Electronic Colloquium on Computational Complexity, Report No. 172 (2021)

Ideals, Determinants, and Straightening: Proving and Using Lower Bounds for Polynomial Ideals

Robert Andrews* Michael A. Forbes†

December 1, 2021

Abstract

We show that any nonzero polynomial in the ideal generated by the $r \times r$ minors of an $n \times n$ matrix X can be used to efficiently approximate the determinant. Specifically, for any nonzero polynomial f in this ideal, we construct a small depth-three f-oracle circuit that approximates the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant in the sense of border complexity. For many classes of algebraic circuits, this implies that every nonzero polynomial in the ideal generated by $r \times r$ minors is at least as hard to approximately compute as the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant. We also prove an analogous result for the Pfaffian of a $2n \times 2n$ skew-symmetric matrix and the ideal generated by Pfaffians of $2r \times 2r$ principal submatrices.

This answers a recent question of Grochow [Gro20, Conjecture 6.3] about complexity in polynomial ideals in the setting of border complexity. Leveraging connections between the complexity of polynomial ideals and other questions in algebraic complexity, our results provide a generic recipe that allows lower bounds for the determinant to be applied to other problems in algebraic complexity. We give several such applications, two of which are highlighted below.

- We prove new lower bounds for the Ideal Proof System of Grochow and Pitassi. Specifically, we give super-polynomial lower bounds for refutations computed by low-depth circuits. This extends the recent breakthrough low-depth circuit lower bounds of Limaye, Srinivasan, and Tavenas [LST21] to the setting of proof complexity. Moreover, we show that for many natural circuit classes, the approximative proof complexity of our hard instance is governed by the approximative circuit complexity of the determinant.
- We construct new hitting set generators for the closure of low-depth circuits. For any $\varepsilon > 0$, we construct generators with seed length $O(n^\varepsilon)$ that hit n-variate low-depth circuits. Our generators attain a near-optimal tradeoff between their seed length and degree, and are computable by low-depth circuits of near-linear size (with respect to the size of their output). This matches the seed length of the generators recently obtained by Limaye, Srinivasan, and Tavenas [LST21], but improves on the degree and circuit complexity of the generator.

*Department of Computer Science, University of Illinois Urbana-Champaign. Email: rgandre2@illinois.edu. Supported by NSF grants CCF-1755921 and CCF-1814788.

†Department of Computer Science, University of Illinois Urbana-Champaign. Email: miforbes@illinois.edu. Supported by NSF grants CCF-1755921, CCF-1814788, and CAREER award 2047310.

ISSN 1433-8092

Contents

1 Introduction

A central goal of algebraic complexity theory is to understand the resources needed to compute multivariate polynomials in algebraic models of computation. Typically, one attempts to determine the complexity of a single family of polynomials $\{f_n(\overline{x}) : n \in \mathbb{N}\}$, such as the $n \times n$ determinant or permanent. A generalization of this task is to examine the complexity of a family of *ideals* $\{I_n \subseteq \mathbb{F}[\overline{x}] : n \in \mathbb{N}\}$ of polynomials. Recall that in a commutative ring R, an ideal $I \subseteq R$ is a subset of R such that (1) if $a, b \in I$, then $a + b \in I$, and (2) if $a \in I$ and $r \in R$, then $ar \in I$. Ideals naturally arise in commutative algebra and algebraic geometry; for example, the set of polynomials that vanish on a subset $V \subseteq \mathbb{F}^n$ is an ideal. Closer to computer science and algebraic complexity, ideals appear in the study of polynomial identity testing, polynomial factorization, and algebraic proof complexity, though these appearances are not always made explicit. Due to the prominence of ideals in algebra and algebraic complexity, it is both natural and worthwhile to study them from a complexity-theoretic perspective.

Every nonzero ideal contains polynomials of arbitrarily large circuit complexity. This is a straightforward consequence of the fact that ideals are closed under multiplication by arbitrary polynomials. A more interesting task, then, is to determine the minimum possible complexity of a nonzero polynomial in an ideal.

Unfortunately, little is known about the complexity of ideals aside from what is implicit in their connection to other problems of algebraic complexity. A recent column by Grochow [Gro20] surveyed these connections and posed some open questions, both general and concrete, about the complexity of ideals. In particular, he raised the following question regarding an explicit family of ideals.

Conjecture ([Gro20, Conjecture 6.3])**.** *Let X be a $n \times n$ matrix of variables and let I_n be the ideal generated by the $n/2 \times n/2$ minors of X. For every nonzero polynomial $f(X) \in I_n$, there is a small algebraic circuit with f-oracle gates that computes the $m \times m$ determinant for some $m = n^{\Theta(1)}$.*

Due to the close relationship between the non-vanishing of minors and matrix rank, it is natural to conjecture that such a circuit exists. If the oracle circuit is not restricted in any manner, then the desired circuit exists simply because the determinant can be computed efficiently by algebraic circuits. However, if the oracle circuit is required to be, for example, a formula, then this question becomes nontrivial, as the determinant is not known to be computable by small formulas.

The main contribution of our work is to resolve this conjecture in the setting of approximate algebraic computation.

Theorem. *Grochow's conjecture is true (with respect to border complexity).*

Specifically, we show that for any nonzero polynomial $f \in I_n$, the $\Theta(n^{1/3}) \times \Theta(n^{1/3})$ determinant can be approximately computed by a small depth-three f-oracle circuit with a single oracle gate. A direct consequence of this is that for many circuit classes $\mathcal{C}$, if the determinant cannot be approximated by polynomial-size $\mathcal{C}$-circuits, then neither can any polynomial in the ideal I_n. Naturally, this has applications to polynomial identity testing and algebraic proof complexity by employing the supporting role played by the complexity of ideals in those areas.

Before describing our results in more detail, we briefly survey what is known about the complexity of ideals and its connections to polynomial identity testing and algebraic proof complexity.

1.1 The Complexity of Ideals

Most of what is known about the complexity of ideals is limited to ideals generated by a single polynomial. The ideal $\langle f \rangle$ generated by a polynomial $f(\overline{x})$ consists of all multiples of f, so questions

about the complexity of this ideal become questions about the complexity of f and its multiples. Determining the minimum complexity of a polynomial in $\langle f \rangle$ amounts to determining whether there is a multiple of f that is significantly easier to compute than f itself. This leads to the question of factoring algebraic circuits: given a small circuit computing a polynomial $g(\overline{x})$, can the factors of $g(\overline{x})$ be computed by small circuits?

This question was addressed in a celebrated result of Kaltofen [Kal87] (with alternate proofs by Bürgisser [Bür00, Theorem 2.21] and Chou, Kumar, and Solomon [CKS19a]), who showed that factors (of low multiplicity) of small circuits can be computed by small circuits. Taking the contrapositive, if $f(\overline{x})$ cannot be computed by small circuits, then neither can any polynomial $g \in \langle f \rangle$ which has f as a factor of low multiplicity. Polynomial factorization has since been studied in restricted algebraic circuit classes, including low-depth circuits [DSY09; CKS19b], formulas [Oli16; DSS18], algebraic branching programs [DSS18; ST20], and sparse polynomials [BSV20]. This is motivated in part by the use of Kaltofen's theorem to establish hardness-to-pseudorandomness results for polynomial identity testing, as done in the work of Kabanets and Impagliazzo [KI04].

Kaltofen's result gives us a strong understanding of the complexity of the low-degree polynomials in a principal ideal. Because algebraic complexity theory is primarily interested in the computation of low-degree polynomials, this suffices for most applications. However, the situation would be cleaner if lower bounds on the complexity of a polynomial f implied comparable lower bounds on the complexity of all polynomials in the ideal $\langle f \rangle$, not just for those polynomials $g \in \langle f \rangle$ for which f is a factor of low multiplicity. Kaltofen [Kal87] asked in the language of factorization whether this is the case; this question remains open and is now known as the Factor Conjecture. In the setting of approximative algebraic computation, the analogue of the Factor Conjecture was proved by Bürgisser [Bür04]. It is interesting to note that, coincidentally, we also make essential use of approximative computation in our work.

For non-principal ideals, much less is known. What knowledge we do have stems from connections to polynomial identity testing and the Ideal Proof System. We defer our explanation of these connections to Subsection 1.2 and Subsection 1.3, respectively.

Approximate algebraic computation will play a key role in our work, so we briefly discuss it here. For simplicity, we will focus on circuits and polynomials defined over the complex numbers; for more details, including a field-independent definition of approximate computation, see Subsection 2.1. We say that a polynomial $f(\overline{x})$ can be approximately computed by small algebraic circuits if there is a collection of polynomials $\{f_\varepsilon : \varepsilon > 0\}$ such that (1) for all $\varepsilon > 0$, the polynomial f_ε can be computed by a small circuit, and (2) we have $\lim_{\varepsilon \to 0} f_\varepsilon = f$, where convergence is coefficient-wise. Over the complex numbers, this can be interpreted as saying that f lies in the closure (with respect to the Euclidean topology) of the set of polynomials computable by small circuits. If f can be approximated well by polynomials from a circuit class $\mathcal{C}$, then we say that f is in $\overline{\mathcal{C}}$, the closure of $\mathcal{C}$. The circuit complexity of the approximating polynomials f_ε is referred to as the *border complexity* of f. Naturally, one can also consider border complexity with respect to other classes of algebraic circuits, such as formulas or branching programs.

Border complexity appeared as early as the late 1970s, when Bini, Capovani, Romani, and Lotti [BCRL79] and Bini [Bin80] improved upon the state-of-the-art algorithms for matrix multiplication by considering an approximative version of the problem. The notion of border complexity also plays a prominent role in the geometric complexity theory program of Mulmuley and Sohoni [MS01]. Roughly speaking, the goal of that program is to prove super-polynomial lower bounds on the border complexity of the permanent using techniques from algebraic geometry and representation theory.

In general, the relationship between exact and border complexity is not well-understood. Forbes [For16] (see also Bläser, Dörfler, and Ikenmeyer [BDI21]) observed that exact and border complexity are equivalent for read-one oblivious algebraic branching programs. Dutta, Dwivedi, and Saxena

[DDS21a] recently showed that polynomials in the border of depth-three circuits of bounded top fan-in can be computed exactly by small algebraic branching programs. However, for classes like VP and VNP (the algebraic analogues of P and NP), it is not clear how they relate to their closure.

Returning to the complexity of ideals, if we are content to operate in the setting of border complexity, then the work of Bürgisser [Bür04] shows that up to polynomial factors, the complexity of a principal ideal $\langle f \rangle$ is governed by the border complexity of its generator f. Unfortunately, this seems to be where our understanding of the complexity of ideals stops. Even ideals generated by two polynomials are not well-understood structurally from the viewpoint of complexity theory. There are examples of explicit ideals, coming from polynomial identity testing, that are not principal and for which we can prove lower bounds; see Subsection 1.2 below for more.

1.2 Polynomial Identity Testing

Polynomial identity testing (which we abbreviate as PIT) is the algorithmic problem of testing whether an algebraic circuit computes the zero polynomial. Typically, one assumes that the circuit computes a polynomial of degree at most $n^{O(1)}$, where n is the number of input variables. A simple coRP algorithm for this problem follows from the Schwartz–Zippel lemma [Zip79; Sch80]. When the input is allowed to be an algebraic circuit without further structural restrictions, no deterministic algorithm is known that improves on the naïve derandomization of this randomized algorithm. In fact, even obtaining a nondeterministic algorithm running in subexponential time is known to imply circuit lower bounds that lie beyond the reach of current techniques [KI04].

More is known for many restricted classes of circuits, including sparse polynomials [KS01], depth-three [DS07; KS07; KS09; KS11; SS11; SS12; SS13] and depth-four [Shp19; PS20; PS21; DDS21b] circuits of bounded top fan-in, read-once formulas [SV15; MV18], read-once oblivious algebraic branching programs [FS13; FSS14; AGKS15; GKS17; GKST17; AFSSV18; GG20; BS21], low-depth multilinear circuits [KMSV13; AvMV15; OSV16; SV18], and low-depth circuits [LST21]. In general, algorithms for PIT are designed by giving an efficient construction of a *hitting set generator*. That is, we construct a low-degree polynomial map $\mathcal{G} : \mathbb{F}^\ell \to \mathbb{F}^n$ with $\ell \ll n$ such that if $f(\overline{x})$ is a nonzero polynomial computable by a small circuit, then $f(\mathcal{G}(\overline{y})) \neq 0$. This reduces the number of variables in the circuit without increasing the degree too much. We then obtain a faster deterministic algorithm by using the brute-force derandomization of the Schwartz–Zippel lemma to test $f(\mathcal{G}(\overline{y}))$.

In fact, constructing such a generator $\mathcal{G}$ corresponds to proving lower bounds against a polynomial ideal. Fix a circuit class $\mathcal{C}$ (for example, the class of n^2-size circuits) and let $\mathcal{G}$ be a hitting set generator for $\mathcal{C}$. Let $\mathcal{G}(\overline{y}) = (\mathcal{G}_1(\overline{y}), \ldots, \mathcal{G}_n(\overline{y}))$ and consider the ideal of polynomials $f(\overline{x})$ that vanish on $\mathcal{G}(\overline{y})$, i.e., polynomials such that $f(\mathcal{G}(\overline{y})) = 0$. This ideal can be written as the intersection

$$I_{\mathcal{G}} \coloneqq \langle x_i - \mathcal{G}_i(\overline{y}) : i \in [n] \rangle \cap \mathbb{F}[\overline{x}],$$

and in general is not generated by a single polynomial. Suppose f is a nonzero polynomial in the ideal $I_{\mathcal{G}}$. Because we assumed $\mathcal{G}$ to be a hitting set generator for the circuit class $\mathcal{C}$, this means that f cannot be computed by circuits from $\mathcal{C}$. That is, proving that $\mathcal{G}$ is a generator for $\mathcal{C}$ is equivalent to proving that no element of $I_{\mathcal{G}}$ can be computed by a circuit from $\mathcal{C}$. To the best of our knowledge, this connection accounts for all known examples of lower bounds for non-principal ideals. We remark that this approach can prove lower bounds against "natural" non-principal ideals. For example, [FSTW16, Corollary 6.7] easily generalizes to prove lower bounds against determinantal ideals for weak circuit classes. However, this approach does not necessarily allow one to choose an ideal and subsequently prove a lower bound against that particular ideal.

One can also construct hitting set generators using lower bounds for ideals. Kabanets and Impagliazzo [KI04] used Kaltofen's factorization result to show that circuit lower bounds for explicit

families of polynomials can be used to derandomize PIT. In the analysis of the Kabanets–Impagliazzo generator, what is really needed is a lower bound for all low-degree multiples of a polynomial f, which is exactly what Kaltofen's theorem provides if f is assumed to be hard to compute. Further work on the algebraic hardness-randomness paradigm in the setting of low-depth circuits [DSY09; CKS19b] followed the approach of Kabanets and Impagliazzo [KI04], proving analogues of Kaltofen's factoring result for bounded-depth circuits.

One can also consider PIT for polynomials of small border complexity. Even in the randomized setting, the complexity of this problem is unclear, as it is not obvious how to evaluate a polynomial $f(\overline{x})$ given only a circuit that approximates $f(\overline{x})$, nor is it clear that such an approximating circuit even has a succinct description. However, one can still try to construct hitting set generators for polynomials of small border complexity. Forbes and Shpilka [FS18] and Guo, Saxena, and Sinhababu [GSS19] gave PSPACE constructions of hitting set generators for polynomials with small border circuit complexity. One of the primary conceptual contributions of Forbes and Shpilka [FS18] was the definition of a *robust* hitting set generator. Roughly, a generator $\mathcal{G}$ for a class $\mathcal{C}$ is robust if for every nonzero polynomial $f \in \mathcal{C}$, the composition $f(\mathcal{G}(\overline{y}))$ is "far" from the zero polynomial (after f has been suitably normalized). It is not hard to show that, over a field of characteristic zero, a generator $\mathcal{G}$ for $\mathcal{C}$ is robust if and only if $\mathcal{G}$ hits the closure $\overline{\mathcal{C}}$ of $\mathcal{C}$. Over an arbitrary field, one can likewise consider the problem of constructing hitting set generators for the closures of circuit classes, although the notion of $f(\mathcal{G}(\overline{y}))$ being far from the zero polynomial is not as clear. In this setting we drop the adjective "robust" and focus simply on hitting sets for the closure of a circuit class. The preceding discussion on the relationship between PIT and the complexity of ideals extends to border complexity.

Designing hitting sets for the closures of circuit classes has been explored as a possible avenue towards resolving grand challenges in polynomial identity testing. Recent work by Medini and Shpilka [MS21] and Saha and Thankey [ST21a] studied PIT for *orbits* of various classes $\mathcal{C}$. The orbit $\mathrm{orb}(\mathcal{C})$ of a class $\mathcal{C}$ corresponds to polynomials of the form $f(A\overline{x} + \overline{b})$, where $f(\overline{x}) \in \mathcal{C}$ and A is an invertible $n \times n$ matrix. Studying PIT for orbits is motivated by the fact that for many simple classes $\mathcal{C}$, there is a far richer class $\mathcal{D}$ such that $\overline{\mathrm{orb}(\mathcal{C})} = \overline{\mathcal{D}}$. That is, in order to derandomize PIT for a powerful class $\mathcal{D}$, it suffices to construct hitting set generators for the closure of the much simpler class $\mathrm{orb}(\mathcal{C})$. Unfortunately, this is not always feasible; for example, Medini and Shpilka [MS21] showed that at least one instantiation of their hitting sets does not extend to the closure of the circuit class it hits.

1.3 The Ideal Proof System

A central question of proof complexity is the following: given an unsatisfiable CNF formula φ, what is the length of the shortest proof of the unsatisfiability of φ? This question can be instantiated with a myriad of different proof systems rooted in logic, algebra, and geometry. Our focus in this work will be on a proof system based in algebra, namely the Ideal Proof System of Grochow and Pitassi [GP18]. For a more comprehensive treatment of other proof systems (and proof complexity in general), see the recent book of Krajíček [Kra19].

Let φ be an unsatisfiable 3CNF formula. One way to prove that φ is unsatisfiable is to translate φ into a system of polynomial equations, swapping the roles of 0 and 1, as follows. The literals x and $\neg x$ are translated into the polynomials $1 - x$ and x, respectively. A clause $\ell_1 \vee \ell_2 \vee \ell_3$ becomes the polynomial $p_{\ell_1} p_{\ell_2} p_{\ell_3}$, where p_{ℓ_i} is the polynomial corresponding to the literal ℓ_i. Let $f_1, \ldots, f_m$ be the polynomials obtained from the clauses of φ. It is not hard to see that φ is satisfiable if and only if there is a $\{0,1\}$-valued solution to the system of equations $f_1 = \cdots = f_m = 0$; equivalently, φ is satisfiable if and only if there is a solution to the system $f_1 = \cdots = f_m = x_1^2 - x_1 = \cdots = x_n^2 - x_n = 0$.

Thus, to show that φ is unsatisfiable, it suffices to prove that a system of polynomial equations is unsatisfiable. This can be done by finding polynomials $g_1(\overline{x}), \ldots, g_m(\overline{x})$ and $h_1(\overline{x}), \ldots, h_n(\overline{x})$ such that $\sum_{i=1}^{m} g_i(\overline{x}) f_i(\overline{x}) + \sum_{i=1}^{n} h_i(\overline{x})(x_i^2 - x_i) = 1$, or more succinctly, by showing that 1 is in the ideal generated by $\{f_1, \ldots, f_m, x_1^2 - x_1, \ldots, x_n^2 - x_n\}$. As a consequence of Hilbert's Nullstellensatz, such a refutation always exists, provided the system is unsatisfiable. These refutations and various notions of their complexity give rise to the Nullstellensatz [BIKPP96] and Polynomial Calculus [CEI96] proof systems, both of which are well-studied and for which lower bounds are known [BIKPP96; BIK+96; Raz98; IPS99].

The recent Ideal Proof System (abbreviated as IPS) of Grochow and Pitassi [GP18] measures the complexity of a refutation by the algebraic circuit complexity of the certificate $\sum_i g_i f_i + \sum_i h_i(x_i^2 - x_i)$ when the f_i and $x_i^2 - x_i$ are provided as part of the input to the circuit. Because a refutation in the IPS is written as an algebraic circuit, there are connections between algebraic circuit lower bounds and lower bounds for the IPS. Grochow and Pitassi [GP18] proved that super-polynomial lower bounds on the size of IPS refutations of a family of CNF formulas imply $\mathsf{VP} \neq \mathsf{VNP}$. As a proof system, the IPS is very powerful: Grochow and Pitassi [GP18] showed that the IPS polynomially simulates Extended Frege, itself a strong logic-based proof system. This simulation also behaves nicely if we consider IPS refutations coming from a restricted circuit class $\mathcal{C}$. For example, over a field of characteristic $p > 0$, the constant-depth version of the IPS polynomially simulates $\mathsf{AC}^0[p]$-Frege, a proof system notorious for its current lack of super-polynomial lower bounds.

Lower bounds, both conditional and unconditional, are known for the IPS. Conditionally, Alekseev, Grigoriev, Hirsch, and Tzameret [AGHT20] showed that the Shub–Smale hypothesis implies super-polynomial lower bounds on the size of IPS refutations of a particular instance of subset sum. Later work by Santhanam and Tzameret [ST21b] showed that over finite fields, if there is an explicit family of polynomials that cannot be computed by polynomial-size algebraic circuits, then a particular family of CNF formulas cannot be refuted by polynomial-size IPS refutations. Combined with earlier work by Grochow and Pitassi [GP18], this establishes that over finite fields, proving super-polynomial lower bounds for the IPS is equivalent to proving super-polynomial lower bounds for algebraic circuits. Forbes, Shpilka, Tzameret, and Wigderson [FSTW16] used techniques from algebraic circuit complexity to prove unconditional lower bounds for restricted subsystems of the IPS, including those computed by depth-three powering formulas, read-once algebraic branching programs, and multilinear formulas.

The Ideal Proof System is defined in terms of algebraic circuits, so it is natural to expect progress on IPS lower bounds to mirror progress on lower bounds for algebraic circuits. Empirically, this has been the case, although additional effort is required to translate circuit lower bounds into IPS lower bounds. To prove circuit lower bounds, one only needs to show that a single polynomial cannot be computed by small circuits. In contrast, to prove lower bounds on the circuit size of IPS refutations of a system of polynomials, it is necessary to show that small circuits cannot compute any valid refutation.

Luckily, the set of IPS refutations of a fixed system of equations exhibits some algebraic structure: all refutations of a fixed system of polynomials lie in a coset of a particular ideal, as observed by Grochow and Pitassi [GP18, Section 6]. Thus, one can try to prove lower bounds for the IPS by proving circuit lower bounds for nonzero cosets of ideals. To the best of our knowledge, the only known lower bounds for nonzero cosets of ideals are those that follow from previously-mentioned lower bounds on the IPS. Notably, these proofs do not directly establish lower bounds for cosets of ideal, but rather reduce the task of proving IPS lower bounds to the more-tractable task of proving algebraic circuit lower bounds. One could hope that by better understanding the complexity of (cosets of) ideals, this progress could be used to prove lower bounds for IPS and restricted variants thereof. We refer the interested reader to Grochow and Pitassi [GP18] and Grochow [Gro20] for

further details.

For more on the Ideal Proof System, see the recent survey of Pitassi and Tzameret [PT16].

1.4 Our Results

We now describe our results in more detail. Throughout this subsection, we let X denote an $n \times m$ matrix of variables and $I^{\det}_{n,m,r} \subseteq \mathbb{F}[X]$ the ideal generated by the $r \times r$ minors of X. For simplicity, we state our results over fields of characteristic zero (such as the rational or complex numbers).

1.4.1 Complexity of Determinantal Ideals

Our main theorem constructs, for any nonzero polynomial $f(X) \in I^{\det}_{n,m,r}$, a small f-oracle circuit that approximately computes the $s \times s$ determinant for $s = \Theta(r^{1/3})$. This answers a question of Grochow [Gro20, Conjecture 6.3] in the setting of border complexity.

Theorem 1.1 (Informal version of Theorem 3.8 and Corollary 3.9)**.** *Let $\mathbb{F}$ be a field of characteristic zero. Let X be an $n \times m$ matrix of variables and let $I^{\det}_{n,m,r} \subseteq \mathbb{F}[X]$ be the ideal generated by the $r \times r$ minors of X. Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial. Then there is a depth-three f-oracle circuit of size $O(n^2m^2)$ that approximately computes the $s \times s$ determinant for $s = \Theta(r^{1/3})$.*

More generally, the conclusion of Theorem 1.1 holds if the determinant is replaced by any polynomial g that can be approximately computed by an algebraic branching program with r vertices. The conclusion of Theorem 1.1 also holds if we have oracle gates that approximately compute f instead of oracles that compute f exactly.

An immediate consequence of Theorem 1.1 is that for formulas and low-depth circuits, the border complexity of any nonzero polynomial in $I^{\det}_{n,m,r}$ is at least as large as the border complexity of the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant, up to polynomial factors. To the best of our knowledge, the only complexity lower bounds for the ideal $I^{\det}_{n,m,r}$ known prior to this work are due to Wiersig [Wie20] and Forbes, Shpilka, Tzameret, and Wigderson [FSTW16, Corollary 6.7], who showed that every nonzero polynomial in $I^{\det}_{n,m,r}$ is $\exp(\Omega(r))$-hard for several weak circuit classes.

To prove Theorem 1.1, we have to reason about arbitrary polynomials in $I^{\det}_{n,m,r}$. That is, if $\{g_1, \ldots, g_N\}$ are the $r \times r$ minors of X, we have to consider all nonzero polynomials of the form $\sum_{i=1}^{N} f_i g_i$, where the f_i are arbitrary polynomials. This is difficult in part because if we apply a linear change of variables $X \mapsto L(X)$, it is not clear how to control the behavior of the f_i. To circumvent this, we use an alternate basis for $\mathbb{F}[X]$ instead of the monomial basis. This alternate basis consists of products of minors (of possibly different sizes) of X that satisfy a particular combinatorial condition; these products are known as *standard bideterminants*. Working in this basis, we gain a better understanding of how the multiplicands f_i behave under a change of variables.

The proof of Theorem 1.1 then proceeds in two steps. First, we find a change of variables that takes a polynomial $f \in I^{\det}_{n,m,r}$ to an approximation (in the border complexity sense) of a standard bideterminant $h(X)$ in the support of f. The analysis of this step crucially relies on the use of the standard bideterminant basis and its properties, which we describe in Subsection 2.5. Because f lies in the ideal $I^{\det}_{n,m,r}$, one can show that $h(X)$ is divisible by a $t \times t$ minor of X for some $t \geqslant r$. The second step is to find a projection of $h(X)$ to the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant. Since h may be a product of minors of varying sizes, we need to find a projection that (1) behaves nicely on small minors of X and (2) allows us to deal with the possibility that h may be a large power of a minor. We accomplish this by modifying an argument of Valiant [Val79].

1.4.2 Complexity of Pfaffian Ideals

Let Y be a $2n \times 2n$ skew-symmetric matrix. It is well-known that the determinant of Y is the square of another polynomial, the *Pfaffian* $\mathrm{Pf}(Y)$ of Y. Let $I^{\mathrm{pfaff}}_{2n,2n} \subseteq \mathbb{F}[Y]$ be the ideal generated by the Pfaffians of the $2r \times 2r$ principal submatrices of Y. Our next result is an analogue of Theorem 1.1 for the ideal $I^{\mathrm{pfaff}}_{2n,2r}$.

Theorem 1.2 (Informal version of Theorem 4.4 and Corollary 4.5)**.** *Let $\mathbb{F}$ be a field of characteristic zero. Let Y be a $2n \times 2n$ skew-symmetric matrix of variables and let $I^{\mathrm{pfaff}}_{2n,2r} \subseteq \mathbb{F}[Y]$ be the ideal generated by the Pfaffians of the $2r \times 2r$ principal submatrices of Y. Let $f(Y) \in I^{\mathrm{pfaff}}_{2n,2r}$ be a nonzero polynomial. Then there is a depth-three f-oracle circuit of size $O(n^4)$ that approximately computes the $s \times s$ Pfaffian for $s = \Theta(r^{1/3})$.*

The proof of Theorem 1.2 is similar to that of Theorem 1.1. The primary difference is that we now express polynomials in $I^{\mathrm{pfaff}}_{2n,2r}$ in an alternate basis consisting of products of Pfaffians of principal submatrices of Y. Along the way, we modify some of the technical details of the construction to accommodate for Pfaffians instead of determinants.

We remark that because the Pfaffian is the square root of the skew-symmetric determinant (in the sense that $\mathrm{Pf}(Y)^2 = \det(Y)$), it is natural to attempt proving Theorem 1.2 using Theorem 1.1. For any polynomial $f(\overline{x})$, one can use the Taylor series expansion of $\sqrt{1+x^2}$ to construct a small $f(\overline{x})^2$-oracle circuit that computes $f(\overline{x})$. Combining this with Theorem 1.1, one obtains an analogue of Theorem 1.1 for the ideal generated by the squares of sub-Pfaffians of Y, which is weaker than Theorem 1.2 above.

1.4.3 The Space of Partial Derivatives in Determinantal Ideals

The remainder of our work consists of three applications of Theorem 1.1 and its proof, the first of which is to algebraic circuit complexity. For a polynomial $f \in \mathbb{F}[X]$, let $\partial_{<\infty}(f)$ denote the span of the partial (Hasse) derivatives of f. The dimension of $\partial_{<\infty}(f)$ and related spaces has been used successfully as a complexity measure in proving lower bounds for restricted classes of algebraic circuits (see the survey of Saptharishi [Sap19] for more on this). While Theorem 1.1 shows that computing a polynomial in $I^{\mathrm{det}}_{n,m,r}$ is not much harder than computing the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant, it is natural to ask if there are polynomials in $I^{\mathrm{det}}_{n,m,r}$ that are "simpler" than the $r \times r$ determinant with respect to complexity measures like $\dim(\partial_{<\infty}(\bullet))$. Our next result shows that among nonzero polynomials in the ideal $I^{\mathrm{det}}_{n,m,r}$, the $r \times r$ determinant in fact minimizes the value of $\dim(\partial_{<\infty}(\bullet))$.

Theorem 1.3 (Informal version of Theorem 5.4)**.** *For every nonzero $f(X) \in I^{\mathrm{det}}_{n,m,r}$, we have* $\dim(\partial_{<\infty}(f)) \geqslant \dim(\partial_{<\infty}(\det_r)) = \binom{2r}{r}$.

Using tools developed in the proof of Theorem 1.1, we can easily reduce the task of proving Theorem 1.3 to the case where $f(X)$ is a product of minors of X. As f is in the ideal $I^{\mathrm{det}}_{n,m,r}$, at least one factor of f must be an $s \times s$ minor of X for some $s \geqslant r$. We can then directly bound $\dim(\partial_{<\infty}(f))$ from below by a slight generalization of the argument used to bound $\dim(\partial_{<\infty}(\det_s))$.

We note that one can easily prove a lower bound of $\dim(\partial_{<\infty}(f)) \geqslant 2^r$ using observations due to Forbes, Shpilka, Tzameret, and Wigderson [FSTW16] (see Section 5 for details). Our result improves on this, obtaining an optimal bound of $\binom{2r}{r} = \Theta(4^r/\sqrt{r})$.

1.4.4 Polynomial Identity Testing for Low-Depth Circuits and Formulas

Next, we use Theorem 1.1 to derandomize special cases of polynomial identity testing. It is a straightforward consequence of Theorem 1.1 that for circuit classes like low-depth circuits and formulas, computing any nonzero element of $I^{\det}_{n,m,r}$ is effectively as hard as computing the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant. Over an algebraically closed field, the ideal $I^{\det}_{n,m,r}$ can be equivalently described as the ideal of polynomials that vanish on matrices of rank less than r. Using this alternate description, we construct hitting set generators that unconditionally hit the closure of small low-depth circuits and conditionally hit the closure of small formulas.

Theorem 1.4 (Informal version of Theorem 6.8 and Theorem 7.3)**.** *Let $\mathbb{F}$ be a field of characteristic zero. For every $k \in \mathbb{N}$, there is a hitting set generator $\mathcal{G}_k$ with seed length $n^{1/2^k+o(1)}$ and degree 2^k that hits the closure of polynomial-size low-depth algebraic circuits. The generator $\mathcal{G}_k$ can be computed by either (1) a circuit of product-depth k and size $n^{1+o(1)}$, or (2) a formula of size $n^{1+o(1)}$. Assuming the border formula complexity of the determinant is super-polynomial, the generator $\mathcal{G}_k$ is also a hitting set generator for the closure of polynomial-size algebraic formulas.*

Our hitting set generators are very simple to describe. For $k = 1$, our generator takes as input two matrices of variables Y and Z, where Y is a $\sqrt{n} \times n^{o(1)}$ matrix and Z is an $n^{o(1)} \times \sqrt{n}$ matrix, and outputs the product YZ. For $k \geqslant 2$, we construct the generator $\mathcal{G}_k$ by arranging the input variables of $\mathcal{G}_{k-1}$ into a square matrix and replacing them with the product of an $n^{1/2^k+o(1)} \times n^{o(1)}$ matrix and an $n^{o(1)} \times n^{1/2^k+o(1)}$ matrix.

To prove that our generators correctly hit polynomial-size low-depth circuits, we must show that every small low-depth circuit does not vanish on the output of our generator. Using the description of $I^{\det}_{n,m,r}$ as the ideal of polynomials vanishing on matrices of rank at most r, establishing the correctness of our generators equates to proving that no small low-depth circuit can compute a polynomial in the ideal $I^{\det}_{\sqrt{n},\sqrt{n},n^{o(1)}}$. Such a lower bound follows in a straightforward manner by combining our Theorem 1.1 with the recent breakthrough lower bounds of Limaye, Srinivasan, and Tavenas [LST21].

In the regime of $n^{\Theta(1)}$ seed length, our generators attain a near-optimal tradeoff between seed length and degree. It is not hard to show that a generator of seed length $n^{1/2^k+o(1)}$ must be of degree at least 2^k, and conversely that any generator of degree 2^k must have seed length at least $\Omega(n^{1/2^k})$ (see Lemma 2.6). We also note that the circuit complexity of our generators is near-optimal, as any function with n outputs necessarily requires size $\Omega(n)$ to compute.

Prior to this, the best-known hitting set generator for low-depth circuits was given by Limaye, Srinivasan, and Tavenas [LST21], using the hardness-randomness results of Chou, Kumar, and Solomon [CKS19b]. They obtained, for all fixed $\varepsilon > 0$, a generator with seed length $O(n^{\varepsilon})$ and degree $O(\log n / \log\log n)$. Our construction attains the same seed length, but improves on the degree (as remarked above) and the circuit complexity of the generator. When instantiated to hit circuits of size s, the generator of Limaye, Srinivasan, and Tavenas [LST21] necessarily has circuit complexity $\Omega(s)$. In contrast, our generator can be computed by a constant-depth circuit or formula of size $n^{1+o(1)}$, even when hitting low-depth circuits of size $O(n^{10^{100}})$.

For formulas, the best-known (conditional) constructions of hitting set generators prior to our work are due to Dvir, Shpilka, and Yehudayoff [DSY09] and Chou, Kumar, and Solomon [CKS19b]. Both works yield generators with parameters similar to the low-depth generator of Limaye, Srinivasan, and Tavenas [LST21] mentioned above (although the generator of [DSY09] can only hit formulas of small individual degree). While our construction has better parameters, we use a stronger hardness assumption than what is needed by prior work. The constructions of Dvir, Shpilka, and Yehudayoff [DSY09] and Chou, Kumar, and Solomon [CKS19b] can be instantiated with any explicit family of polynomials that requires formulas of super-polynomial size. In contrast, our construction depends

crucially on super-polynomial lower bounds on the border formula complexity of the determinant. This is a stronger assumption, as the determinant is computable by polynomial-size branching programs and circuits, a fact which likely does not hold for all explicit families of polynomials.

1.4.5 Lower Bounds for the Ideal Proof System

Finally, we use Theorem 1.1 to prove lower bounds for the Ideal Proof System. Let X and Y be $n \times n$ matrices of variables and let I_n be the $n \times n$ identity matrix. Consider the system of polynomial equations given by $\{\det_n(X) = 0, XY - I_n = 0\}$. This system is unsatisfiable, as $\det_n(X) = 0$ if and only if X is non-invertible, while $XY - I_n = 0$ implies that X is invertible with inverse Y. We show that the constant-depth version of the Ideal Proof System cannot efficiently refute this system. Assuming lower bounds on the border formula complexity of the determinant, we also show that formula-IPS cannot efficiently refute this system. We remark that our lower bounds also hold when the boolean axioms $x_{i,j}^2 - x_{i,j} = 0$ are included in the system of equations, but we suppress these here for brevity.

Theorem 1.5 (Informal version of Corollary 8.2 and Theorem 8.4)**.** *Let $\mathbb{F}$ be a field of characteristic zero. Let X and Y be $n \times n$ matrices of variables and let I_n be the $n \times n$ identity matrix. Then any IPS refutation of the system $\{\det(X) = 0, XY - I_n = 0\}$ cannot be approximately computed by a constant-depth circuit of polynomial size. Assuming the border formula complexity of the determinant is super-polynomial, then any IPS refutation of this system cannot be approximately computed by a formula of polynomial size.*

We do this by following the approach of Forbes, Shpilka, Tzameret, and Wigderson [FSTW16], who showed that lower bounds for the IPS can be derived from circuit lower bounds for multiples of a polynomial. Our choice of the system $\{\det_n(X) = 0, XY - I_n = 0\}$ is motivated by the fact that, using the techniques of [FSTW16], the desired IPS lower bounds follow from circuit lower bounds for multiples of the determinant. We can obtain the necessary lower bounds by combining our Theorem 1.1 with lower bounds against the determinant. In the case of low-depth circuits, our IPS lower bounds are unconditional thanks to the recent breakthrough circuit lower bounds of Limaye, Srinivasan, and Tavenas [LST21]. For formula-IPS, our lower bounds remain conditional.

We also show that computing an IPS refutation of our hard instance $\{\det_n(X) = 0, XY - I_n = 0\}$ reduces to computing the determinant. Namely, we give a small depth-three circuit with $\det_n$-oracle gates that computes an IPS refutation of our hard instance. Passing to border complexity (using Lemma 2.3), this shows that the approximative complexity of the smallest IPS refutation of $\{\det_n(X) = 0, XY - I_n = 0\}$ is sandwiched between the approximative complexity of the $\Theta(n^{1/3}) \times \Theta(n^{1/3})$ and $n \times n$ determinants.

The strongest unconditional lower bounds for the IPS prior to our work are due to Forbes, Shpilka, Tzameret, and Wigderson [FSTW16], who proved lower bounds for subsystems of the IPS computed by restricted classes of circuits, including read-once oblivious algebraic branching programs and multilinear formulas. Impagliazzo, Mouli, and Pitassi [IMP20] showed that the constant-depth version of Polynomial Calculus (PC) over finite fields is surprisingly strong. The size of a constant-depth IPS refutation is essentially the number of lines in a constant-depth PC refutation, so lower bounds for constant-depth IPS over finite fields imply comparable lower bounds for constant-depth PC. However, our lower bounds do not extend to finite fields, nor do our lower bounds hold for refutations of an unsatisfiable CNF, so we are unable to conclude lower bounds for constant-depth PC and related proof systems.

We also mention a recent work of Alekseev [Ale21], who proved lower bounds on the bit-size of refutations in a version of PC augmented with an extension rule. This is somewhat incomparable

to our result: Alekseev's proof system allows for proofs of arbitrary depth, but must pay to use constants of large bit complexity; on the other hand, we work with a low-depth proof system that can use arbitrary rational numbers (or even arbitrary complex numbers) for free. Our lower bound is on circuit size, which is analogous to the number of lines in PC, whereas Alekseev's lower bound is on the number of bits needed to write down a refutation, which does not necessarily imply a lower bound on the number of proof lines.

2 Preliminaries

For a natural number $n \in \mathbb{N}$, we write $[n] \coloneqq \{1, 2, \ldots, n\}$. We use $\overline{x} = (x_1, \ldots, x_n)$ to denote a vector of variables and $X = (x_{i,j})_{i \in [n], j \in [m]}$ to denote a matrix of variables. For a matrix $A \in \mathbb{F}^{n \times m}$ and sets $R \subseteq [n]$, $C \subseteq [m]$, we denote by $A_{R,C}$ the submatrix of A whose rows and columns are taken from the sets R and C, respectively. A submatrix $A_{R,C}$ is *principal* if $R = C$. Given a polynomial $f(\overline{x}) \in \mathbb{F}[\overline{x}]$, it will often be useful to view the variables $\overline{x}$ as the entries of a matrix, typically of size $\lceil \sqrt{n} \rceil \times \lceil \sqrt{n} \rceil$. The precise way in which the variables $\overline{x}$ are arranged into a matrix will not matter, so we will perform this rearrangement implicitly without specifying the details. If X is an $n \times m$ matrix of variables, then for $r \leqslant \min(n, m)$ we denote by $I^{\det}_{n,m,r} \subseteq \mathbb{F}[X]$ the ideal of $\mathbb{F}[X]$ generated by the $r \times r$ minors of X.

We endow $\mathbb{F}[X]$ with a $(\mathbb{N}^n \oplus \mathbb{N}^m)$-grading in the following way. Let $\overline{e}_i \in \mathbb{N}^n$ denote the element of $\mathbb{N}^n$ with 1 in the i^{th} position and zeroes elsewhere. By abuse of notation, we also use $\overline{e}_i$ to denote the corresponding element of $\mathbb{N}^m$. We assign degree $\overline{e}_i \oplus \overline{e}_j$ to the variable $x_{i,j}$ and extend this to $\mathbb{F}[X]$ in the natural way. The degree of an element $f \in \mathbb{F}[X]$ with respect to this grading is called the *multidegree* of f, written $\operatorname{multideg}(f)$. We say an element of $\mathbb{F}[X]$ is *multihomogeneous* if it is homogeneous with respect to this grading.

Recall that given a field $\mathbb{F}$ and an indeterminate x, we write

- $\mathbb{F}[x]$ for the ring of polynomials in x with coefficients from $\mathbb{F}$,
- $\mathbb{F}(x)$ for the field of rational functions in x with $\mathbb{F}$-coefficients,
- $\mathbb{F}[\![x]\!]$ for the ring of formal power series in x over $\mathbb{F}$, and
- $\mathbb{F}((x))$ for the field of formal Laurent series in x over $\mathbb{F}$ (equivalently, the field of fractions of $\mathbb{F}[\![x]\!]$).

We assume familiarity with the basic notion of an algebraic circuit and restricted classes thereof, including formulas, branching programs, and bounded-depth circuits. The interested reader may consult the surveys of Shpilka and Yehudayoff [SY10] and Saptharishi [Sap19] or the text of Bürgisser, Clausen, and Shokrollahi [BCS97] for more on algebraic circuits.

2.1 Border Complexity

We now define border complexity, a modification of the standard notion of algebraic complexity.

Definition 2.1. Let $\mathbb{F}$ be any field and let ε be an indeterminate. Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$. We say that an algebraic circuit C *border computes* f if C is defined over $\mathbb{F}((\varepsilon))$ and computes a polynomial in $\mathbb{F}[\![\varepsilon]\!][\overline{x}]$ such that

$$C(\overline{x}) = f(\overline{x}) + \varepsilon g(\overline{x})$$

for some $g(\overline{x}) \in \mathbb{F}[\![\varepsilon]\!][\overline{x}]$. We abbreviate this as $C(\overline{x}) = f(\overline{x}) + O(\varepsilon)$. The *border complexity of* f is the size of the smallest circuit C that border computes f. $\lozenge$

If $\mathcal{C} \subseteq \mathbb{F}[\overline{x}]$ is a set of polynomials computed by some class of circuits, we denote by $\overline{\mathcal{C}} \subseteq \mathbb{F}[\overline{x}]$ the set of polynomials computed by the border of this same set of circuits. For example, VP denotes the class of n-variate polynomials that have $n^{O(1)}$ degree and can be computed by circuits of $n^{O(1)}$ size, while $\overline{\mathsf{VP}}$ denotes n-variate polynomials of degree $n^{O(1)}$ that can be border computed by circuits of $n^{O(1)}$ size.

Over fields of characteristic zero, one can interpret border complexity as a notion of approximate computation. In this case, if $C(\overline{x}) = f(\overline{x}) + O(\varepsilon)$, then $\lim_{\varepsilon \to 0} C(\overline{x}) = f(\overline{x})$, so C computes a polynomial that coefficient-wise approximates f arbitrarily well as ε goes to zero. Since the circuit C is defined over $\mathbb{F}((\varepsilon))$, it may be the case that C is not well-defined when $\varepsilon = 0$, as intermediate computations may involve division by ε. This prohibits setting $\varepsilon = 0$ in order to obtain a circuit that computes f exactly.

When the underlying field $\mathbb{F}$ has positive characteristic (for example, when $\mathbb{F}$ is finite), this notion of approximation breaks down. However, we can consider "approximate" computation in the symbolic sense defined above, which is still meaningful.

Alternatively, one can define border complexity using only the polynomial ring $\mathbb{F}[\varepsilon]$, avoiding the use of $\mathbb{F}[\![\varepsilon]\!]$ and $\mathbb{F}((\varepsilon))$. In this modified definition, we say that a circuit C border computes $f(\overline{x})$ if C is defined over $\mathbb{F}[\varepsilon]$ and there is a polynomial $g(\overline{x}) \in \mathbb{F}[\varepsilon][\overline{x}]$ and a natural number $q \in \mathbb{N}$ such that

$$C(\overline{x}) = \varepsilon^q f(\overline{x}) + \varepsilon^{q+1} g(\overline{x}).$$

We abbreviate this as $C(\overline{x}) = \varepsilon^q f(\overline{x}) + O(\varepsilon^{q+1})$. It turns out that these notions are equivalent, as one can translate between them by appropriately modifying the constants appearing in the circuit; see Bürgisser [Bür04, Lemma 5.6(1)] for a proof. (Note that the statement of [Bür04, Lemma 5.6(1)] only claims equivalence up to a factor of 2 in complexity. This arises due to the fact that the model of straight-line programs used in [Bür04] charges for scalar multiplications, whereas we allow multiplication by scalars for free.)

Given a set of polynomials $F := \{f_1, \ldots, f_k\} \subseteq \mathbb{F}[\overline{x}]$, one can also define the border complexity of F to be the size of the smallest multi-output circuit $C(\overline{x})$ over $\mathbb{F}((\varepsilon))$ such that C outputs $\{f_1 + O(\varepsilon), \ldots, f_k + O(\varepsilon)\}$. Naturally, one can also consider (single- or multi-output) border complexity with respect to subclasses of algebraic circuits, such as formulas, branching programs, or constant-depth circuits.

It will be useful to make the dependence of a polynomial on the approximation parameter ε explicit. In this case, we may write $f(\overline{x}, \varepsilon)$ for a polynomial in $\mathbb{F}[\![\varepsilon]\!][\overline{x}]$ or $\mathbb{F}[\varepsilon][\overline{x}]$, even though ε is regarded as an element of the underlying ring and is not a variable. This affords convenient notation for applying the map $\varepsilon \mapsto \varepsilon^N$ for some $N \in \mathbb{N}$ or the map $\delta \mapsto \varepsilon^N$ for a second indeterminate δ. We can use this to compose approximations as in the lemma below.

Lemma 2.2 ([Bür04, Lemma 2.3(1)])**.** *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$. Suppose*

1. *Φ is a circuit over $\mathbb{F}((\varepsilon))[\overline{x}]$ such that $\Phi(\overline{x}, \varepsilon) = f(\overline{x}) + O(\varepsilon) \in \mathbb{F}[\![\varepsilon]\!][\overline{x}]$, and*
2. *Ψ is a circuit over $\mathbb{F}((\delta))((\varepsilon))[\overline{x}]$ such that $\Psi(\overline{x}, \varepsilon, \delta) = \Phi(\overline{x}, \varepsilon) + O(\delta) \in \mathbb{F}[\![\delta]\!]((\varepsilon))[\overline{x}]$.*

Then there is some sufficiently large $N \in \mathbb{N}$ such that $\Psi(\overline{x}, \varepsilon, \varepsilon^N) = f(\overline{x}) + O(\varepsilon) \in \mathbb{F}[\![\varepsilon]\!][\overline{x}]$.

It is tempting to prove the preceding lemma by setting $\delta = \varepsilon$ and concluding that $\Psi(\overline{x}, \varepsilon, \varepsilon) = f(\overline{x}) + O(\varepsilon)$. This is incorrect, as the $O(\delta)$ error term in $\Psi(\overline{x}, \varepsilon, \delta)$ may involve division by ε, so setting $\delta = \varepsilon$ may introduce erroneous terms to the output of $\Psi(\overline{x}, \varepsilon, \delta)$. By setting $\delta = \varepsilon^N$ for sufficiently large $N \in \mathbb{N}$, this problem is avoided.

Let $f(\overline{x}), g(\overline{x}) \in \mathbb{F}[\overline{x}]$ be polynomials such that $f(\overline{x}) + O(\varepsilon)$ can be computed by a circuit with g-oracle gates. Suppose we want to replace the g-oracle gates with oracles that approximately

compute $g(\overline{x})$, i.e., oracle gates that compute some $h(\overline{x}, \delta) = g(\overline{x}) + O(\delta)$. As a consequence of the preceding lemma, we can obtain a circuit that computes $f(\overline{x}) + O(\varepsilon)$ by using $h(\overline{x}, \varepsilon^N)$-oracles for some sufficiently large N.

Lemma 2.3. *Let $f(\overline{x}), g(\overline{x}) \in \mathbb{F}[\overline{x}]$ be polynomials. Suppose $f(\overline{x}) + O(\varepsilon)$ can be computed by a circuit of size s with g-oracle gates. Let $h(\overline{x}, \delta) \in \mathbb{F}[\![\delta]\!][\overline{x}]$ be a polynomial such that $h(\overline{x}, \delta) = g(\overline{x}) + O(\delta)$. Then there is some $N \in \mathbb{N}$ such that $f(\overline{x}) + O(\varepsilon)$ can be computed by a circuit of size s with $h(\overline{x}, \varepsilon^N)$-oracle gates.*

Proof. Let $\Phi(\overline{x}, \varepsilon)$ be a g-oracle circuit that computes $f(\overline{x}) + O(\varepsilon)$ over $\mathbb{F}((\varepsilon))[\overline{x}]$. Let $\Psi(\overline{x}, \varepsilon, \delta)$ be the circuit over $\mathbb{F}((\delta))((\varepsilon))[\overline{x}]$ obtained by replacing each g-oracle gate with an $h(\overline{x}, \delta)$ oracle. Since $h(\overline{x}, \delta) = g(\overline{x}) + O(\delta)$, we have

$$\Psi(\overline{x}, \varepsilon, \delta) = \Phi(\overline{x}, \varepsilon) + O(\delta) \in \mathbb{F}[\![\delta]\!]((\varepsilon))[\overline{x}].$$

Applying Lemma 2.2 yields an $N \in \mathbb{N}$ such that $\Psi(\overline{x}, \varepsilon, \varepsilon^N) = f(\overline{x}) + O(\varepsilon)$ as desired. □

2.2 Polynomial Identity Testing

When designing deterministic algorithms for polynomial identity testing (PIT), our focus will be on the black-box regime, where we are given access to a circuit Φ through an evaluation oracle. Derandomizing PIT in this setting is equivalent to giving an explicit construction of a hitting set, defined below, for the set of polynomials computed by small circuits.

Definition 2.4. Let $\mathcal{C} \subseteq \mathbb{F}[\overline{x}]$ be a set of polynomials. A set $\mathcal{H} \subseteq \mathbb{F}^n$ is a *hitting set for $\mathcal{C}$* if for every nonzero $f \in \mathcal{C}$, there is some $\overline{\alpha} \in \mathcal{H}$ such that $f(\overline{\alpha}) \neq 0$. ◊

Alternatively, one can try to find an explicit, low-degree map $\mathcal{G} : \mathbb{F}^\ell \to \mathbb{F}^n$ with $\ell \ll n$ such that $f(\mathcal{G}(\overline{y})) \neq 0$ if f is a nonzero polynomial computed by a small circuit.

Definition 2.5. Let $\mathcal{C} \subseteq \mathbb{F}[\overline{x}]$ be a set of polynomials. A polynomial map $\mathcal{G} : \mathbb{F}^\ell \to \mathbb{F}^n$ is a *hitting set generator for $\mathcal{C}$* if for every nonzero $f \in \mathcal{C}$, we have $f(\mathcal{G}(\overline{y})) \neq 0$. We call ℓ the *seed length* of the generator. The *degree* of the generator, denoted by $\deg(\mathcal{G})$, is given by $\max_{i \in [n]} \deg(\mathcal{G}_i)$. ◊

Small hitting sets (and hitting set generators with small seed length and low degree) are known to exist non-constructively. In derandomizing PIT, one seeks efficient uniform constructions of these objects. One can show that the notions of hitting sets and generators are essentially equivalent using polynomial interpolation (see, e.g., Shpilka and Volkovich [SV15, Section 4]). In this work, we will prefer the language of generators, as they are more amenable to composition than are hitting sets.

It is natural to extend the definition of a hitting set to the setting of border complexity. Over fields of characteristic zero, Forbes and Shpilka [FS18] defined a notion of a *robust hitting set* for a class $\mathcal{C}$. Using continuity, one can easily show that if $\mathcal{H}$ is a robust hitting set for a class $\mathcal{C}$, then $\mathcal{H}$ is also a hitting set for the closure $\overline{\mathcal{C}}$. In this work, we will be concerned with hitting sets for the closures of circuit classes, but we will not pay particular attention to the robustness parameter, as some of our constructions take place in characteristic $p > 0$.

We note that a generator cannot simultaneously have very small seed length and very low degree. In particular, a generator of degree $\Theta(1)$ must have seed length $n^{\Theta(1)}$.

Lemma 2.6. *Let $\mathcal{C} \subseteq \mathbb{F}[\overline{x}]$ be a set of polynomials such that $\mathcal{C}$ contains all linear polynomials. Suppose $\mathcal{G} : \mathbb{F}^\ell \to \mathbb{F}^n$ is a hitting set generator for $\mathcal{C}$ of degree d. Then we must have $\binom{\ell+d}{d} \geqslant n$. In particular, if d is a fixed constant independent of n, then $\ell \geqslant \Omega(n^{1/d})$.*

Proof. For $i \in [n]$, let $\mathcal{G}_i(\overline{y})$ be the i^{th} coordinate of $\mathcal{G}$. Observe that each $\mathcal{G}_i(\overline{y})$ is a polynomial in ℓ variables of degree at most d. The space of ℓ-variate polynomials of degree at most d is a vector space of dimension $\binom{\ell+d}{d}$. Suppose for the sake contradiction that $\binom{\ell+d}{d} < n$. Then there is a non-trivial linear relation among the n coordinates of $\mathcal{G}$. That is, there is a linear polynomial $L(x_1, \ldots, x_n) \neq 0$ such that

$$L(\mathcal{G}_1(\overline{y}), \ldots, \mathcal{G}_n(\overline{y})) = 0.$$

Since $L(\overline{x})$ is linear, we have $L \in \mathcal{C}$. This contradicts the assumption that $\mathcal{G}$ is a hitting set generator for $\mathcal{C}$. □

2.3 Matrix Rank

We will frequently make use of the fact that the rank of a matrix can be characterized by the (non-)vanishing of its minors. This is a straightforward consequence of the fact that the row rank and column rank of a matrix coincide.

Lemma 2.7. *Let $A \in \mathbb{F}^{n \times m}$. Then $\operatorname{rank}(A) \geqslant r$ if and only if some $r \times r$ minor of A does not vanish. Equivalently, $\operatorname{rank}(A) < r$ if and only if every $r \times r$ minor of A vanishes.*

We now define the hitting set generator which will be the focus of our work on PIT.

Construction 2.8. *Let $n, m, r \in \mathbb{N}$ with $r \leqslant \min(n, m)$. Define the map $\mathcal{G}_{n,m,r} : \mathbb{F}^{n \times r} \times \mathbb{F}^{r \times m} \to \mathbb{F}^{n \times m}$ via*

$$\mathcal{G}_{n,m,r}(Y, Z)_{i,j} = (YZ)_{i,j}.$$

The following are immediate consequences of the definition of $\mathcal{G}_{n,m,r}(Y, Z)$.

Lemma 2.9. *Let $\mathcal{G}_{n,m,r} : \mathbb{F}^{n \times r} \times \mathbb{F}^{r \times m} \to \mathbb{F}^{n \times m}$ be defined as in Construction 2.8.*

1. *The image of $\mathcal{G}_{n,m,r}$ contains all $n \times m$ matrices of rank at most r.*
2. *Each coordinate of $\mathcal{G}_{n,m,r}(Y, Z)$ is a $2r$-sparse degree-2 polynomial in the variables $Y \cup Z$.*
3. *The map $\mathcal{G}_{n,m,r}(Y, Z)$ can be computed by a multi-output algebraic circuit of size $2nmr$ and product-depth 1. Additionally, each coordinate of the output can be computed by a homogeneous formula of size $2r$.*

In order to prove that $\mathcal{G}_{n,m,r}$ is a hitting set generator for a class of circuits $\mathcal{C}$, it will be useful to understand which polynomials vanish when composed with $\mathcal{G}_{n,m,r}$. If $f(X) \in \mathbb{F}[X]$ is a nonzero polynomial such that $f(\mathcal{G}_{n,m,r}(Y, Z)) = 0$, then f necessarily vanishes on all $n \times m$ matrices of rank at most r. The ideal of polynomials which vanish on matrices of rank at most r is well-understood from the viewpoint of mathematics.

Let $I^{\det}_{n,m,r}$ be the ideal generated by the $r \times r$ minors of a generic $n \times m$ matrix and let $J_{n,m,r}$ be the ideal of polynomials which vanish on all $n \times m$ matrices of rank at most r. It is clear that $I^{\det}_{n,m,r+1} \subseteq J_{n,m,r}$. When the field $\mathbb{F}$ is algebraically closed, we in fact have the equality $I^{\det}_{n,m,r+1} = J_{n,m,r}$. This follows from Hilbert's Nullstellensatz and the fact that $I^{\det}_{n,m,r}$ is radical (see, for example, [BV88, Theorem 2.10 and Remark 2.12]). This implies that if $f(X)$ is nonzero and $f(\mathcal{G}_{n,m,r}(Y, Z)) = 0$, then $f \in J_{n,m,r} = I^{\det}_{n,m,r+1}$.

In the case where $\mathbb{F}$ is not algebraically closed, we can still conclude that $f \in I^{\det}_{n,m,r+1}$ if $f(\mathcal{G}_{n,m,r}(Y, Z)) = 0$. This follows from the fact that if $f(\mathcal{G}_{n,m,r}(Y, Z)) = 0$, then f vanishes on matrices of rank at most r with entries in any extension $\mathbb{K} \supseteq \mathbb{F}$. In particular, f vanishes on matrices of rank at most r with entries in $\overline{\mathbb{F}}$, the algebraic closure of $\mathbb{F}$.

We record the preceding observations as a lemma.

Lemma 2.10. *Let $\mathbb{F}$ be any field and let $n, m, r \in \mathbb{N}$ with $r \leqslant \min(n, m)$. Let $I^{\det}_{n,m,r}$ denote the ideal of $\mathbb{F}[X]$ generated by the $r \times r$ minors of a generic $n \times m$ matrix and let $f(X) \in \mathbb{F}[X]$. Then $f(\mathcal{G}_{n,m,r-1}(Y, Z)) = 0$ if and only if $f(X) \in I^{\det}_{n,m,r}$.*

2.4 Hasse Derivatives

In this work, we use Hasse derivatives in place of the standard partial derivative. Originally defined by Hasse [Has36], Hasse derivatives are a notion of derivative that is more well-behaved over fields of small positive characteristic. For a more thorough treatment of Hasse derivatives and their properties, see, for example, the thesis of Forbes [For14, Appendix C].

Definition 2.11. Let $\mathbb{F}$ be a field and let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$. For $\overline{a} \in \mathbb{N}^n$, we define the $\overline{a}^{th}$ *Hasse derivative* of $f(\overline{x})$ to be

$$\frac{\partial}{\partial \overline{x}^{\overline{a}}}(f) \coloneqq \operatorname{Coeff}_{\overline{y}^{\overline{a}}}(f(\overline{x} + \overline{y})),$$

where $f(\overline{x} + \overline{y})$ is viewed as a polynomial in $\mathbb{F}[\overline{x}][\overline{y}]$. ◊

Equivalently, one can define Hasse derivatives in terms of their action on monomials.

Lemma 2.12. *Let $\overline{a}, \overline{b} \in \mathbb{N}^n$. Then*

$$\frac{\partial}{\partial \overline{x}^{\overline{a}}}(\overline{x}^{\overline{b}}) = \prod_{i=1}^{n} \binom{b_i}{a_i} x_i^{b_i - a_i},$$

where we use the convention that $\binom{b}{a} = 0$ if $b < a$.

A straightforward consequence of the preceding lemma is that Hasse derivatives interact nicely with degree.

Lemma 2.13. *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ and let $\overline{a} \in \mathbb{N}^n$. Then*

$$\deg\left(\frac{\partial}{\partial \overline{x}^{\overline{a}}}(f)\right) \leqslant \deg(f) - \|\overline{a}\|_1,$$

with equality if $\frac{\partial}{\partial \overline{x}^{\overline{a}}}(f) \neq 0$.

Hasse derivatives also respect the multigrading on $\mathbb{F}[X]$.

Lemma 2.14. *Let $f \in \mathbb{F}[X]$ be a multihomogeneous polynomial and let $A \in \mathbb{N}^{n \times m}$. Write $X^A \coloneqq \prod_{i=1}^{n} \prod_{j=1}^{m} x_{i,j}^{a_{i,j}}$ for the monomial with powers given by the matrix A. If $\frac{\partial f}{\partial X^A} \neq 0$, then*

$$\operatorname{multideg}\left(\frac{\partial f}{\partial X^A}\right) = \operatorname{multideg}(f) - \operatorname{multideg}(X^A).$$

Just like standard partial derivatives, Hasse derivatives commute with one another.

Lemma 2.15 (see, e.g., [For14, Lemma C.1.4(5)])**.** *Let $f \in \mathbb{F}[\overline{x}]$ and let $\overline{a}, \overline{b} \in \mathbb{N}^n$. Then*

$$\frac{\partial}{\partial \overline{x}^{\overline{a}}}\left(\frac{\partial}{\partial \overline{x}^{\overline{b}}}(f)\right) = \frac{\partial}{\partial \overline{x}^{\overline{b}}}\left(\frac{\partial}{\partial \overline{x}^{\overline{a}}}(f)\right).$$

Hasse derivatives obey a modified form of the product rule.

Lemma 2.16 (see, e.g., [For14, Lemma C.1.7]). *Let $f_1, \dots, f_m \in \mathbb{F}[\overline{x}]$. For any $i \in [n]$ and $a \in \mathbb{N}$, we have*

$$\frac{\partial}{\partial x_i^a}(f_1 \cdots f_m) = \sum_{a_1+\cdots+a_m=a} \frac{\partial}{\partial x_i^{a_1}}(f_1) \cdots \frac{\partial}{\partial x_i^{a_m}}(f_m).$$

We now define the space of (d^{th} order) partial derivatives of a polynomial. The dimension of this space (and related spaces, like the space of shifted partial derivatives [Kay12]) is a useful complexity measure within algebraic circuit complexity.

Definition 2.17. Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$. The *space of partial derivatives of f*, denoted $\partial_{<\infty}(f)$, is defined as

$$\partial_{<\infty}(f) \coloneqq \operatorname{span}_{\mathbb{F}}\left\{ \frac{\partial f}{\partial \overline{x}^{\overline{a}}} : \overline{a} \in \mathbb{N}^n \right\}.$$

The *space of d^{th}-order partial derivatives of f*, written $\partial_d(f)$, is given by

$$\partial_d(f) \coloneqq \operatorname{span}_{\mathbb{F}}\left\{ \frac{\partial f}{\partial \overline{x}^{\overline{a}}} : \overline{a} \in \mathbb{N}^n, \|\overline{a}\|_1 = d \right\}.$$

We also write

$$\partial_{\leqslant d}(f) \coloneqq \operatorname{span}_{\mathbb{F}} \bigcup_{i=0}^{d} \partial_i(f)$$

for the space of partial derivatives of order at most d. ◊

We will need the following lemma relating the dimension of the space of partial derivatives of a polynomial $f(\overline{x})$ and a linear projection $f(A\overline{x})$.

Lemma 2.18. *Let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ and let $A \in \mathbb{F}^{n\times n}$. Then for every $d \in \mathbb{N}$, we have* $\dim(\partial_{\leqslant d}(f(A\overline{x}))) \leqslant \dim(\partial_{\leqslant d}(f(\overline{x})))$. *In particular, if A is invertible, then* $\dim(\partial_{\leqslant d}(f(A\overline{x}))) = \dim(\partial_{\leqslant d}(f(\overline{x})))$.

Proof. Using the chain rule for Hasse derivatives, one can show (see, e.g., [For14, Corollary C.2.7]) that for all $\overline{e} \in \mathbb{N}^n$ with $\|\overline{e}\|_1 \leqslant d$, we have

$$\frac{\partial}{\partial \overline{x}^{\overline{e}}}(f(A\overline{x})) \in \operatorname{span}_{\mathbb{F}}\{g(A\overline{x}) : g(\overline{x}) \in \partial_{\leqslant d}(f(\overline{x}))\}.$$

Let $V \coloneqq \operatorname{span}_{\mathbb{F}}\{g(A\overline{x}) : g(\overline{x}) \in \partial_{\leqslant d}(f(\overline{x}))\}$. This implies

$$\partial_{\leqslant d}(f(A\overline{x})) \subseteq V,$$

so

$$\dim \partial_{\leqslant d}(f(A\overline{x})) \leqslant \dim V.$$

We now show that $\dim V$ bounded by $\dim \partial_{\leqslant d}(f(\overline{x}))$. Let $g_1(\overline{x}), \dots, g_k(\overline{x}) \in \partial_{\leqslant d}(f(\overline{x}))$ and suppose that $g_1(A\overline{x}), \dots, g_k(A\overline{x})$ are linearly independent. This implies that $g_1(\overline{x}), \dots, g_k(\overline{x})$ are linearly independent, as any linear relation satisfied by $g_1(\overline{x}), \dots, g_k(\overline{x})$ will also be satisfied by $g_1(A\overline{x}), \dots, g_k(A\overline{x})$. If we select the g_i such that $\{g_1(A\overline{x}), \dots, g_k(A\overline{x})\}$ forms a basis of V, then we have

$$\dim V = k \leqslant \dim \partial_{\leqslant d}(f(\overline{x})).$$

Combining this with the previous inequality completes the proof.

In the case where A is invertible, we use the fact that $\overline{x} = A^{-1}A\overline{x}$ to obtain

$$\dim \partial_{\leqslant d}(f(\overline{x})) \leqslant \dim \partial_{\leqslant d}(f(A\overline{x})) \leqslant \dim \partial_{\leqslant d}(f(\overline{x})),$$

so equality holds. □

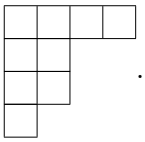
We note that by taking $d \geqslant \deg(f)$ in Lemma 2.18, one can replace $\partial_{\leqslant d}(\bullet)$ with $\partial_{<\infty}(\bullet)$.

2.5 Bideterminants and the Straightening Law

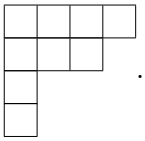
The proof of Theorem 3.8 relies on understanding how a polynomial $f \in I^{\det}_{n,m,r}$ behaves under the map $X \mapsto AXB$ for invertible matrices A and B. For example, it is easy to see that $f(AXB)$ also lies in $I^{\det}_{n,m,r}$. However, it is not clear if there is other structure we may take advantage of. By working in a different basis of $\mathbb{F}[X]$, we can better understand how $f(AXB)$ relates to $f(X)$. Before describing this basis, we recall the notions of a Young diagram and Young tableau.

Definition 2.19. A *partition* $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_k)$ is a non-increasing sequence of natural numbers. If $\sum_{i=1}^{k} \sigma_i = n$, we write $\sigma \vdash n$. The *transpose* of σ, denoted $\hat{\sigma}$, is the partition given by $\hat{\sigma}_i = |\{j : \sigma_j \geqslant i\}|$. Associated with a partition σ is its *Young diagram* $D_\sigma \subseteq \mathbb{N} \times \mathbb{N}$, given by $D_\sigma = \{(i, j) : j \leqslant \sigma_i\}$. $\Diamond$

Note that $\hat{\sigma}_1$ counts the number of rows in the Young diagram of σ. We graphically depict the Young diagram of a partition as a collection of boxes. For example, the Young diagram of the partition $(4, 2, 2, 1)$ is



This partition has transpose $(4, 3, 1, 1)$, with Young diagram given by



The lexicographic ordering on integer sequences induces an ordering on partitions, which we denote by $<_{\text{lex}}$.

We now define Young tableaux, which can be obtained by writing a number in each cell of the Young diagram of some partition σ.

Definition 2.20. Given a partition σ, a *Young tableau* T *of shape* σ is a map $T : D_\sigma \to \mathbb{N}$ assigning a natural number to each cell of the Young diagram of σ. We denote the i^{th} row of T by $T(i, \bullet)$, which we will view as either a set or a one-row Young tableau depending on context. A Young tableau is *standard* if its entries are strictly increasing along each column and along each row. A Young tableau is *semistandard* if its entries are strictly increasing along each column and are nondecreasing along each row. If $T : D_\sigma \to \mathbb{N}$ is a Young tableau, its *conjugate tableau* $\hat{T} : D_{\hat{\sigma}} \to \mathbb{N}$ is given by $\hat{T}(i, j) = T(j, i)$. $\Diamond$

Continuing the example above, one Young tableau (of many) of shape $(4, 2, 2, 1)$ is given by

1	2	4	3
1	2		
4	1		
3			

.

Next, we introduce bitableaux and bideterminants. A bitableau is simply a pair of Young tableau of the same shape, while a bideterminant is a natural polynomial associated to this pair of tableaux.

Definition 2.21. Let $X = (x_{1,1}, \ldots, x_{n,n})$ be an $n \times n$ matrix of variables. A *bitableau* (S, T) is a pair of Young tableaux of the same shape σ. If the entries of S and T are from $[n]$, we associate to (S, T) the *bideterminant* $(S|T)(X)$, defined as

$$(S|T)(X) \coloneqq \prod_{i=1}^{\hat{\sigma}_1} \det \begin{pmatrix} x_{S(i,1),T(i,1)} & x_{S(i,1),T(i,2)} & \cdots & x_{S(i,1),T(i,\sigma_i)} \\ x_{S(i,2),T(i,1)} & x_{S(i,2),T(i,2)} & \cdots & x_{S(i,2),T(i,\sigma_i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{S(i,\sigma_i),T(i,1)} & x_{S(i,\sigma_i),T(i,2)} & \cdots & x_{S(i,\sigma_i),T(i,\sigma_i)} \end{pmatrix}.$$

The i^{th} term in this product is the determinant of the submatrix whose rows and columns are listed in the i^{th} row of the tableaux S and T, respectively. The *width* of the bideterminant $(S|T)$ is given by σ_1. We say that the bitableau (S, T) and bideterminant $(S|T)$ are *standard* if, as tableaux, both S and T are increasing along each row and nondecreasing along each column (equivalently, that S and T are both the transpose of a semistandard Young tableau). ◊

For example, associated to the bitableau

$$\left(\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 1 & 3 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array} \;,\; \begin{array}{|c|c|c|} \hline 1 & 3 & 4 \\ \hline 2 & 4 \\ \cline{1-2} 3 \\ \cline{1-1} \end{array} \right)$$

is the bideterminant

$$\det \begin{pmatrix} x_{1,1} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,3} & x_{3,4} \end{pmatrix} \det \begin{pmatrix} x_{1,2} & x_{1,4} \\ x_{3,2} & x_{3,4} \end{pmatrix} \det\big(x_{4,3}\big).$$

Note that a bideterminant $(S|T)$ is multihomogeneous of degree $(s_1\overline{e}_1 + \cdots + s_n\overline{e}_n) \oplus (t_1\overline{e}_1 + \cdots + t_n\overline{e}_n)$, where s_i and t_i count the number of occurrences of i in S and T, respectively.

It is easy to see that the bideterminants span $\mathbb{F}[X]$, since a monomial $\prod_{i=1}^{d} x_{r_i,c_i}$ is the bideterminant corresponding to the bitableau

$$\left(\begin{array}{|c|} \hline r_1 \\ \hline r_2 \\ \hline \ldots \\ \hline r_d \\ \hline \end{array} \;,\; \begin{array}{|c|} \hline c_1 \\ \hline c_2 \\ \hline \ldots \\ \hline c_d \\ \hline \end{array} \right).$$

Perhaps surprisingly, there is a natural subset of the bideterminants which form a basis of $\mathbb{F}[X]$.

Theorem 2.22 ([DRS74])**.** *The standard bideterminants form a basis of* $\mathbb{F}[X]$.

To show $\mathbb{F}[X]$ is spanned by standard bideterminants, it suffices to express non-standard bideterminants as linear combinations of standard bideterminants. The fact that this can be done, along with some additional structural information, is known as the straightening law. For more on the straightening law, including its history and its applications to invariant theory, see the introduction of Désarménien, Kung, and Rota [DKR78].

Theorem 2.23 ([DRS74], see also [DKR78; dCEP80])**.** *Let* $(S|T)(X)$ *be a bideterminant of shape* σ*. Then* $(S|T)(X)$ *can be expressed as a linear combination*

$$(S|T)(X) = \sum_{(A,B)} c_{A,B}(A|B)(X),$$

where the $c_{A,B}$ *are integers and the sum ranges over all standard bitableaux* (A, B) *of shape* τ *such that* $\tau \geqslant_{\text{lex}} \sigma$*.*

One immediate corollary of this is a characterization of polynomials in the ideal $I^{\det}_{n,m,r}$ by their support in the standard bideterminant basis.

Corollary 2.24. *A polynomial $f \in \mathbb{F}[X]$ is an element of the ideal $I^{\det}_{n,m,r}$ if and only if f is supported on bideterminants of width at least r.*

2.6 Pfaffians

This subsection departs slightly from the setting of the previous subsections. Let X be a $2n \times 2n$ skew-symmetric matrix of variables. That is, the (i,j) entry of X is the variable $x_{i,j}$ and the variables $x_{i,j}$ and $x_{j,i}$ satisfy the relation $x_{i,j} = -x_{j,i}$. It is well-known that the determinant of X is the square of a polynomial; this square root of the determinant is the *Pfaffian* of X. Formally, one can define the Pfaffian $\mathrm{Pf}(X)$ as

$$\mathrm{Pf}(X) = \frac{1}{2^n n!} \sum_{\sigma \in S_{2n}} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} x_{\sigma(2i-1),\sigma(2i)},$$

where S_{2n} is the group of all permutations on $[2n] = \{1, \ldots, 2n\}$. Each monomial in the above sum appears $2^n n!$ times, so every monomial in the support of the Pfaffian has a coefficient of 1 or -1. In particular, the Pfaffian is well-defined even over fields of small characteristic.

As remarked above, we have $\mathrm{Pf}(X)^2 = \det(X)$ when X is a skew-symmetric matrix. If X is an $m \times m$ skew-symmetric matrix for odd m, then $\det(X) = 0$, so we restrict our attention to matrices of even order. The equation $\mathrm{Pf}(X)^2 = \det(X)$ relates the Pfaffian and determinant of a skew-symmetric matrix. For general matrices, we can relate Pfaffians and determinants via the following lemma.

Lemma 2.25. *Let A be a $2n \times 2n$ skew-symmetric matrix and let B be an arbitrary $2n \times 2n$ matrix. Then $BAB^\top$ is skew-symmetric and $\mathrm{Pf}(BAB^\top) = \det(B)\,\mathrm{Pf}(A)$.*

We will also make use of the symmetries of the Pfaffian as described in the next lemma.

Lemma 2.26. *Let A be an $n \times n$ matrix. Then*

$$\mathrm{Pf}\begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix} = (-1)^{\binom{n}{2}} \det(A).$$

As with determinants, one can consider the ideal generated by sub-Pfaffians of the same size of a skew-symmetric matrix. To ensure that the Pfaffian of a submatrix of X is well-defined, we restrict our attention to *principal submatrices*. Recall that a submatrix $X_{R,C}$ of X is principal if $R = C$. If X is skew-symmetric, then so is any principal submatrix of X. Throughout this work, we will use $I^{\mathrm{pfaff}}_{2n,2r}$ to denote the ideal of $\mathbb{F}[X]$ generated by the Pfaffians of the $2r \times 2r$ principal submatrices of X.

Much like the case with determinants, one can understand the ideal $I^{\mathrm{pfaff}}_{2n,2r}$ using an analogous straightening law for Pfaffians. To do this, we begin by defining the analogues of standard bideterminants for Pfaffian ideals.

Definition 2.27. Let T be a conjugate semistandard Young tableau of shape σ such that every row of T has even length. We associate to T the *standard monomial* $[T](X)$, which is a polynomial defined as the product of Pfaffians

$$[T](X) \coloneqq \prod_{i=1}^{\hat{\sigma}_1} \mathrm{Pf}\begin{pmatrix} x_{T(i,1),T(i,1)} & x_{T(i,1),T(i,2)} & \cdots & x_{T(i,1),T(i,\sigma_i)} \\ x_{T(i,2),T(i,1)} & x_{T(i,2),T(i,2)} & \cdots & x_{T(i,2),T(i,\sigma_i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T(i,\sigma_i),T(i,1)} & x_{T(i,\sigma_i),T(i,2)} & \cdots & x_{T(i,\sigma_i),T(i,\sigma_i)} \end{pmatrix}.$$

That is, the i^{th} polynomial in the above product is the Pfaffian of the submatrix of X whose rows and columns are listed in the i^{th} row of the tableau T. The *width* of $[T](X)$ is σ_1, the size of the largest Pfaffian in the above product. $\lozenge$

If we were to extend the above definition to all Young tableaux, it is clear that the resulting set of polynomials spans $\mathbb{F}[X]$, since

$$\left[\boxed{i}\,\boxed{j}\right](X) = \operatorname{Pf}\begin{pmatrix} 0 & x_{i,j} \\ -x_{i,j} & 0 \end{pmatrix} = x_{i,j}.$$

However, we do not lose much by ignoring these non-standard monomials. In a manner analogous to the determinantal case, de Concini and Procesi [dCP76] proved that the standard monomials form a basis of $\mathbb{F}[X]$.

Theorem 2.28 ([dCP76, Theorem 6.5])**.** *For any commutative ring R with unity, the standard monomials form a basis of $R[X]$.*

To prove this, de Concini and Procesi [dCP76] showed that the standard monomials span $R[X]$ and that any non-standard monomial can be written as a linear combination of standard monomials. The expression of a non-standard monomial as a linear combination of standard monomials is, as in the determinantal case, known as the *straightening law*. Using the straightening law of de Concini and Procesi [dCP76, Lemmas 6.1 and 6.2], one can show (following Doubilet, Rota, and Stein [DRS74, Section 8]) that a non-standard monomial of width $2r$ is supported only on standard monomials of width at least $2r$. A straightforward corollary of this is that every polynomial in the ideal generated by the Pfaffians of the principal $2r \times 2r$ submatrices of a matrix X is supported on standard monomials of width at least $2r$.

Corollary 2.29. *Let X be a generic $2n \times 2n$ skew-symmetric matrix. Let $I^{\text{pfaff}}_{2n,2r}$ be the ideal generated by the Pfaffians of the $2r \times 2r$ principal submatrices of X. Then any $f \in I^{\text{pfaff}}_{2n,2r}$ is supported on standard monomials of width at least $2r$.*

2.7 Monomial Orders

Our use of border complexity stems from the need to construct circuits that compute only a particular subset of the monomials appearing in the support of a polynomial f. To do this, we make use of monomial orders and leading monomials, which we now define.

Definition 2.30. A *monomial order* $\prec$ is a total order on the monomials of $\mathbb{F}[\overline{x}]$ which satisfies

1. $1 \prec \overline{x}^{\overline{a}}$ for all nonzero $\overline{a} \in \mathbb{N}^n$, and
2. if $\overline{x}^{\overline{a}} \prec \overline{x}^{\overline{b}}$, then $\overline{x}^{\overline{a}+\overline{c}} \prec \overline{x}^{\overline{b}+\overline{c}}$ for all $\overline{a}, \overline{b}, \overline{c} \in \mathbb{N}^n$. $\lozenge$

Definition 2.31. Let $\prec$ be a monomial order and let $f(\overline{x}) \in \mathbb{F}[\overline{x}]$ be a nonzero polynomial. The *leading monomial of f with respect to $\prec$*, written $\operatorname{LM}_{\prec}(f)$, is the $\prec$-maximal monomial appearing in the support of f. The *leading coefficient of f with respect to $\prec$*, denoted $\operatorname{LC}_{\prec}(f)$, is the coefficient of $\operatorname{LM}_{\prec}(f)$ when f is written as a sum of monomials. $\lozenge$

We may write $\operatorname{LM}(f)$ and $\operatorname{LC}(f)$ for the leading monomial and coefficient of f, respectively, if the order $\prec$ is clear from context. A useful property of leading monomials is that taking the leading monomial commutes with products of polynomials.

Lemma 2.32. *Let $\prec$ be a monomial order and let $f, g \in \mathbb{F}[\overline{x}]$ be nonzero polynomials. Then* $\mathrm{LM}_\prec(fg) = \mathrm{LM}_\prec(f) \cdot \mathrm{LM}_\prec(g)$.

We will primarily be interested in lexicographic orders, which are a special case of weight orders. To specify a weight order, we are given some weight vector $\overline{u} \in \mathbb{R}^n$, and we order two monomials $\overline{x}^{\overline{a}}$ and $\overline{x}^{\overline{b}}$ by comparing the inner products $\langle \overline{u}, \overline{a} \rangle$ and $\langle \overline{u}, \overline{b} \rangle$. To obtain a total order on the set of monomials, ties must be broken. This is done by choosing another weight vector $\overline{w} \in \mathbb{R}^n$ and breaking ties by comparing $\langle \overline{w}, \overline{a} \rangle$ and $\langle \overline{w}, \overline{b} \rangle$. If ties are still possible, we continue choosing new weight vectors until all ties are broken. It turns out that *every* monomial order can be obtained from such a collection of weight vectors.

Theorem 2.33 ([Rob86, Theorem 2.5], see [Rob85] for a proof)**.** *Let $\prec$ be a monomial ordering on $\mathbb{F}[\overline{x}]$. Denote by $\langle \bullet, \bullet \rangle$ the standard inner product on $\mathbb{R}^n$. There is an integer $s \in [n]$ and vectors $\overline{u}^{(1)}, \ldots, \overline{u}^{(s)} \in \mathbb{R}^n$ such that $\overline{x}^{\overline{a}} \prec \overline{x}^{\overline{b}}$ if and only if there is some $j \in [s]$ such that*

1. $\left\langle \overline{a}, \overline{u}^{(i)} \right\rangle = \left\langle \overline{b}, \overline{u}^{(i)} \right\rangle$ *for all $i < j$, and*
2. $\left\langle \overline{a}, \overline{u}^{(j)} \right\rangle < \left\langle \overline{b}, \overline{u}^{(j)} \right\rangle$.

Our focus will be on monomial orders specified by integral weight vectors, which includes all lexicographic orders.

Fact 2.34. Any lexicographic monomial ordering can be specified by a collection of integral weight vectors. ◊

Let $f(\overline{x}) \in R[\overline{x}]$ be a polynomial over a commutative ring R and let $\prec$ be a monomial order that corresponds to a collection of integral weights. It will be useful later on to find an assignment $x_i \mapsto \varepsilon^{d_i}$ of the variables to powers of ε such that $f(\overline{x})$ evaluates to $\varepsilon^m \mathrm{LC}_\prec(f) + O(\varepsilon^{m+1})$ for some integer m. As a first step, we record as a lemma an argument of Bürgisser [Bür04, Example 2.2] on degenerating a polynomial to a face of its Newton polytope.

Lemma 2.35 ([Bür04, Example 2.2])**.** *Let R be a commutative ring and let $f \in R[\overline{x}]$ be given by*

$$f(\overline{x}) = \sum_{\overline{a} \in \mathrm{supp}(f)} \alpha_{\overline{a}} \overline{x}^{\overline{a}}.$$

Let $\overline{u} \in \mathbb{Z}^n$, let $\lambda = \max_{\overline{a} \in \mathrm{supp}(f)} \langle \overline{a}, \overline{u} \rangle$, and let $H = \{\overline{a} \in \mathrm{supp}(f) : \langle \overline{a}, \overline{u} \rangle = \lambda\}$. Then

$$\varepsilon^\lambda f(\varepsilon^{-u_1} x_1, \ldots, \varepsilon^{-u_n} x_n) = \sum_{\overline{a} \in \mathrm{supp}(f) \cap H} \alpha_{\overline{a}} \overline{x}^{\overline{a}} + O(\varepsilon).$$

One can iteratively apply this lemma, further restricting the monomials of f to have exponents that lie in the intersection of multiple hyperplanes.

Lemma 2.36. *Let R be a commutative ring and let $f \in R[\overline{x}]$ be given by*

$$f(\overline{x}) = \sum_{\overline{a} \in \mathrm{supp}(f)} \alpha_{\overline{a}} \overline{x}^{\overline{a}}.$$

Let $\overline{u}^{(1)}, \ldots, \overline{u}^{(k)} \in \mathbb{R}^n$ be vectors. For each $i \in [k]$, let

$$\lambda_i \coloneqq \max_{\overline{a} \in \mathrm{supp}(f) \cap H_1 \cap \cdots \cap H_{i-1}} \left\langle \overline{a}, \overline{u}^{(i)} \right\rangle$$
$$H_i \coloneqq \left\{ \overline{a} \in \mathrm{supp}(f) \cap H_1 \cap \cdots \cap H_{i-1} : \left\langle \overline{a}, \overline{u}^{(i)} \right\rangle = \lambda \right\}.$$

Then there are integers $d_1, \dots, d_n$ and m such that

$$\varepsilon^m f(\varepsilon^{d_1} x_1, \dots, \varepsilon^{d_n} x_n) = \sum_{\overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_k} \alpha_{\overline{a}} \overline{x}^{\overline{a}} + O(\varepsilon).$$

Proof. We proceed by induction on k, noting that the case of $k = 1$ exactly corresponds to Lemma 2.35. When $k \geqslant 2$, by induction we have integers $d'_1, \dots, d'_n$ and m' such that

$$\varepsilon^{m'} f(\varepsilon^{d'_1} x_1, \dots, \varepsilon^{d'_n} x_n) = \sum_{\overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_{k-1}} \alpha_{\overline{a}} \overline{x}^{\overline{a}} + \varepsilon \cdot g(\overline{x}, \varepsilon),$$

where $g(\overline{x}, \varepsilon) \in \mathbb{F}[\varepsilon][\overline{x}]$. By Lemma 2.35, we have

$$\begin{aligned} &\delta^{\lambda_k} \varepsilon^{m'} f(\varepsilon^{d'_1} \delta^{-\overline{u}_1^{(k)}} x_1, \dots, \varepsilon^{d'_n} \delta^{-\overline{u}_n^{(k)}} x_n) \\ &\qquad = \sum_{\overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_k} \alpha_{\overline{a}} \overline{x}^{\overline{a}} + \delta^{\lambda_k} \varepsilon \cdot g(\delta^{-\overline{u}_1^{(k)}} x_1, \dots, \delta^{-\overline{u}_n^{(k)}} x_n, \varepsilon) + O(\delta). \end{aligned}$$

The expression $\delta^{\lambda_k} \varepsilon \cdot g(\delta^{-\overline{u}_1^{(k)}} x_1, \dots, \delta^{-\overline{u}_n^{(k)}} x_n, \varepsilon)$ lies in the ring $\varepsilon \mathbb{F}[\delta, \delta^{-1}, \varepsilon][\overline{x}]$ and may have terms whose coefficient involves a negative power of δ. Let M be the largest power of δ appearing in the denominator of the coefficient of a monomial in $\delta^{\lambda_k} \varepsilon \cdot g(\delta^{-\overline{u}_1^{(k)}} x_1, \dots, \delta^{-\overline{u}_n^{(k)}} x_n, \varepsilon)$. Then under the substitution

$$\begin{aligned} \varepsilon &\mapsto \varepsilon^{M+1} \\ \delta &\mapsto \varepsilon, \end{aligned}$$

every monomial of $\varepsilon^{\lambda_k + M + 1} g(\varepsilon^{-\overline{u}_1^{(k)}} x_1, \dots, \varepsilon^{-\overline{u}_n^{(k)}} x_n, \varepsilon)$ has a coefficient in $\varepsilon \mathbb{F}[\varepsilon]$. In particular, we have

$$\varepsilon^{\lambda_k + (M+1)m'} f(\varepsilon^{d'_1 (M+1) - \overline{u}_1^{(k)}} x_1, \dots, \varepsilon^{d'_n (M+1) - \overline{u}_n^{(k)}} x_n) = \sum_{\overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_k} \alpha_{\overline{a}} \overline{x}^{\overline{a}} + O(\varepsilon).$$

This completes the proof of the inductive step. □

By applying Lemma 2.36 to a polynomial and subsequently setting $x_i \mapsto 1$ for all $i \in [n]$, we can approximate the leading coefficient of f in the sense of border complexity. If the ring R is a field, then this is not necessarily useful. However, we will apply this result when the ring R is a polynomial ring in another set of variables, which makes this lemma useful.

Lemma 2.37. *Let R be a commutative ring. Let $f(\overline{x}) \in R[\overline{x}]$ and let $\prec$ be a monomial order on $\overline{x}$. Suppose that the ordering $\prec$ can be specified by a collection of integral weight vectors $\overline{u}^{(1)}, \dots, \overline{u}^{(s)} \in \mathbb{N}^n$. Then there is some $m \in \mathbb{Z}$ and a collection of nonzero integers $\{d_1, \dots, d_n\}$ such that the mapping*

$$x_i \mapsto \varepsilon^{d_i}$$

sends $f(\overline{x})$ to

$$\varepsilon^m \cdot \operatorname{LC}(f) + O(\varepsilon^{m+1}).$$

Proof. As in the statement of Lemma 2.36, for $i \in [k]$ let

$$\lambda_i \coloneqq \max_{\overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_{i-1}} \left\langle \overline{a}, \overline{u}^{(i)} \right\rangle$$
$$H_i \coloneqq \left\{ \overline{a} \in \operatorname{supp}(f) \cap H_1 \cap \cdots \cap H_{i-1} : \left\langle \overline{a}, \overline{u}^{(i)} \right\rangle = \lambda \right\}.$$

Let $\overline{x}^{\overline{e}} = \operatorname{LM}(f)$. Since $\overline{u}^{(1)}, \ldots, \overline{u}^{(k)}$ are weight vectors specifying a monomial order, it follows from the definition of such an order that $H_k = \{\overline{e}\}$. Applying Lemma 2.36 yields integers $d_1, \ldots, d_n$ and m such that

$$f(\varepsilon^{d_1} x_1, \ldots, \varepsilon^{d_n} x_n) = \varepsilon^m \operatorname{LC}(f) \operatorname{LM}(f) + O(\varepsilon^{m+1}).$$

Setting $x_i \mapsto 1$ for all $i \in [n]$ yields

$$f(\varepsilon^{d_1}, \ldots, \varepsilon^{d_n}) = \varepsilon^m \operatorname{LC}(f) + O(\varepsilon^{m+1})$$

as claimed. □

2.8 The Ideal Proof System

The ideal proof system of Grochow and Pitassi [GP18] is an algebraic proof system used to refute unsatisfiable systems of polynomial equations. The complexity of a proof in this system is measured by the size of the smallest algebraic circuit representing that proof.

Definition 2.38 ([GP18])**.** Let $\mathbb{F}$ be a field and let $f_1(\overline{x}), \ldots, f_m(\overline{x}) \in \mathbb{F}[\overline{x}]$. An *ideal proof system (IPS) certificate* that the system $f_1(\overline{x}) = \cdots = f_m(\overline{x}) = 0$ is unsatisfiable over the algebraic closure $\overline{\mathbb{F}}$ is a polynomial $C(\overline{x}, \overline{y}) \in \mathbb{F}[\overline{x}, \overline{y}]$ such that

1. $C(\overline{x}, \overline{0}) = 0$, and
2. $C(\overline{x}, f_1(\overline{x}), \ldots, f_m(\overline{x})) = 1$. ◊

The first condition equates to requiring that $C(\overline{x}, \overline{y})$ is in the ideal generated by $y_1, \ldots, y_m$. This, along with the second condition, implies that $C(\overline{x}, \overline{y})$ is a certificate for the fact $1 \in \langle f_1(\overline{x}), \ldots, f_m(\overline{x}) \rangle$, hence that $f_1 = \cdots = f_m = 0$ is unsatisfiable.

For a class of algebraic circuits $\mathcal{C}$, one can also consider the $\mathcal{C}$-IPS proof system wherein we require the IPS certificate be computed by a circuit from $\mathcal{C}$. We will primarily be concerned with IPS certificates computable by formulas or low-depth circuits.

As mentioned in the introduction, proving lower bounds on the complexity of IPS refutations is *a priori* more difficult than proving lower bounds for algebraic circuits. This is due to the fact that there may be infinitely many IPS certificates for a single system of equations, so we are faced with proving lower bounds for an infinite family of polynomials. However, these certificates all lie in a coset of an ideal, so one could hope to understand this ideal well enough to prove lower bounds for the relevant coset. See Grochow and Pitassi [GP18, Section 6] for more on the difference between lower bounds for algebraic circuits and IPS.

The following lemma establishes a connection between lower bounds for multiples and lower bounds for IPS. Forbes, Shpilka, Tzameret, and Wigderson [FSTW16] originally stated and proved this lemma with $\{x_i^2 - x_i : i \in [n]\}$ as an additional set of axioms, but these are not necessary. We will make use of this lemma when proving lower bounds for IPS.

Lemma 2.39 ([FSTW16, Lemma 7.1])**.** *Let $f(\overline{x}), g_1(\overline{x}), \ldots, g_k(\overline{x}) \in \mathbb{F}[\overline{x}]$ be an unsatisfiable system of equations where $g_1(\overline{x}), \ldots, g_k(\overline{x})$ is satisfiable. Let $C \in \mathbb{F}[\overline{x}, y, \overline{z}]$ be an IPS refutation of $f, g_1, \ldots, g_k$. Then $1 - C(\overline{x}, 0, g_1(\overline{x}), \ldots, g_k(\overline{x}))$ is a nonzero multiple of $f(\overline{x})$.*

3 Hardness of Determinantal Ideals

Recall that X denotes an $n \times m$ matrix of variables and $I^{\det}_{n,m,r} \subseteq \mathbb{F}[X]$ is the ideal generated by the $r \times r$ minors of X. In this section, we study the minimum possible border complexity of a nonzero polynomial in $I^{\det}_{n,m,r}$. Our main result is that, up to polynomial factors, there is no polynomial $f \in I^{\det}_{n,m,r}$ that is easier to compute than the $r \times r$ determinant. We do this by constructing, for every nonzero $f \in I^{\det}_{n,m,r}$, a depth-three f-oracle circuit that border computes the $\Theta(r^{1/3}) \times \Theta(r^{1/3})$ determinant.

The argument proceeds in two steps. First, we show that for every $f(X) \in I^{\det}_{n,m,r}$, there is a linear change of variables that takes $f(X)$ to $(S|T)(X)+O(\varepsilon)$ for some bideterminant $(S|T)$ of width at least r. The analysis of this step crucially relies on the straightening law (Theorem 2.23). Second, for any $g(\overline{y})$ computed by an ABP of size at most r and any bideterminant $(S|T)(X)$ of width r, we construct a depth-three $(S|T)$-oracle circuit computing $g(\overline{y}) + O(\varepsilon)$. As the determinant can be efficiently computed by ABPs, composing these steps yields an f-oracle circuit for $\det_{\Theta(r^{1/3})}(X) + O(\varepsilon)$.

3.1 Computing a Single Bideterminant

For $i, j \in [n]$ with $i \neq j$, we define the *substitution operator* $\mathrm{Sub}_{i\to j}$ acting on a transpose semistandard Young tableau T as follows: for every row in T containing i but not j, substitute i with j and re-order the row to be in increasing order. Let $h^j_i(T)$ denote the number of rows of T changed by applying $\mathrm{Sub}_{i\to j}$ to T. In general, the map $T \mapsto (\mathrm{Sub}_{i\to j}(T), h^j_i(T))$ may not be injective. However, the following lemma shows that mapping is injective when restricted to tableaux satisfying a particular property.

Lemma 3.1 ([dCEP80, Proposition 1.6])**.** *Let $i, j \in [n]$. Suppose T is a conjugate semistandard tableau with entries in $[n]$ with the property that if a row of T contains an integer $k \leqslant i$, then that row contains all integers in $\{i, i+1, \ldots, j-1\}$. Then $\mathrm{Sub}_{i\to j}(T)$ is also a conjugate semistandard tableau and T is determined by $\mathrm{Sub}_{i\to j}(T)$ and $h^j_i(T)$.*

While the condition in the above lemma seems strange at first, it arises in a natural way when one repeatedly applies the $\mathrm{Sub}_{i\to j}$ operators as described by the next claim. For the sake of completeness, we provide a proof.

Claim 3.2 (implicit in proof of [dCEP80, Corollary 1.7])**.** *Let T be a conjugate semistandard tableau with entries in $[n]$. Let*

$$(1,2) \prec (1,3) \prec \cdots \prec (1,n) \prec (2,3) \prec \cdots \prec (n-2,n-1) \prec (n-2,n) \prec (n-1,n)$$

be a partial order on $[n]^2$. Let $i, j \in [n]$ be such that $i < j$ and let (i', j') be the immediate predecessor of (i, j) in the $\prec$ order. Then the tableau

$$T' \coloneqq \mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{1\to 3} \circ \mathrm{Sub}_{1\to 2}(T)$$

satisfies the hypothesis of Lemma 3.1 for (i, j). In other words, if a row of T' contains an integer $k \leqslant i$, then that row contains all integers in $\{i, i+1, \ldots, j-1\}$.

Proof. The case of $(i, j) = (1, 2)$ is vacuously true. Suppose $(i, j) \succ (1, 2)$ and that some row r of T' contains an integer $k \leqslant i$.

- If $k = i$, then it must be the case that the operator $\mathrm{Sub}_{i\to j-1} \circ \cdots \circ \mathrm{Sub}_{i\to i+1}$ did not replace the i in row r. This implies that the tableau $\mathrm{Sub}_{i-1\to n} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(T)$ contains every element

of $\{i, i+1, \ldots, j-1\}$ in row r. Because of this, the operator $\mathrm{Sub}_{i\to j-1} \circ \cdots \circ \mathrm{Sub}_{i\to i+1}$ does not modify any of the entries in row r coming from the set $\{i, i+1, \ldots, j-1\}$, so row r of T' contains every element of $\{i, \ldots, j-1\}$.

- If $k < i$, then the application of the composite operator $\mathrm{Sub}_{k\to n} \circ \mathrm{Sub}_{k\to n-1} \circ \cdots \circ \mathrm{Sub}_{k\to k+1}$ in the definition of T' did not replace the k appearing in row r of T. This means that every element of $\{k, \ldots, n\}$ appears in row r of the tableau $\mathrm{Sub}_{k-1\to n} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(T)$. Applying the operator $\mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{k\to k+1}$ will not change this, so row r of T' contains every element of $\{k, \ldots, n\}$. In particular, every element of $\{i, \ldots, j-1\}$ appears in this row. □

For a partition σ and natural number $n \in \mathbb{N}$, we let K_σ and $\overline{K}_\sigma$ denote the conjugate semistandard tableaux whose i^{th} row has entries $(1, \ldots, \sigma_i)$ and $(n-i+1, n-i+2, \ldots, n)$, respectively. For example, if $\sigma = (4,3,1)$ and $n = 5$, we have

$$K_{(4,3,1)} = \begin{array}{|c|c|c|c|}\hline 1&2&3&4\\\hline 1&2&3\\\cline{1-3} 1\\\cline{1-1}\end{array} \qquad \overline{K}_{(4,3,1)} = \begin{array}{|c|c|c|c|}\hline 2&3&4&5\\\hline 3&4&5\\\cline{1-3} 5\\\cline{1-1}\end{array}.$$

The operators $\mathrm{Sub}_{i\to j}$ provide a convenient way to transform an arbitrary conjugate semistandard tableau into $\overline{K}_\sigma$.

Lemma 3.3 ([dCEP80, Corollary 1.7])**.** *Let T be a conjugate semistandard tableau of shape σ. Then*

$$(\mathrm{Sub}_{n-1\to n} \circ \mathrm{Sub}_{n-2\to n} \circ \cdots \circ \mathrm{Sub}_{2\to 3} \circ \mathrm{Sub}_{1\to n} \circ \cdots \circ \mathrm{Sub}_{1\to 3} \circ \mathrm{Sub}_{1\to 2})(T) = \overline{K}_\sigma.$$

Moreover, if we denote by h_i^j the number of times i is replaced by j in the application of $\mathrm{Sub}_{i\to j}$ above, then T is determined by σ and the h_i^j.

We are now ready to progress towards the main result of this section. Namely, for any nonzero $f \in I^{\det}_{n,m,r}$, we will find a linear change of variables that sends f to $(K_\sigma | K_\sigma) + O(\varepsilon)$ where σ is the shape of some standard bideterminant in the support of f when f is written in the standard bideterminant basis. For comparison, it is easy to do something similar in the monomial basis: given a polynomial $f(\overline{x})$ of degree d, there is some $m \in \mathbb{N}$ such that

$$\varepsilon^m f(\varepsilon^{-(d+1)} x_1, \varepsilon^{-(d+1)^2} x_2, \ldots, \varepsilon^{-(d+1)^n} x_n) = \mathrm{LC}_{\mathrm{lex}}(f)\,\mathrm{LM}_{\mathrm{lex}}(f) + O(\varepsilon)$$

where we take the lexicographic monomial order induced by $x_1 \succ x_2 \succ \cdots \succ x_n$. To some extent, we are constructing an analogous change of variables in the bideterminant basis.

The main difficulty lies in finding a useful change of variables. In the monomial basis, individual terms can be distinguished by their degree, so it suffices to use a change of variables that only involves multiplying each x_i by some power of ε. However, in the bideterminant basis, multidegree is too coarse a notion to distinguish between bideterminants, so it seems that finding a clever substitution $x_{i,j} \mapsto \varepsilon^{d_{i,j}} x_{i,j}$ will not be enough.

We start by working in a larger polynomial ring $\mathbb{F}[X, \Lambda, \Xi]$. We will give two changes of variables: one that enforces structure on the tableaux encoding the rows of the bideterminants in the support of a polynomial f, and another that handles the tableaux encoding the columns of the bideterminants. The proof of this lemma is inspired by and borrows ideas from the proof of [dCEP80, Theorem 3.3].

Lemma 3.4. *Let $\Lambda = (\lambda_{i,j})$ be an $n \times n$ matrix of variables and let $\prec_\Lambda$ be the lexicographic monomial order on $\mathbb{F}[\Lambda]$ induced by the order $\lambda_{i,j} \succ \lambda_{k,\ell}$ if $i < k$ or $i = k$ and $j < \ell$. Likewise, let $\Xi = (\xi_{i,j})$ be an $m \times m$ matrix of variables and let $\prec_\Xi$ be the corresponding lexicographic monomial order on $\mathbb{F}[\Xi]$.*

Then there are matrices $M \in \mathbb{F}[\Lambda]^{n\times n}$ and $N \in \mathbb{F}[\Xi]^{m\times m}$ with $\det(M) = \pm 1$ and $\det(N) = \pm 1$ such that the following holds.

Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial and let $f(X) = \sum_{k\in[s]} \alpha_k(S_k|T_k)(X)$ be the expansion of f in the standard bideterminant basis. For $k \in [s]$, let σ_k be the shape of the bideterminant $(S_k|T_k)$. Then there are nonempty sets $A, B \subseteq [s]$ such that

$$\mathrm{LC}_{\prec_\Lambda}(f(MX)) = \sum_{k\in A} \alpha_k(K_{\sigma_k}|T_k)(X)$$
$$\mathrm{LC}_{\prec_\Xi}(f(XN)) = \sum_{k\in B} \alpha_k(S_k|K_{\sigma_k})(X),$$

where we take leading coefficients in the rings $\mathbb{F}[X][\Lambda]$ and $\mathbb{F}[X][\Xi]$, respectively.

Proof. We first construct the matrix M and prove the corresponding claim. For $i, j \in [n]$ with $i \neq j$, let $E_{i,j}(z)$ be the $n \times n$ matrix with ones on the diagonal and z in the (i, j) entry. Let J_n be the $n \times n$ matrix whose (i, j) entry is 1 if $i + j = n + 1$ and zero otherwise. We define the matrix M as

$$M \coloneqq E_{1,2}(\lambda_{1,2})E_{1,3}(\lambda_{1,3}) \cdots E_{1,n}(\lambda_{1,n})E_{2,3}(\lambda_{2,3}) \cdots E_{n-1,n}(\lambda_{n-1,n})J_n.$$

Since $\det(J_n) = \pm 1$ and $\det(E_{i,j}(z)) = 1$ for $i \neq j$, it follows that $\det(M) = \pm 1$.

We now analyze the polynomial $f(MX)$. Recall that for a tableau S, we denote by $h^j_i(S)$ the number of entries changed from i to j when we apply the operator $\mathrm{Sub}_{i\to j}$ to S. Observe that for a bideterminant $(S|T)$, it follows from properties of the determinant that

$$(S|T)(E_{i,j}(z)X) = z^{h^j_i(S)}(\mathrm{Sub}_{i\to j}(S)|T)(X) + O(z^{h^j_i(S)-1}),$$

where $O(z^{h^j_i(S)-1})$ denotes a polynomial in $\mathbb{F}[X][z]$ of degree at most $h^j_i(S) - 1$. For $i, j \in [n]$ with $i \neq j$, define

$$f_{i,j}(X, \Lambda) \coloneqq f(E_{1,2}(\lambda_{1,2})E_{1,3}(\lambda_{1,3}) \cdots E_{1,n}(\lambda_{1,n})E_{2,3}(\lambda_{2,3}) \cdots E_{i,j}(\lambda_{i,j})X).$$

Note that $f(MX) = f_{n-1,n}(J_nX, \Lambda)$.

We claim that for every $i, j \in [n]$ with $i < j$, there is a non-empty set $A_{i,j} \subseteq [s]$ such that

$$\mathrm{LC}_{\prec_\Lambda}(f_{i,j}(X, \Lambda)) = \sum_{k\in A_{i,j}} \alpha_k(\mathrm{Sub}_{i\to j} \circ \cdots \circ \mathrm{Sub}_{2\to 3} \circ \mathrm{Sub}_{1\to n} \circ \cdots \circ \mathrm{Sub}_{1\to 3} \circ \mathrm{Sub}_{1\to 2}(S_k)|T_k)(X).$$

By Lemma 3.3, this implies

$$\mathrm{LC}_{\prec_\Lambda}(f_{n-1,n}(X, \Lambda)) = \sum_{k\in A_{n-1,n}} \alpha_k(\overline{K}_{\sigma_k}|T_k)(X).$$

Using the fact that $(\overline{K}_{\sigma_k}|T)(J_nX) = (K_{\sigma_k}|T_k)(X)$, this yields

$$\mathrm{LC}_{\prec_\Lambda}(f(MX)) = \mathrm{LC}_{\prec_\Lambda}(f_{n-1,n}(J_nX, \Lambda)) = \sum_{k\in A_{n-1,n}} \alpha_k(K_{\sigma_k}|T_k)(X)$$

as claimed.

We now prove the claim by induction on (i, j) in the order $(1, 2) \prec (1, 3) \prec \cdots \prec (1, n) \prec (2, 3) \prec \cdots \prec (n - 1, n)$. Let (i', j') be the predecessor of (i, j) in the $\prec$ order. In the case that $(i, j) = (1, 2)$, we abuse notation and set $f_{i',j'} \coloneqq f$ and $A_{i',j'} \coloneqq [s]$. Let

$$H^j_i \coloneqq \max_{k\in A_{i',j'}} h^j_i(\mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k))$$

and

$$A_{i,j} = \{k \in A_{i',j'} : h_i^j(\mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k)) = H_i^j\}.$$

Note that $A_{i,j}$ is necessarily non-empty, as H_i^j is a maximum over a finite nonempty set. By induction, there is some $\overline{e} \in \mathbb{N}^{n\times n}$ such that

$$f_{i',j'}(X,\Lambda) = \Lambda^{\overline{e}} \sum_{k\in A_{i',j'}} \alpha_k(\mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k)|T_k)(X) + g(X,\Lambda),$$

where $g(X,\Lambda) \in \mathbb{F}[X][\Lambda]$ is a polynomial in which every monomial is smaller than $\Lambda^{\overline{e}}$ in the $\prec_\Lambda$ order. Since $f_{i',j'}$ only depends on $\lambda_{1,2},\ldots,\lambda_{i',j'}$, it follows that $\Lambda^{\overline{e}}$ is a monomial in only these variables. We then apply the definition of $f_{i,j}$ to obtain

$$\begin{aligned}
f_{i,j}(X,\Lambda) &= f_{i',j'}(E_{i,j}(\lambda_{i,j})X,\Lambda) \\
&= \Lambda^{\overline{e}} \sum_{k\in A_{i',j'}} \alpha_k(\mathrm{Sub}_{i'\to j'} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k)|T_k)(E_{i,j}(\lambda_{i,j})X) + g(E_{i,j}(\lambda_{i,j})X,\Lambda) \\
&= \Lambda^{\overline{e}}\lambda_{i,j}^{H_i^j} \sum_{k\in A_{i,j}} \alpha_k(\mathrm{Sub}_{i\to j} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k)|T_k)(X) + \Lambda^{\overline{e}} p(X,\lambda_{i,j}) + g(E_{i,j}(\lambda_{i,j})X,\Lambda),
\end{aligned}$$

where $p(X,\lambda_{i,j}) \in \mathbb{F}[X][\Lambda]$ is a polynomial of degree at most $H_i^j - 1$ in $\lambda_{i,j}$. This implies that every monomial of $\Lambda^{\overline{e}}p(X,\Lambda)$ is smaller than $\Lambda^{\overline{e}}\lambda_{i,j}^{H_i^j}$ in the $\prec_\Lambda$ order. Observe that the substitution $X \mapsto E_{i,j}(\lambda_{i,j})X$ only changes the $\lambda_{i,j}$-degree of any Λ-monomial in $g(X,\Lambda)$. In particular, because every monomial of $g(X,\Lambda)$ is smaller than $\Lambda^{\overline{e}}$ in the $\prec_\Lambda$ order, the same holds true for every Λ-monomial of $g(E_{i,j}(\lambda_{i,j})X,\Lambda)$. This implies that

$$\mathrm{LC}_{\prec_\Lambda}(f_{i,j}) = \sum_{k\in A_{i,j}} \alpha_k(\mathrm{Sub}_{i\to j} \circ \cdots \circ \mathrm{Sub}_{1\to 2}(S_k)|T_k)(X)$$

as claimed. This establishes the claimed properties of M.

To construct the matrix N, we overload notation and let $E_{i,j}(z)$ be the $m \times m$ matrix with ones on the diagonal and z in the (i,j) entry. Just as the matrix M consisted of a sequence of row operations, the matrix N will be composed of a sequence of column operations. We define N as

$$N \coloneqq J_m E_{m-1,m}(\xi_{m-1,m}) \cdots E_{2,3}(\xi_{2,3})E_{1,m}(\xi_{1,m}) \cdots E_{1,3}(\xi_{1,3})E_{1,2}(\xi_{1,2}).$$

Since $\det(J_m) = \pm 1$ and $\det(E_{i,j}(z)) = 1$ for $i < j$, we get that $\det(N) = \pm 1$.

As in the previous case, it follows from properties of the determinant that for a bideterminant $(S|T)$, we have

$$(S|T)(XE_{i,j}(z)) = z^{h_i^j(T)}(S|\mathrm{Sub}_{i\to j}(T))(X) + O(z^{h_i^j(T)-1}).$$

Using this, the analysis of the leading coefficient of $f(XN) \in \mathbb{F}[X][\Xi]$ proceeds in a manner analogous to the case of $f(MX)$, so we omit the details. □

We now come to the main result of this subsection: a change of variables that sends a polynomial $f(X)$ to $(K_\sigma|K_\sigma)(X) + O(\varepsilon)$ where σ is the shape of some standard bideterminant in the support of f.

Proposition 3.5. *Let $f(X) \in I_{n,m,r}^{\det}$ be nonzero. There is a collection of nm linearly independent linear functions $\ell_{i,j}(X,\varepsilon) \in \mathbb{F}(\varepsilon)[X]$ indexed by $(i,j) \in [n]\times[m]$, an integer $q \in \mathbb{Z}$, a nonzero $\alpha \in \mathbb{F}$, and a partition σ with $\sigma_1 \geqslant r$ such that*

$$f(\ell_{1,1}(X,\varepsilon),\ldots,\ell_{n,m}(X,\varepsilon)) = \varepsilon^q \alpha(K_\sigma|K_\sigma)(X) + O(\varepsilon^{q+1}).$$

Proof. Let $f = \sum_{k\in[s]} \alpha_k(S_k|T_k)$ be the expansion of f in the standard bideterminant basis. Let M and N be the matrices constructed in Lemma 3.4. Let $\prec$ denote the lexicographic order on $\mathbb{F}[X][\Lambda,\Xi]$ induced by $\lambda_{1,2} \succ \lambda_{1,3} \succ \cdots \succ \lambda_{n-1,n} \succ \xi_{1,2} \succ \cdots \succ \xi_{m-1,m}$. Lemma 3.4 implies that there is a non-empty set $A \subseteq [s]$ such that

$$g(X) \coloneqq \mathrm{LC}_\prec(f(MX)) = \sum_{k\in A} \alpha_k(K_{\sigma_k}|T_k)(X),$$

and likewise that there is a non-empty set $B \subseteq A$ such that

$$\mathrm{LC}_\prec(g(XN)) = \sum_{k\in B} \alpha_k(K_{\sigma_k}|K_{\sigma_k})(X).$$

This implies that

$$\mathrm{LC}_\prec(f(MXN)) = \sum_{k\in B} \alpha_k(K_{\sigma_k}|K_{\sigma_k})(X),$$

where σ_k denotes the shape of the bideterminant $(S_k|T_k)$. By Corollary 2.24, each bideterminant in the above sum has width at least r, so $(\sigma_k)_1 \geqslant r$ for all $k \in A$.

Let y and z be new indeterminates and let $D \coloneqq \deg(f(X))$. Consider the change of variables

$$x_{i,j} \mapsto y^{(D+1)^i} z^{(D+1)^j} x_{i,j}.$$

Let $h(X,\Lambda,\Xi,y,z)$ be the image of $f(MXN)$ under this map. By construction, an X-monomial of multidegree $(\sum_i a_i\overline{e}_i) \oplus (\sum_i b_i\overline{e}_i)$ is multiplied by a factor of $y^{\sum_i a_i(D+1)^i} z^{\sum_j b_j(D+1)^j}$. In particular, since $\max_i a_i \leqslant D$ and $\max_i b_i \leqslant D$, X-monomials of distinct multidegree have distinct (y,z)-degree under this mapping. Observe that $\mathrm{multideg}((K_\sigma|K_\sigma)(X)) \neq \mathrm{multideg}((K_\tau|K_\tau)(X))$ for distinct partitions $\sigma \neq \tau$. Since each bideterminant $(K_\sigma|K_\sigma)(X)$ is mapped to a unique (y,z)-degree under this substitution, we get that the polynomial

$$p(X) = \mathrm{LC}_{(y,z)}(\mathrm{LC}_{(\Lambda,\Xi)}(h(X,\Lambda,\Xi,y,z)))$$

is a nonzero multiple of the bideterminant $(K_{\sigma_k}|K_{\sigma_k})(X)$ for some $k \in B$. If we augment the monomial order $\prec$ by setting $\Lambda \succ \Xi \succ y \succ z$ and taking the corresponding lexicographic order, we then have

$$\mathrm{LC}_\prec(h(X,\Lambda,\Xi,y,z)) = \alpha_k(K_{\sigma_k}|K_{\sigma_k})(X)$$

for some $k \in B$.

Applying Lemma 2.37 to $h(X,\Lambda,\Xi,y,z)$ viewed as an element of $\mathbb{F}[X][\Lambda,\Xi,y,z]$, we get a map $\varphi : (\Lambda \cup \Xi \cup \{y,z\}) \to \{\varepsilon^d : d \in \mathbb{Z}\}$ such that

$$\varphi(h(X,\Lambda,\Xi,y,z)) = \varepsilon^q \alpha_k(K_{\sigma_k}|K_{\sigma_k})(X) + O(\varepsilon^{q+1})$$

for some integer q.

Note that $h(X,\Lambda,\Xi,y,z)$ was obtained from $f(X)$ by an invertible linear transformation of the X variables. That is, there are nm linearly independent linear polynomials $\ell'_{1,1}(X),\ldots,\ell'_{n,m}(X) \in \mathbb{F}[\Lambda,\Xi,y,z][X]$ such that

$$h(X,\Lambda,\Xi,y,z) = f(\ell'_{1,1}(X),\ldots,\ell'_{n,m}(X)).$$

Set $\ell_{i,j}(X,\varepsilon) \coloneqq \varphi(\ell'_{i,j}(X)) \in \mathbb{F}(\varepsilon)[X]$ for each $(i,j) \in [n] \times [m]$. Since the transformation $x_{i,j} \mapsto \ell'_{i,j}(X)$ is invertible as long as $y \neq 0$ and $z \neq 0$, the transformation $x_{i,j} \mapsto \ell_{i,j}(X,\varepsilon)$ remains invertible under φ. Finally, it follows from the definition of φ that

$$\begin{aligned} f(\ell_{1,1}(X,\varepsilon),\dots,\ell_{n,m}(X,\varepsilon)) &= f(\varphi(\ell'_{1,1}(X)),\dots,\varphi(\ell'_{n,m}(X))) \\ &= \varphi(f(\ell'_{1,1}(X),\dots,\ell'_{n,m}(X))) \\ &= \varphi(h(X,\Lambda,\Xi,y,z)) \\ &= \varepsilon^q \alpha_k(K_{\sigma_k}|K_{\sigma_k})(X) + O(\varepsilon^{q+1}). \end{aligned}$$

□

3.2 Projecting to the Determinant

So far, we have constructed a linear change of variables taking a polynomial $f \in I^{\det}_{n,m,r}$ to $(K_\sigma|K_\sigma) + O(\varepsilon)$ for a bideterminant $(K_\sigma|K_\sigma)$ of width at least r. Next, we show that a $(K_\sigma|K_\sigma)$-oracle can be used to compute $g(\overline{y}) + O(\varepsilon)$, where g is any polynomial computable by an algebraic branching program on r vertices. Ideally, one would like to appeal to the VBP-completeness of the determinant, which gives a projection from $\det_r(X)$ to $g(\overline{y})$, to prove such a result. The difficulty lies in the fact that a bideterminant may be a product of multiple determinants of varying sizes. Because of this, we need a projection that behaves well on proper minors of X and also allows us to deal with the possibility that we may be projecting from a power of the determinant as opposed to the determinant itself. We almost construct such a projection, but we will need some post-processing in the form of an extra addition gate in order to handle powers of the determinant.

Let $g(\overline{y})$ be computable by a small algebraic branching program. We begin by describing a projection $\varphi : X \to \overline{y} \cup \mathbb{F}$ of a generic matrix X such that $\det(\varphi(X)) = 1 + g(\overline{y})$ and the leading principal minors of $\varphi(X)$ have determinant 1. This is a small modification of an argument due to Valiant [Val79, Theorem 1]; we include a proof for the sake of completeness.

Lemma 3.6. *Let $g(\overline{y}) \in \mathbb{F}[\overline{y}]$ and suppose g can be computed by a layered algebraic branching program on m vertices. Then there is an $m \times m$ matrix $A \in \mathbb{F}[\overline{y}]^{m\times m}$ whose entries are linear polynomials in $\overline{y}$ such that*

1. *$\det(A) = 1 + g(\overline{y})$, and*
2. *for every $k \in [m-1]$, we have $\det(A_{[k],[k]}) = 1$.*

Proof. We first recall the correspondence between cycle covers in graphs and the determinant. Let G be a weighted directed graph on m vertices and denote the weight of the edge (i,j) by $w(i,j)$. Let $A(G) = (a_{i,j})$ be the $m \times m$ matrix given by

$$a_{i,j} = \begin{cases} w(i,j) & (i,j) \in E(G) \\ 0 & (i,j) \notin E(G). \end{cases}$$

Recall that a *cycle cover* C of G is a collection of vertex-disjoint cycles in G which span the vertices of G. Let $CC(G)$ denote the collection of all cycle covers of G. Given a cycle cover C of G, let $\pi(C)$ denote the product of the edge weights in C. If every cycle cover of G consists of odd-length cycles, then the definitions of $A(G)$ and the determinant imply that

$$\det(A(G)) = \sum_{C \in CC(G)} \pi(C).$$

We now proceed with the proof of Lemma 3.6. Suppose $g(\overline{y})$ can be computed by a layered algebraic branching program on m nodes. Let s and t be the start and end nodes of this branching program, respectively. Since the program is layered, every s-t path has the same length. If the length of each s-t path is even, we add an edge of weight 1 from t to s and a self-loop of weight 1 to every vertex (including s and t); if the length of each s-t path is odd, we identify the vertices s and t with one another (resulting in a graph on $m-1$ nodes), add an isolated vertex r, and then add a self-loop to every vertex. Denote the resulting graph by G. In both cases, G has one cycle cover for every s-t path in the branching program, as well as a single cycle cover corresponding to the set of self-loops in the graph. Moreover, every cycle cover in G consists solely of odd-length cycles.

For a cycle cover C corresponding to an s-t path P in the branching program, it follows from the definition of G that $\pi(C) = \pi(P)$, where $\pi(P)$ is the product of the weights on the edges of P. If C is the all-self-loops cycle cover, then $\pi(C) = 1$. Since every cycle cover in G consists of odd-length cycles, we have

$$\det(A(G)) = \sum_{C \in CC(G)} \pi(C) = 1 + \sum_{P} \pi(P) = 1 + g(\overline{y}),$$

where the second summation is over all s-t paths P in the branching program. This proves the first part of the lemma.

To prove the second part, let $v_1, \ldots, v_m$ be a topological ordering of the vertices in the algebraic branching program. Note that $v_1 = s$ and $v_m = t$. If every s-t path in the branching program has even length, we order the rows and columns of $A(G)$ such that

$$A(G)_{i,j} = w(v_i, v_j).$$

If instead every s-t path in the branching program has odd length, we set

$$A(G)_{i,j} = \begin{cases} w(r, v_j) & i = 1 \\ w(v_i, r) & j = 1 \\ w(v_i, v_j) & \text{otherwise,} \end{cases}$$

where r is the isolated vertex with a self-loop. In either case, note that if $i > j$ and $A(G)_{i,j} \neq 0$, then we must have $i = m$. This implies that for every $k \in [m-1]$, the matrix $A(G)_{[k],[k]}$ is upper-triangular with ones along the diagonal. Thus $\det(A(G)_{[k],[k]}) = 1$ as desired. □

Although we want to construct an $(K_\sigma | K_\sigma)$-oracle circuit that computes any polynomial $g(\overline{y})$ that is computable by a small layered algebraic branching program, it will be convenient for us to assume that g is homogeneous. This is not restrictive, as one can always introduce a new variable z and consider the homogeneous polynomial $\hat{g}(\overline{y}, z) \coloneqq z^{\deg(g)} g(y_1/z, \ldots, y_n/z)$, which specializes to $g(\overline{y})$ under the map $z \mapsto 1$. One needs to show that $\hat{g}(\overline{y}, z)$ is as easy to compute as $g(\overline{y})$. Below, we provide a proof that this can be done for layered ABPs, although we technically show that this is the case for $z^d g(y_1/z, \ldots, y_n/z)$ for some $d \geqslant \deg(g)$.

Lemma 3.7. *Let $g(\overline{y}) \in \mathbb{F}[\overline{y}]$ be a polynomial and suppose that g can be computed by a layered algebraic branching program on m vertices. Let z be a new variable. Then there is a homogeneous polynomial $\hat{g}(\overline{y}, z) \in \mathbb{F}[\overline{y}, z]$ such that $\hat{g}$ can be computed by a layered algebraic branching program on m vertices and that $\hat{g}(\overline{y}, 1) = g(\overline{y})$.*

Proof. Let $G = (V = V_0 \sqcup V_1 \sqcup \cdots \sqcup V_k, E)$ be an m-vertex ABP that computes $g(\overline{y})$, where the V_i are the layers of the ABP. Without loss of generality, we assume that no vertex of G computes the zero polynomial; if this is the case, we simply remove such a vertex. We relabel the edges of G as

follows: if an edge $e \in E$ is labeled by the polynomial $\ell_e(\overline{y}) = \alpha_0 + \sum_{i=1}^n \alpha_i y_i$, we relabel the edge e with $\hat{\ell}_e(\overline{y}, z) = \alpha_0 z + \sum_{i=1}^n \alpha_i y_i$. Let $\hat{G}$ denote the relabeled ABP.

It is clear that $\hat{G}$ is an m-vertex layered ABP. For each vertex $v \in V$, let $g_v(\overline{y})$ be the polynomial computed by v in G, and let $\hat{g}_v(\overline{y}, z)$ be the polynomial computed at v in $\hat{G}$. We claim that for each $i \in \{0, 1, \ldots, k\}$ and $v \in V_i$, the polynomial $\hat{g}_v(\overline{y}, z)$ is homogeneous of degree i and that $\hat{g}_v(\overline{y}, 1) = g_v(\overline{y})$. We prove this by induction on the depth of the vertex v in G, i.e., the layer of V containing v.

If $v \in V_0$, then $\hat{g}_v(\overline{y}, z) = g_v(\overline{y}) = 1$ and we are done. Otherwise, we have $v \in V_i$ for some $i \geqslant 1$. By definition, we have

$$\hat{g}_v(\overline{y}, z) = \sum_{u \in V_{i-1}} \hat{\ell}_{u \to v}(\overline{y}, z) \cdot \hat{g}_u(\overline{y}, z).$$

By induction, for every $u \in V_{i-1}$, the polynomial $\hat{g}_u(\overline{y}, z)$ is a homogeneous degree-$(i-1)$ polynomial that satisfies $\hat{g}_u(\overline{y}, 1) = g_u(\overline{y})$. Furthermore, each nonzero $\hat{\ell}_{u \to v}(\overline{y}, z)$ is a homogeneous degree-1 polynomial, so it follows that $\hat{g}_v(\overline{y}, z)$ is a homogeneous degree-i polynomial. Setting $z \mapsto 1$, we have

$$\begin{aligned}
\hat{g}_v(\overline{y}, 1) &= \sum_{u \in V_{i-1}} \hat{\ell}_{u \to v}(\overline{y}, 1) \cdot \hat{g}_u(\overline{y}, 1) \\
&= \sum_{u \in V_{i-1}} \ell_{u \to v}(\overline{y}) \cdot g_u(\overline{y}) \\
&= g_v(\overline{y}).
\end{aligned}$$

Thus, the polynomial $\hat{g}_v(\overline{y}, z)$ is as claimed.

To finish the proof of the lemma, observe that if v is the output vertex of G, then $\hat{g}_v(\overline{y}, z)$ is the desired polynomial. □

Given a nonzero $f(X) \in I^{\det}_{n,m,r}$, we will use the preceding lemmas together with Proposition 3.5 to construct a depth-three f-oracle circuit computing $\det_{\Theta(r^{1/3})}(X) + O(\varepsilon)$. In fact, for any polynomial $g(\overline{y})$ computable by a layered algebraic branching program on r vertices, we can construct an f-oracle circuit computing g.

Theorem 3.8. *Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial and let $h(X, \varepsilon) \in \mathbb{F}[\![\varepsilon]\!][X]$ be any polynomial such that $h(X, \varepsilon) = f(X) + O(\varepsilon)$. Let $g(\overline{y}) \in \mathbb{F}[\overline{y}]$ be a polynomial in the border of layered algebraic branching programs with at most r vertices. Then there is a depth-three h-oracle circuit Φ defined over $\mathbb{F}(\varepsilon)$ such that the following hold.*

1. *Φ has nm addition gates at the bottom layer, a single h-oracle gate in the middle layer, and a single addition gate at the top layer.*
2. *If $\operatorname{char}(\mathbb{F}) = 0$, then Φ computes $g(\overline{y}) + O(\varepsilon)$.*
3. *If $\operatorname{char}(\mathbb{F}) = p > 0$, then Φ computes $g(\overline{y})^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.*

Proof. By Lemma 2.3, it suffices to prove the theorem in the case where the oracle gates compute f exactly. By assumption, there is a polynomial $\widetilde{g}(\overline{y}, \varepsilon) \in \mathbb{F}[\varepsilon][\overline{y}]$ such that $\widetilde{g}(\overline{y}, \varepsilon) = g(\overline{y}) + O(\varepsilon)$ and $\widetilde{g}(\overline{y}, \varepsilon)$ can be computed by a layered algebraic branching program on at most r vertices. Lemma 3.7 implies that there is a homogeneous polynomial $\hat{g}(\overline{y}, \varepsilon, z) \in \mathbb{F}[\varepsilon][\overline{y}, z]$ computable by a layered algebraic branching program on at most r vertices such that $\hat{g}(\overline{y}, \varepsilon, 1) = \widetilde{g}(\overline{y}, \varepsilon)$.

Applying Proposition 3.5 to $f(X)$, we obtain linear functions $\ell_{1,1}(X, \varepsilon), \ldots, \ell_{n,m}(X, \varepsilon)$, a nonzero $\alpha \in \mathbb{F}$, and some $q \in \mathbb{Z}$ such that

$$f(\ell_{1,1}(X, \varepsilon), \ldots, \ell_{n,m}(X, \varepsilon)) = \varepsilon^q \alpha (K_\sigma | K_\sigma)(X) + O(\varepsilon^{q+1})$$

for some partition σ of width at least r. Since $\hat{g}(\overline{y}, \varepsilon, z)$ can be computed by a layered algebraic branching program on at most r vertices, we can obtain a layered ABP on exactly r vertices computing $\hat{g}(\overline{y}, \varepsilon, z)$ by adding isolated vertices. Let $A(\overline{y}, z) \in \mathbb{F}[\varepsilon][\overline{y}, z]^{r \times r}$ be the matrix obtained by applying Lemma 3.6 to $\hat{g}(\overline{y}, \varepsilon, z)$. Extend $A(\overline{y}, z)$ to an $n \times m$ matrix by adding ones along the main diagonal and zeroes elsewhere. Then we have

$$\begin{aligned}
f(\ell_{1,1}(A(\overline{y}, z), \varepsilon), \ldots, \ell_{n,m}(A(\overline{y}, z), \varepsilon)) &= \varepsilon^q \alpha (K_\sigma | K_\sigma)(A(\overline{y}, z)) + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha \prod_{i=1}^{\hat{\sigma}_1} \det_{\sigma_i}(A(\overline{y}, z)_{[\sigma_i],[\sigma_i]}) + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha \prod_{i:\sigma_i \geqslant r} \det_{\sigma_i}(A(\overline{y}, z)_{[\sigma_i],[\sigma_i]}) \cdot \prod_{i:\sigma_i < r} \det_{\sigma_i}(A(\overline{y}, z)_{[\sigma_i],[\sigma_i]}) + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha \prod_{i:\sigma_i \geqslant r} (1 + \hat{g}(\overline{y}, \varepsilon, z)) + O(\varepsilon^{q+1}).
\end{aligned}$$

Let $h(\overline{y}, \varepsilon, z) \coloneqq f(\ell_{1,1}(A(\overline{y}, z), \varepsilon), \ldots, \ell_{n,m}(A(\overline{y}, z), \varepsilon))$ and let $t = |\{i : \sigma_i \geqslant r\}|$. The above establishes $h(\overline{y}, \varepsilon, z) = \varepsilon^q \alpha (1 + \hat{g}(\overline{y}, \varepsilon, z))^t + O(\varepsilon^{q+1})$.

Suppose $\mathrm{char}(\mathbb{F}) = 0$. Under the substitution $y_i \mapsto \delta \cdot y_i$ and $z \mapsto \delta$, we have

$$\begin{aligned}
h(\delta \cdot \overline{y}, \varepsilon, \delta) &= \varepsilon^q \alpha (1 + \hat{g}(\delta \cdot \overline{y}, \varepsilon, \delta))^t + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha (1 + \delta^{\deg(\hat{g})} \hat{g}(\overline{y}, \varepsilon, 1))^t + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha (1 + \delta^{\deg(\hat{g})} g(\overline{y}) + O(\varepsilon))^t + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha \sum_{i=0}^{t} \binom{t}{i} \delta^{i \cdot \deg(\hat{g})} g(\overline{y})^i + O(\varepsilon^{q+1}) \\
&= \varepsilon^q \alpha + \varepsilon^q \delta^{\deg(\hat{g})} \alpha t g(\overline{y}) + O(\varepsilon^q \delta^{2 \deg(\hat{g})}) + O(\varepsilon^{q+1}).
\end{aligned}$$

Performing the substitution

$$\begin{aligned}
\varepsilon &\mapsto \varepsilon^N \\
\delta &\mapsto \varepsilon
\end{aligned}$$

for N sufficiently large yields

$$h(\varepsilon \cdot \overline{y}, \varepsilon^N, \varepsilon) = \varepsilon^{qN} \alpha + \varepsilon^{qN + \deg(\hat{g})} \alpha t g(\overline{y}) + O(\varepsilon^{qN + \deg(\hat{g}) + 1}).$$

The desired f-oracle circuit for g is then given by

$$\Phi(\overline{y}) \coloneqq \frac{h(\varepsilon \cdot \overline{y}, \varepsilon^N, \varepsilon) - \varepsilon^{qN} \alpha}{\varepsilon^{qN + \deg(\hat{g})} \alpha t} = g(\overline{y}) + O(\varepsilon).$$

If instead $\mathrm{char}(\mathbb{F}) = p > 0$, the above proof only needs to be modified in the case that p divides t. Let $k \in \mathbb{N}$ be the largest natural number such that p^k divides t and write $t = p^k b$. In this case, we instead get

$$h(\delta \cdot \overline{y}, \varepsilon, \delta) = \varepsilon^q \alpha + \varepsilon^q \delta^{\deg(\hat{g}) p^k} \alpha b g(\overline{y})^{p^k} + O(\varepsilon^q \delta^{2 \deg(\hat{g}) p^k}) + O(\varepsilon^{q+1}).$$

Again, for N sufficiently large, we obtain an f-oracle circuit for g via

$$\Phi(\overline{y}) \coloneqq \frac{h(\varepsilon \cdot \overline{y}, \varepsilon^N, \varepsilon) - \varepsilon^{qN} \alpha}{\varepsilon^{qN + \deg(\hat{g}) p^k} \alpha b} = g(\overline{y})^{p^k} + O(\varepsilon). \qquad \square$$

We now instantiate Theorem 3.8 with the determinant and iterated matrix multiplication polynomials. These corollaries are essentially obvious, but seem interesting in their own right and will be of use in later sections.

Corollary 3.9. *Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial and let $h(X,\varepsilon) \in \mathbb{F}[\![\varepsilon]\!][X]$ be any polynomial such that $h(X,\varepsilon) = f(X) + O(\varepsilon)$. Let $t \leqslant O(r^{1/3})$. Then there is a depth-three h-oracle circuit Φ defined over $\mathbb{F}(\varepsilon)$ with the following properties.*

1. *The bottom layer of Φ consists of nm addition gates, the middle layer has a single h-oracle gate, and the top layer has a single addition gate.*
2. *If $\mathrm{char}(\mathbb{F}) = 0$, then Φ computes $\det_t(Y) + O(\varepsilon)$.*
3. *If $\mathrm{char}(\mathbb{F}) = p > 0$, then Φ computes $\det_t(Y)^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.*

Proof. Mahajan and Vinay [MV97, Theorem 2] constructed a layered ABP on $O(t^3) \leqslant r$ vertices that computes $\det_t(Y)$. The corollary then follows from Theorem 3.8. □

Corollary 3.10. *Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial and let $h(X,\varepsilon) \in \mathbb{F}[\![\varepsilon]\!][X]$ be any polynomial such that $h(X,\varepsilon) = f(X) + O(\varepsilon)$. Let $w, d \in \mathbb{N}$ satisfy $w(d-1)+2 \leqslant r$. Then there is a depth-three h-oracle circuit Φ defined over $\mathbb{F}(\varepsilon)$ with the following properties.*

1. *The bottom layer of Φ consists of nm addition gates, the middle layer has a single h-oracle gate, and the top layer has a single addition gate.*
2. *If $\mathrm{char}(\mathbb{F}) = 0$, then Φ computes $\mathrm{IMM}_{w,d}(\overline{y}) + O(\varepsilon)$.*
3. *If $\mathrm{char}(\mathbb{F}) = p > 0$, then Φ computes $\mathrm{IMM}_{w,d}(\overline{y})^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.*

Proof. It is clear that $\mathrm{IMM}_{w,d}(\overline{y})$ is computable by a layered algebraic branching program on $w(d-1)+2 \leqslant r$ vertices. Theorem 3.8 completes the proof. □

We conclude this section with a remark on the fact that in characteristic $p > 0$, we only obtain an oracle circuit for a p^{th} power of the target polynomial $g(\overline{y})$.

Remark 3.11. Let $\mathbb{F}$ be a field of characteristic $p > 0$. If we interpret Theorem 3.8 as a result on "factoring" a polynomial $I^{\det}_{n,m,r}$, then the appearance of p^{th} powers in the "factors" is not too surprising. Most results on polynomial factorization [Kal87; DSY09; KSS15; CKS19b] only guarantee a circuit that computes a p^{th} power of a factor if the multiplicity of this factor is a multiple of p^k for some $k > 0$. In fact, if $f(\overline{x})^p$ can be computed by a size s circuit, it is open whether $f(\overline{x})$ can be computed by a circuit of size $\mathrm{poly}(n, \deg(f), s)$, although some results are known when n is small compared to s [And20]. ◊

4 Hardness of Pfaffian Ideals

This section proves an analogue of Theorem 3.8 for ideals generated by sub-Pfaffians of a skew-symmetric matrix. The outline of the proof is similar to that of Theorem 3.8, but some technical details must be modified to accommodate the change to Pfaffians.

4.1 Computing a Standard Monomial

In this subsection, we construct, for any nonzero $f \in I^{\mathrm{pfaff}}_{2n,2r}$, a change of variables that takes f to $[K_\sigma](X) + O(\varepsilon)$ for some partition σ with $\sigma_1 \geqslant 2r$. The outline of the proof is the same as the proof of Lemma 3.4, replacing the straightening law for bideterminants with the corresponding straightening law for Pfaffians.

The following lemma finds a change of variables that takes f to a sum of standard monomials of the form $[K_\sigma](X)$. This is the Pfaffian analogue of Lemma 3.4 and borrows ideas from the proof of Abeasis and Del Fra [AD80, Lemmas 2.1 and 2.2] in a manner analogous to the use of [dCEP80, Theorem 3.3] in proving Lemma 3.4.

Lemma 4.1. *Let $\Lambda = (\lambda_{i,j})$ be a $2n \times 2n$ matrix of variables and let $\prec_\Lambda$ be the lexicographic monomial order on $\mathbb{F}[\Lambda]$ induced by the order $\lambda_{i,j} \succ \lambda_{k,\ell}$ if $i < k$ or $i = k$ and $j < \ell$. Then there is a matrix $M \in \mathbb{F}[\Lambda]^{2n\times 2n}$ with $\det(M) = \pm 1$ such that the following holds.*

Let $f(X) \in I^{\mathrm{pfaff}}_{2n,2r}$ be a nonzero polynomial and let $f(X) = \sum_{k\in[s]} \alpha_k [S_k](X)$ be the expansion of f as a sum of standard monomials. For $k \in [s]$, let σ_k be the shape of the tableau S_k. Then there is a nonempty set $A \subseteq [s]$ such that

$$\mathrm{LC}_{\prec_\Lambda}(f(MXM^\top)) = \sum_{k\in A} \alpha_k [K_{\sigma_k}](X)$$

where we take the leading coefficient in the ring $\mathbb{F}[X][\Lambda]$.

Proof. We begin with the construction of the matrix M. For $i, j \in [2n]$ with $i < j$, let $E_{i,j}(z)$ denote the matrix which has ones on the diagonal and z in the (i, j) entry. We then let $M_{i,j}(\Lambda) \in \mathbb{F}[\Lambda]^{2n\times 2n}$ be the matrix

$$M_{i,j}(\Lambda) \coloneqq E_{1,2}(\lambda_{1,2})E_{1,3}(\lambda_{1,3}) \cdots E_{1,n}(\lambda_{1,n})E_{2,3}(\lambda_{2,3}) \cdots E_{i,j}(\lambda_{i,j}).$$

Letting J_{2n} denote the $2n \times 2n$ matrix with ones on the anti-diagonal and zeroes elsewhere, we then define $M = M_{n-1,n}(\Lambda)J_n$. It is clear from the definition of M that $\det(M) = \pm 1$.

We now show that the polynomial $f(MXM^\top)$ behaves as claimed. Recall that if S is a Young tableau, we let $h^j_i(S)$ denote the number of entries changed from i to j when the operator $\mathrm{Sub}_{i\to j}$ is applied to S. Observe that if S is a one-row tableau, then the multilinearity of the Pfaffian and Lemma 2.25 imply

$$[S](E_{i,j}(z)XE_{i,j}(z)^\top) = \begin{cases} [S](X) + z[\mathrm{Sub}_{i\to j}(S)](X) & \text{if } i \text{ appears in } S \text{ but } j \text{ does not} \\ [S](X) & \text{otherwise.} \end{cases}$$

Note that if both i and j appear in S or if neither appear in S, then $S = \mathrm{Sub}_{i\to j}(S)$. Thus, viewing the above as a polynomial in $\mathbb{F}[X][z]$, we see that the leading term is $z^{h^j_i(S)}[\mathrm{Sub}_{i\to j}(S)](X)$. This extends to a multi-row tableau S via

$$[S](E_{i,j}(z)XE_{i,j}(z)^\top) = z^{h^j_i(S)}[\mathrm{Sub}_{i\to j}(S)](X) + O(z^{h^j_i(S)-1}),$$

where $O(z^{h^j_i(S)-1})$ denotes a polynomial in $\mathbb{F}[X][z]$ of degree at most $h^j_i(S) - 1$.

For $i, j \in [2n]$ with $i < j$, let

$$f_{i,j}(X, \Lambda) \coloneqq f(M_{i,j}(\Lambda)XM_{i,j}(\Lambda)^\top).$$

Note that $f(MXM^\top) = f_{n-1,n}(J_nXJ_n^\top)$. We claim that for every $i,j \in [2n]$ with $i<j$, there is a nonempty set $A_{i,j} \subseteq [s]$ such that

$$\mathrm{LC}_{\prec_\Lambda}(f_{i,j}(X,\Lambda)) = \sum_{k\in A_{i,j}} \alpha_k[\mathrm{Sub}_{i\to j}\circ\cdots\mathrm{Sub}_{2\to 3}\circ\mathrm{Sub}_{1\to n}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k)](X).$$

Assuming this, Lemma 3.3 implies

$$\mathrm{LC}_{\prec_\Lambda}(f_{n-1,n}(X,\Lambda)) = \sum_{k\in A_{n-1,n}} \alpha_k[\overline{K}_{\sigma_k}](X).$$

From this, we use the fact that $[\overline{K}_\sigma](J_nXJ_n^\top) = [K_\sigma](X)$ to obtain

$$\begin{aligned}\mathrm{LC}_{\prec_\Lambda}(f(MXM^\top)) &= \mathrm{LC}_{\prec_\Lambda}(f_{n-1,n}(J_nXJ_n^\top))\\ &= \sum_{k\in A_{n-1,n}} \alpha_k[\overline{K}_{\sigma_k}](J_nXJ_n^\top)\\ &= \sum_{k\in A_{n-1,n}} \alpha_k[K_{\sigma_k}](X)\end{aligned}$$

as desired.

It remains to prove the claim about $\mathrm{LC}_{\prec_\Lambda}(f_{i,j}(X,\Lambda))$. We proceed by induction on (i,j) in the order $(1,2)\prec(1,3)\prec\cdots\prec(1,n)\prec(2,3)\prec\cdots\prec(n-1,n)$. Let (i',j') be the predecessor of (i,j) in the $\prec$ order. If $(i,j)=(1,2)$, we set $f_{i',j'}(X,\Lambda)=f(X)$ and $A_{i',j'}=[s]$. Let

$$H_i^j \coloneqq \max_{k\in A_{i',j'}} h_i^j(\mathrm{Sub}_{i'\to j'}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k))$$

and

$$A_{i,j} = \{k\in A_{i',j'} : h_i^j(\mathrm{Sub}_{i'\to j'}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k)) = H_i^j\}.$$

The set $A_{i,j}$ is necessarily nonempty, as H_i^j is obtained by maximizing over a finite nonempty set. By induction, there is some $\overline{e}\in\mathbb{N}^{2n\times 2n}$ such that

$$f_{i',j'}(X,\Lambda) = \Lambda^{\overline{e}}\sum_{k\in A_{i',j'}}\alpha_k[\mathrm{Sub}_{i'\to j'}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k)](X) + g(X,\Lambda),$$

where $g(X,\Lambda)\in\mathbb{F}[X][\Lambda]$ is a polynomial supported on monomials that are smaller than $\Lambda^{\overline{e}}$ in the $\prec_\Lambda$ order. Because $f_{i',j'}$ only depends on $\lambda_{1,2},\ldots,\lambda_{i',j'}$, we know that $\Lambda^{\overline{e}}$ is a monomial consisting of only these variables. Applying the definition of $f_{i,j}$, we then have

$$\begin{aligned}f_{i,j}(X,\Lambda) &= f_{i',j'}(E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top,\Lambda)\\ &= \Lambda^{\overline{e}}\sum_{k\in A_{i',j'}}\alpha_k[\mathrm{Sub}_{i'\to j'}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k)](E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top) + g(E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top,\Lambda)\\ &= \Lambda^{\overline{e}}\lambda_{i,j}^{H_i^j}\sum_{\alpha\in A_{i,j}}\alpha_k[\mathrm{Sub}_{i\to j}\circ\cdots\circ\mathrm{Sub}_{1\to 2}(S_k)](X) + \Lambda^{\overline{e}}p(X,\lambda_{i,j}) + g(E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top,\Lambda),\end{aligned}$$

where $p(X,\lambda_{i,j})\in\mathbb{F}[X][\Lambda]$ is a polynomial of degree at most H_i^j-1 in $\lambda_{i,j}$. Because of this, every monomial of $\Lambda^{\overline{e}}p(X,\lambda_{i,j})$ is smaller than $\Lambda^{\overline{e}}\lambda_{i,j}^{H_i^j}$ in the $\prec_\Lambda$ order. The same holds true for

$g(E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top, \Lambda)$, as the substitution $X \mapsto E_{i,j}(\lambda_{i,j})XE_{i,j}(\lambda_{i,j})^\top$ only changes the $\lambda_{i,j}$-degree of a monomial in $g(X, \Lambda)$ and every monomial of $g(X, \Lambda)$ is already smaller than $\Lambda^{\overline{e}}$ in the $\prec_\Lambda$ order. This implies that

$$\mathrm{LC}_{\prec_\Lambda}(f_{i,j}(X, \Lambda)) = \sum_{k \in A_{i,j}} \alpha_k [\mathrm{Sub}_{i \to j} \circ \cdots \circ \mathrm{Sub}_{1 \to 2}](X)$$

as claimed. □

We now use the result of Lemma 4.1 to construct a change of variables that takes a nonzero $f \in I^{\mathrm{pfaff}}_{2n,2r}$ to $[K_\sigma](X) + O(\varepsilon)$ for a partition σ of width at least $2r$. This is the analogue of Proposition 3.5 for Pfaffians. The proof is similar to that of Proposition 3.5: after applying Lemma 4.1, we scale the rows and columns of X by powers of a new variable y to isolate a single standard monomial $[K_\sigma](X)$.

Proposition 4.2. *Let $f(X) \in I^{\mathrm{pfaff}}_{2n,2r}$ be nonzero. There is a collection of $4n^2$ linearly independent linear functions $\ell_{i,j}(X, \varepsilon) \in \mathbb{F}(\varepsilon)[X]$ indexed by $(i, j) \in [2n] \times [2n]$, an integer $q \in \mathbb{Z}$, a nonzero $\alpha \in \mathbb{F}$, and a partition σ with $\sigma_1 \geqslant 2r$ such that*

$$f(\ell_{1,1}(X, \varepsilon), \ldots, \ell_{2n,2n}(X, \varepsilon)) = \varepsilon^q \alpha [K_\sigma](X) + O(\varepsilon^{q+1}).$$

Proof. Let $M \in \mathbb{F}[\Lambda]^{2n \times 2n}$ be the matrix constructed in Lemma 4.1. Let $f(X) = \sum_{k \in [s]} \alpha_k [S_k](X)$ be the expansion of f as a sum of standard monomials. Then Lemma 4.1 implies

$$\mathrm{LC}_{\prec_\Lambda}(f(MXM^\top)) = \sum_{k \in A} \alpha_k [K_{\sigma_k}](X),$$

where $A \subseteq [s]$ is nonempty and σ_k is the shape of the tableau S_k. From Corollary 2.29, we know that $(\sigma_k)_1 \geqslant 2r$ for all $k \in A$.

Let $d \coloneqq \deg(f(X))$. Let y be a new indeterminate and let $D \in \mathbb{F}[y]^{2n \times 2n}$ be the diagonal matrix given by $D_{i,i} = (d+1)^i$. Observe that $(DXD^\top)_{[k],[k]} = D_{[k],[k]} X_{[k],[k]} D^\top_{[k],[k]}$. Using this and Lemma 2.25, we have

$$\begin{aligned}
\mathrm{Pf}_k(DXD^\top) &= \mathrm{Pf}(D_{[k],[k]} X_{[k],[k]} D^\top_{[k],[k]}) \\
&= \det(D_{[k],[k]}) \, \mathrm{Pf}_k(X) \\
&= y^{\sum_{i=1}^k (d+1)^i} \, \mathrm{Pf}_k(X).
\end{aligned}$$

It then follows that for a partition σ, we have

$$\begin{aligned}
[K_\sigma](DXD^\top) &= \prod_{i=1}^{\hat{\sigma}_1} \mathrm{Pf}_{\sigma_i}(DXD^\top) \\
&= \prod_{i=1}^{\hat{\sigma}_1} y^{\sum_{j=1}^{\sigma_i} (d+1)^j} \, \mathrm{Pf}_{\sigma_i}(X) \\
&= y^{\sum_{i=1}^{\hat{\sigma}_1} \sum_{j=1}^{\sigma_i} (d+1)^j} [K_\sigma](X) \\
&= y^{\sum_{i=1}^{\sigma_1} \hat{\sigma}_i (d+1)^i} [K_\sigma](X).
\end{aligned}$$

Suppose σ and τ are distinct partitions with $\max(\hat{\sigma}_1, \hat{\tau}_1) \leqslant d$. Then we can interpret $\deg_y([K_\sigma](DXD^\top))$ and $\deg_y([K_\tau](DXD^\top))$ as numbers in base $d+1$. Because these numbers differ in at least one place

value, we have $\deg_y([K_\sigma](DXD^\top)) \neq \deg_y([K_\tau](DXD^\top))$. In particular, if σ and τ are distinct shapes of tableaux appearing in the support of $f(X)$, then by our choice of d we have $\max(\hat{\sigma}_1, \hat{\tau}_1) \leqslant d$, so $\deg_y([K_\sigma](DXD^\top)) \neq \deg_y([K_\tau](DXD^\top))$.

Consider the polynomial $f(MDXD^\top M^\top)$. The preceding discussion implies

$$\begin{aligned}\mathrm{LC}_y(\mathrm{LC}_{\prec_\Lambda}(f(MDXD^\top M^\top))) &= \mathrm{LC}_y\left(\sum_{k\in A} \alpha_k [K_{\sigma_k}](DXD^\top)\right) \\ &= \mathrm{LC}_y\left(\sum_{k\in A} y^{\sum_{i=1}^{(\hat{\sigma_k})_1} (\hat{\sigma_k})_i\,(d+1)^i} [K_{\sigma_k}](X)\right) \\ &= \alpha_k [K_{\sigma_k}](X)\end{aligned}$$

for some fixed $k \in A$.

By taking leading coefficients in the ring $\mathbb{F}[X][\Lambda, y]$ with respect to the lexicographic order that sets $\Lambda \succ y$, we then have

$$\mathrm{LC}(f(MDXD^\top M^\top)) = \alpha_k [K_{\sigma_k}](X).$$

Invoking Lemma 2.37 yields a map $\varphi : \Lambda \cup \{y\} \to \{\varepsilon^i : i \in \mathbb{Z}, i \neq 0\}$ that, when extended to a homomorphism $\varphi : \mathbb{F}[X, \Lambda, y] \to \mathbb{F}(\varepsilon)[X]$, gives us

$$\varphi(f(MDXD^\top M^\top)) = \varepsilon^q \alpha_k [K_{\sigma_k}](X) + O(\varepsilon^{q+1})$$

for some $q \in \mathbb{Z}$. Finally, the transformation $X \mapsto \varphi(MD)X\varphi(D^\top M^\top)$ is linear and invertible, since $\det(\varphi(M)) = \pm 1$ and $\det(\varphi(D)) = \varepsilon^m$ for some nonzero $m \in \mathbb{Z}$. □

4.2 Projecting to the Pfaffian

The previous subsection yields a change of variables that takes any nonzero $f \in I^{\mathrm{pfaff}}_{2n,2r}$ to $[K_\sigma](X) + O(\varepsilon)$ for some partition σ of width at least $2r$. As in the case of the determinant, we now want to find a projection of X that takes $[K_\sigma](X)$ to $\mathrm{Pf}_m(X)$ for m as large as possible. Naïvely, we would like to combine Lemma 3.6 with Lemma 2.26 to achieve this. This nearly works, but suffers from the drawback that for a matrix A, the Pfaffians of the leading principal submatrices of

$$\begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix}$$

do not correspond to minors of the leading principal submatrices of A. However, we can amend this by suitably permuting the rows and columns of the above matrix to obtain a new matrix whose leading principal sub-Pfaffians do correspond to minors of leading principal submatrices of A.

Lemma 4.3. *Let A be an $n \times n$ matrix. Then there is a $2n \times 2n$ skew-symmetric matrix M such that for every $k \in [n]$, we have $\mathrm{Pf}(M_{[2k],[2k]}) = \pm \det(A_{[k],[k]})$.*

Proof. Let $\sigma \in S_{2n}$ be the permutation sending $(1, 2, \ldots, 2n)$ to $(1, n+1, 2, n+2, \ldots, n, 2n)$. Let

$$B = \begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix}$$

and let C be the permutation matrix corresponding to σ, i.e., $c_{i,j} = 1$ if and only if $j = \sigma(i)$. Then $M \coloneqq CBC^\top$ is the matrix whose i^{th} row (respectively j^{th} column) is row $\sigma(i)$ (respectively column $\sigma(j)$) of B. We claim that for all $k \in [n]$, we have $\mathrm{Pf}(M_{[2k],[2k]}) = \pm \det(A_{[k],[k]})$.

To see this, let $k \in [n]$ be arbitrary. Let $\tau \in S_{2k}$ be the permutation sending $(1, k+1, 2, k+2, \ldots, k, 2k)$ to $(1, 2, 3, \ldots, 2k)$ and let D be the corresponding permutation matrix. We will show that $\operatorname{Pf}(DM_{[2k],[2k]}D^\top) = \pm\det(A_{[k],[k]})$. By Lemma 2.25, this implies $\operatorname{Pf}(M_{[2k],[2k]}) = \pm\det(A_{[k],[k]})$, so M behaves as desired.

It remains to show that $\operatorname{Pf}(DM_{[2k],[2k]}D^\top) = \pm\det(A_{[k],[k]})$. Note that for $i \in [2k]$, we have

$$\sigma(\tau(i)) = \begin{cases} i & \text{if } i \leqslant k \\ i - k + n & \text{if } i > k. \end{cases}$$

For $i, j \in [2k]$, we have, by definition,

$$\begin{aligned}
(DM_{[2k],[2k]}D^\top)_{i,j} &= (M_{[2k],[2k]})_{\tau(i),\tau(j)} \\
&= M_{\tau(i),\tau(j)} \\
&= (CBC^\top)_{\tau(i),\tau(j)} \\
&= B_{\sigma(\tau(i)),\sigma(\tau(j))} \\
&= \begin{cases} 0 & \text{if } i \leqslant k \text{ and } j \leqslant k \\ A_{i,j} & \text{if } i \leqslant k \text{ and } j > k \\ -A_{j,i} & \text{if } i > k \text{ and } j \leqslant k \\ 0 & \text{if } i > k \text{ and } j > k. \end{cases}
\end{aligned}$$

Thus, the matrix $DM_{[2k],[2k]}D^\top$ is the $2k \times 2k$ matrix given by

$$DM_{[2k],[2k]}D^\top = \begin{pmatrix} 0 & A_{[k],[k]} \\ -A_{[k],[k]}^\top & 0 \end{pmatrix}.$$

It follows from Lemma 2.26 that

$$\operatorname{Pf}(DM_{[2k],[2k]}D^\top) = (-1)^{\binom{k}{2}}\det(A_{[k],[k]})$$

as needed. □

We are now ready to conclude our main result for Pfaffian ideals, an analogue of Theorem 3.8 for Pfaffians. The proof is similar to the proof of Theorem 3.8, but augments the use of Lemma 3.6 with Lemma 4.3.

Theorem 4.4. *Let X be a $2n \times 2n$ generic skew-symmetric matrix. Let $f(X)$ be a nonzero polynomial in the ideal generated by the Pfaffians of the principal $2r \times 2r$ submatrices of X. Let $h(X, \varepsilon) \in \mathbb{F}[\![\varepsilon]\!][X]$ be any polynomial such that $h(X, \varepsilon) = f(X) + O(\varepsilon)$. Let $g(\overline{y}) \in \mathbb{F}[\overline{y}]$ be a polynomial in the border of layered algebraic branching programs with at most r vertices. Then there is a depth-three h-oracle circuit Φ defined over $\mathbb{F}(\varepsilon)$ such that the following hold.*

1. *Φ has nm addition gates at the bottom layer, a single h-oracle gate in the middle layer, and a single addition gate at the top layer.*
2. *If $\operatorname{char}(\mathbb{F}) = 0$, then Φ computes $g(\overline{y}) + O(\varepsilon)$.*
3. *If $\operatorname{char}(\mathbb{F}) = p > 0$, then Φ computes $g(\overline{y})^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.*

Proof. Using Lemma 2.3, we only need to consider the case where the oracle gates compute f exactly. By assumption, there is a polynomial $\widetilde{g}(\overline{y},\varepsilon) \in \mathbb{F}[\varepsilon][\overline{y}]$ such that $\widetilde{g}(\overline{y},\varepsilon) = g(\overline{y}) + O(\varepsilon)$ and $\widetilde{g}(\overline{y},\varepsilon)$ can be computed by a layered algebraic branching program on at most r vertices. Further, Lemma 3.7 yields a homogeneous polynomial $\hat{g}(\overline{y},\varepsilon,z) \in \mathbb{F}[\varepsilon][\overline{y},z]$ computable by a layered algebraic branching program on at most r vertices such that $\hat{g}(\overline{y},\varepsilon,1) = \widetilde{g}(\overline{y},\varepsilon)$. In what follows, we work with $\hat{g}(\overline{y},\varepsilon,z)$.

Applying Proposition 4.2 to $f(X)$ gives us linear functions $\ell_{i,j}(X,\varepsilon) \in \mathbb{F}(\varepsilon)[X]$, an integer $q \in \mathbb{Z}$, and a nonzero $\alpha \in \mathbb{F}$ such that

$$f(\ell_{1,1}(X,\varepsilon),\dots,\ell_{2n,2n}(X,\varepsilon)) = \varepsilon^q \alpha[K_\sigma](X) + O(\varepsilon^{q+1})$$

for some partition σ with $\sigma_1 \geqslant 2r$.

Because $\hat{g}(\overline{y},\varepsilon,z)$ can be computed by a layered ABP on at most r vertices, we can obtain a layered ABP with exactly r vertices that computes $\hat{g}(\overline{y},\varepsilon,z)$ by adding dummy vertices if necessary. Let $A(\overline{y},z) \in \mathbb{F}[\varepsilon][\overline{y},z]^{r\times r}$ be the matrix obtained by applying Lemma 3.6 to $\hat{g}(\overline{y},\varepsilon,z)$. Extend $A(\overline{y},z)$ to an $n \times n$ matrix by adding ones along the diagonal and zeroes elsewhere.

Let $M(\overline{y},z)$ be the $2n \times 2n$ matrix obtained by applying Lemma 4.3 to $A(\overline{y},z)$. Let $\varphi : X \to \mathbb{F}[\varepsilon][\overline{y},z]$ be the substitution given by $\varphi(X) = M(\overline{y},z)$. Under this substitution, we have

$$\begin{aligned}
&f(\ell_{1,1}(\varphi(X),\varepsilon),\dots,\ell_{2n,2n}(\varphi(X),\varepsilon)) \\
&\quad= \varepsilon^q \alpha[K_\sigma](\varphi(X)) + O(\varepsilon^{q+1}) \\
&\quad= \varepsilon^q \alpha \prod_{i=1}^{\hat{\sigma}_1} \mathrm{Pf}_{\sigma_i}(\varphi(X)_{[\sigma_i],[\sigma_i]}) + O(\varepsilon^{q+1}) \\
&\quad= \pm\varepsilon^q \alpha \prod_{i=1}^{\hat{\sigma}_1} \det{}_{\sigma_i/2}(A(\overline{y},z)_{[\sigma_i/2],[\sigma_i/2]}) + O(\varepsilon^{q+1}) \\
&\quad= \pm\varepsilon^q \alpha \prod_{i:\sigma_i \geqslant 2r} \det{}_{\sigma_i/2}(A(\overline{y},z)_{[\sigma_i/2],[\sigma_i/2]}) \cdot \prod_{i:\sigma_i < 2r} \det{}_{\sigma_i/2}(A(\overline{y},z)_{[\sigma_i/2],[\sigma_i/2]}) + O(\varepsilon^{q+1}) \\
&\quad= \pm\varepsilon^q \alpha \prod_{i:\sigma_i \geqslant 2r} (1 + \hat{g}(\overline{y},\varepsilon,z)) + O(\varepsilon^{q+1}).
\end{aligned}$$

Let $h(\overline{y},\varepsilon,z) \coloneqq f(\ell_{1,1}(\varphi(X),\varepsilon),\dots,\ell_{2n,2n}(\varphi(X),\varepsilon))$ and let $t \coloneqq |\{i : \sigma_i \geqslant 2r\}|$. In this notation, the above establishes that $h(\overline{y},\varepsilon,z) = \pm\varepsilon^q\alpha(1+\hat{g}(\overline{y},\varepsilon,z))^t + O(\varepsilon^{q+1})$.

Suppose $\mathrm{char}(\mathbb{F}) = 0$. Let δ be a new indeterminate. By performing the substitutions $y_i \mapsto \delta y_i$ and $z \mapsto \delta$, we obtain

$$\begin{aligned}
h(\delta y_1,\dots,\delta y_m,\varepsilon,\delta) &= \pm\varepsilon^q\alpha(1+\hat{g}(\delta\cdot\overline{y},\varepsilon,\delta))^t + O(\varepsilon^{q+1}) \\
&= \pm\varepsilon^q\alpha\Big(1+\delta^{\deg(\hat{g})}\hat{g}(\overline{y},\varepsilon,1)\Big)^t + O(\varepsilon^{q+1}) \\
&= \pm\varepsilon^q\alpha\Big(1+\delta^{\deg(\hat{g})}g(\overline{y})+O(\varepsilon)\Big)^t + O(\varepsilon^{q+1}) \\
&= \pm\varepsilon^q\alpha\sum_{i=0}^{t}\binom{t}{i}\delta^{t\cdot\deg(\hat{g})}g(\overline{y})^t + O(\varepsilon^{q+1}) \\
&= \pm\varepsilon^q\alpha \pm \varepsilon^q\delta^{\deg(\hat{g})}\alpha t g(\overline{y}) + O(\varepsilon^q\delta^{2\deg(\hat{g})}) + O(\varepsilon^{q+1}).
\end{aligned}$$

Setting

$$\begin{aligned}
\varepsilon &\mapsto \varepsilon^N \\
\delta &\mapsto \varepsilon
\end{aligned}$$

for N sufficiently large yields

$$h(\varepsilon y_1, \ldots, \varepsilon y_m, \varepsilon^N, \varepsilon) = \pm\varepsilon^{qN}\alpha \pm \varepsilon^{qN+\deg(\hat{g})}\alpha t g(\overline{y}) + O(\varepsilon^{qN+\deg(\hat{g})+1}).$$

The claimed f-oracle circuit is then given by

$$\Phi(\overline{y}) \coloneqq \frac{h(\varepsilon y_1, \ldots, \varepsilon y_m, \varepsilon^N, \varepsilon) \mp \varepsilon^{qN}\alpha}{\pm\varepsilon^{qN+\deg(\hat{g})}\alpha t} = g(\overline{y}) + O(\varepsilon).$$

In the case that $\mathrm{char}(\mathbb{F}) = p > 0$, we need to modify the above argument in the case that p divides t. Let $k \in \mathbb{N}$ be such that p^k is the largest power of p that divides t. Write $t = p^k b$. We then have

$$h(\delta y_1, \ldots, \delta y_m, \varepsilon, \delta) = \pm\varepsilon^q\alpha \pm \varepsilon^q\delta^{p^k \deg(\hat{g})}\alpha b g(\overline{y})^{p^k} + O(\varepsilon^q\delta^{2p^k \deg(\hat{g})}) + O(\varepsilon^{q+1}).$$

Again, for sufficiently large N, we can construct an f-oracle circuit that approximately computes g via

$$\Phi(\overline{y}) \coloneqq \frac{h(\varepsilon y_1, \ldots, \varepsilon y_m, \varepsilon^N, \varepsilon) \mp \varepsilon^{qN}\alpha}{\pm\varepsilon^{qN+p^k \deg(\hat{g})}\alpha b} = g(\overline{y})^{p^k} + O(\varepsilon). \qquad \square$$

Since the Pfaffian can be computed efficiently by algebraic branching programs, we immediately obtain the following corollary of Theorem 4.4.

Corollary 4.5. *Let $f(X) \in I^{\mathrm{pfaff}}_{2n,2r}$ be a nonzero polynomial, let $h(X, \varepsilon) \in \mathbb{F}[\![\varepsilon]\!][X]$ be any polynomial such that $h(X, \varepsilon) = f(X) + O(\varepsilon)$, and let $t \leqslant O(r^{1/3})$. Then there is a depth-three h-oracle circuit Φ defined over $\mathbb{F}(\varepsilon)$ with the following properties.*

1. *The bottom layer of Φ consists of $4n^2$ addition gates, the middle layer has a single h-oracle gate, and the top layer has a single addition gate.*
2. *If $\mathrm{char}(\mathbb{F}) = 0$, then Φ computes $\mathrm{Pf}_t(X) + O(\varepsilon)$.*
3. *If $\mathrm{char}(\mathbb{F}) = p > 0$, then Φ computes $\mathrm{Pf}_t(X)^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.*

Proof. Mahajan, Subramanya, and Vinay [MSV04, Theorem 12] constructed a layered algebraic branching program of size $O(n^3)$ that computes the $2n \times 2n$ Pfaffian. Combining this with Theorem 4.4 completes the proof. $\square$

5 Partial Derivatives in Determinantal Ideals

We now proceed to our applications of Theorem 3.8. Our first such application is the determination of the minimum possible value of $\dim(\partial_{<\infty}(f))$ for a nonzero $f \in I^{\mathrm{det}}_{n,m,r}$. The dimension of the space of partial derivatives (and variants thereof) has been successfully used as a complexity measure in proving algebraic circuit lower bounds. Though Theorem 3.8 gives us a tool to prove circuit lower bounds for any nonzero polynomial $f(X) \in I^{\mathrm{det}}_{n,m,r}$, there may be instances where the f-oracle circuit is too costly to implement. For example, if f is computed by a homogeneous or read-once circuit, these properties are not inherited by the oracle circuit. In such cases, it may be useful to have direct estimates for $\dim(\partial_{<\infty}(f))$.

For notational convenience, let

$$\dim(\partial_{<\infty}(I^{\mathrm{det}}_{n,m,r})) \coloneqq \min_{f \in I^{\mathrm{det}}_{n,m,r} \setminus \{0\}} \dim(\partial_{<\infty}(f)).$$

Since $\det_r(X) \in I^{\det}_{n,m,r}$ and $\dim(\partial_{<\infty}(\det_r)) = \binom{2r}{r}$, we clearly have $\dim(\partial_{<\infty}(I^{\det}_{n,m,r})) \leqslant \binom{2r}{r}$. Combining Corollary 3.9 with Lemma 2.18 establishes the existence of a universal constant $c > 0$ such that $\dim(\partial_{<\infty}(I^{\det}_{n,m,r})) \geqslant \dim(\partial_{<\infty}(\det_{cr^{1/3}})) = \binom{2cr^{1/3}}{cr^{1/3}}$.

Alternatively, one can use the observation of Forbes, Shpilka, Tzameret, and Wigderson [FSTW16, Lemma 6.4] that the set of rank-r matrices contains all r-sparse vectors in $\mathbb{F}^{n\times m}$. This implies that rank-r matrices are a hitting set for all polynomials that have a monomial supported on at most r variables. For any $f \in I^{\det}_{n,m,r+1}$, it follows by definition that $f(X)$ vanishes on matrices of rank r. This implies that the leading monomial of $f(X)$ is supported on at least $r+1$ variables. From here, it is straightforward to conclude that there are at least 2^{r+1} distinct leading monomials among the partial derivatives of $f(X)$, which implies the stronger lower bound $\dim(\partial_{<\infty}(I^{\det}_{n,m,r+1})) \geqslant 2^{r+1}$.

In this section, we will show that the naïve upper bound on $\dim(\partial_{<\infty}(I^{\det}_{n,m,r}))$ is tight. That is,

$$\dim(\partial_{<\infty}(I^{\det}_{n,m,r})) = \dim(\partial_{<\infty}(\det_r)) = \binom{2r}{r}.$$

If one interprets $\dim(\partial_{<\infty}(f))$ as a measure of the complexity of f, then this says the $r\times r$ determinant $\det_r(X)$ is of minimal complexity in $I^{\det}_{n,m,r}$.

We will show that $\dim(\partial_{<\infty}((S|T))) \geqslant \binom{2r}{r}$ for any bideterminant $(S|T) \in I^{\det}_{n,m,r}$ and then extend this to all nonzero polynomials in $I^{\det}_{n,m,r}$ using Proposition 3.5. We start by considering partial derivatives with respect to a single variable. Recall that the operator $\frac{\partial^d}{\partial x^d_{i,j}}$ refers to the order-d Hasse derivative with respect to $x_{i,j}$. In the lemma below, we abuse notation and allow a bitableau to have rows whose lengths are not necessarily nonincreasing.

Lemma 5.1. *Let X be an $n \times m$ matrix of variables and let $(S|T)(X)$ be a nonzero bideterminant of shape σ. Let $(i,j) \in [n] \times [m]$ and let $d \coloneqq \operatorname{ideg}_{x_{i,j}}(S|T)(X)$. Then*

$$\frac{\partial^d}{\partial x^d_{i,j}}(S|T)(X) = \pm(S'|T')(X),$$

where (S', T') is the bitableau whose k^{th} row $(S'(k,\bullet), T'(k,\bullet))$ is given by

$$(S'(k,\bullet), T'(k,\bullet)) = \begin{cases} (S(k,\bullet) \setminus \{i\}, T(k,\bullet) \setminus \{j\}) & \text{if } (i,j) \in S(k,\bullet) \times T(k,\bullet) \\ (S(k,\bullet), T(k,\bullet)) & \text{if } (i,j) \notin S(k,\bullet) \times T(k,\bullet). \end{cases}$$

Proof. By definition, we have

$$(S|T) = \prod_{k=1}^{\hat{\sigma}_1} (S(k,\bullet)|T(k,\bullet)).$$

Let $A \subseteq [\hat{\sigma}_1]$ be the set of indices given by

$$A \coloneqq \{k : i \in S(k,\bullet) \text{ and } j \in T(k,\bullet)\}.$$

For $k \in [\hat{\sigma}_1]$, we have

$$\operatorname{ideg}_{x_{i,j}}(S(k,\bullet)|T(k,\bullet)) = \begin{cases} 1 & k \in A, \\ 0 & \text{otherwise.} \end{cases}$$

If $\operatorname{ideg}_{x_{i,j}}(S(k,\bullet)|T(k,\bullet)) = 1$, then expanding the determinant $(S(k,\bullet)|T(k,\bullet))$ by minors gives us

$$\frac{\partial}{\partial x_{i,j}}(S(k,\bullet)|T(k,\bullet)) = \pm(S'(k,\bullet)|T'(k,\bullet)),$$

where $S'(k,\bullet)$ and $T'(k,\bullet)$ are the one-row tableaux obtained by removing i from $S(k,\bullet)$ and j from $T(k,\bullet)$, respectively. Note that for $\ell > \operatorname{ideg}_{x_{i,j}}(S(k,\bullet)|T(k,\bullet))$, we have

$$\frac{\partial^\ell}{\partial x_{i,j}^\ell}(S(k,\bullet)|T(k,\bullet)) = 0.$$

Using the product rule (Lemma 2.16), we then have

$$\begin{aligned}\left(\frac{\partial^d}{\partial x_{i,j}^d}\right)(S|T) &= \left(\frac{\partial^d}{\partial x_{i,j}^d}\right)\left(\prod_{k=1}^{\hat{\sigma}_1}(S(k,\bullet)|T(k,\bullet))\right)\\ &= \sum_{d_1+\cdots+d_{\hat{\sigma}_1}=d}\prod_{k=1}^{\hat{\sigma}_1}\frac{\partial^{d_k}}{\partial x_{i,j}^{d_k}}(S(k,\bullet)|T(k,\bullet))\\ &= \prod_{k=1}^{\hat{\sigma}_1}(\pm(S'(k,\bullet)|T'(k,\bullet)))\\ &= \pm(S'|T') \neq 0.\end{aligned}$$

□

We now extend the preceding lemma to partial derivatives with respect to multiple variables.

Lemma 5.2. *Let $(S|T)$ be a nonzero bideterminant of shape σ. Let $R \subseteq S(1,\bullet)$ and $C \subseteq T(1,\bullet)$ be subsets of the entries in the first row of S and T, respectively, such that $|R| = |C|$. Write $R = \{r_1,\ldots,r_\ell\}$ and $C = \{c_1,\ldots,c_\ell\}$. Then there are positive integers $\{d_1,\ldots,d_\ell\}$ such that*

$$\left(\prod_{i=1}^{\ell}\frac{\partial^{d_i}}{\partial x_{r_i,c_i}^{d_i}}\right)((S|T)) \neq 0.$$

In the case $\operatorname{char}(\mathbb{F}) = 0$*, we may take* $d_1 = \cdots = d_\ell = 1$.

Proof. We prove this via induction on ℓ. The case $\ell = 1$ follows from Lemma 5.1. When $\ell \geqslant 2$, let $d_1 \coloneqq \operatorname{ideg}_{x_{r_1,c_1}}(S|T)$. Lemma 5.1 implies

$$\frac{\partial^{d_1}}{\partial x_{r_1,c_1}^{d_1}}(S|T) = \pm(S'|T'),$$

where (S',T') is the bitableau obtained from (S,T) as in the statement of Lemma 5.1. Let $R' \coloneqq R \setminus \{r_1\}$ and $C' \coloneqq C \setminus \{c_1\}$. Since $S(1,\bullet) \subseteq S'(1,\bullet) \cup \{r_1\}$ and $R \subseteq S(1,\bullet)$, it follows that $R' \subseteq S'(1,\bullet)$. Similarly, we have $C' \subseteq T'(1,\bullet)$. By induction, there are positive integers $d_2,\ldots,d_\ell$ such that

$$\left(\prod_{i=2}^{\ell}\frac{\partial^{d_i}}{\partial x_{r_i,c_i}^{d_i}}\right)((S'|T')) \neq 0.$$

This implies

$$\left(\prod_{i=2}^{\ell}\frac{\partial^{d_i}}{\partial x_{r_i,c_i}^{d_i}}\right)\left(\frac{\partial^{d_1}}{\partial x_{r_1,c_1}^{d_1}}((S|T))\right) \neq 0,$$

so the fact that partial derivatives commute (Lemma 2.15) yields

$$\left(\prod_{i=1}^{\ell}\frac{\partial^{d_i}}{\partial x_{r_i,c_i}^{d_i}}\right)((S|T)) \neq 0.$$

If $\operatorname{char}(\mathbb{F}) = 0$, we also obtain

$$\left(\prod_{i=1}^{\ell} \frac{\partial}{\partial x_{r_i,c_i}}\right)((S|T)) \neq 0. \qquad \square$$

We now use Lemma 5.2 to lower bound the dimension of the space of partial derivatives of any bideterminant.

Proposition 5.3. *Let $(S|T)$ be a nonzero bideterminant of width r. Then $\dim(\partial_{<\infty}((S|T))) \geqslant \binom{2r}{r}$. If $\operatorname{char}(\mathbb{F}) = 0$, then we also have $\dim(\partial_{\leqslant d}((S|T))) \geqslant \sum_{i=0}^{d} \binom{r}{i}^2$.*

Proof. Recall that because $(S|T)$ is of width r, we have $|S(1,\bullet)| = |T(1,\bullet)| = r$. For sets $R \subseteq S(1,\bullet)$ and $C \subseteq T(1,\bullet)$ with $|R| = |C|$, let $R = \{r_1, \ldots, r_\ell\}$ and $C = \{c_1, \ldots, c_\ell\}$ and define

$$\frac{\partial}{\partial x_{R,C}} \coloneqq \prod_{i=1}^{\ell} \frac{\partial^{d_i}}{\partial x_{r_i,c_i}^{d_i}},$$

where $d_1, \ldots, d_\ell$ are obtained by applying Lemma 5.2 to $(S|T)$, R, and C. We will show that

$$D \coloneqq \left\{ \frac{\partial}{\partial x_{R,C}}((S|T)) : R \subseteq S(1,\bullet), C \subseteq T(1,\bullet), |R| = |C| \right\}$$

is a set of linearly independent partial derivatives of $(S|T)$. From this, it follows immediately that

$$\dim(\partial_{<\infty}((S|T))) \geqslant |D| = \sum_{i=0}^{r} \binom{r}{i}^2 = \binom{2r}{r}$$

and, in the case $\operatorname{char}(\mathbb{F}) = 0$,

$$\dim(\partial_{\leqslant d}((S|T))) \geqslant \sum_{i=0}^{d} \binom{r}{i}^2.$$

It remains to show that the elements of D are linearly independent. From Lemma 5.2, we know that $\frac{\partial}{\partial x_{R,C}}((S|T)) \neq 0$. It follows from Lemma 2.14 that

$$\operatorname{multideg}\left(\frac{\partial}{\partial x_{R,C}}((S|T))\right) = \operatorname{multideg}((S|T)) - \left(\sum_{i=1}^{\ell} d_i \overline{e}_{r_i}\right) \oplus \left(\sum_{i=1}^{\ell} d_i \overline{e}_{c_i}\right).$$

Let $R' \subseteq S(1,\bullet)$ and $C' \subseteq T(1,\bullet)$ be such that $(R,C) \neq (R',C')$. From the above, we have

$$\begin{aligned} &\operatorname{multideg}\left(\frac{\partial}{\partial x_{R,C}}((S|T))\right) - \operatorname{multideg}\left(\frac{\partial}{\partial x_{R',C'}}((S|T))\right) \\ &= \left(\sum_{i=1}^{\ell} d'_i \overline{e}_{r'_i} - \sum_{i=1}^{\ell} d_i \overline{e}_{r_i}\right) \oplus \left(\sum_{i=1}^{\ell} d'_i \overline{e}_{c'_i} - \sum_{i=1}^{\ell} d_i \overline{e}_{c_i}\right). \end{aligned}$$

Suppose without loss of generality that $R \neq R'$ and that $r_1 \in R \setminus R'$. Then the r_1 coordinate of $\sum_{i=1}^{\ell} d_i \overline{e}_{r_i} - \sum_{i=1}^{\ell} d'_i \overline{e}_{r'_i}$ is nonzero, so

$$\operatorname{multideg}\left(\frac{\partial}{\partial x_{R,C}}((S|T))\right) \neq \operatorname{multideg}\left(\frac{\partial}{\partial x_{R',C'}}((S|T))\right).$$

The argument when $C \neq C'$ is analogous. Thus, the elements of D are nonzero, multihomogeneous, and of distinct multidegree. Polynomials of differing multidegree are linearly independent, so this immediately implies that the elements of D are linearly independent as desired. $\square$

We now use Proposition 3.5 to extend Proposition 5.3 to all nonzero polynomials in I_r.

Theorem 5.4. *For every nonzero $f \in I^{\det}_{n,m,r}$, we have $\dim(\partial_{<\infty}(f)) \geqslant \binom{2r}{r} = \dim(\partial_{<\infty}(\det_r))$. In the case $\mathrm{char}(\mathbb{F}) = 0$, we also have $\dim(\partial_{\leqslant d}(f)) \geqslant \sum_{i=0}^{d} \binom{r}{i}^2$.*

Proof. Apply Proposition 3.5 to f to obtain linear functions $\ell_{1,1}(X,\varepsilon), \ldots, \ell_{n,m}(X,\varepsilon) \in \mathbb{F}(\varepsilon)[X]$ such that

$$g(X,\varepsilon) \coloneqq \frac{1}{\varepsilon^q} f(\ell_{1,1}(X,\varepsilon), \ldots, \ell_{n,m}(X,\varepsilon)) = \alpha(K_\sigma | K_\sigma)(X) + O(\varepsilon)$$

for some $q \in \mathbb{Z}$, a nonzero $\alpha \in \mathbb{F}$, and a partition σ with $\sigma_1 \geqslant r$.

Note that $g(X,0) = \alpha(K_\sigma|K_\sigma)(X)$. By Proposition 5.3, we have $\dim(\partial_{<\infty}(g(X,0))) \geqslant \binom{2\sigma_1}{\sigma_1} \geqslant \binom{2r}{r}$. This implies that $\dim(\partial_{<\infty}(g(X,\varepsilon)) \geqslant \binom{2r}{r}$. Since the change of variables $x_{i,j} \mapsto \ell_{i,j}(X,\varepsilon)$ is an invertible linear transformation over $\mathbb{F}(\varepsilon)$, Lemma 2.18 implies

$$\dim(\partial_{<\infty}(f)) = \dim(\partial_{<\infty}(g(X,\varepsilon)) \geqslant \binom{2r}{r}.$$

When $\mathrm{char}(\mathbb{F}) = 0$, Proposition 5.3 also yields $\dim(\partial_{\leqslant d}(g(X,0))) \geqslant \sum_{i=0}^{d} \binom{2\sigma_1}{i}^2 \geqslant \sum_{i=0}^{d} \binom{r}{i}^2$. As above, this extends to a lower bound on $\dim(\partial_{\leqslant d}(g(X,\varepsilon)))$, so using Lemma 2.18 we get

$$\dim(\partial_{\leqslant d}(f)) = \dim(\partial_{\leqslant d}(g(X,\varepsilon)) \geqslant \sum_{i=0}^{d} \binom{r}{i}^2. \qquad \square$$

Remark 5.5. The hypothesis $\mathrm{char}(\mathbb{F}) = 0$ in the second part of Theorem 5.4 cannot be avoided in general. If $\mathrm{char}(\mathbb{F}) = p > 0$ and $f \in I^{\det}_{n,m,r} \setminus \{0\}$, then $\frac{\partial}{\partial x_i}(f^p) = 0$ for all i, so $\dim(\partial_{\leqslant 1}(f^p)) = 1 < 1 + r^2$. $\lozenge$

6 Hardness Versus Randomness I: Low-Depth Circuits

A recent breakthrough of Limaye, Srinivasan, and Tavenas [LST21] obtained super-polynomial lower bounds for low-depth algebraic circuits. Combining their result with the hardness-randomness result of Chou, Kumar, and Solomon [CKS19b] yields a deterministic algorithm for identity testing of low-depth algebraic circuits. Specifically, for every fixed $\varepsilon > 0$, they construct an explicit hitting set generator with seed length $O(n^\varepsilon)$ and degree $O(\log n / \log\log n)$ that hits polynomial-size $o(\log\log\log n)$-depth circuits.

In this section, we give an improved construction of a hitting set generator for low-depth circuits. For every $k \in \mathbb{N}$, we construct a generator with seed length $n^{1/2^k + o(1)}$ and degree 2^k that hits polynomial-size $o(\log\log\log n)$-depth circuits. It follows from Lemma 2.6 that the tradeoff between the seed length and degree of our generator is optimal up to the $n^{o(1)}$ factor in the seed length. Our generator is also computable by a circuit of product-depth k and size $n^{1+o(1)}$. As remarked in the introduction, existing techniques in algebraic hardness-randomness produce generators that cannot be computed by circuits smaller than those they hit. Additionally, our generator hits the closure of small low-depth circuits. We note that the generator of Chou, Kumar, and Solomon [CKS19b], when instantiated with a polynomial hard for the border of low-depth circuits, can also be shown to hit the closure of low-depth circuits.

Our result can be interpreted as a hardness-randomness framework for low-depth circuits in an aggressive setting of parameters. In order to instantiate our generator, we need lower bounds on the size of low-depth circuits that compute the determinant, which itself can be computed by

small algebraic branching programs. In contrast, typical hardness-randomness results only require lower bounds for a family of polynomials whose coefficients can be computed explicitly, but the polynomials themselves need not be efficiently computable. In return for these strong lower bound assumptions, we obtain a generator with parameters that improve on known constructions and are near-optimal in the regime of $n^{\Theta(1)}$ seed length.

6.1 Making [LST21] Robust

In this subsection, we establish that the lower bound of Limaye, Srinivasan, and Tavenas [LST21] extends to the border of low-depth circuits. This essentially follows from the fact that they use a rank-based measure to prove their lower bound. The extension of lower bounds based on rank measures to the setting of border complexity is a standard observation in algebraic circuit complexity, but we make this explicit for the sake of completeness. Throughout this subsection, we assume familiarity with the notation and definitions of [LST21].

The proof of Limaye, Srinivasan, and Tavenas [LST21] proceeds in two steps. They first establish a lower bound against low-depth set-multilinear circuits. They then show that a low-depth circuit computing a low-degree set-multilinear polynomial can be made set-multilinear without increasing the depth or size too much. Combined, this establishes a lower bound against general low-depth circuits.

We first observe that the lower bound against set-multilinear circuits is robust. To prove their lower bound, they construct from a given polynomial $f(\overline{x})$ a matrix M_f such that M_f has small rank if f can be computed by a small set-multilinear circuit of low depth.

Lemma 6.1 ([LST21, Claim 16])**.** *Let $k \geqslant 10d$ and let w be any word of length d such that the entries of w are $\lfloor \alpha k \rfloor$ and $-k$ where $\alpha = 1/\sqrt{2}$. Then for any $\Delta \geqslant 1$, any set-multilinear formula Φ of product-depth Δ and size s satisfies*

$$\mathrm{relrk}_w(\Phi) \leqslant s \cdot 2^{-\frac{kd^{1/(2^\Delta - 1)}}{20}}.$$

Next, they show that M_f has large rank when f corresponds to the iterated matrix multiplication polynomial $\mathrm{IMM}_{n,d}(\overline{x})$. To do this, they show that a set-multilinear projection of $\mathrm{IMM}_{n,d}(\overline{x})$ has large rank. This projection behaves nicely in the setting of border complexity, as we describe below.

Lemma 6.2 (cf. [LST21, Lemma 8])**.** *Let $w \in A^d$ be any word which is b-unbiased. If there is a set-multilinear circuit computing $\mathrm{IMM}_{2^b,d}(\overline{x}) + O(\varepsilon)$ of size s and product-depth Δ, then there is also a set-multilinear circuit of size s and product-depth Δ computing $P_w + O(\varepsilon)$ for a polynomial $P_w \in \mathbb{F}_{sm}[\overline{X}(w)]$ such that $\mathrm{relrk}_w(P_w + O(\varepsilon)) \geqslant 2^{-b/2}$.*

Proof. [LST21, Lemma 8] establishes that such a polynomial P_w can be obtained as a set-multilinear projection of $\mathrm{IMM}_{2^b,d}$. Since a nonzero projection of any polynomial in $\varepsilon\mathbb{F}[\varepsilon][\overline{x}]$ remains in $\varepsilon\mathbb{F}[\varepsilon][\overline{x}]$, the same projection takes $\mathrm{IMM}_{2^b,d} + O(\varepsilon)$ to $P_w + O(\varepsilon)$. It is clear that such a projection does not increase the size or product-depth of a circuit. Further, since this projection is set-multilinear, the set-multilinearity of the circuit is preserved.

It remains to show that $\mathrm{relrk}_w(P_w + O(\varepsilon)) \geqslant 2^{-b/2}$. Limaye, Srinivasan, and Tavenas [LST21] show that $\mathrm{relrk}_w(P_w) \geqslant 2^{-b/2}$. Since P_w can be obtained as a projection of $P_w + O(\varepsilon)$ by setting $\varepsilon = 0$, the lower bound on relative rank extends to $P_w + O(\varepsilon)$ for any error term $O(\varepsilon)$. □

Given the preceding lemmas, we now establish lower bounds on the size of low-depth set-multilinear circuits computing $\mathrm{IMM}_{n,d}(\overline{x})$ in the setting of border complexity. The proof is analogous to that of [LST21, Lemma 15].

Lemma 6.3 (cf. [LST21, Lemma 15]). *Let $n, d, \Delta \in \mathbb{N}\setminus\{0\}$ such that $n \geqslant 4^{10d+1}$. Any set-multilinear circuit Φ of product-depth Δ that computes $\mathrm{IMM}_{n,d}(\overline{x}) + O(\varepsilon)$ must have size*

$$n^{\Omega\left(\frac{d^{1/(2^{\Delta}-1)}}{\Delta}\right)}.$$

Given this lower bound, we now implement the second step of [LST21] by lifting this lower bound to general low-depth circuits. Let $f(\overline{x})$ be a set-multilinear polynomial. Limaye, Srinivasan, and Tavenas [LST21] lift their lower bound from set-multilinear circuits to general circuits by giving a non-trivial simulation of low-depth circuits by set-multilinear circuits. Our goal is to perform this same lifting in the border setting: given a low-depth circuit computing $f(\overline{x}) + O(\varepsilon)$, we want to find a low-depth set-multilinear circuit that also computes $f(\overline{x}) + O(\varepsilon)$.

There is a subtle issue in that the error term $O(\varepsilon)$ may not correspond to a set-multilinear polynomial, so we cannot immediately conclude the existence of a low-depth set-multilinear circuit computing $f(\overline{x}) + O(\varepsilon)$. However, if we allow the error term to change, such a transformation is possible. Given a low-depth circuit computing $f(\overline{x}) + O(\varepsilon)$, the set-multilinearization procedure of Limaye, Srinivasan, and Tavenas [LST21] in fact yields a small, low-depth circuit computing the set-multilinear part of $f(\overline{x}) + O(\varepsilon)$. This only modifies the error term, which is permissible in our setting.

Lemma 6.4 (cf. [LST21, Proposition 9]). *Let s, N, and d, be growing parameters with $s \geqslant Nd$ and let $\Delta \in \mathbb{N}$. Assume that $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > d$. If Φ is a circuit of size at most s and product-depth at most Δ computing $P + O(\varepsilon)$ for a set-multilinear polynomial P over the sets of variables $(X_1, \ldots, X_d)$ (with $|X_i| \leqslant N$), then there is a set-multilinear circuit $\widetilde{\Phi}$ of size $d^{O(d)}\mathrm{poly}(s)$ and product-depth at most 2Δ computing $P + O(\varepsilon)$.*

Using Lemma 6.4, we now lift Lemma 6.3 to a lower bound against low-depth circuits without the set-multilinear restriction. The proof is identical to that of [LST21, Corollary 4].

Corollary 6.5 (cf. [LST21, Corollary 4]). *Let $d \leqslant (\log n)/100$ and suppose either $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > d$. Any algebraic circuit of product-depth Δ which computes $\mathrm{IMM}_{n,d}(\overline{x}) + O(\varepsilon)$ must have size at least $n^{d^{\exp(-O(\Delta))}}$.*

6.2 Constructing a Hitting Set Generator

We now use the lower bound of Corollary 6.5 to design hitting set generators for the closure of small low-depth circuits. Of course, a generator with improved parameters can be constructed if one assumes an even stronger lower bound on the size of low-depth circuits needed to compute $\mathrm{IMM}_{n,d}(\overline{x}) + O(\varepsilon)$. For ease of exposition, we directly instantiate our generator with the lower bound of Corollary 6.5.

The generator of Construction 2.8 will act as a basic building block in our construction. In order to make use of this generator, we need to extend Corollary 6.5 to a lower bound for any non-zero polynomial in the ideal $I^{\det}_{n,m,r}$. This essentially follows by combining Corollary 6.5 with Corollary 3.10.

Lemma 6.6. *There is a universal constant $c_{6.6} > 0$ such that the following holds. Let $f(X) \in I^{\det}_{n,m,r}$ be a nonzero polynomial. Assume that either $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > \deg(f)$. Then any circuit of product-depth Δ which computes $f(X) + O(\varepsilon)$ must be of size*

$$r^{(\log r)^{\exp(-c_{6.6}\Delta)}}.$$

Proof. Without loss of generality, we assume that there is no $n' < n$ such that $f \in I^{\det}_{n',m,r}$ and that there is no $m' < m$ such that $f \in I^{\det}_{n,m',r}$. If there is such an n' or m', we may zero out the n^{th} row (respectively m^{th} column) of X without affecting the polynomial f. In particular, we may assume that f depends on at least one variable in each row and column of X, so f depends on at least $\max(n, m)$ variables. This implies that any circuit computing $f + O(\varepsilon)$ must have size at least $s \geqslant \max(n, m)$.

Let Φ be a circuit of size s and product-depth Δ that computes $f(X)+O(\varepsilon)$. Let $d := (\log r)/1000$ and $w := r/\log r$. Using Corollary 3.10, we obtain a circuit $\Psi(\overline{y})$ of size $s + O(n^2m^2)$ and product-depth Δ that computes

$$\Psi(\overline{y}) = \begin{cases} \mathrm{IMM}_{w,d}(\overline{y}) + O(\varepsilon) & \text{if } \mathrm{char}(\mathbb{F}) = 0 \\ \mathrm{IMM}_{w,d}(\overline{y})^{p^k} + O(\varepsilon) & \text{if } \mathrm{char}(\mathbb{F}) = p > 0. \end{cases}$$

In the case $\mathrm{char}(\mathbb{F}) = p > 0$, the fact that Ψ is obtained from Φ by adding a layer of addition gates above and below Φ implies

$$\deg(f) \geqslant \deg(\mathrm{IMM}^{p^k}_{w,d}) = dp^k.$$

By assumption, we have $p > \deg(f)$, so $k = 0$. That is, we have the equality

$$\Psi(\overline{y}) = \mathrm{IMM}_{w,d}(\overline{y}) + O(\varepsilon)$$

both when $\mathrm{char}(\mathbb{F}) = 0$ or when $\mathrm{char}(\mathbb{F}) > \deg(f)$. When r is sufficiently large, we have

$$d = (\log r)/1000 \leqslant (\log r - \log\log r)/100 = (\log w)/100.$$

Corollary 6.5 then implies

$$s + O(n^2m^2) \geqslant w^{d^{\exp(-O(\Delta))}} = r^{(\log r)^{\exp(-O(\Delta))}}.$$

Since $n^2m^2 \leqslant O(s^4)$, we conclude the desired lower bound on s. □

Having established border complexity lower bounds against low-depth circuits for all nonzero polynomials in $I^{\det}_{n,m,r}$, we now turn to polynomial identity testing. Using Lemma 6.6, we show that matrices of low rank are a hitting set for the closure of low-depth circuits.

Lemma 6.7. *Let $\mathbb{F}$ be a field of characteristic zero or characteristic larger than s^Δ. Let $\mathcal{G}_{n,m,r}(Y, Z)$ be the generator defined in Construction 2.8. There is a universal constant $c_{6.7} > 0$ such that for $r = 2^{(\log s)^{1-\exp(-c_{6.7}\Delta)}}$, the map $\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y, Z)$ is a hitting set generator for the closure of n-variate circuits of size s and product-depth Δ.*

Proof. Let $c_{6.6}$ be the constant from Lemma 6.6 and choose $k > 0$ large enough so that $\exp(-c_{6.6}) < (1/2)^{\frac{1}{k-1}}$. Suppose for the sake of contradiction that the statement of the lemma fails for $c_{6.7} = kc_{6.6}$. Then there is a circuit Φ of size s and product-depth Δ which computes $f(X) + O(\varepsilon)$ for some nonzero $f(X)$ such that $f(\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y, Z)) = 0$. By Lemma 2.10, we have $f(X) \in I^{\det}_{\sqrt{n},\sqrt{n},r}$. Since f is computed by a circuit of size s and product-depth Δ, we have $\deg(f) \leqslant s^\Delta$, so either $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > \deg(f)$. The lower bound of Lemma 6.6 implies

$$s \geqslant 2^{(\log r)^{1+\exp(-c_{6.6}\Delta)}} = 2^{(\log s)^{(1-\exp(-c_{6.7}\Delta))(1+\exp(-c_{6.6}\Delta)}}.$$

We claim that

$$(1 - \exp(-c_{6.7}\Delta))(1 + \exp(-c_{6.6}\Delta)) > 1.$$

This would imply $s > s$, a contradiction, which would in turn prove that $\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y,Z)$ is a hitting set generator for the closure.

To prove this inequality, we first note that it suffices to prove the equivalent

$$\exp(-c_{6.6}\Delta) > \exp(-c_{6.7}\Delta) + \exp(-(c_{6.6}+c_{6.7})\Delta).$$

By our choice of $c_{6.7}$ and the fact that $\Delta \geqslant 1$, we have

$$\begin{aligned}\exp(-c_{6.7}\Delta) + \exp(-(c_{6.6}+c_{6.7})\Delta) &< 2\exp(-c_{6.7}\Delta)\\ &= 2\exp(-kc_{6.6}\Delta)\\ &\leqslant 2\exp(-c_{6.6}\Delta)\exp(-(k-1)c_{6.6})\\ &< \exp(-c_{6.6}\Delta),\end{aligned}$$

where the last step follows from our choice of k so that $\exp(-c_{6.6}) \leqslant 2^{\frac{1}{k-1}}$. This establishes the claimed inequality and completes the proof of the lemma. □

Lemma 6.7 constructs a hitting set generator for polynomial-size low-depth circuits with seed length $n^{1/2+o(1)}$ and degree 2. By Lemma 2.6, this seed length is near-optimal for a degree-two generator. To obtain hitting set generators with better seed length, we recursively apply the generator of Lemma 6.7.

Theorem 6.8. *Let $\mathbb{F}$ be a field of characteristic zero. For every fixed $k \in \mathbb{N}$, there is an explicit hitting set generator $\mathcal{G}_k$ for the closure of n-variate, size-s, product-depth $\Delta \leqslant o(\log\log\log n)$ circuits such that $\mathcal{G}_k$ has the following properties.*

1. *$\mathcal{G}_k$ has seed length $n^{1/2^k}s^{o(1)}$.*
2. *$\deg(\mathcal{G}_k) = 2^k$.*
3. *$\mathcal{G}_k$ can be computed by a circuit of product-depth k and size $ns^{o(1)}$. Moreover, each product gate in this circuit has fan-in 2.*

Proof. We proceed via induction on k. Observe that Lemma 6.7 establishes the theorem in the case $k = 1$. When $k \geqslant 2$, let $\mathcal{G}_{k-1}(\overline{w})$ be the generator given by induction and let Φ be a nonzero circuit of size s and product-depth Δ over $\mathbb{F}(\varepsilon)$. By induction, we have that $\mathcal{G}_{k-1}(\overline{w})$ hits Φ even when $\varepsilon = 0$, so $\Phi(\mathcal{G}_{k-1}(\overline{w})) \neq 0$ and $\Phi(\mathcal{G}_{k-1}(\overline{w})) \notin \varepsilon\mathbb{F}[\varepsilon][\overline{w}]$. Further, the composition $\Phi(\mathcal{G}_{k-1}(\overline{w}))$ can be computed by a circuit of product-depth $\Delta + k - 1$ and size $s + ns^{o(1)} \leqslant s^{1+o(1)}$.

Let n_{k-1} and d_{k-1} be the seed length and degree, respectively, of $\mathcal{G}_{k-1}(\overline{w})$. Arrange the variables $\overline{w}$ into a $\sqrt{n_{k-1}} \times \sqrt{n_{k-1}}$ matrix and let

$$\mathcal{G}_k(Y,Z) \coloneqq \mathcal{G}_{k-1}(\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y,Z)),$$

where

$$r_k = 2^{\log(s^{1+o(1)})^{1-\exp(-c_{6.7}(\Delta+k-1))}}.$$

Lemma 6.7 implies that $\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y,Z)$ hits $\Phi(\mathcal{G}_{k-1}(\overline{w}))$ even when $\varepsilon = 0$. Equivalently, the composition $\mathcal{G}_k(Y,Z)$ hits Φ even when $\varepsilon = 0$. We now analyze the parameters of $\mathcal{G}_k(Y,Z)$.

Seed length By definition, the seed length n_k of $\mathcal{G}_k(Y,Z)$ is bounded by

$$n_k \leqslant 2\sqrt{n_{k-1}}r_k.$$

It follows from induction that $n_{k-1} \leqslant n^{1/2^{k-1}} s^{o(1)}$, so we bound the above as

$$n_k \leqslant 2n^{1/2^k} s^{o(1)} r_k.$$

We now bound r_k. As k is fixed and $\Delta \leqslant o(\log\log\log n) \leqslant o(\log\log\log s)$, we have $k + \Delta \leqslant o(\log\log\log s)$. This implies

$$\exp(-c_{6.7}(\Delta + k - 1)) \geqslant \frac{1}{\exp(o(\log\log\log s))} \geqslant \omega\left(\frac{1}{\log\log s}\right).$$

From this, we obtain

$$(\log(s^{1+o(1)}))^{1-\exp(-c_{6.7}(\Delta+k-1))} \leqslant (\log(s^{1+o(1)}))^{1-\omega(\frac{1}{\log\log s})} \leqslant \frac{\log s^{1+o(1)}}{\omega(1)} \leqslant o(\log s).$$

By definition, we have

$$r_k = 2^{\log(s^{1+o(1)})^{1-\exp(-c_{6.7}(\Delta+k-1))}} \leqslant 2^{o(\log s)} \leqslant s^{o(1)}.$$

Thus $n_k \leqslant n^{1/2^k} s^{o(1)}$.

Degree Clearly, we have $\deg(\mathcal{G}_k) = 2\deg(\mathcal{G}_{k-1})$. By induction, $\deg(\mathcal{G}_{k-1}) = 2^{k-1}$, so $\deg(\mathcal{G}_k) = 2^k$.

Circuit size We can compute $\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y, Z)$ with a circuit of product-depth 1 and size $O(n_{k-1}r_k)$. Using induction to bound n_{k-1} and the analysis of the seed length to bound r_k, we have $O(n_{k-1}r_k) \leqslant ns^{o(1)}$. By induction, we can compute $\mathcal{G}_{k-1}(\overline{w})$ with a circuit of product-depth $k-1$ and size $ns^{o(1)}$. Composing these circuits yields a circuit computing $\mathcal{G}_k(Y, Z)$ of product-depth k and size $ns^{o(1)}$. □

Remark 6.9. While Lemma 6.7 holds over fields of sufficiently large positive characteristic, this is not true of Theorem 6.8. This occurs because in our construction of the generator $\mathcal{G}_k$, we apply Lemma 6.7 to a polynomial of degree $s^{\Delta}2^k$. Doing so requires $\mathrm{char}(\mathbb{F}) > s^{\Delta}2^k$ for all k, which is not possible for fields of non-zero characteristic. Of course, for any fixed k, the generator $\mathcal{G}_k$ can be constructed over fields of sufficiently large characteristic. ◊

7 Hardness Versus Randomness II: Formulas

One can mimic the results of Section 6 in the setting of algebraic formulas. While we still lack strong lower bounds for formulas, it seems reasonable to conjecture that neither iterated matrix multiplication nor the determinant can be computed by polynomial-size algebraic formulas. If we strengthen this assumption to a lower bound against border formula complexity, then we can obtain hitting set generators for the closure of small formulas just as in Lemma 6.7 and Theorem 6.8. In this section, we describe this construction.

We start by constructing a generator whose correctness is conditional on the hardness of bideterminants for border formulas. By Theorem 3.8, such lower bounds are implied by lower bounds on the border formula size of any family of polynomials computable by small ABPs, including iterated matrix multiplication and the determinant. Phrasing our results in terms of the border formula complexity of bideterminants allows us to derive hardness-to-randomness results for homogeneous formulas as well as general formulas.

First, we show that lower bounds for bideterminants imply the generator $\mathcal{G}_{n,m,r}$ of Construction 2.8, with appropriate parameters, hits the closure of small formulas. Recall that for a partition σ, the i^{th} row of the tableau K_σ consists of $(1, 2, \ldots, \sigma_i)$.

Lemma 7.1. *Let $\mathbb{F}$ be an arbitrary field. Let $t : \mathbb{N} \to \mathbb{N}$ be a function such that for every partition σ, the border formula complexity of $(K_\sigma|K_\sigma)(X)$ is bounded from below by $t(\sigma_1)$. Let $\mathcal{G}_{n,m,r}(Y,Z)$ be the generator defined in Construction 2.8. Then $\mathcal{G}_{\sqrt{n},\sqrt{n},t^{-1}(2sn)-1}(Y,Z)$ is a hitting set generator for the closure of n-variate formulas of size s.*

If $t : \mathbb{N} \to \mathbb{N}$ instead lower bounds the size of homogeneous formulas computing $(K_\sigma|K_\sigma)(X)+O(\varepsilon)$, then $\mathcal{G}_{\sqrt{n},\sqrt{n},t^{-1}(2sn)-1}$ hits the closure of n-variate size-s homogeneous formulas.

Proof. We first consider non-homogeneous formulas. Let $r \coloneqq t^{-1}(2sn)$. Suppose for the sake of contradiction that $\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y,Z)$ is a not a hitting set generator for the closure of size-s formulas. Then there is some nonzero polynomial $f(X)$ such that $f(\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y,Z)) = 0$ and $f(X)+O(\varepsilon)$ can be computed by a formula of size s. Since $f(\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}) = 0$, Lemma 2.10 implies that $f \in I^{\det}_{\sqrt{n},\sqrt{n},r}$. By Proposition 3.5, there are linear forms $\ell_{1,1}(X,\varepsilon),\dots,\ell_{\sqrt{n},\sqrt{n}}(X,\varepsilon)$, some nonzero $\alpha \in \mathbb{F}$, an integer q, and a partition σ with $\sigma_1 \geqslant r$ such that

$$\frac{1}{\alpha\varepsilon^q} f(\ell_{1,1}(X,\varepsilon),\dots,\ell_{\sqrt{n},\sqrt{n}}(X,\varepsilon)) = (K_\sigma|K_\sigma)(X) + O(\varepsilon).$$

This yields a formula of size sn that computes $(K_\sigma|K_\sigma)(X)+O(\varepsilon)$. This contradicts the assumption that any such formula must be of size at least $t(\sigma_1) \geqslant t(r) \geqslant 2sn$. Thus $\mathcal{G}_{\sqrt{n},\sqrt{n},r-1}(Y,Z)$ is a hitting set generator for the closure of n-variate size-s formulas.

The homogeneous case is analogous. The only difference is that if $f(X)$ is computed by a size-s homogeneous formula, we need to establish that $\frac{1}{\alpha} f(\ell_{1,1}(X,\varepsilon),\dots,\ell_{\sqrt{n},\sqrt{n}}(X,\varepsilon))$ is computable by a homogeneous formula of size sn. This follows immediately from the fact that the $\ell_{i,j}(X,\varepsilon) \in \mathbb{F}(\varepsilon)[X]$ are homogeneous linear polynomials in X. □

Assuming super-polynomial lower bounds on the border formula complexity of bideterminants, we can recursively apply the generator of Lemma 7.1 to obtain generators with smaller seed length. This is analogous to the derivation of Theorem 6.8 from Lemma 6.7. The only difference is in the analysis, as we now have to compute the generator using formulas, not low-depth circuits.

Proposition 7.2. *Let $\mathbb{F}$ be an arbitrary field. Let $t : \mathbb{N} \to \mathbb{N}$ be a function such that for every partition σ, the border formula complexity of $(K_\sigma|K_\sigma)(X)$ is bounded from below by $t(\sigma_1)$. Assume $t(r) \geqslant r^{\omega(1)}$. Then for every fixed $k \in \mathbb{N}$, there is an explicit hitting set generator $\mathcal{G}_k$ for the closure of n-variate size-s (homogeneous) formulas with the following properties.*

1. *$\mathcal{G}_k$ has seed length $n^{1/2^k} s^{o(1)}$.*
2. *$\deg(\mathcal{G}_k) = 2^k$.*
3. *$\mathcal{G}_k$ can be computed by a homogeneous formula of size $ns^{o(1)}$.*

Proof. We use induction on k, noting that the case $k = 1$ follows immediately from Lemma 7.1. When $k \geqslant 2$, let $\mathcal{G}_{k-1}(\overline{w})$ be the generator given by induction and let Φ be a nonzero (homogeneous) formula of size s. By induction, $\mathcal{G}_{k-1}$ hits Φ even when $\varepsilon = 0$, so $\Phi(\mathcal{G}_{k-1}(\overline{w})) \neq 0$ and $\Phi(\mathcal{G}_{k-1}(\overline{w})) \notin \varepsilon\mathbb{F}[\varepsilon][\overline{w}]$. Furthermore, the composition $\Phi(\mathcal{G}_{k-1}(\overline{w}))$ can be computed by a (homogeneous) formula of size $ns^{1+o(1)}$.

Let n_{k-1} and d_{k-1} be the seed length and degree, respectively, of $\mathcal{G}_{k-1}$. Arrange the variables of $\overline{w}$ into a $\sqrt{n_{k-1}} \times \sqrt{n_{k-1}}$ matrix and let

$$\mathcal{G}_k(Y,Z) \coloneqq \mathcal{G}_{k-1}(\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y,Z)),$$

where $r_k \coloneqq t^{-1}(ns^{1+o(1)})$ and $\mathcal{G}_{n,m,r}(Y,Z)$ is the generator of Construction 2.8. By Lemma 7.1, the generator $\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y,Z)$ hits the composition $\Phi(\mathcal{G}_{k-1}(\overline{w}))$ even when $\varepsilon = 0$. Equivalently, $\mathcal{G}_k(Y,Z)$ hits Φ, even when $\varepsilon = 0$. We now analyze the parameters of $\mathcal{G}_k$.

Seed length By construction, $\mathcal{G}_k$ has seed length $2\sqrt{n_{k-1}}r_k$. It follows from induction that $n_{k-1} \leqslant n^{1/2^{k-1}} s^{o(1)}$. By assumption, we have

$$r_k = t^{-1}(ns^{1+o(1)}) \leqslant (ns)^{o(1)} \leqslant s^{o(1)}.$$

This lets us bound the seed length of $\mathcal{G}_k$ by

$$2\sqrt{n_{k-1}}r_k \leqslant n^{1/2^k} s^{o(1)}$$

as claimed.

Degree Clearly $\deg(\mathcal{G}_k) = 2\deg(\mathcal{G}_{k-1})$. By induction, we have $\deg(\mathcal{G}_{k-1}) = 2^{k-1}$, so $\deg(\mathcal{G}_k) = 2^k$.

Formula size Each coordinate of $\mathcal{G}_{\sqrt{n_{k-1}},\sqrt{n_{k-1}},r_k}(Y, Z)$ can be computed by a homogeneous formula of size $2r_k \leqslant s^{o(1)}$. By induction, the generator $\mathcal{G}_{k-1}$ can be computed by a homogeneous formula of size $ns^{o(1)}$. Composing these formulas gives a homogeneous formula of size $ns^{o(1)}$ that computes $\mathcal{G}_k$. □

We now relax the hardness assumption of Proposition 7.2 using Theorem 3.8. This allows us to construct hitting set generators for the closure of small formulas using lower bounds on the border formula complexity of any family of polynomials that can be computed efficiently by algebraic branching programs, including the determinant and iterated matrix multiplication.

Theorem 7.3. *Let $\mathbb{F}$ be a field of characteristic zero. Let $\{f_n(\overline{x}) : n \in \mathbb{N}\}$ be a family of $n^{\Theta(1)}$-variate polynomials such that (1) $f_n(\overline{x})$ is computable by algebraic branching programs of size $n^{\Theta(1)}$, and (2) the border formula complexity of $f_n(\overline{x})$ is bounded from below by $n^{\omega(1)}$. Then the conclusion of Proposition 7.2 holds for formulas; that is, for every fixed $k \in \mathbb{N}$, there is an explicit hitting set generator $\mathcal{G}_k$ for the closure of n-variate size-s formulas with the following properties.*

1. *$\mathcal{G}_k$ has seed length $n^{1/2^k} s^{o(1)}$.*
2. *$\deg(\mathcal{G}_k) = 2^k$.*
3. *$\mathcal{G}_k$ can be computed by a homogeneous formula of size $ns^{o(1)}$.*

Proof. Because the determinant is VBP-complete, the assumed lower bound on the border formula complexity of f_n implies that the border formula complexity of $\det_n(X)$ is bounded from below by $n^{\omega(1)}$. Thus, it suffices to extend this to a lower bound on the border formula complexity of bideterminants as in the hypothesis of Proposition 7.2.

Let X be an $n \times m$ generic matrix and let σ be a partition. Let Φ be a formula of size s which computes $(K_\sigma|K_\sigma)(X) + O(\varepsilon)$. By using Corollary 3.9 and converting the resulting circuit into a formula, we obtain a formula of size $O(sn^2m^2) \leqslant O(s^3)$ which computes $\det_r(X) + O(\varepsilon)$ for $r = \Theta(\sigma_1^{1/3})$. Such a formula must be of size $r^{\omega(1)}$. This implies $s \geqslant r^{\omega(1)}$, which in turn yields $s \geqslant \sigma_1^{\omega(1)}$. Hence the hypothesis of Proposition 7.2 holds, so we obtain the claimed family of generators. □

8 Lower Bounds for the Ideal Proof System

Our final application of Theorem 3.8 is to proof complexity. We construct an unsatisfiable system of equations $\mathcal{F}$ such that no IPS refutation of $\mathcal{F}$ can be computed by a low-depth circuit of polynomial

size. We also show that if the border formula complexity of the determinant is super-polynomial, then polynomial-size formulas cannot refute $\mathcal{F}$.

In general, one cannot immediately transfer circuit lower bounds to proof complexity lower bounds. The difficulty in proving lower bounds on the size of IPS refutations lies in the fact that for a given system of equations $\mathcal{F}$, there are many possible refutations of $\mathcal{F}$ and we must prove a lower bound for each of them. The set of IPS refutations of $\mathcal{F}$ in fact has useful algebraic structure (see [GP18, Section 6]), but to the best of our knowledge this has not been used successfully in proving IPS lower bounds.

Forbes, Shpilka, Tzameret, and Wigderson [FSTW16] developed machinery to derive IPS lower bounds from stronger notions of circuit lower bounds. Specifically, they showed that circuit lower bounds can be lifted to IPS lower bounds if one can prove circuit lower bounds on either (a) circuits that compute a polynomial $f(\overline{x})$ as a function over the boolean hypercube or (b) circuits that compute any multiple of $f(\overline{x})$. Using this approach, they proved $\mathcal{C}$-IPS lower bounds for various restricted circuit classes $\mathcal{C}$.

Recent work by Santhanam and Tzameret [ST21b] constructed a family of CNF formulas that require IPS refutations of super-polynomial size if and only if $\mathsf{VP} \neq \mathsf{VNP}$. To the best of our knowledge, this is the first instance where an algebraic circuit lower bound (without further assumptions) is known to imply a lower bound for IPS. Their result requires the underlying field to be finite; in contrast, we work with fields of characteristic zero, which are necessarily infinite.

Recall that Theorem 3.8 extends circuit lower bounds for $\det(X)$ to circuit lower bounds for the ideal $I^{\det}_{n,m,r}$. Since $I^{\det}_{n,m,r}$ is closed under multiplication by arbitrary polynomials, it is natural to follow the strategy of [FSTW16] and attempt to lift the lower bound for $I^{\det}_{n,m,r}$ to an IPS lower bound. To do this, we need a system of polynomials $f, g_1, \ldots, g_k$ that satisfies the hypothesis of Lemma 2.39 with the additional property that $f \in I^{\det}_{n,m,r}$, where r is not too small compared to n and m. Fortunately, such a system is easy to construct. Recall that for two matrices $A, B \in \mathbb{F}^{n \times m}$, their Hadamard product $A \odot B$ is given by $(A \odot B)_{i,j} \coloneqq a_{i,j} b_{i,j}$. Let X and Y be two $n \times n$ matrices of variables and I_n be the $n \times n$ identity matrix. Consider the system

$$\begin{aligned} \det_n(X) &= 0 \\ XY - I_n &= 0 \\ X \odot X - X &= 0 \\ Y \odot Y - Y &= 0. \end{aligned}$$

This system is unsatisfiable, since $\det_n(X) = 0$ implies that X is not invertible, while $XY - I_n = 0$ implies that X is invertible. However, removing the equation $\det_n(X) = 0$ results in a satisfiable system as witnessed by $X = Y = I_n$. Thus, this system satisfies the hypotheses of Lemma 2.39 and is a natural candidate for IPS lower bounds.

Note the equations $X \odot X - X = 0$ and $Y \odot Y - Y = 0$ can be removed without affecting the hardness of this system. These equations enforce boolean constraints on the variables $x_{i,j}$ and $y_{i,j}$, which is the typical setting of proof complexity.

We now show that lower bounds for $\det(X)$ can be lifted to IPS lower bounds in the setting of low-depth circuits.

Theorem 8.1. *Let X and Y be $n \times n$ matrices of variables. Assume that*

1. *if $\mathrm{char}(\mathbb{F}) = 0$, any product-depth Δ circuit which computes $\det_n(X) + O(\varepsilon)$ must be of size at least $t(n, \Delta)$ for some function $t : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$; and*

2. *if $\mathrm{char}(\mathbb{F}) = p > 0$, any product-depth Δ circuit which computes $\det_n(X)^{p^k} + O(\varepsilon)$ for any $k \in \mathbb{N}$ must be of size at least $t(n, \Delta)$ for some function $t : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$.*

Let C be an IPS refutation of

$$\begin{aligned} \det_n(X) &= 0 \\ XY - I_n &= 0 \\ X \odot X - X &= 0 \\ Y \odot Y - Y &= 0. \end{aligned}$$

Then any product-depth Δ circuit that computes $C(X, Y, z, W, U, V)+O(\varepsilon)$ must be of size $t(\Omega(n^{1/3}), \Delta+1) - O(n^4)$.

Proof. Suppose C can be computed by a circuit of size s and product-depth Δ. The system above is clearly unsatisfiable, as $\det_n(X) = 0$ implies that X is not invertible, whereas $XY - I_n = 0$ implies that X is invertible. Observe that if we omit the equation $\det_n(X) = 0$, then this system becomes satisfiable (take $X = Y = I_n$). Lemma 2.39 implies that

$$1 - C(X, Y, 0, XY - I_n, X \odot X - X, Y \odot Y - Y) = f(X, Y) \det_n(X)$$

for some nonzero $f(X, Y) \in \mathbb{F}[X, Y]$. The coordinates of $XY - I_n$, $X \odot X - X$, and $Y \odot Y - Y$ can be computed by a multi-output circuit of size $O(n^3)$ and product-depth 1. This yields a circuit of size $s + O(n^3)$ and product-depth $\Delta + 1$ that computes $f(X, Y) \det_n(X) + O(\varepsilon)$. It is clear that $f(X, Y) \det_n(X) \in I^{\det}_{2n,n,n}$, where we view $X \cup Y$ as a $2n \times n$ matrix of variables. Using Corollary 3.9, we obtain a product-depth $\Delta + 1$ circuit Φ of size $s + O(n^4)$ such that

1. if $\mathrm{char}(\mathbb{F}) = 0$, then Φ computes $\det_{\Theta(n^{1/3})}(X) + O(\varepsilon)$; and
2. if $\mathrm{char}(\mathbb{F}) = p > 0$, then Φ computes $\det_{\Theta(n^{1/3})}(X)^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.

In both cases, we must have $s + O(n^4) \geqslant t(\Omega(n^{1/3}), \Delta + 1)$, which completes the proof. □

Since Corollary 6.5 establishes unconditional lower bounds on the size of low-depth circuits that border compute $\det_n(X)$ over fields of characteristic zero, we obtain corresponding lower bounds for low-depth IPS.

Corollary 8.2. *Let $\mathbb{F}$ be a field of characteristic zero. Let X and Y be $n \times n$ matrices of variables. Let C be an IPS refutation of*

$$\begin{aligned} \det_n(X) &= 0 \\ XY - I_n &= 0 \\ X \odot X - X &= 0 \\ Y \odot Y - Y &= 0. \end{aligned}$$

Then any product-depth Δ circuit that computes $C(X, Y, z, W, U, V)+O(\varepsilon)$ must be of size $n^{(\log n)^{\exp(-O(\Delta))}}$.

Proof. This follows immediately from Theorem 8.1 and Corollary 6.5. □

Remark 8.3. Over fields of characteristic $p > 0$, bounded-depth IPS can efficiently simulate $\mathsf{AC}^0[p]$-Frege [GP18, Theorem 3.5]. Proving super-polynomial lower bounds on the length of $\mathsf{AC}^0[p]$-Frege proofs is a longstanding open problem in proof complexity. Corollary 8.2 can be seen as a step towards resolving this problem. In order to obtain $\mathsf{AC}^0[p]$-Frege lower bounds, two obstacles must be overcome. First, one must extend the lower bound of Limaye, Srinivasan, and Tavenas [LST21]

to hold over fields of small characteristic and to hold for p^{th} powers of the determinant. Second, it is necessary to prove an IPS lower bound for a system of equations that arises from the encoding of a CNF formula. Our system is not the encoding of a CNF; the IPS lower bounds of Forbes, Shpilka, Tzameret, and Wigderson [FSTW16] also suffer from this drawback. ◊

We can also carry out the reasoning of Theorem 8.1 with formulas instead of low-depth circuits. The resulting formula-IPS lower bound is conditional, as we currently lack good lower bounds on the formula size of any explicit polynomial, let alone the determinant.

Theorem 8.4. *Let X and Y be $n \times n$ matrices of variables. Assume that*

1. *if $\operatorname{char}(\mathbb{F}) = 0$, any formula which computes $\det_n(X) + O(\varepsilon)$ must be of size at least $t(n)$ for some function $t : \mathbb{N} \to \mathbb{N}$; and*
2. *if $\operatorname{char}(\mathbb{F}) = p > 0$, any formula which computes $\det_n(X)^{p^k} + O(\varepsilon)$ for any $k \in \mathbb{N}$ must be of size at least $t(n)$ for some function $t : \mathbb{N} \to \mathbb{N}$.*

Let C be an IPS refutation of

$$\begin{aligned} \det_n(X) &= 0 \\ XY - I_n &= 0 \\ X \odot X - X &= 0 \\ Y \odot Y - Y &= 0. \end{aligned}$$

Then any formula that computes $C(X, Y, z, W, U, V) + O(\varepsilon)$ must be of size $\Omega\left(\frac{t(\Omega(n^{1/3}))}{n^3}\right)$.

Proof. Suppose C can be computed by a formula of size s. As in the proof of Theorem 8.1, we deduce from Lemma 2.39 that

$$1 - C(X, Y, 0, XY - I_n, X \odot X - X, Y \odot Y - Y) = f(X, Y) \det_n(X)$$

for some nonzero $f(X, Y) \in \mathbb{F}[X, Y]$. The coordinates of $XY - I_n$, $X \odot X - X$, and $Y \odot Y - Y$ can each be computed by a formula of size $O(n)$. This yields a formula of size $O(sn)$ that computes $f(X, Y) \det_n(X) + O(\varepsilon)$. From Corollary 3.9, we obtain a formula Φ of size $O(sn^3)$ such that

1. if $\operatorname{char}(\mathbb{F}) = 0$, then Φ computes $\det_{\Theta(n^{1/3})}(X) + O(\varepsilon)$; and
2. if $\operatorname{char}(\mathbb{F}) = p > 0$, then Φ computes $\det_{\Theta(n^{1/3})}(X)^{p^k} + O(\varepsilon)$ for some $k \in \mathbb{N}$.

By assumption, we must have $O(sn^3) \geqslant t(\Omega(n^{1/3}))$, which implies the desired lower bound on s. □

The previous results show that, in the setting of border complexity, the task of computing the $\Theta(n^{1/3}) \times \Theta(n^{1/3})$ determinant can be reduced to computing any IPS refutation of the system $\{\det_n(X) = 0, XY - I_n = 0\}$. We complement this by constructing a depth-three $\det_n$-oracle circuit that computes a refutation of this system. Together with our previous results, this shows that for sufficiently well-behaved circuit classes, the approximate complexity of refuting $\{\det_n(X) = 0, XY - I_n = 0\}$ is bounded from below and above by the approximate complexity of computing the $\Theta(n^{1/3}) \times \Theta(n^{1/3})$ and $n \times n$ determinants, respectively.

Proposition 8.5. *Let $\mathbb{F}$ be any field and let X and Y be $n \times n$ matrices of variables. Then the following hold.*

1. *There is an $O(n^2)$-size depth-three circuit with $\det_n$-oracle gates that computes an IPS refutation of the system $\{\det_n(X) = 0, XY - I_n = 0\}$.*

2. *There is an $O(n^2)$-size depth-three circuit with $(\det_n + O(\varepsilon))$-oracle gates that approximately computes an IPS refutation of the system $\{\det_n(X) = 0, XY - I_n = 0\}$.*

Proof. Item (2) follows immediately from (1) using Lemma 2.3, so it suffices to prove (1). Let Z be an $n \times n$ matrix of variables and let w be an additional variable. We claim that

$$C(X, Y, w, Z) \coloneqq 1 - \det_n(Z + I_n) + w \cdot \det_n(Y)$$

is an IPS refutation of the system $\{\det_n(X) = 0, XY - I_n = 0\}$. It is clear from the expression above that $C(X, Y, w, Z)$ can be computed by a depth-three circuit with $\det_n$-oracle gates. To see that $C(X, Y, w, Z)$ is a valid IPS refutation, we have

$$C(X, Y, 0, 0) = 1 - \det_n(I_n) + 0 = 0$$

and

$$\begin{aligned} C(X, Y, \det_n(X), XY - I_n) &= 1 - \det_n(XY - I_n + I_n) + \det_n(X)\det_n(Y) \\ &= 1 - \det_n(XY) + \det_n(XY) \\ &= 1. \end{aligned}$$

Thus $C(X, Y, w, Z)$ is an IPS refutation of the system $\{\det_n(X) = 0, XY - I_n = 0\}$. □

We end with a brief discussion on the hard instance used in this section.

Remark 8.6. Grochow and Pitassi [GP18, Example A.6] showed that a short IPS refutation of $\{\det_n(X) = 0, XY - I_n = 0\}$ can be used to construct a short IPS proof of the inversion principle $XY = I_n \implies YX = I_n$. The inversion principle is one of the "hard matrix identities" of Soltys and Cook [SC04], which are four tautologies proposed as candidates for separating the Frege and Extended Frege proof systems. Unfortunately, our methods are not able to prove lower bounds, conditional or otherwise, on the size of IPS proofs of the hard matrix identities. ◊

References

[AD80] Silvana Abeasis and Alberto Del Fra. "Young diagrams and ideals of Pfaffians". In: *Adv. in Math.* 35.2 (1980), pp. 158–178 (cit. on p. 33).

[AFSSV18] 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". In: *ACM Trans. Comput. Theory* 10.1 (2018) (cit. on p. 3).

[AGHT20] 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: *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC 2020)*. Chicago, IL, USA: Association for Computing Machinery, 2020, pp. 54–67 (cit. on p. 5).

[AGKS15] Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena. "Hitting-Sets for ROABP and Sum of Set-Multilinear Circuits". In: *SIAM J. Comput.* 44.3 (2015), pp. 669–697 (cit. on p. 3).

[Ale21] Yaroslav Alekseev. "A Lower Bound for Polynomial Calculus with Extension Rule". In: *36th Computational Complexity Conference (CCC 2021)*. Ed. by Valentine Kabanets. Vol. 200. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 21:1–21:18 (cit. on p. 9).

[And20] Robert Andrews. "Algebraic Hardness Versus Randomness in Low Characteristic". In: *35th Computational Complexity Conference (CCC 2020)*. Ed. by Shubhangi Saraf. Vol. 169. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, 37:1–37:32 (cit. on p. 32).

[AvMV15] Matthew Anderson, Dieter van Melkebeek, and Ilya Volkovich. "Deterministic polynomial identity tests for multilinear bounded-read formulae". In: *Computational Complexity* 24 (2015), pp. 695–776 (cit. on p. 3).

[BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. "$O(n^{2.7799})$ complexity for $n \times n$ approximate matrix multiplication". In: *Information Processing Letters* 8.5 (1979), pp. 234–235 (cit. on p. 2).

[BCS97] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. "Algebraic complexity theory". Vol. 315. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. With the collaboration of Thomas Lickteig. Springer-Verlag, Berlin, 1997, pp. xxiv+618 (cit. on p. 10).

[BDI21] Markus Bläser, Julian Dörfler, and Christian Ikenmeyer. "On the Complexity of Evaluating Highest Weight Vectors". In: *36th Computational Complexity Conference (CCC 2021)*. Ed. by Valentine Kabanets. Vol. 200. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 29:1–29:36 (cit. on p. 2).

[BIK+96] Sam Buss, Russell Impagliazzo, Jan Krajíček, Pavel Pudlák, Alexander A. Razborov, and Jiři Sgall. "Proof complexity in algebraic systems and bounded depth Frege systems with modular counting". In: *Computational Complexity* 6 (1996), pp. 256–298 (cit. on p. 5).

[BIKPP96] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. "Lower bounds on Hilbert's Nullstellensatz and propositional proofs". In: *Proceedings of the London Mathematical Society* 73.3 (1996). Preliminary version in the *35th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1994)*, pp. 1–26 (cit. on p. 5).

[Bin80] Dario Bini. "Relations between exact and approximate bilinear algorithms. Applications". In: *Calcolo* 17 (1980), pp. 87–97 (cit. on p. 2).

[BS21] Pranav Bisht and Nitin Saxena. "Blackbox identity testing for sum of speacial ROABPs and its border class". In: *Computational Complexity* 30.8 (2021), pp. 1–48 (cit. on p. 3).

[BSV20] Vishwas Bhargava, Shubhangi Saraf, and Ilya Volkovich. "Deterministic Factorization of Sparse Polynomials with Bounded Individual Degree". In: *J. ACM* 67.2 (2020). Preliminary version in the *59th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2018)*, 8:1–8:28 (cit. on p. 2).

[Bür00] Peter Bürgisser. "Completeness and Reduction in Algebraic Complexity Theory". Springer-Verlag Berlin Heidelberg, 2000 (cit. on p. 2).

[Bür04] Peter Bürgisser. "The complexity of factors of multivariate polynomials". In: *Foundations of Computational Mathematics* 4.4 (2004), pp. 369–396 (cit. on pp. 2, 3, 11, 20).

[BV88] Winfried Bruns and Udo Vetter. "Determinantal rings". Vol. 1327. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1988, pp. viii+236 (cit. on p. 13).

[CEI96] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. "Using the Groebner Basis Algorithm to Find Proofs of Unsatisfiability". In: *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC 1996)*. Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, pp. 174–183 (cit. on p. 5).

[CKS19a] Chi-Ning Chou, Mrinal Kumar, and Noam Solomon. "Closure of VP under taking factors: a short and simple proof". `arXiv:1903.02366`. 2019 (cit. on p. 2).

[CKS19b] Chi-Ning Chou, Mrinal Kumar, and Noam Solomon. "Closure Results for Polynomial Factorization". In: *Theory of Computing* 15.13 (2019). Preliminary version in the *33rd Annual Computational Complexity Conference (CCC 2018)*, pp. 1–34 (cit. on pp. 2, 4, 8, 32, 43).

[dCEP80] Corrado de Concini, David Eisenbud, and Claudio Procesi. "Young diagrams and determinantal varieties". In: *Invent. Math.* 56.2 (1980), pp. 129–165 (cit. on pp. 17, 23, 24, 33).

[dCP76] Corrado de Concini and Claudio Procesi. "A characteristic free approach to invariant theory". In: *Advances in Math.* 21.3 (1976), pp. 330–354 (cit. on p. 19).

[DDS21a] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. "Demystifying the border of depth-3 algebraic circuits". In: *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2021)*. 2021 (cit. on p. 2).

[DDS21b] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. "Deterministic Identity Testing Paradigms for Bounded Top-Fanin Depth-4 Circuits". In: *36th Computational Complexity Conference (CCC 2021)*. Ed. by Valentine Kabanets. Vol. 200. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 11:1–11:27 (cit. on p. 3).

[DKR78] Jacques Désarménien, Joseph P. S. Kung, and Gian-Carlo Rota. "Invariant theory, Young bitableaux, and combinatorics". In: *Advances in Math.* 27.1 (1978), pp. 63–92 (cit. on p. 17).

[DRS74] Peter Doubilet, Gian-Carlo Rota, and Joel Stein. "On the foundations of combinatorial theory. IX. Combinatorial methods in invariant theory". In: *Studies in Applied Mathematics* 53 (1974), pp. 185–216 (cit. on pp. 17, 19).

[DS07] Zeev Dvir and Amir Shpilka. "Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits". In: *SIAM J. Comput.* 36.5 (2007), pp. 1404–1434. Preliminary version in the *37th Annual ACM Symposium on Theory of Computing (STOC 2005)* (cit. on p. 3).

[DSS18] Pranjal Dutta, Nitin Saxena, and Amit Sinhababu. "Discovering the roots: uniform closure results for algebraic classes under factoring". In: *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC 2018)*. 2018, pp. 1152–1165 (cit. on p. 2).

[DSY09] Zeev Dvir, Amir Shpilka, and Amir Yehudayoff. "Hardness-Randomness Tradeoffs for Bounded Depth Arithmetic Circuits". In: *SIAM J. Comput.* 39.4 (2009), pp. 1279–1293 (cit. on pp. 2, 4, 8, 32).

[For14] Michael A. Forbes. "Polynomial identity testing of read-once oblivious algebraic branching programs". PhD thesis. Massachusetts Institute of Technology, Cambridge, MA, USA, 2014 (cit. on pp. 14, 15).

[For16] Michael A. Forbes. "Some concrete questions on the border complexity of polynomials". Talk presented at the Workshop on Algebraic Complexity Theory (WACT), Tel Aviv. 2016 (cit. on p. 2).

[FS13] Michael A. Forbes and Amir Shpilka. "Quasipolynomial-Time Identity Testing of Noncommutative and Read-Once Oblivious Algebraic Branching Programs". In: *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)*. 2013, pp. 243–252 (cit. on p. 3).

[FS18] Michael A. Forbes and Amir Shpilka. "A PSPACE Construction of a Hitting Set for the Closure of Small Algebraic Circuits". In: *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC 2018)*. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 1180–1192 (cit. on pp. 4, 12).

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. "Hitting sets for multilinear read-once algebraic branching programs, in any order". In: *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*. 2014, pp. 867–875 (cit. on p. 3).

[FSTW16] Michael A. Forbes, Amir Shpilka, Iddo Tzameret, and Avi Wigderson. "Proof Complexity Lower Bounds from Algebraic Circuit Complexity". In: *Proceedings of the 31st Annual Computational Complexity Conference (CCC 2016)*. 2016, 32:1–32:17 (cit. on pp. 3, 5–7, 9, 22, 40, 51, 53).

[GG20] Zeyu Guo and Rohit Gurjar. "Improved Explicit Hitting-Sets for ROABPs". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Ed. by Jarosław Byrka and Raghu Meka. Vol. 176. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, 4:1–4:16 (cit. on p. 3).

[GKS17] Rohit Gurjar, Arpita Korwar, and Nitin Saxena. "Identity Testing for Constant-Width, and Any-Order, Read-Once Oblivious Arithmetic Branching Programs". In: *Theory of Computing* 13.1 (2017), pp. 1–21 (cit. on p. 3).

[GKST17] Rohit Gurjar, Arpita Korwar, Nitin Saxena, and Thomas Thierauf. "Deterministic Identity Testing for Sum of Read-Once Oblivious Arithmetic Branching Programs". In: *Computational Complexity* 26.4 (2017), pp. 835–880 (cit. on p. 3).

[GP18] Joshua A. Grochow and Toniann Pitassi. "Circuit Complexity, Proof Complexity, and Polynomial Identity Testing: The Ideal Proof System". In: *J. ACM* 65.6 (Nov. 2018), 37:1–37:59 (cit. on pp. 4, 5, 22, 51, 52, 54).

[Gro20] Joshua A. Grochow. "Complexity in ideals of polynomials: questions on algebraic complexity of circuits and proofs". In: *Bull. EATCS* 130 (2020) (cit. on pp. 1, 5, 6).

[GSS19] Zeyu Guo, Nitin Saxena, and Amit Sinhababu. "Algebraic Dependencies and PSPACE Algorithms in Approximative Complexity over Any Field". In: *Theory of Computing* 15.16 (2019), pp. 1–30 (cit. on p. 4).

[Has36] Helmut Hasse. "Theorie der höheren Differentiale in einem algebraischen Funktionenkörper mit vollkommenem Konstantenkörper bei beliebiger Charakteristik". In: *J. Reine Angew. Math.* 175 (1936), pp. 50–54 (cit. on p. 14).

[IMP20] Russell Impagliazzo, Sasank Mouli, and Toniann Pitassi. "The Surprising Power of Constant Depth Algebraic Proofs". In: *Proceedings of the Thirty fifth Annual IEEE Symposium on Logic in Computer Science (LICS 2020)*. Saarbrucken, Germany: IEEE Computer Society Press, July 2020, pp. 591–603 (cit. on p. 9).

[IPS99] Russell Impagliazzo, Pavel Pudlák, and Jiří Sgall. "Lower bounds for the polynomial calculus and the Gröbner basis algorithm". In: *Computational Complexity* 8 (1999), pp. 127–144 (cit. on p. 5).

[Kal87] Erich Kaltofen. "Single-Factor Hensel Lifting and its Application to the Straight-Line Complexity of Certain Polynomials". In: *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*. 1987, pp. 443–452 (cit. on pp. 2, 32).

[Kay12] Neeraj Kayal. "An exponential lower bound for the sum of powers of bounded degree polynomials". Electronic Colloquium on Computational Complexity (ECCC), Technical Report TR12-081. 2012 (cit. on p. 15).

[KI04] Valentine Kabanets and Russell Impagliazzo. "Derandomizing Polynomial Identity Tests Means Proving Circuit Lower Bounds". In: *Computational Complexity* 13.1-2 (2004), pp. 1–46 (cit. on pp. 2–4).

[KMSV13] Zohar S. Karnin, Partha Mukhopadhyay, Amir Shpilka, and Ilya Volkovich. "Deterministic Identity Testing of Depth-4 Multilinear Circuits with Bounded Top Fan-in". In: *SIAM Journal on Computing* 42.6 (2013), pp. 2114–2131 (cit. on p. 3).

[Kra19] Jan Krajíček. "Proof Complexity". Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2019 (cit. on p. 4).

[KS01] Adam R. Klivans and Daniel Spielman. "Randomness Efficient Identity Testing of Multivariate Polynomials". In: *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)*. Hersonissos, Greece: Association for Computing Machinery, 2001, pp. 216–223 (cit. on p. 3).

[KS07] Neeraj Kayal and Nitin Saxena. "Polynomial identity testing for depth 3 circuits". In: *Comput. Complexity* 16.2 (2007), pp. 115–138. Preliminary version in the *21st Annual IEEE Conference on Computational Complexity (CCC 2006)* (cit. on p. 3).

[KS09] Neeraj Kayal and Shubhangi Saraf. "Blackbox polynomial identity testing for depth 3 circuits". In: *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*. IEEE Computer Soc., Los Alamitos, CA, 2009, pp. 198–207 (cit. on p. 3).

[KS11] Zohar S. Karnin and Amir Shpilka. "Black box polynomial identity testing of generalized depth-3 arithmetic circuits with bounded top fan-in". In: *Combinatorica* 31.3 (2011), pp. 333–364. Preliminary version in the *23rd Annual IEEE Conference on Computational Complexity (CCC 2008)* (cit. on p. 3).

[KSS15] Swastik Kopparty, Shubhangi Saraf, and Amir Shpilka. "Equivalence of Polynomial Identity Testing and Polynomial Factorization". In: *Computational Complexity* 24.2 (2015), pp. 295–331 (cit. on p. 32).

[LST21] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. "Superpolynomial Lower Bounds Against Low-Depth Algebraic Circuits". In: *Proceedings of the 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2021)*. Preliminary version in the Electronic Colloquium on Computational Complexity (ECCC), Technical Report TR21-081. 2021 (cit. on pp. 1, 3, 8, 9, 43–45, 52).

[MS01] Ketan Mulmuley and Milind A. Sohoni. "Geometric Complexity Theory I: An Approach to the P vs. NP and Related Problems". In: *SIAM J. Comput.* 31.2 (2001), pp. 496–526 (cit. on p. 2).

[MS21] Dori Medini and Amir Shpilka. "Hitting Sets and Reconstruction for Dense Orbits in VP_e and $\Sigma\Pi\Sigma$ Circuits". In: *36th Computational Complexity Conference (CCC 2021)*. Ed. by Valentine Kabanets. Vol. 200. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 19:1–19:27 (cit. on p. 4).

[MSV04] Meena Mahajan, P. R. Subramanya, and V. Vinay. "The combinatorial approach yields an NC algorithm for computing Pfaffians". In: *Discrete Appl. Math.* 143.1-3 (2004), pp. 1–16 (cit. on p. 39).

[MV18] Daniel Minahan and Ilya Volkovich. "Complete Derandomization of Identity Testing and Reconstruction of Read-Once Formulas". In: *ACM Trans. Comput. Theory* 10.3 (2018) (cit. on p. 3).

[MV97] Meena Mahajan and V. Vinay. "Determinant: Combinatorics, Algorithms, and Complexity". In: *Chicago Journal of Theoretical Computer Science* 1997.5 (1997) (cit. on p. 32).

[Oli16] Rafael Oliveira. "Factors of low individual degree polynomials". In: *Computational Complexity* 25.2 (2016), pp. 507–561. Preliminary version in the *30th Annual Computational Complexity Conference (CCC 2015)* (cit. on p. 2).

[OSV16] Rafael Oliveira, Amir Shpilka, and Ben Lee Volk. "Subexponential Size Hitting Sets for Bounded Depth Multilinear Formulas". In: *Computational Complexity* 25 (2016), pp. 455–505 (cit. on p. 3).

[PS20] Shir Peleg and Amir Shpilka. "A Generalized Sylvester-Gallai Type Theorem for Quadratic Polynomials". In: *35th Computational Complexity Conference (CCC 2020)*. Ed. by Shubhangi Saraf. Vol. 169. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, 8:1–8:33 (cit. on p. 3).

[PS21] Shir Peleg and Amir Shpilka. "Polynomial Time Deterministic Identity Testing Algorithm for $\Sigma[3]\Pi\Sigma\Pi[2]$ Circuits via Edelstein–Kelly Type Theorem for Quadratic Polynomials". In: *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC 2021)*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 259–271 (cit. on p. 3).

[PT16] Toniann Pitassi and Iddo Tzameret. "Algebraic Proof Complexity: Progress, Frontiers and Challenges". In: *ACM SIGLOG News* 3.3 (Aug. 2016), pp. 21–43 (cit. on p. 6).

[Raz98] Alexander A. Razborov. "Lower bounds for the polynomial calculus". In: *Computational Complexity* 7 (1998), pp. 291–324 (cit. on p. 5).

[Rob85] Lorenzo Robbiano. "Term orderings on the polynomial ring". In: *EUROCAL '85*. Ed. by Bob F. Caviness. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985, pp. 513–517 (cit. on p. 20).

[Rob86] Lorenzo Robbiano. "On the theory of graded structures". In: *Journal of Symbolic Computation* 2.2 (1986), pp. 139–170 (cit. on p. 20).

[Sap19] Ramprasad Saptharishi. "A survey of lower bounds in arithmetic circuit complexity". `https://github.com/dasarpmar/lowerbounds-survey`. 2019 (cit. on pp. 7, 10).

[SC04] Michael Soltys and Stephen Cook. "The proof complexity of linear algebra". In: *Annals of Pure and Applied Logic* 130.1 (2004), pp. 277–323 (cit. on p. 54).

[Sch80] Jacob T. Schwartz. "Fast Probabilistic Algorithms for Verification of Polynomial Identities". In: *J. ACM* 27.4 (1980), pp. 701–717 (cit. on p. 3).

[Shp19] Amir Shpilka. "Sylvester-Gallai Type Theorems for Quadratic Polynomials". In: *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC 2019)*. Phoenix, AZ, USA: Association for Computing Machinery, 2019, pp. 1203–1214 (cit. on p. 3).

[SS11] Nitin Saxena and C. Seshadhri. "An almost optimal rank bound for depth-3 identities". In: *SIAM J. Comput.* 40.1 (2011), pp. 200–224. Preliminary version in the *24th Annual IEEE Conference on Computational Complexity (CCC 2009)* (cit. on p. 3).

[SS12] Nitin Saxena and C. Seshadhri. "Blackbox identity testing for bounded top-fanin depth-3 circuits: the field doesn't matter". In: *SIAM J. Comput.* 41.5 (2012), pp. 1285–1298. Preliminary version in the *43rd Annual ACM Symposium on Theory of Computing (STOC 2011)* (cit. on p. 3).

[SS13] Nitin Saxena and C. Seshadhri. "From Sylvester-Gallai configurations to rank bounds: improved blackbox identity test for depth-3 circuits". In: *J. ACM* 60.5 (2013), 33:1–33:33. Preliminary version in the *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)* (cit. on p. 3).

[ST20] Amit Sinhababu and Thomas Thierauf. "Factorization of Polynomials Given By Arithmetic Branching Programs". In: *Proceedings of the 35th Annual Computational Complexity Conference (CCC 2020)*. Ed. by Shubhangi Saraf. Vol. 169. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, 33:1–33:19 (cit. on p. 2).

[ST21a] Chandan Saha and Bhargav Thankey. "Hitting Sets for Orbits of Circuit Classes and Polynomial Families". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Ed. by Mary Wootters and Laura Sanità. Vol. 207. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 50:1–50:26 (cit. on p. 4).

[ST21b] Rahul Santhanam and Iddo Tzameret. "Iterated Lower Bound Formulas: A Diagonalization-Based Approach to Proof Complexity". In: *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC 2021)*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 234–247 (cit. on pp. 5, 51).

[SV15] Amir Shpilka and Ilya Volkovich. "Read-once polynomial identity testing". In: *Computational Complexity* 27 (2015), pp. 477–532 (cit. on pp. 3, 12).

[SV18] Shubhangi Saraf and Ilya Volkovich. "Black-Box Identity Testing of Depth-4 Multilinear Circuits". In: *Combinatorica* 38 (2018), pp. 1205–1238 (cit. on p. 3).

[SY10] Amir Shpilka and Amir Yehudayoff. "Arithmetic Circuits: A survey of recent results and open questions". In: *Foundations and Trends in Theoretical Computer Science* 5.3-4 (2010), pp. 207–388 (cit. on p. 10).

[Val79] Leslie G. Valiant. "Completeness Classes in Algebra". In: *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC 1979)*. Atlanta, Georgia, USA: Association for Computing Machinery, 1979, pp. 249–261 (cit. on pp. 6, 28).

[Wie20] Finn Wiersig. "Sparse Polynomials in Polynomial Ideals". Bachelor's thesis. Otto von Guericke University of Magdeburg, 2020 (cit. on p. 6).

[Zip79] Richard Zippel. "Probabilistic algorithms for sparse polynomials". In: *Proceedings of the International Symposium on Symbolic and Algebraic Computation, EUROSAM 1979*. 1979, pp. 216–226 (cit. on p. 3).

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il