mathrm{LD}(fg) = \mathrm{LD}(f)\,\mathrm{LD}(g)$ and $\mathrm{TD}(fg) = \mathrm{TD}(f)\,\mathrm{TD}(g)$.*

Proof: As $\mathrm{LD}(f) = \mathrm{LC}_{\overline{x},\overline{y}|\overline{z}}(f(\overline{x}\circ\overline{z},\overline{y}\circ\overline{z}))$, where this leading coefficient is taken in the ring $\mathbb{F}[\overline{x},\overline{y}][\overline{z}]$, this automatically follows from the fact that leading coefficients are homomorphic with respect to multiplication (Theorem 3.18). The result for trailing diagonals is symmetric. $\square$

We now show how to relate the leading monomials of the coefficient space of f to the respective monomials associated to the leading diagonal of f.

Proposition 6.17. *Let $f \in \mathbb{F}[x_1,\dots,x_n,y_1,\dots,y_n]$. For any $\overline{b}$, if $\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(\mathrm{LD}(f)) \neq 0$, then*

$$\mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(\mathrm{LD}(f))\right) = \mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)\right) .$$

The respective trailing statement also holds.

Proof: We prove the leading statement, the trailing version is symmetric. Let $f = \sum_{\overline{a},\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}$. We can then expand $f(\overline{x}\circ\overline{z},\overline{y}\circ\overline{z})$ as follows.

$$f(\overline{x}\circ\overline{z},\overline{y}\circ\overline{z}) = \sum_{\overline{c}}\left(\sum_{\overline{a}+\overline{b}=\overline{c}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}\right)\overline{z}^{\overline{c}}$$

choose $\overline{c}_0$ so that $\mathrm{LC}_{\overline{x},\overline{y}|\overline{z}}(f) = \mathrm{Coeff}_{\overline{x},\overline{y}|\overline{z}^{\overline{c}_0}}(f)$, we get that

$$= \left(\sum_{\overline{a}+\overline{b}=\overline{c}_0} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}\right)\overline{z}^{\overline{c}_0} + \sum_{\overline{c}\prec\overline{c}_0}\left(\sum_{\overline{a}+\overline{b}=\overline{c}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}\right)\overline{z}^{\overline{c}} ,$$

where $\mathrm{LD}(f) = \sum_{\overline{a}+\overline{b}=\overline{c}_0} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}$ and $\sum_{\overline{a}+\overline{b}=\overline{c}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}} = 0$ for $\overline{c} \succ \overline{c}_0$. In particular, this means that for any $\overline{b}$ we have that $\alpha_{\overline{a},\overline{b}} = 0$ for $\overline{a} \succ \overline{c}_0 - \overline{b}$.

Thus we have that

$$\begin{aligned}
\mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(\mathrm{LD}(f))\right) &= \mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}\left(\sum_{\overline{a}+\overline{b}=\overline{c}_0} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}\right)\right)\\
&= \mathrm{LM}\left(\alpha_{\overline{c}_0-\overline{b},\overline{b}} \overline{x}^{\overline{c}_0-\overline{b}}\right)\\
&= \overline{x}^{\overline{c}_0-\overline{b}} ,
\end{aligned}$$

as we assume this leading monomial exists, which is equivalent here to $\alpha_{\overline{c}_0-\overline{b},\overline{b}} \neq 0$.

In comparison,

$$\begin{aligned}
\mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)\right) &= \mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}\left(\sum_{\overline{a},\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}\right)\right)\\
&= \mathrm{LM}\left(\sum_{\overline{a}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}}\right)\\
&= \mathrm{LM}\left(\sum_{\overline{a}\succ\overline{c}_0-\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} + \alpha_{\overline{c}_0-\overline{b},\overline{b}} \overline{x}^{\overline{c}_0-\overline{b}} + \sum_{\overline{a}\prec\overline{c}_0-\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}}\right)
\end{aligned}$$

as $\alpha_{\overline{a},\overline{b}} = 0$ for $\overline{a} \succ \overline{c}_0 - \overline{b}$,

$$\begin{aligned}
&= \mathrm{LM}\left(\alpha_{\overline{c}_0-\overline{b},\overline{b}} \overline{x}^{\overline{c}_0-\overline{b}} + \sum_{\overline{a}\prec\overline{c}_0-\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}}\right)\\
&= \overline{x}^{\overline{c}_0-\overline{b}} ,
\end{aligned}$$

where in the last step we again used that $\alpha_{\overline{c}_0-\overline{b},\overline{b}} \neq 0$. This establishes the desired equality. □

We now relate the extremal monomials of the coefficient space of f to the monomials of the coefficient space of the extremal diagonals of f.

Corollary 6.18. *Let $f \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$. Then*

$$\mathrm{LM}(\mathbf{Coeff}_{\overline{x}|\overline{y}}(f)) \supseteq \mathrm{LM}(\mathbf{Coeff}_{\overline{x}|\overline{y}}(\mathrm{LD}(f))) \ .$$

The respective trailing statement also holds.

Proof: This follows as $\mathrm{LM}(\mathbf{Coeff}_{\overline{x}|\overline{y}}(\mathrm{LD}(f)))$ is equal to

$$\left\{ \mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}\left(\mathrm{LD}(f)\right)\right) \,\middle|\, \mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}\left(\mathrm{LD}(f)\right) \neq 0 \right\} \ ,$$

but by Theorem 6.17 this set equals

$$\left\{ \mathrm{LM}\left(\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(f)\right) \,\middle|\, \mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}\left(\mathrm{LD}(f)\right) \neq 0 \right\} \ ,$$

which is clearly contained in $\mathrm{LM}(\mathbf{Coeff}_{\overline{x}|\overline{y}}(f))$. □

We now observe that the number of leading monomials of the coefficient space of a leading diagonal is equal to its sparsity.

Lemma 6.19. *Let $f \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$. For a polynomial g, let $|g|_0$ denotes its sparsity. Then*

$$\left|\mathrm{LM}\left(\mathbf{Coeff}_{\overline{x}|\overline{y}}\left(\mathrm{LD}(f)\right)\right)\right| = |\mathrm{LD}(f)|_0 \ .$$

The respective trailing statement also holds.

Proof: We prove the claim for the leading diagonal, the trailing statement is symmetric. Note that the claim is a vacuous "0=0" if f is zero. For nonzero f, express it as $f = \sum_{\overline{a},\overline{b}} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}}$ so that $\mathrm{LD}(f) = \sum_{\overline{a}+\overline{b}=\overline{c}_0} \alpha_{\overline{a},\overline{b}} \overline{x}^{\overline{a}} \overline{y}^{\overline{b}} = \sum_{\overline{b}} \alpha_{\overline{c}_0-\overline{b},\overline{b}} \overline{x}^{\overline{c}_0-\overline{b}} \overline{y}^{\overline{b}}$ for some $\overline{c}_0 \in \mathbb{N}^n$. Then $\mathrm{Coeff}_{\overline{x}|\overline{y}^{\overline{b}}}(\mathrm{LD}(f)) = \alpha_{\overline{c}_0-\overline{b},\overline{b}} \overline{x}^{\overline{c}_0-\overline{b}}$. As the monomials $\overline{x}^{\overline{c}_0-\overline{b}}$ are distinct and hence linearly independent for distinct $\overline{b}$, it follows that $\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(\mathrm{LD}(f)) = |\{\overline{b} | \alpha_{\overline{c}_0-\overline{b},\overline{b}} \neq 0\}|$, which is equal the sparsity $|\mathrm{LD}(f)|_0$. □

Finally, we now lower bound the coefficient dimension of a polynomial by the sparsity of its extremal diagonals.

Corollary 6.20. *Let $f \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$. Then*

$$\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f) \geq |\mathrm{LD}(f)|_0, |\mathrm{TD}(f)|_0 \ ,$$

where for a polynomial g, $|g|_0$ denotes its sparsity. □

Proof: We give the proof for the leading diagonal, the trailing diagonal is symmetric. By the above,

$$\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(f) \geq \left|\mathrm{LM}\left(\mathbf{Coeff}_{\overline{x}|\overline{y}}(f)\right)\right| \geq \left|\mathrm{LM}\left(\mathbf{Coeff}_{\overline{x}|\overline{y}}\left(\mathrm{LD}(f)\right)\right)\right| = |\mathrm{LD}(f)|_0 \ ,$$

where we passed from span to number of leading monomials (Lemma 3.19), and then passed to the leading monomials of the leading diagonal (Theorem 6.18), and then passed to sparsity of the leading diagonal (Theorem 6.19). □

6.5.2 Lower bounds for multiples for read-once and read-twice ABPs

Having developed the theory of leading diagonals in the previous section, we now turn to using this theory to obtain explicit polynomials whose nonzero multiples all require large roABPs. We also generalize this to read-$O(1)$ oblivious ABPs, but only state the results for $k=2$ as this has a natural application to proof complexity (Section 7). As the restricted computations considered above ($\sum\bigwedge\sum$ formulas and sparse polynomials) have small roABPs, the hard polynomials in this section will also have multiples requiring large complexity in these models as well and thus qualitatively reprove some of the above results. However, we included the previous sections as the hard polynomials there are simpler (being monomials or translations of monomials), and more importantly we will need those results for the proofs below.

The proofs will use the characterization of roABPs by their coefficient dimension (Lemma 3.7), the lower bound for coefficient dimension in terms of the sparsity of the extremal diagonals (Corollary 6.20), and polynomials whose multiples are all non-sparse (Theorem 6.13).

Proposition 6.21. *Let $f(\overline{x},\overline{y}) := \prod_{i=1}^{n}(x_i+y_i+\alpha_i) \in \mathbb{F}[x_1,\ldots,x_n,y_1,\ldots,y_n]$, for $\alpha_i \in \mathbb{F}$. Then for any $0 \neq g \in \mathbb{F}[\overline{x},\overline{y}]$,*

$$\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(g\cdot f) \geq 2^n\ .$$

In particular, all nonzero multiples of f require width at least 2^n to be computed by an roABP in any order of the variables where $\overline{x} \prec \overline{y}$.

Proof: Observe that the leading diagonal of f is insensitive to the α_i. That is, $\mathrm{LD}(x_i+y_i+\alpha_i) = x_i+y_i$, so by multiplicativity of the leading diagonal (Theorem 6.16) we have that $\mathrm{LD}(f) = \prod_i(x_i+y_i)$. Thus, appealing to Corollary 6.20 and Theorem 6.13,

$$\begin{aligned}\dim \mathbf{Coeff}_{\overline{x}|\overline{y}}(g\cdot f) &\geq |\mathrm{LD}(g\cdot f)|_0\\ &= |\mathrm{LD}(g)\cdot\mathrm{LD}(f)|_0\\ &= |\mathrm{LD}(g)\cdot\textstyle\prod_i(x_i+y_i)|_0\\ &\geq 2^n\ .\end{aligned}$$

The claim about roABP width follows from Lemma 3.7. □

Note that this lower bound actually works in the "monotone" setting (if we replace Theorem 6.13 with the monotone Theorem 6.14), as the result only uses the zero/nonzero pattern of the coefficients.

The above result gives lower bounds for coefficient dimension in a *fixed* variable partition. We now symmetrize this construction to get lower bounds for coefficient dimension in *any* variable partition. We proceed as in Section 5.5, where we plant the fixed-partition lower bound into an arbitrary partition. Note that unlike that construction, we will not need auxiliary variables here.

Corollary 6.22. *Let $f \in \mathbb{F}[x_1,\ldots,x_n]$ be defined by $f(\overline{x}) := \prod_{i<j}(x_i+x_j+\alpha_{i,j})$ for $\alpha_{i,j}\in\mathbb{F}$. Then for any partition $\overline{x} = (\overline{u},\overline{v},\overline{w})$ with $m := |\overline{u}| = |\overline{v}|$, and any $0 \neq g \in \mathbb{F}[\overline{x}]$,*

$$\dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}(g\cdot f) \geq 2^m\ ,$$

where we treat $g\cdot f$ as a polynomial in $\mathbb{F}(\overline{w})[\overline{u},\overline{v}]$. In particular, all nonzero multiples of f require width $\geq 2^{\lfloor n/2\rfloor}$ to be computed by an roABP in any order of the variables.

Proof: We can factor f into a copy of the hard polynomial from Theorem 6.21, and the rest. That is,

$$f(\overline{x}) = \prod_{i<j}(x_i + x_j + \alpha_{i,j}) = \prod_{i=1}^{m}(u_i + v_i + \beta_i) \cdot f'(\overline{u}, \overline{v}, \overline{w}) ,$$

for some $\beta_i \in \mathbb{F}$ and nonzero $f'(\overline{u}, \overline{v}, \overline{w}) \in \mathbb{F}[\overline{u}, \overline{v}, \overline{w}]$. Thus,

$$g \cdot f = \big(g(\overline{u}, \overline{v}, \overline{w}) \cdot f'(\overline{u}, \overline{v}, \overline{w})\big) \cdot \prod_{i=1}^{m}(u_i + v_i + \beta_i) .$$

Noting that g, f' are nonzero in $\mathbb{F}[\overline{u}, \overline{v}, \overline{w}]$, they are also nonzero in $\mathbb{F}(\overline{w})[\overline{u}, \overline{v}]$, so that $g \cdot f$ is nonzero multiple of $\prod_{i=1}^{m}(u_i + v_i + \beta_i)$ in $\mathbb{F}(\overline{w})[\overline{u}, \overline{v}]$. Appealing to our lower bound for (nonzero) multiples of coefficient dimension (Theorem 6.21), we have that

$$\dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}(g \cdot f) = \dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}\left(g \cdot f' \cdot \prod_{i=1}^{m}(u_i + v_i + \beta_i)\right) \geq 2^m .$$

The statement about roABPs follows from Lemma 3.7. □

We briefly remark that the above bound does not match the naive bound achieved by writing the polynomial $\prod_{i<j}(x_i + x_j + \alpha_{i,j})$ in its sparse representation, which has $2^{\Theta(n^2)}$ terms. The gap between the lower bound ($2^{\Omega(n)}$) and the upper bound ($2^{O(n^2)}$) is explained by our use of a complete graph to embed the lower bounds of Theorem 6.21 into an arbitrary partition. As discussed after Theorem 5.13 one can use expander graphs to essentially close this gap.

We now observe that the above lower bounds for coefficient dimension suffices to obtain lower bounds for read-twice oblivious ABPs, as we can appeal to the structural result of Anderson, Forbes, Saptharishi, Shpilka and Volk [6] (Theorem 3.9). This result shows that for any read-twice oblivious ABP that (after discarding some variables) there is a partition of the variables across which has small coefficient dimension, which is in contrast to the above lower bound.

Corollary 6.23. *Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be defined by $f(\overline{x}) := \prod_{i<j}(x_i + x_j + \alpha_{i,j})$ for $\alpha_{i,j} \in \mathbb{F}$. Then for any $0 \neq g \in \mathbb{F}[\overline{x}]$, $g \cdot f$ requires width-$2^{\Omega(n)}$ as a read-twice oblivious ABP.*

Proof: Suppose that $g \cdot f$ has a read-twice oblivious ABP of width-w. By the lower-bound of Anderson, Forbes, Saptharishi, Shpilka and Volk [6] (Theorem 3.9), there exists a partition $\overline{x} = (\overline{u}, \overline{v}, \overline{w})$ where $|\overline{u}|, |\overline{v}| \geq \Omega(n)$, and such that $\dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}(g \cdot f) \leq w^4$ (where we treat $g \cdot f$ as a polynomial in $\mathbb{F}(\overline{w})[\overline{u}, \overline{v}]$). Note that we can enforce that the partition obeys $m := |\overline{u}| = |\overline{v}| \geq \Omega(n)$, as we can balance $\overline{u}$ and $\overline{v}$ by pushing variables into $\overline{w}$, as this cannot increase the coefficient dimension (Theorem 3.8). However, appealing to our coefficient dimension bound (Theorem 6.22)

$$w^4 \geq \dim_{\mathbb{F}(\overline{w})} \mathbf{Coeff}_{\overline{u}|\overline{v}}(g \cdot f) \geq 2^m \geq 2^{\Omega(n)} ,$$

so that $w \geq 2^{\Omega(n)}$ as desired. □

7 IPS lower bounds via lower bounds for multiples

In this section we use the lower bounds for multiples of Section 6 to derive lower bounds for $\mathcal{C}$-IPS proofs for various restricted algebraic circuit classes $\mathcal{C}$. The advantage of this approach over the functional lower bounds strategy of Section 5 is that we derive lower bounds for the general IPS system, not just its subclass linear-IPS. While our equivalence (Theorem 4.4) of $\mathcal{C}$-IPS and $\mathcal{C}$-IPS$_{\mathrm{LIN}}$ holds for any strong-enough class $\mathcal{C}$, the restricted classes we consider here (depth-3 powering formulas and roABPs) [18] are not strong enough to use Theorem 4.4 to lift the results of Section 5 to lower bounds for the full IPS system. However, as discussed in the introduction, the techniques of this section can only yield lower bounds for $\mathcal{C}$-IPS refutations of systems of equations which are hard to compute within $\mathcal{C}$ (though our examples are computable by small (general) circuits).

We begin by first detailing the relation between IPS refutations and multiples. We then use our lower bounds for multiples (Section 6) to derive as corollaries lower bounds for $\sum\bigwedge\sum$-IPS and roABP-IPS refutations.

Lemma 7.1. *Let $f,\overline{g},\overline{x}^2-\overline{x}\in\mathbb{F}[x_1,\dots,x_n]$ be an unsatisfiable system of equations, where $\overline{g},\overline{x}^2-\overline{x}$ is satisfiable. Let $C\in\mathbb{F}[\overline{x},y,\overline{z},\overline{w}]$ be an IPS refutation of $f,\overline{g},\overline{x}^2-\overline{x}$. Then*

$$1-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})$$

is a nonzero multiple of f.

Proof: That C is an IPS refutation means that

$$C(\overline{x},f,\overline{g},\overline{x}^2-\overline{x})=1, \qquad C(\overline{x},0,\overline{0},\overline{0})=0\,.$$

We first show that $1-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})$ is a multiple of f, using the first condition on C. Expand $C(\overline{x},y,\overline{z},\overline{w})$ as a univariate in y, so that

$$C(\overline{x},y,\overline{z},\overline{w})=\sum_{i\geq 0}C_i(\overline{x},\overline{z},\overline{w})y^i\,,$$

for $C_i\in\mathbb{F}[\overline{x},\overline{z},\overline{w}]$. In particular, $C_0(\overline{x},\overline{z},\overline{w})=C(\overline{x},0,\overline{z},\overline{w})$. Thus,

$$\begin{aligned}1-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})&=C(\overline{x},f,\overline{g},\overline{x}^2-\overline{x})-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})\\&=\left(\textstyle\sum_{i\geq 0}C_i(\overline{x},\overline{g},\overline{x}^2-\overline{x})f^i\right)-C_0(\overline{x},\overline{g},\overline{x}^2-\overline{x})\\&=\textstyle\sum_{i\geq 1}C_i(\overline{x},\overline{g},\overline{x}^2-\overline{x})f^i\\&=\left(\textstyle\sum_{i\geq 1}C_i(\overline{x},\overline{g},\overline{x}^2-\overline{x})f^{i-1}\right)\cdot f\,.\end{aligned}$$

Thus, $1-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})$ is a multiple of f as desired.

We now show that this is a *nonzero* multiple, using the second condition on C and the satisfiability of $\overline{g},\overline{x}^2-\overline{x}$. That is, the second condition implies that $0=C(\overline{x},0,\overline{0},\overline{0})=C_0(\overline{x},\overline{0},\overline{0})$. If $1-C(\overline{x},0,\overline{g},\overline{x}^2-\overline{x})$ is zero, then by the above we have that $C_0(\overline{x},\overline{g},\overline{x}^2-\overline{x})=1$, so that $C_0(\overline{x},\overline{z},\overline{w})$ is an IPS refutation of

[18] As in Section 6, we will not treat multilinear formulas in this section as they are less natural for the techniques under consideration. Further, IPS lower bounds for multilinear formulas *can* be obtained via functional lower bounds (Theorem 5.15).

$\overline{g}, \overline{x}^2 - \overline{x}$, which contradicts the satisfiability of $\overline{g}, \overline{x}^2 - \overline{x}$ as IPS is a sound proof system. So it must then be that $1 - C(\overline{x}, 0, \overline{g}, \overline{x}^2 - \overline{x})$ is nonzero.

That is, take an $\overline{\alpha}$ satisfying $\overline{g}, \overline{x}^2 - \overline{x}$ so that $\overline{g}(\overline{\alpha}) = \overline{0}, \overline{\alpha}^2 - \overline{\alpha} = \overline{0}$. Substituting this $\overline{\alpha}$ into $C_0(\overline{x}, \overline{g}, \overline{x}^2 - \overline{x})$, we have that $C_0(\overline{x}, \overline{g}, \overline{x}^2 - \overline{x})|_{\overline{x} \leftarrow \overline{\alpha}} = C_0(\overline{\alpha}, \overline{0}, \overline{0})$, and because $C_0(\overline{x}, \overline{0}, \overline{0}) \equiv 0$ in $\mathbb{F}[\overline{x}]$ via the above we have that $C_0(\overline{\alpha}, \overline{0}, \overline{0}) = 0$. Thus, we have that $1 - C(\overline{x}, 0, \overline{g}, \overline{x}^2 - \overline{x}) = 1 - C_0(\overline{x}, \overline{g}, \overline{x}^2 - \overline{x})$ is a nonzero polynomial as its evaluation at $\overline{x} \leftarrow \overline{\alpha}$ is 1. □

The above lemma thus gives a template for obtaining lower bounds for IPS. First, obtain a "hard" polynomial f whose nonzero multiples are hard for $\mathcal{C}$, where f is hopefully also computable by small (general) circuits. Then find additional (simple) polynomials $\overline{g}$ such that $\overline{g}, \overline{x}^2 - \overline{x}$ is satisfiable yet $f, \overline{g}, \overline{x}^2 - \overline{x}$ is unsatisfiable. By the above lemma one then has the desired IPS lower bound for refuting $f, \overline{g}, \overline{x}^2 - \overline{x}$, assuming that $\mathcal{C}$ is sufficiently general. However, for our results we need to be more careful as even though $C(\overline{x}, y, \overline{z}, \overline{w})$ is from the restricted class $\mathcal{C}$, the derived polynomial $C(\overline{x}, 0, \overline{g}, \overline{x}^2 - \overline{x})$ may not be, and thus we will need to appeal to lower bounds for stronger classes.

We now instantiate this template, first for depth-3 powering formulas, where we use lower bounds for multiples of the stronger $\sum \bigwedge \sum \prod^2$ model.

Corollary 7.2. *Let $\mathbb{F}$ be a field with* $\mathrm{char}(\mathbb{F}) = 0$. *Let $f := x_1 \cdots x_n$ and $g := x_1 + \cdots + x_n - n$ with $f, g \in \mathbb{F}[x_1, \ldots, x_n]$. Then $f, g, \overline{x}^2 - \overline{x}$ are unsatisfiable and any $\sum \bigwedge \sum$-IPS refutation requires size at least* $\exp(\Omega(n))$.

Proof: The hypothesis $\mathrm{char}(\mathbb{F}) = 0$ implies that $\{0, \ldots, n\}$ are distinct numbers. In particular, the system $g(\overline{x}) = 0$ and $\overline{x}^2 - \overline{x} = \overline{0}$ is satisfiable and has the unique satisfying assignment $\overline{1}$. However, this single assignment does not satisfy f, as $f(\overline{1}) = \prod_{i=1}^n 1 = 1 \neq 0$, so the entire system is unsatisfiable. Thus, applying our strategy (Theorem 7.1), we see that for any $\sum \bigwedge \sum$-IPS refutation $C(\overline{x}, y, z, \overline{w})$ of $f, g, \overline{x}^2 - \overline{x}$, the polynomial $1 - C(\overline{x}, 0, g, \overline{x}^2 - \overline{x})$ is a nonzero multiple of f.

Let s be the size of C as a $\sum \bigwedge \sum$ formula. As g is linear and the Boolean axioms $\overline{x}^2 - \overline{x}$ are quadratic, it follows that $1 - C(\overline{x}, 0, g, \overline{x}^2 - \overline{x})$ is a sum of powers of quadratics ($\sum \bigwedge \sum \prod^2$) of size $\mathsf{poly}(s)$. As nonzero multiples of f requires $\exp(\Omega(n))$ size as a $\sum \bigwedge \sum \prod^2$ formula (Corollary 6.11) it follows that $\mathsf{poly}(s) \geq \exp(\Omega(n))$, so that $s \geq \exp(\Omega(n))$ as desired. □

We similarly obtain a lower bound for roABP-IPS, where here we use lower bounds for multiples of read-*twice* oblivious ABPs.

Corollary 7.3. *Let $\mathbb{F}$ be a field with* $\mathrm{char}(\mathbb{F}) > n$. *Let $f := \prod_{i<j}(x_i + x_j - 1)$ and $g := x_1 + \cdots + x_n - n$ with $f, g \in \mathbb{F}[x_1, \ldots, x_n]$. Then $f, g, \overline{x}^2 - \overline{x}$ are unsatisfiable and any roABP-*IPS *refutation (in any order of the variables) requires size $\geq$* $\exp(\Omega(n))$.

Proof: The hypothesis $\mathrm{char}(\mathbb{F}) > n$ implies that $\{0, \ldots, n\}$ are distinct numbers. In particular, the system $g(\overline{x}) = 0$ and $\overline{x}^2 - \overline{x} = \overline{0}$ is satisfiable and has the unique satisfying assignment $\overline{1}$. However, this single assignment does not satisfy f as $f(\overline{1}) = \prod_{i<j}(1 + 1 - 1) = 1 \neq 0$, so the entire system is unsatisfiable. Thus, applying our strategy (Theorem 7.1), we see that for any roABP-IPS refutation $C(\overline{x}, y, z, \overline{w})$ of $f, g, \overline{x}^2 - \overline{x}$ that $1 - C(\overline{x}, 0, g, \overline{x}^2 - \overline{x})$ is a nonzero multiple of f.

Let s be the size of C as an roABP. We now argue that $1 - C(\overline{x}, 0, g, \overline{x}^2 - \overline{x})$ has a small read-*twice* oblivious ABP. First, note that we can expand $C(\overline{x}, 0, z, \overline{w})$ into powers of z, so that $C(\overline{x}, 0, z, \overline{w}) =$

$\sum_{0\le i\le s} C_i(\overline{x},\overline{w})z^i$ (where we use that s bounds the width and *degree* of the roABP C). Each $C_i(\overline{x},\overline{w})$ has a $\mathsf{poly}(s)$-size roABP (in the order of the variables of C where z is omitted) as we can compute C_i via interpolation over z, using that each evaluation preserves roABP size (Theorem 3.8). Further, as g is linear, for any i we see that g^i can be computed by a $\mathsf{poly}(n,i)$-size roABP (in any order of the variables) (Theorem 3.15). Combining these facts using closure properties of roABPs under addition and multiplication (Theorem 3.8), we see that $C(\overline{x},0,g,\overline{w})$, and hence $1-C(\overline{x},0,g,\overline{w})$, has a $\mathsf{poly}(s,n)$-size roABP in the order of the variables that C induces on $\overline{x},\overline{w}$. Next observe, that as each Boolean axiom $x_i^2-x_i$ only refers to a single variable, substituting $\overline{w}\leftarrow\overline{x}^2-\overline{x}$ in the roABP for $1-C(\overline{x},0,g,\overline{w})$ will preserve obliviousness of the ABP, but now each variable will be read twice, so that $1-C(\overline{x},0,g,\overline{x}^2-\overline{x})$ has a $\mathsf{poly}(s,n)$-size read-twice oblivious ABP.

Now, using that nonzero multiples of f requires $\exp(\Omega(n))$ size to be computed as read-twice oblivious ABPs (Theorem 6.23) it follows that $\mathsf{poly}(s,n)\ge\exp(\Omega(n))$, so that $s\ge\exp(\Omega(n))$ as desired. □

We note that the above lower bound is for the *size* of the roABP. One can also obtain the stronger result (for similar but less natural axioms) showing that the *width* (and hence also the size) of the roABP must be large, but we do not pursue this as it does not qualitatively change the result.

8 Discussion

In this work we proved new lower bounds for various natural restricted versions of the Ideal Proof System (IPS) of Grochow and Pitassi [35]. While existing work in algebraic proof complexity showed limitations of weak measures of complexity such as the degree and sparsity of a polynomial, our lower bounds are for stronger measures of circuit size that match many of the frontier lower bounds in algebraic circuit complexity. However, our work leaves several open questions and directions for further study, which we now list.

1. Can one obtain proof complexity lower bounds from the recent techniques for lower bounds for depth-4 circuits, such as the results of Gupta, Kamath, Kayal and Saptharishi [36]? Neither of our approaches (functional lower bounds or lower bounds for multiples) currently extend to their techniques.

2. Many proof complexity lower bounds are for refuting unsatisfiable k-CNFs, where $k=O(1)$, which can be encoded as systems of polynomial equations where each equation involves $O(1)$ variables. Can one obtain interesting IPS lower bounds for such systems? Our techniques only establish exponential lower bounds where there is at least one axiom involving $\Omega(n)$ variables.

3. Given an equation $f(\overline{x})=0$ where f has a size-s circuit, there is a natural way to convert this equation to $\mathsf{poly}(s)$-many equations on $O(1)$ *extension variables* by tracing through the computation of f. Can one understand how introducing extension variables affects the complexity of refuting polynomial systems of equations? This seems a viable approach to the previous question when applied to our technique of using lower bounds for multiples.

4. We have shown various lower bounds for multiples by invoking the hardness of the determinant (Theorem 6.7), but this does not lead to satisfactory proof lower bounds as the axioms are complicated. Can one *implicitly* invoke the hardness of the determinant? For example, consider the hard matrix identities suggested by Cook and Rackoff (see for example the survey of Beame and Pitassi [10]) and later studied by Soltys and Cook [80]. That is, consider unsatisfiable equations such as $XY - \mathbf{I}_n, YX - 2 \cdot \mathbf{I}_n$, where X and Y are symbolic $n \times n$ matrices and $\mathbf{I}_n$ is the $n \times n$ identity matrix. The simplest refutations known involve the determinant (see Hrubeš-Tzameret [40], and the discussion in Grochow-Pitassi [35]), can one provide evidence that computing the determinant is intrinsic to such refutations?

5. The lower bounds of this paper are for the *static* IPS system, where one cannot simplify intermediate computations. There are also *dynamic* algebraic proof systems (see Appendix A), can one extend our techniques to that setting?

A Relating IPS to other proof systems

In this section we summarize some existing work on algebraic proof systems and how these other proof systems compare to IPS. In particular, we define the (dynamic) *Polynomial Calculus* refutation system over circuits (related to but slightly different than the system of Grigoriev and Hirsch [32]) and relate it to the (static) IPS system ([59, 35]) considered in this paper. We then examine the roABP-PC system, essentially considered by Tzameret [83], and its separations from sparse-PC. Finally, we consider multilinear-formula-PC as studied by Raz and Tzameret [66, 65] and show that its tree-like version simulates multilinear-formula-IPS, and is hence separated from sparse-PC.

A.1 Polynomial calculus refutations

A substantial body of prior work considers *dynamic* proof systems, which are systems that allow simplification of intermediate polynomials in the proof. In contrast, IPS is a *static* system where the proof is single object with no "intermediate" computations to simplify. We now define the principle dynamic system of interest, the *Polynomial Calculus* system. We give a definition over an arbitrary circuit class, which generalizes the definition of the system as introduced by Clegg, Edmonds, and Impagliazzo [13]. In particular, the notion of proof size in the definition below is made to accommodate very weak circuit classes, while being comparable to standard definitions of size for sufficiently strong circuit measures. For example, in depth-3 powering formulas unbounded multiplications (such as $x_1 \cdots x_n$) are expensive and hence we must charge for the circuit size of the result of a product $v = g \cdot u$ as opposed to the size of just g.

Definition A.1. Let $f_1(\overline{x}), \ldots, f_m(\overline{x}) \in \mathbb{F}[x_1, \ldots, x_n]$ be a system of polynomials. A **Polynomial Calculus (PC) proof** for showing that $p \in \mathbb{F}[\overline{x}]$ is in the ideal generated by $\overline{f}, \overline{x}^2 - \overline{x}$ is a directed acyclic graph with a single sink, where

- Leaves are labelled with an equation from $\overline{f}, \overline{x}^2 - \overline{x}$.

- An internal node v with children $u_1,\dots,u_k$ for $k>1$ is labelled with a linear combination $v = \alpha_1 u_1 + \cdots + \alpha_k u_k$ for $\alpha_i \in \mathbb{F}$.
- An internal node v with a single child u is labelled with the product $g \cdot u$ for some $g \in \mathbb{F}[\overline{x}]$.

The **value** of a node in the proof is defined inductively via the above labels interpreted as equations, and the value of the output node is required to be the desired polynomial p. The proof is **tree-like** if the underlying graph is a tree, and is otherwise **dag-like**. A **PC refutation** of $\overline{f},\overline{x}^2-\overline{x}$ is a proof that 1 is in the ideal of $\overline{f},\overline{x}^2-\overline{x}$ so that $\overline{f},\overline{x}^2-\overline{x}$ is unsatisfiable.

The **size** of each node is defined inductively as follows.

- The size of a leaf v is the size of the minimal circuit agreeing with the value of v.
- The size of an addition node $v = \alpha_1 u_1 + \cdots + \alpha_k u_k$ is k plus the size of each child u_i, plus the size of the minimal circuit agreeing with the value of v.
- The size of a product node $v = g \cdot u$ is the size of the child u plus the size of the minimal circuit agreeing with the value of v.

The size of the proof is the sum of the sizes of each node in the proof. For a restricted algebraic circuit class $\mathcal{C}$, a **$\mathcal{C}$-PC proof** is a PC proof where the circuits are measured as their size coming from the restricted class $\mathcal{C}$.

As with IPS, one can show this is a sound and complete proof system for unsatisfiability of equations. Also as with IPS, in our definition of PC we included the Boolean axioms $\overline{x}^2-\overline{x}$ as this in the most common regime.

An important aspect of the above proof system is that it is *semantic*, as the polynomials derived in the proof are simplified to their smallest equivalent algebraic circuit. This is valid in that such simplifications can be efficiently verified (with randomness) using polynomial identity testing (which can sometimes be derandomized, see Section 3). In contrast, one could instead require a *syntactic* proof system, which would have to provide a proof via syntactic manipulation of algebraic circuits that such simplifications are valid. We will focus on semantic systems as they more naturally compare with IPS, which also requires polynomial identity testing for verification.

While many prior work ([59, 66, 65, 83, 35]) considered algebraic proof systems whose verification relied on polynomial identity testing (because of semantic simplification or otherwise), we note that the system of Grigoriev and Hirsch [32] (which they called “formula-$\mathcal{PC}$”) is actually a *syntactic* system and as such is deterministically checkable. Despite their system being restricted to being syntactic, it is still strong enough to simulate Frege and obtain small-depth refutations of the subset-sum axiom, the pigeonhole principle, and Tseitin tautologies.

Remark A.2. Note that our definition here varies slightly from the definition of Clegg, Edmonds, and Impagliazzo [13], in that we allow products by an arbitrary polynomial g instead of only allowing products of a single variable x_i. For some circuit classes $\mathcal{C}$ these two definitions are polynomially equivalent (see for example the discussion in Raz and Tzameret [66]). In general however, using the product rule $f \vdash x_i \cdot f$ in a *tree-like* proof can only yield $g \cdot f$ where g is a small formula. However, we will be interested in algebraic circuit classes not known to be simulated by small formulas (such as roABPs, which can compute iterated matrix products which are believed to require superpolynomial-size formulas) and thus will consider this stronger product rule.

We now observe that tree-like $\mathcal{C}$-PC can simulate $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}}$ for natural restricted circuit classes $\mathcal{C}$.

Lemma A.3. *Let $\mathcal{C}$ be a restricted class of circuits computing polynomials in $\mathbb{F}[x_1,\ldots,x_n]$, and suppose that $\mathcal{C}$-circuits grow polynomially in size under multiplication and addition, that is,*

- $\mathrm{size}_{\mathcal{C}}(f\cdot g)\le \mathsf{poly}(\mathrm{size}_{\mathcal{C}}(f),\mathrm{size}_{\mathcal{C}}(g))$.
- $\mathrm{size}_{\mathcal{C}}(f+g)\le \mathsf{poly}(\mathrm{size}_{\mathcal{C}}(f))+\mathsf{poly}(\mathrm{size}_{\mathcal{C}}(g))$.

In particular, one can take $\mathcal{C}$ to be sparse polynomials, depth-3 powering formulas (in characteristic zero), or roABPs.

Then if $\overline{f},\overline{x}^2-\overline{x}$ are computable by size-t $\mathcal{C}$-circuits and have a $\mathcal{C}$-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation of size-s, then $\overline{f},\overline{x}^2-\overline{x}$ have a tree-like $\mathcal{C}$-PC refutation of size-$\mathsf{poly}(s,t,n)$, which is $\mathsf{poly}(s,t,n)$-explicit given the IPS refutation.

Proof: That the relevant classes obey these closure properties is mostly immediate. See for example Theorem 3.8 for roABPs. For depth-3 powering formulas, the closure under addition is immediate and for multiplication it follows from Fischer [19].

Turning to the simulation, such an IPS refutation is an equation of the form $\sum_j g_j f_j+\sum_i h_i\cdot(x_i^2-x_i)=1$. Using the closure properties of $\mathcal{C}$, one can compute the expression $\sum_j g_j f_j+\sum_i h_i\cdot(x_i^2-x_i)$ in the desired size, which yields the required (explicit) derivation of 1. □

Note that the above claim does *not* work for multilinear formulas, as multilinear polynomials are not closed under multiplication. That tree-like multilinear-formula-PC simulates multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ is more intricate, and is given in Theorem A.11.

The Polynomial Calculus proof system has received substantial attention since its introduction by Clegg, Edmonds, and Impagliazzo [13], typically when the complexity of the proofs is measured in terms of the number of monomials. In particular, Impagliazzo, Pudlák and Sgall [41] showed an exponential lower bound for the subset-sum axiom.

Theorem A.4 (Impagliazzo, Pudlák and Sgall [41])**.** *Let $\mathbb{F}$ be a field of characteristic zero. Let $\overline{\alpha}\in\mathbb{F}^n$, $\beta\in\mathbb{F}$ and $A:=\{\sum_{i=1}^n\alpha_i x_i:\overline{x}\in\{0,1\}^n\}$ be so that $\beta\notin A$. Then $\alpha_1x_1+\cdots+\alpha_nx_n-\beta,\overline{x}^2-\overline{x}$ is unsatisfiable and any PC refutation requires degree $\ge\lceil n/2\rceil+1$ and $\exp(\Omega(n))$-many monomials.* □

A.2 roABP-PC

The class of roABPs are a natural restricted class of algebraic computation that non-trivially goes beyond sparse polynomials. In proof complexity, roABP-PC was explored by Tzameret [83] (under the name of *ordered formulas*, a formula-variant of roABPs, but the results there apply to roABPs as well). In particular, Tzameret [83] observed that roABP-PC can be deterministically checked using the efficient PIT algorithm for roABPs due to Raz and Shpilka [64].

Given the Impagliazzo, Pudlák and Sgall [41] lower bound for the subset-sum axiom (Theorem A.4), our roABP-IPS upper bound for this axiom (Corollary 4.14), and the relation between IPS_{LIN} and tree-like PC (Theorem A.3), we can conclude the following exponential separation.

Corollary A.5. *Let $\mathbb{F}$ be a field of characteristic zero. Then $x_1+\cdots+x_n+1,\overline{x}^2-\overline{x}$ is unsatisfiable, requires sparse-PC refutations of size $\exp(\Omega(n))$, but has $\mathsf{poly}(n)$-explicit $\mathsf{poly}(n)$-size roABP-IPS_{LIN} and tree-like roABP-PC refutations.* □

This strengthens a result of Tzameret [83], who separated *dag*-like roABP-PC from sparse-PC. However, we note that it is not clear whether sparse-PC can be efficiently simulated by roABP-IPS_{LIN}.

A.3 Multilinear formula PC

We now proceed to study algebraic proofs defined in terms of multilinear formulas, as explored by Raz and Tzameret [66, 65]. We seek to show that the tree-like version of this system can simulate multilinear-formula-IPS_{LIN}. While tree-like $\mathcal{C}$-PC can naturally simulate $\mathcal{C}$-IPS_{LIN} if $\mathcal{C}$ is closed under multiplication (Theorem A.3), the product of two multilinear polynomials may not be multilinear. Therefore, the simulation we derive is more intricate, and is similar to the efficient multilinearization results for multilinear formulas from Section 4.3. We first define the Raz-Tzameret [66, 65] system (which they called *fMC*).

Definition A.6. Let $f_1(\overline{x}),\ldots,f_m(\overline{x})\in\mathbb{F}[x_1,\ldots,x_n]$ be a system of polynomials. A **multilinear-formula-PC$^\neg$ refutation** for showing that the system $\overline{f},\overline{x}^2-\overline{x}$ is unsatisfiable is a multilinear-formula-PC refutation of $\overline{f}(\overline{x}),\overline{x}^2-\overline{x},\overline{\neg x}^2-\overline{\neg x},\overline{x}\circ\overline{\neg x}$ in the ring $\mathbb{F}[x_1,\ldots,x_n,\neg x_1,\ldots,\neg x_n]$, where '$\circ$' denotes the entrywise product so that $\overline{x}\circ\overline{\neg x}=(x_1\neg x_1,\ldots,x_n\neg x_n)$.

That is, a multilinear-formula-PC$^\neg$ refutation of $\overline{f},\overline{x}^2-\overline{x}$ is a multilinear-formula-PC refutation with the additional variables $\overline{\neg x}:=(\neg x_1,\ldots,\neg x_n)$ which are constrained so that $\neg x_i=1-x_i$ (so that '$\neg$' here is simply a modifier of the symbol 'x' as opposed to being imbued with mathematical meaning). These additional variables are important, as without them the system is not complete. For example, in attempting to refute the subset-sum axiom $\sum_i x_i+1,\overline{x}^2-\overline{x}$ in multilinear-formula-PC alone, one can never multiply the axiom $\sum_i x_i+1$ by another (non-constant) polynomial as it would ruin multilinearity. However, in multilinear-formula-PC$^\neg$ one can instead multiply by polynomials in $\overline{\neg x}$ and appropriately simplify. We now formalize this by showing that tree-like multilinear-formula-PC$^\neg$ can simulate multilinear-formula-$\text{IPS}_{\text{LIN}'}$ (which is complete (Theorem 4.12)).

We begin by proving a lemma on how the $\overline{\neg x}$ variables can help multilinearize products. In particular, if we have a monomial $(\overline{1}-\overline{\neg x})^{\overline{a}}$ (which is meant to be equal to $\overline{x}^{\overline{a}}$) and multiply by $\overline{x}^{\overline{1}}$ we should be able to prove that this product equals $\overline{x}^{\overline{1}}$ modulo the axioms.

Lemma A.7. *Working in the ring $\mathbb{F}[x_1,\ldots,x_d,\neg x_1,\ldots,\neg x_d]$, and for $\overline{0}\le\overline{a}\le\overline{1}$,*

$$(\overline{1}-\overline{\neg x})^{\overline{a}}\overline{x}^{\overline{1}}-\overline{x}^{\overline{1}}=C(\overline{x},\overline{x}\circ\overline{\neg x})\ ,$$

for $C(\overline{x},\overline{z})\in\mathbb{F}[\overline{x},\overline{z}]$ where $C(\overline{x},\overline{x}\circ\overline{\neg x})$ can be $\mathsf{poly}(2^d)$-explicitly derived from the axioms $\overline{x}\circ\overline{\neg x}$ in $\mathsf{poly}(2^d)$ steps using tree-like multilinear-formula-PC.

Proof:

$$\begin{aligned}
(\overline{1}-\overline{\neg x})^{\overline{a}}\overline{x}^{\overline{1}} &= \overline{x}^{\overline{1}-\overline{a}}\cdot(\overline{x}-\overline{x}\circ\overline{\neg x})^{\overline{a}} \\
&= \overline{x}^{\overline{1}-\overline{a}}\cdot\left(\sum_{\overline{0}\le\overline{b}\le\overline{a}}\overline{x}^{\overline{a}-\overline{b}}(-\overline{x}\circ\overline{\neg x})^{\overline{b}}\right) \\
&= \overline{x}^{\overline{1}-\overline{a}}\cdot\left(\overline{x}^{\overline{a}}+\sum_{\overline{0}<\overline{b}\le\overline{a}}\overline{x}^{\overline{a}-\overline{b}}(-\overline{x}\circ\overline{\neg x})^{\overline{b}}\right) \\
&= \overline{x}^{\overline{1}}+\sum_{\overline{0}<\overline{b}\le\overline{a}}\overline{x}^{\overline{1}-\overline{b}}(-\overline{x}\circ\overline{\neg x})^{\overline{b}} \\
&= \overline{x}^{\overline{1}}+C(\overline{x},\overline{x}\circ\overline{\neg x})\ ,
\end{aligned}$$

where $C(\overline{x},\overline{z})$ is defined by

$$C(\overline{x},\overline{z}) := \sum_{\overline{0}<\overline{b}\le\overline{a}}\overline{x}^{\overline{1}-\overline{b}}(-\overline{z})^{\overline{b}}\ .$$

Now note that $C(\overline{x},\overline{x}\circ\overline{\neg x})$ can be derived by tree-like multilinear-formula-PC. That is, the expression $\overline{x}^{\overline{1}-\overline{b}}(-\overline{x}\circ\overline{\neg x})^{\overline{b}}$ is multilinear (as the product is variable disjoint) and in the ideal of $\overline{x}\circ\overline{\neg x}$ as $\overline{b}>\overline{0}$, and is clearly a $\mathsf{poly}(d)$-size explicit multilinear formula. Summing over the 2^d-1 relevant $\overline{b}$ gives the result. □

We now show how to prove the equivalence of $g(\overline{x})$ and $g(\overline{1}-\overline{\neg x})$ modulo $\overline{x}+\overline{\neg x}-\overline{1}$, if g is computable by a small multilinear formula, where we proceed variable by variable.

Lemma A.8. *Working in the ring $\mathbb{F}[x_1,\ldots,x_n,\neg x_1,\ldots,\neg x_n]$, if $g\in\mathbb{F}[\overline{x}]$ is computable by a size-t multilinear formula, than*

$$g(\overline{x})-g(\overline{1}-\overline{\neg x}) = C(\overline{x},\overline{x}+\overline{\neg x}-\overline{1})\ ,$$

for $C(\overline{x},\overline{z})\in\mathbb{F}[\overline{x},\overline{z}]$ where $C(\overline{x},\overline{x}+\overline{\neg x}-\overline{1})$ is derivable from $\overline{x}+\overline{\neg x}-\overline{1}$ in size-$\mathsf{poly}(t,n)$ tree-like multilinear-formula-PC, which is $\mathsf{poly}(t,n)$-explicit given the formula for g.

Proof: We proceed to replace $\overline{1}-\overline{\neg x}$ with $\overline{x}$ one variable at a time. Using $(\overline{x}_{\le i},(\overline{1}-\overline{\neg x})_{>i})$ to denote $(x_1,\ldots,x_i,1-\neg x_{i+1},\ldots,1-\neg x_n)$, we see that via telescoping that

$$\begin{aligned}
g(\overline{x})-g(\overline{1}-\overline{\neg x}) &= g(\overline{x}_{\le n},(\overline{1}-\overline{\neg x})_{>n})-g(\overline{x}_{<1},(\overline{1}-\overline{\neg x})_{\ge 1}) \\
&= \sum_{i=1}^{n}\Big(g(\overline{x}_{\le i},(\overline{1}-\overline{\neg x})_{>i})-g(\overline{x}_{<i},(\overline{1}-\overline{\neg x})_{\ge i})\Big) \\
&= \sum_{i=1}^{n}\Big(g(\overline{x}_{<i},x_i,(\overline{1}-\overline{\neg x})_{>i})-g(\overline{x}_{<i},1-\neg x_i,(\overline{1}-\overline{\neg x})_{>i})\Big)\ .
\end{aligned}$$

Now note that $g(\overline{x}_{<i},y,(\overline{1}-\overline{\neg x})_{>i})$ is a multilinear polynomial, which as it is linear in y can be written as

$$g(\overline{x}_{<i},y,(\overline{1}-\overline{\neg x})_{>i}) = \Big(g(\overline{x}_{<i},1,(\overline{1}-\overline{\neg x})_{>i})-g(\overline{x}_{<i},0,(\overline{1}-\overline{\neg x})_{>i})\Big)y+g(\overline{x}_{<i},0,(\overline{1}-\overline{\neg x})_{>i})\ .$$

Thus, plugging in x_i and $1-\neg x_i$,

$$\begin{aligned}
&g(\overline{x}_{<i}, x_i, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 1-\neg x_i, (\overline{1}-\overline{\neg x})_{>i}) \\
&\qquad = \Big(\Big(g(\overline{x}_{<i}, 1, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big)x_i + g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big) \\
&\qquad\qquad - \Big(\Big(g(\overline{x}_{<i}, 1, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big)(1-\neg x_i) + g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big) \\
&\qquad = \Big(g(\overline{x}_{<i}, 1, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big)(x_i + \neg x_i - 1)\ .
\end{aligned}$$

Plugging this into the above telescoping equation,

$$\begin{aligned}
g(\overline{x}) - g(\overline{1}-\overline{\neg x}) &= \sum_{i=1}^{n} \Big(g(\overline{x}_{<i}, x_i, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 1-\neg x_i, (\overline{1}-\overline{\neg x})_{>i})\Big) \\
&= \sum_{i=1}^{n} \Big(g(\overline{x}_{<i}, 1, (\overline{1}-\overline{\neg x})_{>i}) - g(\overline{x}_{<i}, 0, (\overline{1}-\overline{\neg x})_{>i})\Big)(x_i + \neg x_i - 1) \\
&=: C(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})\ .
\end{aligned}$$

Clearly each $g(\overline{x}_{<i}, b, (\overline{1}-\overline{\neg x})_{>i})$ for $b \in \{0,1\}$ has a $\mathsf{poly}(t)$-size multilinear algebraic formula, so the entire expression $C(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})$ can be computed by tree-like multilinear-formula-PC from $\overline{x}+\overline{\neg x}-\overline{1}$ explicitly in $\mathsf{poly}(t,n)$ steps. □

Using the above lemma, we now show how to multilinearize a multilinear-formula times a low-degree multilinear monomial.

Lemma A.9. *Let $g, f \in \mathbb{F}[x_1, \ldots, x_n, y_1, \ldots, y_d]$, where g is computable by a size-t multilinear formula and $y = \prod_{i=1}^{d} y_i$. Then*

$$g(\overline{1}-\overline{\neg x}, \overline{1}-\overline{\neg y})\overline{y}^{\overline{1}} - \mathrm{ml}(g(\overline{x}, \overline{y})\overline{y}^{\overline{1}}) = C(\overline{x}, \overline{y}, \overline{x}+\overline{\neg x}-\overline{1}, \overline{y}\circ\overline{\neg y})\ ,$$

where $C(\overline{x}, \overline{y}, \overline{x}+\overline{\neg x}-\overline{1}, \overline{y}\circ\overline{\neg y})$ can be derived from the axioms $\overline{x}+\overline{\neg x}-\overline{1}, \overline{y}\circ\overline{\neg y}$ in $\mathsf{poly}(2^d, t, n)$ steps using tree-like multilinear-formula-PC.

Proof: Express g as $g(\overline{x}, \overline{y}) = \sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}}$ in the ring $\mathbb{F}[\overline{x}][\overline{y}]$, so that each $g_{\overline{a}}$ is multilinear. Then,

$$g(\overline{1}-\overline{\neg x}, \overline{1}-\overline{\neg y}) \cdot \overline{y}^{\overline{1}} = \sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{1}-\overline{\neg x})(\overline{1}-\overline{\neg y})^{\overline{a}} \cdot \overline{y}^{\overline{1}}$$

appealing to Theorem A.7 to obtain $(\overline{1}-\overline{\neg y})^{\overline{a}}\overline{y}^{\overline{1}} = \overline{y}^{\overline{1}} + C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y})$ for $C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y})$ derivable in tree-like multilinear-formula-PC from $\overline{y}\circ\overline{\neg y}$ in $\mathsf{poly}(2^d)$ steps,

$$= \sum_{\overline{a}} g_{\overline{a}}(\overline{1}-\overline{\neg x}) \left(\overline{y}^{\overline{1}} + C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y})\right)$$

appealing to Theorem A.8 to obtain $g_{\overline{a}}(\overline{1}-\overline{\neg x}) = g_{\overline{a}}(\overline{x}) + B_{\overline{a}}(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})$ for $B_{\overline{a}}(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})$ derivable in tree-like multilinear-formula-PC from $\overline{x}+\overline{\neg x}-\overline{1}$ in $\mathsf{poly}(t,n)$ steps,

$$\begin{aligned}
&= \sum_{\overline{a}} \left(g_{\overline{a}}(\overline{x}) + B_{\overline{a}}(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})\right) \cdot \left(\overline{y}^{\overline{1}} + C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y})\right) \\
&= \sum_{\overline{a}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{1}} + \sum_{\overline{a}} \Big(B_{\overline{a}}(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})\overline{y}^{\overline{1}} + g_{\overline{a}}(\overline{x})C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y}) \\
&\qquad\qquad + B_{\overline{a}}(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1})C_{\overline{a}}(\overline{y}, \overline{y}\circ\overline{\neg y})\Big) \\
&= \mathrm{ml}(g(\overline{x},\overline{y})\overline{y}^{\overline{1}}) + C(\overline{x},\overline{y},\overline{x}+\overline{\neg x}-\overline{1}, \overline{y}\circ\overline{\neg y}) ,
\end{aligned}$$

by defining C appropriately, and as

$$\begin{aligned}
\mathrm{ml}(g(\overline{x},\overline{y})\overline{y}^{\overline{1}}) &= \mathrm{ml}\left(\textstyle\sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}} \cdot \overline{y}^{\overline{1}}\right) \\
&= \mathrm{ml}\left(\textstyle\sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{a}+\overline{1}}\right) \\
&= \textstyle\sum_{\overline{0}\le\overline{a}\le\overline{1}} g_{\overline{a}}(\overline{x})\overline{y}^{\overline{1}} .
\end{aligned}$$

By interpolation, it follows that for each exponent $\overline{a}$ there are constants $\overline{\alpha}_{\overline{a},\overline{\beta}}$ such that $g_{\overline{a}}(\overline{x}) = \sum_{\overline{\beta}\in\{0,1\}^d} \alpha_{\overline{a},\overline{\beta}}\, g(\overline{x},\overline{\beta})$. From this it follows that $g_{\overline{a}}$ is computable by a multilinear formula of size-$\mathsf{poly}(t,2^d)$. It thus follows that $C(\overline{x},\overline{y},\overline{x}+\overline{\neg x}-\overline{1},\overline{y}\circ\overline{\neg y})$ is a sum of 2^d terms, each of which is explicitly derivable in $\mathsf{poly}(2^d,t,n)$ steps in tree-like multilinear-formula-PC from $\overline{x}+\overline{\neg x}-\overline{1},\overline{y}\circ\overline{\neg y}$ (as the multiplications are variable-disjoint), and thus $C(\overline{x},\overline{y},\overline{x}+\overline{\neg x}-\overline{1},\overline{y}\circ\overline{\neg y})$ is similar derived by tree-like multilinear-formula-PC. □

By linearity we can extend the above to multilinearization of a multilinear-formula times a sparse low-degree multilinear polynomial.

Corollary A.10. *Let $g, f \in \mathbb{F}[x_1,\ldots,x_n]$ be multilinear, where g is computable by a size-t multilinear formula and f is s-sparse and $\deg f \le d$. Then*

$$g(\overline{1}-\overline{\neg x})\cdot f(\overline{x}) - \mathrm{ml}(g(\overline{x})\cdot f(\overline{x})) = C(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1}, \overline{x}\circ\overline{\neg x}) ,$$

where $C(\overline{x}, \overline{x}+\overline{\neg x}-\overline{1}, \overline{x}\circ\overline{\neg x})$ can be derived from the axioms $\overline{x}+\overline{\neg x}-\overline{1}, \overline{x}\circ\overline{\neg x}$ in $\mathsf{poly}(2^d,s,t,n)$ steps using tree-like multilinear-formula-PC. □

We now conclude by showing that tree-like multilinear-formula-PC$^\neg$ can efficiently simulate multilinear-formula-IPS$_{\mathrm{LIN}'}$. Recall that this proof system simply requires an IPS refutation that is linear in the non-Boolean axioms, so that in particular $\sum_j g_j f_j \equiv 1 \mod \overline{x}^2 - \overline{x}$ for efficiently computable g_j.

Corollary A.11. *Let $f_1,\ldots,f_m \in \mathbb{F}[x_1,\ldots,x_n]$ be degree at most d multilinear s-sparse polynomials which are unsatisfiable over $\overline{x}\in\{0,1\}^n$. Suppose that there are multilinear $g_j \in \mathbb{F}[\overline{x}]$ computable by size-t multilinear formula such that*

$$\sum_{i=1}^{m} g_j(\overline{x}) f_j(\overline{x}) \equiv 1 \mod \overline{x}^2 - \overline{x} .$$

Then there is a tree-like multilinear-formula-PC$^{\neg}$ refutation of $\overline{f},\overline{x}^2-\overline{x}$ of size $\mathsf{poly}(2^d,s,t,n,m)$, which is $\mathsf{poly}(2^d,s,t,n,m)$-explicit given the formulas for the f_j,g_j.

In particular, if there is a size-t multilinear-formula-$\mathrm{IPS}_{LIN'}$ *refutation of $\overline{f},\overline{x}^2-\overline{x}$, then there is a tree-like multilinear-formula-PC$^{\neg}$ refutation of $\overline{f},\overline{x}^2-\overline{x}$ of size $\mathsf{poly}(2^d,s,t,n,m)$ which is $\mathsf{poly}(2^d,s,t,n,m)$-explicit given the refutation of $\overline{f},\overline{x}^2-\overline{x}$ and formulas for the f_j.*

Proof: By the above multilinearization (Corollary A.10), there are $C_j\in\mathbb{F}[\overline{x},\overline{z},\overline{w}]$ such that

$$g_j(\overline{1}-\overline{\neg x})f_j(\overline{x})=\mathrm{ml}(g_j(\overline{x})f_j(\overline{x}))+C_j(\overline{x},\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x})\ .$$

where $C_j(\overline{x},\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x})$ is derivable from $\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x}$ in $\mathsf{poly}(2^d,s,t,n)$ steps of tree-like multilinear-formula-PC. Thus, as $g_j(\overline{1}-\overline{\neg x})$ has a $\mathsf{poly}(t)$-size multilinear formula, in $\mathsf{poly}(2^d,s,t,n,m)$ steps we can derive from $\overline{f}(\overline{x}),\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x}$,

$$\sum_{j=1}^{m}\left(g_j(\overline{1}-\overline{\neg x})f_j(\overline{x})-C_j(\overline{x},\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x})\right)=\sum_{j=1}^{m}\mathrm{ml}(g_j(\overline{x})f_j(\overline{x}))$$

as $\sum_{i=1}^{m}g_j(\overline{x})f_j(\overline{x})\equiv 1 \mod \overline{x}^2-\overline{x}$ we have that

$$\sum_{j=1}^{m}\mathrm{ml}(g_j(\overline{x})f_j(\overline{x}))=\mathrm{ml}\left(\sum_{i=1}^{m}g_j(\overline{x})f_j(\overline{x})\right)=1\ ,$$

where we appealed to linearity of multilinearization (Theorem 3.12), so that

$$\sum_{j=1}^{m}\left(g_j(\overline{1}-\overline{\neg x})f_j(\overline{x})-C_j(\overline{x},\overline{x}+\overline{\neg x}-\overline{1},\overline{x}\circ\overline{\neg x})\right)=1\ ,$$

yielding the desired refutation, where the explicitness is clear.

The claim about multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}'}$ follows, and thus a refutation induces the equation $\sum_{i=1}^{m}g_j(\overline{x})f_j(\overline{x})\equiv 1 \mod \overline{x}^2-\overline{x}$ with the appropriate size bounds. □

Given this efficient simulation of multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}'}$ by tree-like multilinear-formula-PC$^{\neg}$ (Theorem A.11), our multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}}$ refutation of the subset-sum axiom (Theorem 4.15), and the lower bound for sparse-PC of the subset-sum axiom (Theorem A.4), we obtain the following separation result.

Corollary A.12. *Let $\mathbb{F}$ be a field of characteristic zero. Then $x_1+\cdots+x_n+1,\overline{x}^2-\overline{x}$ is unsatisfiable, requires sparse-PC refutations of size $\exp(\Omega(n))$, but has $\mathsf{poly}(n)$-explicit $\mathsf{poly}(n)$-size multilinear-formula-*$\mathrm{IPS}_{\mathrm{LIN}}$ *and tree-like multilinear-formula-PC$^{\neg}$ refutations.* □

This strengthens a results of Raz and Tzameret [66, 65], who separated *dag*-like multilinear-formula-PC$^{\neg}$ from sparse-PC. However, we note that it is not clear whether sparse-PC can be efficiently simulated by multilinear-formula-$\mathrm{IPS}_{\mathrm{LIN}'}$.

B Explicit multilinear polynomial satisfying a functional equation

In Section 5.2 we showed that any polynomial that agrees with the function $\overline{x} \mapsto 1/(\sum_i x_i - \beta)$ on the Boolean cube must have degree $\geq n$. However, as there is a unique multilinear polynomial obeying this functional equation it is natural to ask for an explicit description of this polynomial, which we now give.

Proposition B.1. *Let $n \geq 1$ and $\mathbb{F}$ be a field with $\operatorname{char}(\mathbb{F}) > n$. Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be the unique multilinear polynomial such that*

$$f(\overline{x}) = \frac{1}{\sum_i x_i - \beta} ,$$

for $\overline{x} \in \{0,1\}^n$. Then

$$f(\overline{x}) = -\sum_{k=0}^{n} \frac{k!}{\prod_{j=0}^{k}(\beta - j)} S_{n,k} ,$$

where $S_{n,k} := \sum_{S \subseteq \binom{[n]}{k}} \prod_{i \in S} x_i$ is the k-th elementary symmetric polynomial.

Proof: It follows from the uniqueness of the evaluations of multilinear polynomials over the Boolean cube that

$$f(\overline{x}) = \sum_{T \subseteq [n]} f(\mathbb{1}_T) \prod_{i \in T} x_i \prod_{i \notin T} (1 - x_i)$$

where $\mathbb{1}_T \in \{0,1\}^n$ is the indicator vector of the set T, so that

$$= \sum_{T \subseteq [n]} \frac{1}{|T| - \beta} \prod_{i \in T} x_i \prod_{i \notin T} (1 - x_i) .$$

Using this, let us determine the coefficient of $\prod_{i \in S} x_i$ in $f(\overline{x})$, for $S \subseteq [n]$ with $|S| = k$. First observe that setting $x_i = 0$ for $i \notin S$ preserves this coefficient, so that

$$\operatorname{Coeff}_{\prod_{i \in S} x_i}\Big(f(\overline{x})\Big) = \operatorname{Coeff}_{\prod_{i \in S} x_i}\left(\sum_{T \subseteq [n]} \frac{1}{|T| - \beta} \prod_{i \in T} x_i \prod_{i \notin T} (1 - x_i)\right)\Bigg|_{x_i \leftarrow 0, i \in S}$$

and thus those sets T with $T \not\subseteq S$ are zeroed out,

$$\begin{aligned}
&= \mathrm{Coeff}_{\prod_{i\in S} x_i}\left(\sum_{T\subseteq S} \frac{1}{|T|-\beta} \prod_{i\in T} x_i \prod_{i\in S\setminus T}(1-x_i)\right) \\
&= \sum_{T\subseteq S} \frac{1}{|T|-\beta} \mathrm{Coeff}_{\prod_{i\in S} x_i}\left(\prod_{i\in T} x_i \prod_{i\in S\setminus T}(1-x_i)\right) \\
&= \sum_{T\subseteq S} \frac{1}{|T|-\beta}(-1)^{k-|T|} \\
&= \sum_{j=0}^{k} \binom{k}{j} \frac{1}{j-\beta}(-1)^{k-j} \\
&= -\frac{k!}{\prod_{j=0}^{k}(\beta-j)} ,
\end{aligned}$$

where the last step uses the subclaim below.

Subclaim B.2.

$$\sum_{j=0}^{k} \binom{k}{j} \frac{1}{j-\beta}(-1)^{k-j} = -\frac{k!}{\prod_{j=0}^{k}(\beta-j)} .$$

Sub-Proof: Clearing denominators,

$$\prod_{i=0}^{k}(i-\beta) \cdot \sum_{j=0}^{k} \binom{k}{j} \frac{1}{j-\beta}(-1)^{k-j} = \sum_{j=0}^{k} \binom{k}{j}(-1)^{k-j} \prod_{i\neq j}(i-\beta) .$$

Note that the right hand side is a univariate degree $\le k$ polynomial in β, so it is determined by its value on $\ell \in \{0,\dots,k\}$ (that $\mathbb{F}$ has large characteristic implies that these values are distinct). Note that on these values,

$$\begin{aligned}
\sum_{j=0}^{k} \binom{k}{j}(-1)^{k-j} \prod_{i\neq j}(i-\ell) &= \binom{k}{\ell}(-1)^{k-\ell} \prod_{0\le i<\ell}(i-\ell) \cdot \prod_{\ell<i\le k}(i-\ell) \\
&= \binom{k}{\ell}(-1)^{k-\ell} \cdot (-1)^{\ell}\ell! \cdot (k-\ell)! \\
&= (-1)^k k! .
\end{aligned}$$

Thus by interpolation $\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\prod_{i\neq j}(i-\beta) = (-1)^k k!$ for all β, and thus dividing by $\prod_{i=0}^{k}(i-\beta)$ and clearing -1's yields the claim. □

This then gives the claim as the coefficient of $\prod_{i\in S} x_i$ only depends on $|S| = k$. □

Noting that the above coefficients are all nonzero because $\mathrm{char}(\mathbb{F}) > n$. Thus, we obtain the following corollary by observing that degree and sparsity are non-increasing under multilinearization (Theorem 3.12).

Corollary B.3. *Let $n \geq 1$ and $\mathbb{F}$ be a field with* $\mathrm{char}(\mathbb{F}) > n$. *Suppose that $\beta \in \mathbb{F} \setminus \{0, \ldots, n\}$. Let $f \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial such that*

$$f(\overline{x}) \left(\sum_i x_i - \beta \right) = 1 \mod \overline{x}^2 - \overline{x} \,.$$

Then $\deg f \geq n$*, and f requires $\geq 2^n$ monomials.* □

Acknowledgments

We would like to thank Rafael Oliveira for telling us of Theorem 6.14, Mrinal Kumar and Ramprasad Saptharishi for conversations [23] clarifying the roles of functional lower bounds in this work, as well as Avishay Tal for pointing out how Theorem B.1 implies an optimal functional lower bound for sparsity (Corollary B.3). We would also like to thank Joshua Grochow for helpful discussions regarding this work. We are grateful for the anonymous reviewers for their careful read of the paper and for their comments.

References

[1] MANINDRA AGRAWAL: Proving lower bounds via pseudo-random generators. In *Proc. 25th Found. Softw. Techn. Theoret. Comp. Sci. Conf. (FSTTCS'05)*, pp. 92–105. Springer, 2005. [doi:10.1007/11590156_6] 14, 56, 57

[2] MANINDRA AGRAWAL, ROHIT GURJAR, ARPITA KORWAR, AND NITIN SAXENA: Hitting-sets for ROABP and sum of set-multilinear circuits. *SIAM J. Comput.*, 44(3):669–697, 2015. [doi:10.1137/140975103, arXiv:1406.7535, ECCC:TR14-085] 8, 18

[3] MANINDRA AGRAWAL, CHANDAN SAHA, AND NITIN SAXENA: Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Proc. 45th STOC*, pp. 321–330. ACM Press, 2013. [doi:10.1145/2488608.2488649, arXiv:1209.2333] 7, 58, 60

[4] MICHAEL ALEKHNOVICH AND ALEXANDER A. RAZBOROV: Lower bounds for polynomial calculus: Non-binomial case. *Trudy Mat. Inst. Steklova*, 242:23–43, 2003 (Russian). MathNet.Ru. Preliminary version in FOCS'01. 3

[5] YAROSLAV ALEKSEEV, DIMA GRIGORIEV, EDWARD A. HIRSCH, AND IDDO TZAMERET: Semi-algebraic proofs, IPS lower bounds, and the τ-conjecture: Can a natural number be negative? In *Proc. 52nd STOC*, pp. 54–67. ACM Press, 2020. [doi:10.1145/3357713.3384245, arXiv:1911.06738, ECCC:TR19-142] 16

[6] MATTHEW ANDERSON, MICHAEL A. FORBES, RAMPRASAD SAPTHARISHI, AMIR SHPILKA, AND BEN LEE VOLK: Identity testing and lower bounds for read-k oblivious algebraic branching programs. *ACM Trans. Comput. Theory*, 10(1):1–30, 2018. Preliminary version in CCC'16. [doi:10.1145/3170709, arXiv:1511.07136, ECCC:TR15-184] 8, 18, 20, 66

[7] MATTHEW ANDERSON, DIETER VAN MELKEBEEK, AND ILYA VOLKOVICH: Derandomizing polynomial identity testing for multilinear constant-read formulae. In *Proc. 26th IEEE Conf. on Comput. Complexity (CCC'11)*, pp. 273–282. IEEE Comp. Soc., 2011. [doi:10.1109/CCC.2011.18, ECCC:TR10-188] 57

[8] VIKRAMAN ARVIND, PUSHKAR S. JOGLEKAR, PARTHA MUKHOPADHYAY, AND S. RAJA: Randomized polynomial-time identity testing for noncommutative circuits. *Theory of Computing*, 15(7):1–36, 2019. Preliminary version in STOC'17. [doi:10.4086/toc.2019.v015a007, arXiv:1606.00596, ECCC:TR16-089] 17

[9] PAUL BEAME, RUSSELL IMPAGLIAZZO, JAN KRAJÍČEK, TONIANN PITASSI, AND PAVEL PUDLÁK: Lower bounds on Hilbert's Nullstellensatz and propositional proofs. *Proc. London Math. Soc.*, s3-73(1):1–26, 1996. Preliminary version in FOCS'94. [doi:10.1112/plms/s3-73.1.1] 3, 4

[10] PAUL BEAME AND TONIANN PITASSI: Propositional proof complexity: Past, present and future. *Bull. EATCS*, 65:66–89, 1998. [ECCC:TR98-067] 70

[11] SAMUEL R. BUSS, DIMA GRIGORIEV, RUSSELL IMPAGLIAZZO, AND TONIANN PITASSI: Linear gaps between degrees for the polynomial calculus modulo distinct primes. *J. Comput. System Sci.*, 62(2):267–289, 2001. Preliminary versions in CCC'99 and STOC'99. [doi:10.1006/jcss.2000.1726] 3

[12] SAMUEL R. BUSS, RUSSELL IMPAGLIAZZO, JAN KRAJÍČEK, PAVEL PUDLÁK, ALEXANDER A. RAZBOROV, AND JIŘÍ SGALL: Proof complexity in algebraic systems and bounded depth Frege systems with modular counting. *Comput. Complexity*, 6(3):256–298, 1996. [doi:10.1007/BF01294258] 3

[13] MATTHEW CLEGG, JEFF EDMONDS, AND RUSSELL IMPAGLIAZZO: Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proc. 28th STOC*, pp. 174–183. ACM Press, 1996. [doi:10.1145/237814.237860] 3, 16, 70, 71, 72

[14] STEPHEN A. COOK AND ROBERT A. RECKHOW: Corrections for "On the lengths of proofs in the propositional calculus (Preliminary version)". *SIGACT News*, 6(3):15–22, 1974. Preliminary version in STOC'74. [doi:10.1145/1008311.1008313]

[15] STEPHEN A. COOK AND ROBERT A. RECKHOW: The relative efficiency of propositional proof systems. *J. Symbolic Logic*, 44(1):36–50, 1979. [doi:10.2307/2273702] 2

[16] DAVID COX, JOHN LITTLE, AND DONAL O'SHEA: *Ideals, Varieties, and Algorithms*. Springer, 3rd edition, 2007. [doi:10.1007/978-3-319-16721-3] 22, 23

[17] RICHARD A. DEMILLO AND RICHARD J. LIPTON: A probabilistic remark on algebraic program testing. *Inform. Process. Lett.*, 7(4):193–195, 1978. [doi:10.1016/0020-0190(78)90067-4] 17

[18] ZEEV DVIR, AMIR SHPILKA, AND AMIR YEHUDAYOFF: Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. Comput.*, 39(4):1279–1293, 2010. Preliminary version in STOC'08. [doi:10.1137/080735850] 14, 54, 55

[19] ISMOR FISCHER: Sums of like powers of multivariate linear forms. *Math. Magazine*, 67(1):59–61, 1994. [doi:10.1080/0025570X.1994.11996185] 10, 72

[20] MICHAEL A. FORBES: *Polynomial Identity Testing of Read-Once Oblivious Algebraic Branching Programs*. Ph. D. thesis, MIT, 2014. LINK at handle.net. 17, 18, 19

[21] MICHAEL A. FORBES: Deterministic divisibility testing via shifted partial derivatives. In *Proc. 56th FOCS*, pp. 451–465. IEEE Comp. Soc., 2015. [doi:10.1109/FOCS.2015.35] 7, 14, 18, 56, 57, 58, 60

[22] MICHAEL A. FORBES, ANKIT GUPTA, AND AMIR SHPILKA: Personal Communication to Gupta, Kamath, Kayal, Saptharishi [37], 2013. 22

[23] MICHAEL A. FORBES, MRINAL KUMAR, AND RAMPRASAD SAPTHARISHI: Functional lower bounds for arithmetic circuits and connections to boolean circuit complexity. In *Proc. 31st Comput. Complexity Conf. (CCC'16)*, pp. 33:1–19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.CCC.2016.33, arXiv:1605.04207] 11, 12, 80

[24] MICHAEL A. FORBES, RAMPRASAD SAPTHARISHI, AND AMIR SHPILKA: Hitting sets for multilinear read-once algebraic branching programs, in any order. In *Proc. 46th STOC*, pp. 867–875. ACM Press, 2014. [doi:10.1145/2591796.2591816, arXiv:1309.5668] 7, 8, 18, 57, 58

[25] MICHAEL A. FORBES AND AMIR SHPILKA: On identity testing of tensors, low-rank recovery and compressed sensing. In *Proc. 44th STOC*, pp. 163–172. ACM Press, 2012. [doi:10.1145/2213977.2213995, arXiv:1111.0663, ECCC:TR11-147] 8, 15, 62

[26] MICHAEL A. FORBES AND AMIR SHPILKA: Explicit Noether Normalization for simultaneous conjugation via polynomial identity testing. In *Proc. 17th Internat. Workshop on Randomization and Computation (RANDOM'13)*, pp. 527–542. Springer, 2013. [doi:10.1007/978-3-642-40328-6 37, arXiv:1303.0084, ECCC:TR13-033] 7, 15, 57, 58, 59, 60

[27] MICHAEL A. FORBES AND AMIR SHPILKA: Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. In *Proc. 54th FOCS*, pp. 243–252. IEEE Comp. Soc., 2013. [doi:10.1109/FOCS.2013.34, arXiv:1209.2408, ECCC:TR12-115] 7, 8, 20, 22, 86

[28] MICHAEL A. FORBES AND AMIR SHPILKA: Complexity theory column 88: Challenges in polynomial factorization. *SIGACT News*, 46(4):32–49, 2015. [doi:10.1145/2852040.2852051] 55

[29] MICHAEL A. FORBES, AMIR SHPILKA, IDDO TZAMERET, AND AVI WIGDERSON: Proof complexity lower bounds from algebraic circuit complexity. In *Proc. 31st Comput. Complexity Conf. (CCC'16)*, pp. 32:1–17. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.CCC.2016.32, arXiv:1606.05050, ECCC:TR16-098] 1, 16

[30] JOACHIM VON ZUR GATHEN AND ERICH L. KALTOFEN: Factoring sparse multivariate polynomials. *J. Comput. System Sci.*, 31(2):265–287, 1985. Preliminary version in FOCS'83. [doi:10.1016/0022-0000(85)90044-3] 55

[31] DIMA GRIGORIEV: Tseitin's tautologies and lower bounds for Nullstellensatz proofs. In *Proc. 39th FOCS*, pp. 648–652. IEEE Comp. Soc., 1998. [doi:10.1109/SFCS.1998.743515] 3

[32] DIMA GRIGORIEV AND EDWARD A. HIRSCH: Algebraic proof systems over formulas. *Theoret. Comput. Sci.*, 303(1):83–102, 2003. [doi:10.1016/S0304-3975(02)00446-2, ECCC:TR01-011] 3, 4, 10, 24, 70, 71

[33] DIMA GRIGORIEV AND MAREK KARPINSKI: An exponential lower bound for depth 3 arithmetic circuits. In *Proc. 30th STOC*, pp. 577–582. ACM Press, 1998. [doi:10.1145/276698.276872] 12

[34] DIMA GRIGORIEV AND ALEXANDER A. RAZBOROV: Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000. Preliminary version in FOCS'98. [doi:10.1007/s002009900021] 11, 12

[35] JOSHUA A. GROCHOW AND TONIANN PITASSI: Circuit complexity, proof complexity, and polynomial identity testing: The ideal proof system. *J. ACM*, 65(6):37:1–59, 2018. [doi:10.1145/3230742] 3, 4, 5, 9, 10, 24, 25, 29, 69, 70, 71

[36] ANKIT GUPTA, PRITISH KAMATH, NEERAJ KAYAL, AND RAMPRASAD SAPTHARISHI: Approaching the chasm at depth four. *J. ACM*, 61(6):33:1–16, 2014. Preliminary version in CCC'13. [doi:10.1145/2629541] 7, 12, 60, 69

[37] ANKIT GUPTA, PRITISH KAMATH, NEERAJ KAYAL, AND RAMPRASAD SAPTHARISHI: Arithmetic circuits: A chasm at depth three. *SIAM J. Comput.*, 45(3):1064–1079, 2016. Preliminary version in FOCS'13. [doi:10.1137/140957123, ECCC:TR13-026] 82

[38] ROHIT GURJAR, ARPITA KORWAR, NITIN SAXENA, AND THOMAS THIERAUF: Deterministic identity testing for sum of read-once oblivious arithmetic branching programs. *Comput. Complexity*, 26(4):1–46, 2016. Preliminary version in CCC'15. [doi:10.1007/s00037-016-0141-z, arXiv:1411.7341] 58, 60

[39] JOOS HEINTZ AND CLAUS-PETER SCHNORR: Testing polynomials which are easy to compute (extended abstract). In *Proc. 12th STOC*, pp. 262–272. ACM Press, 1980. [doi:10.1145/800141.804674] 14, 56

[40] PAVEL HRUBEŠ AND IDDO TZAMERET: Short proofs for the determinant identities. *SIAM J. Comput.*, 44(2):340–383, 2015. Preliminary version in STOC'12. [doi:10.1137/130917788, arXiv:1112.6265, ECCC:TR11-174] 70

[41] RUSSELL IMPAGLIAZZO, PAVEL PUDLÁK, AND JIŘÍ SGALL: Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Comput. Complexity*, 8(2):127–144, 1999. [doi:10.1007/s000370050024, ECCC:TR97-042] 3, 9, 11, 12, 24, 45, 47, 72, 73

[42] RUSSELL IMPAGLIAZZO AND AVI WIGDERSON: P=BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proc. 29th STOC*, pp. 220–229. ACM Press, 1997. [doi:10.1145/258533.258590] 54

[43] VALENTINE KABANETS AND RUSSELL IMPAGLIAZZO: Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complexity*, 13(1–2):1–46, 2004. Preliminary version in STOC'03. [doi:10.1007/s00037-004-0182-6] 14, 54, 55, 56

[44] ERICH L. KALTOFEN: Factorization of polynomials given by straight-line programs. In SILVIO MICALI, editor, *Randomness and Computation*, volume 5 of *Adv. Comput. Res.*, pp. 375–412. JAI Press, Inc., 1989. LINK at author's website. 14, 53, 54, 55

[45] NEERAJ KAYAL: Personal Communication to Saxena [71], 2008. 15, 59

[46] NEERAJ KAYAL: An exponential lower bound for the sum of powers of bounded degree polynomials. *Electron. Colloq. Comput. Complexity*, TR12-081(81), 2012. [ECCC] 7, 12, 60

[47] ADAM KLIVANS AND DANIEL A. SPIELMAN: Randomness efficient identity testing of multivariate polynomials. In *Proc. 33rd STOC*, pp. 216–223. ACM Press, 2001. [doi:10.1145/380752.380801] 18

[48] JAN KRAJÍČEK: *Bounded Arithmetic, Propositional Logic, and Complexity Theory*. Volume 60 of *Encycl. Math. Appl.* Cambridge Univ. Press, 1995. [doi:10.1017/CBO9780511529948] 2, 3

[49] MRINAL KUMAR AND RAMPRASAD SAPTHARISHI: An exponential lower bound for homogeneous depth-5 circuits over finite fields. In *Proc. 32nd Comput. Complexity Conf. (CCC'17)*, pp. 31:1–30. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.CCC.2017.31, arXiv:1507.00177, ECCC:TR15-109] 12

[50] FU LI, IDDO TZAMERET, AND ZHENGYU WANG: Characterizing propositional proofs as non-commutative formulas. *SIAM J. Comput.*, 47(4):1424–1462, 2018. [doi:10.1137/16M1107632, arXiv:1412.8746] 5, 6, 8, 10, 24

[51] MEENA MAHAJAN, BV RAGHAVENDRA RAO, AND KARTEEK SREENIVASAIAH: Building above read-once polynomials: Identity testing and hardness of representation. *Algorithmica*, 76(4):890–909, 2016. Preliminary version in COCOON'14. [ECCC:TR15-202] 60

[52] NOAM NISAN: Lower bounds for non-commutative computation. In *Proc. 23rd STOC*, pp. 410–418. ACM Press, 1991. [doi:10.1145/103418.103462] 5, 6, 7, 8, 18, 19

[53] NOAM NISAN AND AVI WIGDERSON: Hardness vs randomness. *J. Comput. System Sci.*, 49(2):149–167, 1994. Preliminary version in FOCS'88. [doi:10.1016/S0022-0000(05)80043-1] 15, 54

[54] NOAM NISAN AND AVI WIGDERSON: Lower bounds on arithmetic circuits via partial derivatives. *Comput. Complexity*, 6(3):217–234, 1996. Preliminary version in FOCS'95. [doi:10.1007/BF01294256] 7, 12, 38, 59

[55] RAFAEL OLIVEIRA: Personal communication, 2015. 61

[56] RAFAEL OLIVEIRA: Factors of low individual degree polynomials. *Comput. Complexity*, 25(2):507–561, 2016. Preliminary version in CCC'15. 14, 54, 55

[57] RAFAEL OLIVEIRA, AMIR SHPILKA, AND BEN LEE VOLK: Subexponential size hitting sets for bounded depth multilinear formulas. *Comput. Complexity*, 25(2):455–505, 2016. Preliminary version in CCC'15. 9

[58] ØYSTEIN ORE: Über höhere Kongruenzen. *Norsk Mat. Forenings Skrifter Ser. I*, 7(15):27, 1922. 17

[59] TONIANN PITASSI: Algebraic propositional proof systems. In *Descriptive Complexity and Finite Models*, volume 31 of *DIMACS Ser. Discr. Math. and Theoret. Comp. Sci.*, pp. 215–244. Amer. Math. Soc., 1997. LINK at author's website. [doi:10.1090/dimacs/031] 3, 4, 5, 70, 71

[60] TONIANN PITASSI AND IDDO TZAMERET: Algebraic proof complexity: progress, frontiers and challenges. *ACM SIGLOG News*, 3(3):21–43, 2016. [doi:10.1145/2984450.2984455] 3

[61] PAVEL PUDLÁK: Lower bounds for resolution and cutting plane proofs and monotone computations. *J. Symbolic Logic*, 62(3):981–998, 1997. [doi:10.2307/2275583] 3

[62] RAN RAZ: Separation of multilinear circuit and formula size. *Theory of Computing*, 2(6):121–135, 2006. Preliminary version in FOCS'04. [doi:10.4086/toc.2006.v002a006] 9, 21

[63] RAN RAZ: Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2):8:1–17, 2009. Preliminary version in STOC'04. [doi:10.1145/1502793.1502797, ECCC:TR03-067] 9, 18, 21, 53

[64] RAN RAZ AND AMIR SHPILKA: Deterministic polynomial identity testing in non-commutative models. *Comput. Complexity*, 14(1):1–19, 2005. Preliminary version in CCC'04. [doi:10.1007/s00037-005-0188-8] 5, 8, 18, 72

[65] RAN RAZ AND IDDO TZAMERET: Resolution over linear equations and multilinear proofs. *Ann. Pure Appl. Logic*, 155(3):194–224, 2008. [doi:10.1016/j.apal.2008.04.001, arXiv:0708.1529, ECCC:TR07-078] 3, 4, 10, 24, 70, 71, 73, 77

[66] RAN RAZ AND IDDO TZAMERET: The strength of multilinear proofs. *Comput. Complexity*, 17(3):407–457, 2008. [doi:10.1007/s00037-008-0246-0, ECCC:TR06-001] 3, 4, 10, 16, 24, 70, 71, 73, 77

[67] RAN RAZ AND AMIR YEHUDAYOFF: Balancing syntactically multilinear arithmetic circuits. *Comput. Complexity*, 17(4):515–535, 2008. [doi:10.1007/s00037-008-0254-0] 21

[68] RAN RAZ AND AMIR YEHUDAYOFF: Lower bounds and separations for constant depth multilinear circuits. *Comput. Complexity*, 18(2):171–207, 2009. Preliminary version in CCC'08. [doi:10.1007/s00037-009-0270-8, ECCC:TR08-006] 9, 21, 53

[69] ALEXANDER A. RAZBOROV: Lower bounds for the polynomial calculus. *Comput. Complexity*, 7(4):291–324, 1998. [doi:10.1007/s000370050013] 3

[70] RAMPRASAD SAPTHARISHI: 2012. Personal communication to Forbes–Shpilka [27]. 20

[71] NITIN SAXENA: Diagonal circuit identity testing and lower bounds. In *Proc. 35th Internat. Colloq. on Automata, Languages, and Programming (ICALP'08)*, pp. 60–71. Springer, 2008. [doi:10.1007/978-3-540-70575-8_6, ECCC:TR07-124] 6, 7, 18, 21, 22, 84

[72] NITIN SAXENA: Progress on polynomial identity testing. *Bull. EATCS*, 99:49–79, 2009. EATCS. [arXiv:1401.0976, ECCC:TR09-101] 17

[73] NITIN SAXENA: Progress on polynomial identity testing - II. In M. AGRAWAL AND V. ARVIND, editors, *Perspectives in Computational Complexity: The Somenath Biswas Anniversary Volume*, pp. 131–146. Springer, 2014. [doi:10.1007/978-3-319-05446-9_7, arXiv:1401.0976, ECCC:TR13-186] 17

[74] JACOB T. SCHWARTZ: Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980. Preliminary version in EUROSAM'79. [doi:10.1145/322217.322225] 17

[75] AMIR SHPILKA: Affine projections of symmetric polynomials. *J. Comput. System Sci.*, 65(4):639–659, 2002. Preliminary version in CCC'01. [doi:10.1016/S0022-0000(02)00021-1, ECCC:TR01-035] 6

[76] AMIR SHPILKA AND ILYA VOLKOVICH: Improved polynomial identity testing for read-once formulas. In *Proc. 13th Internat. Workshop on Randomization and Computation (RANDOM'09)*, volume 5687, pp. 700–713. Springer, 2009. 57, 58, 59, 60

[77] AMIR SHPILKA AND AVI WIGDERSON: Depth-3 arithmetic circuits over fields of characteristic zero. *Comput. Complexity*, 10(1):1–27, 2001. [doi:10.1007/PL00001609] 22

[78] AMIR SHPILKA AND AMIR YEHUDAYOFF: Arithmetic circuits: A survey of recent results and open questions. *Found. Trends Theor. Comp. Sci.*, 5(3–4):207–388, 2010. [doi:10.1561/0400000039] 6, 17, 18, 25

[79] MICHAEL SHUB AND STEVE SMALE: On the intractability of Hilbert's Nullstellensatz and an algebraic version of "NP≠P?". *Duke Math. J.*, 81:47–54, 1995. Available in the Collected papers of Stephen Smale, pp. 1508–1515, World Scientific 2000. 16

[80] MICHAEL SOLTYS AND STEPHEN A. COOK: The proof complexity of linear algebra. *Ann. Pure Appl. Logic*, 130(1–3):277–323, 2004. Preliminary version in LICS'02. [doi:10.1016/j.apal.2003.10.018] 70

[81] VOLKER STRASSEN: Vermeidung von Divisionen. *J. reine angew. Math.*, 264:184–202, 1973. [doi:10.1515/crll.1973.264.184] 24, 25

[82] MADHU SUDAN, LUCA TREVISAN, AND SALIL P. VADHAN: Pseudorandom generators without the XOR lemma. *J. Comput. System Sci.*, 62(2):236–266, 2001. Preliminary version in STOC'99. [doi:10.1006/jcss.2000.1730] 54

[83] IDDO TZAMERET: Algebraic proofs over noncommutative formulas. *Inform. Comput.*, 209(10):1269–1292, 2011. Preliminary version in TAMC'10. [doi:10.1016/j.ic.2011.07.004, arXiv:1004.2159, ECCC:TR10-097] 3, 8, 16, 70, 71, 72, 73

[84] ILYA VOLKOVICH: Computations beyond exponentiation gates and applications. *Electron. Colloq. Comput. Complexity*, TR15-042, 2015. [ECCC] 61

[85] ILYA VOLKOVICH: Deterministically factoring sparse polynomials into multilinear factors and sums of univariate polynomials. In *Proc. 19th Internat. Workshop on Randomization and Computation (RANDOM'15)*, pp. 943–958. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015. [doi:10.4230/LIPIcs.APPROX-RANDOM.2015.943, ECCC:TR14-168] 55, 57

[86] RICHARD ZIPPEL: Probabilistic algorithms for sparse polynomials. In *Proc. Internat. Symp. Symbolic and Algebraic Manipulation (EUROSAM'79)*, pp. 216–226. Springer, 1979. [doi:10.1007/3-540-09519-5_73] 17

AUTHORS

Michael A. Forbes
Assistant professor
University of Illinois at Urbana-Champaign
IL, USA
miforbes@illinois.edu
http://miforbes.cs.illinois.edu/

Amir Shpilka
Professor
Tel Aviv University
Tel Aviv, Israel
shpilka@post.tau.ac.il
www.cs.tau.ac.il/~shpilka

Iddo Tzameret
Professor
Imperial College London
London, UK
iddo.tzameret@gmail.com
https://www.doc.ic.ac.uk/~itzamere/

Avi Wigderson
Institute for Advanced Study, Princeton
avi@ias.edu
www.math.ias.edu/avi/home

ABOUT THE AUTHORS

MICHAEL A. FORBES obtained his Ph. D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 2014, where he was co-advised by Scott Aaronson and Amir Shpilka. In his dissertation, he developed new deterministic algorithms to solve cases of the Polynomial Identity Testing problem. In 2017, he joined the faculty of the University of Illinois at Urbana-Champaign. His research focuses on the interaction of randomness, algebra, and computation.

AMIR SHPILKA obtained his Ph. D. in Computer Science and Mathematics from the Hebrew University in Jerusalem in 2001 under the supervision of Avi Wigderson. From 2005 to 2014 he was a faculty member at the CS department at the Technion. In October 2014 he joined the CS department at Tel Aviv University. His research interests lie in complexity theory, especially in algebraic complexity.

IDDO TZAMERET holds a Chair in Computational Complexity at Imperial College London. Before that he was a professor at the University of London (Royal Holloway), a visiting scholar at Oxford University, an assistant professor at Tsinghua University and a research fellow at the Academy of Sciences in Prague. He received his Ph. D. from Tel Aviv University under the supervision of Ran Raz and Nachum Dershowitz. His research lies in complexity theory, exploring different approaches to the limits of efficient computation and inference, both as a natural and a mathematical phenomenon. This includes algebraic, logical and combinatorial approaches in complexity, lower bounds on concrete computational models, proof complexity and satisfiability. Among his other expertise is the history of rock ’n roll.

AVI WIGDERSON was born in Haifa, Israel in 1956, and received his Ph. D. in 1983 at Princeton University under Dick Lipton. He enjoys and is fascinated with studying the power and limits of efficient computation, and the remarkable impact of this field on understanding our world. Avi’s other major source of fascination and joy are his three kids, Eyal, Einat, and Yuval, and his granddaughter Tamar.