JACOBIAN HITS CIRCUITS: HITTING-SETS, LOWER BOUNDS FOR DEPTH-D OCCUR-K FORMULAS & DEPTH-3 TRANSCENDENCE DEGREE-K CIRCUITS

MANINDRA AGRAWAL†, CHANDAN SAHA‡, RAMPRASAD SAPTHARISHI§, AND NITIN SAXENA¶

Abstract. We present a single common tool to strictly subsume all known cases of polynomial time blackbox polynomial identity testing (PIT), that have been hitherto solved using diverse tools and techniques, over fields of zero or large characteristic. In particular, we show that polynomial time hitting-set generators for identity testing of the two seemingly different and well studied models – depth-3 circuits with bounded top fanin, and constant-depth constant-read multilinear formulas – can be constructed using one common algebraic-geometry theme: Jacobian captures algebraic independence. By exploiting the Jacobian, we design the first efficient hitting-set generators for broad generalizations of the above-mentioned models, namely:

- depth-3 (ΣΠΣ) circuits with constant transcendence degree of the polynomials computed by the product gates (no bounded top fanin restriction), and
- constant-depth constant-occur formulas (no multilinear restriction).

Constant-occur of a variable, as we define it, is a more general concept than constant-read. Also, earlier work on the latter model assumed that the formula is multilinear. Thus, our work goes further beyond the related results obtained by Saxena & Seshadhri (STOC 2011), Saraf & Volkovich (STOC 2011), Anderson et al. (CCC 2011), Beecken et al. (ICALP 2011) and Grenet et al. (FSTTCS 2011), and brings them under one unifying technique.

In addition, using the same Jacobian based approach, we prove exponential lower bounds for the immanant (which includes permanent and determinant) on the same depth-3 and depth-4 models for which we give efficient PIT algorithms. Our results reinforce the intimate connection between identity testing and lower bounds by exhibiting a concrete mathematical tool — the Jacobian — that is equally effective in solving both the problems on certain interesting and previously well-investigated (but not well understood) models of computation.

Key words. algebraic independence, blackbox, circuits, depth, identity testing, immanant, Jacobian, lower bound, Vandermonde

AMS subject classifications. 68W30, 68Q25

1. Introduction and examples. A polynomial in many variables, when written down verbosely as a sum of monomials, might have a humongous expression. Arithmetic circuits, on the other hand, provide a succinct way to represent multivariate polynomials. An arithmetic circuit, consisting of addition (+) and multiplication (×) gates, takes several variables as input and outputs a polynomial in those variables. The study of arithmetic circuits - as to which algorithmic questions on polynomials can be resolved efficiently in this model of computation, and which polynomials do not admit any polynomial-sized circuit representation - form the foundation of algebraic complexity theory.

One particular algorithmic question, the problem of polynomial identity testing (PIT), occupies a pivotal position in the theory of arithmetic circuit complexity. It is the problem of deciding if the output of a given arithmetic circuit is an identically zero polynomial. Being such an elementary problem, identity testing has enjoyed its status of prime importance by appearing in several fundamental results including...
primality testing [AKS04], the PCP theorem [ALM+98] and the IP = PSPACE result [LFKN90, Sha90], among many others like graph matching [Lov79, MVV87], polynomial interpolation [CDGK91], matrix completion [IKS10], polynomial solvability [KY08], factorization [SV10], learning of arithmetic circuits [KS06] and the geometric complexity theory approach [Mul12, Mul11]. What is more intriguing is that there is an intimate connection between identity testing and lower bounds [KI03, HS80, AvM10], especially the problem of separating the complexity classes VP from VNP (which must necessarily be shown before showing P ≠ NP [Val79, SV85]).

Proving VP ≠ VNP amounts to showing that an explicit class of polynomials, like the Permanent, cannot be represented by polynomial-sized arithmetic circuits, which in turn would follow if identity testing can be derandomized using a certain kind of pseudo-random generator [Agr05, KI03]. (Note that identity testing has a simple and efficient randomized algorithm - pick a random point and evaluate the circuit at it [Sch80, Zip79, DL78].)

During the past decade, the quest for derandomization of PIT has yielded several results on restricted models of circuits. But, fortunately, the search has been made more focused by a line of work [CKKS13, Koi12, AV08, VSBR83] which states that a polynomial time blackbox derandomization of identity testing for depth-3 circuits (via a certain pseudo-random generator) implies a quasi-polynomial time derandomization of PIT for poly-degree\footnote{Circuits computing polynomials with degree bounded by a polynomial function in the size of the circuit.} circuits. By polynomial time blackbox test, we mean:

A polynomial time hitting-set generator, which is a boolean Turing machine that produces a set of points with small integer coordinates. These points are then fed (one by one) into the circuit, which internally uses any arithmetic, to output the evaluations of the polynomial at those points. (For characteristic p > 0, one works with a field extension, where each coordinate of a point is a small vector of integers in \{0, ..., p − 1\}.)

With depth-3 as the final frontier, the results that have been achieved so far include polynomial time hitting-set generators for the following models:

- depth-2 (ΣΠ) circuits (equivalently, the class of sparse polynomials) [KS01],
- depth-3 (ΣΠΣ) circuits with constant top fanin [SS11],
- constant-depth constant-read multilinear formulas [AvMV11, SV11] (& their sparse-substituted variants),
- circuits generated by sparse polynomials with constant transcendence degree [BMS11].

To our knowledge, these are the only instances for which polynomial time hitting-set generators are known. The result on depth-3 bounded top fanin circuits is based upon the Chinese Remaindering technique of [KS07] and the ideal-theory framework studied in [SS10]. Their work followed after a sequence of developments in rank bound estimates [DS05, KS08, SS09, SS09b, SS10], some using incidence geometry - although, this result in particular is not rank based. On the other hand, the work on constant-depth multilinear formulas [AvMV11, SV11] is obtained by building upon and extending the techniques of other earlier results [KMSV10, SV09, SV08] on ‘read-once’ models. At a high level, this involved a study of the structure of multilinear formulas under the application of partial derivatives with respect to a carefully chosen set of variables and invoking depth-3 rank bounds (survey [SY10]). More recently, a third technique has emerged in [BMS11] which is based upon the
abstract concept of algebraic independence of polynomials and a related computational handle called the Jacobian. They showed that for any given poly-degree circuit $C$ and sparse polynomials $f_1, \ldots, f_m$ with constant transcendence degree, a hitting-set for $C(f_1, \ldots, f_m)$ can be constructed in polynomial time.

**Our contribution.** With these diverse techniques floating around the study of hitting-set generators, one wonders: Could there be one single tool that is sufficiently powerful to capture all these models? Is there any unique feature underlying these seemingly different models that can lend itself to the conception of such a unifying tool? The answer to both these questions, as we show in this work, is yes. The key to this lies in studying the properties of the Jacobian, a mathematical object lying at the very core of algebraic-geometry. As for the ‘unique feature’, notice that in the above four models some parameter of the circuit is bounded - be it bounded top fanin, bounded read of variables, or bounded transcendence degree. (Bounded depth should not be seen as an extra restriction on the circuit model because of [GKKS13, AV08].) At an intuitive level, it seems to us that it is this ‘bounded parameter’-ness of the circuit that makes the Jacobian perform at its best.

In the process of finding a universal technique, we significantly strengthen the earlier results. We construct hitting-set generators not only for depth-3 circuits with bounded top fanin, but also for circuits of the form $C(T_1, \ldots, T_m)$, where $C$ is a poly-degree circuit and $T_1, \ldots, T_m$ are products, of linear polynomials, with bounded transcendence degree. In case of depth-3 circuits, $C(T_1, \ldots, T_m)$ is simply $T_1 + \ldots + T_m$. Further, we remove the restriction of multilinearity totally from the constant-depth constant-read model and construct the first efficient hitting-set generator for this class. The condition of constant-read is also replaced by the more general notion of constant-occur.

At this point, one is faced with a natural question: how effective is the Jacobian in proving lower bounds? The intimate connection between efficient algorithms and lower bounds has recurrently appeared in various contexts [Wil11, Rag08, Uma03, PSZ00, IW97]. For arithmetic circuits, this link is provably tight [K03, Agr05, AV08]: Derandomizing identity testing is equivalent to proving circuit lower bounds. Which means, one might have to look for techniques that are powerful enough to handle the dual worlds of algorithm design and lower bounds with equal effectiveness - for e.g. the partial derivative technique has been used to prove lower bounds and identity testing (albeit non-blackbox) on restricted models (survey [CKW11]); the $\tau$-conjecture is another such example [GKPS11]. In this work, we demonstrate the utility of the Jacobian in proving exponential lower bounds for the immanant (which includes determinant and permanent) on the same depth-3 and depth-4 models for which we give efficient PIT algorithms. In particular, this includes depth-4 constant-occur formulas, depth-4 circuits with constant transcendence degree of the underlying sparse polynomials (which significantly generalizes the lower bound result in [GKPS11]), and depth-3 circuits with constant transcendence degree of the polynomials computed by the product gates. To our knowledge, all these lower bounds are new and it is not known how to prove them using earlier techniques. A summary of the results in this paper is provided in Figure 1.1.

**Remark -** The algorithms of [SS11], [SV11] and [AvMV11] work over any field, whereas ours work under the assumption that the field characteristic is zero or large. Also, [AvMV11] presents a quasi-polynomial time algorithm for arbitrary depth, constant read, multilinear formula - a result which we do not know yet how to capture
fig. 1.1. comparison with the earlier efficient hitting-sets

<table>
<thead>
<tr>
<th>Previous best</th>
<th>Running time*</th>
<th>Extended Model</th>
<th>This paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΣΠΣΣ(2(k) circuits</td>
<td>( s^k )</td>
<td>( C(T_1, \ldots, T_m) = 0 )</td>
<td>( s^k )</td>
</tr>
<tr>
<td>( T_1 + \cdots + T_k = 0 )</td>
<td>( \text{SS11} )</td>
<td>poly-degree C &amp; \text{trdeg} {T_i} \leq k</td>
<td>( \text{trdeg} {T_i} = \Omega(n) )</td>
</tr>
<tr>
<td>multilinear circuits</td>
<td>( s^k )</td>
<td>( \text{SV11} )</td>
<td>( s^{k^2} )</td>
</tr>
<tr>
<td>depth-D, read-k multilinear formulas</td>
<td>( s^{2^k} )</td>
<td>( \text{SV11} )</td>
<td>( s = 2^{\Omega(n/k^2)} )</td>
</tr>
<tr>
<td>( C(f_1, \ldots, f_m) \neq 0 )</td>
<td>( \text{BMS11} )</td>
<td>poly-degree C, ΣΠΣ circuits ( f_i )'s ( &amp; \text{trdeg {f_i} \leq k} )</td>
<td>( s = 2^{\Omega(n/k)} )</td>
</tr>
<tr>
<td>( \text{poly-degree C, } F \text{ circuits } f_i \text{'s} )</td>
<td>( \text{BMS11} )</td>
<td>( \text{Imm} )</td>
<td>( \text{Imm} ) lower bound</td>
</tr>
</tbody>
</table>

using our technique. Two other quasi-polynomial hitting sets that our work does not capture are the hitting sets for constant-depth set-multilinear circuits [ASS13] and read-once oblivious ABPs [FS12].

1.1. A tale of two PITs (& three lower bounds). A set of polynomials \( f = \{ f_1, \ldots, f_m \} \subset \mathbb{F}[x_1, \ldots, x_n] \) (in short, \( \mathbb{F}[x] \)) is said to be algebraically independent over \( \mathbb{F} \) if there is no nonzero polynomial \( H \in \mathbb{F}[y_1, \ldots, y_m] \) such that \( H(f_1, \ldots, f_m) \) is identically zero. A maximal subset of \( f \) that is algebraically independent is a transcendence basis of \( f \) and the size of such a basis is the transcendence degree\(^3\) of \( f \) (denoted \( \text{trdeg}_F f \)). Our first theorem states:

**Theorem 1.1.** Let \( C \) be a poly-degree circuit of size \( s \) and each of \( T_1, \ldots, T_m \) be a product of \( d \) linear polynomials in \( \mathbb{F}[x_1, \ldots, x_n] \) such that \( \text{trdeg}_F \{T_1, \ldots, T_m\} \leq r \). A hitting-set for such \( C(T_1, \ldots, T_m) \) can be constructed in time polynomial in \( n \) and \( (sd)^r \), assuming \( \text{char}(F) = 0 \) or \( > d^r \).

If \( C \) is a single + gate, we get a hitting-set generator for depth-3 circuits with constant transcendence degree of the polynomials computed by the product gates (there is no restriction on top fanin).

Our second result uses the following generalization of read-k formulas (where every variable appears in at most \( k \) leaf nodes of the formula) to occur-k formulas. Two reasons behind this generalization are: One, to accommodate the power of exponentiation - if we take the \( e \)-th power of a read-k formula using a product gate, the ‘read’ of the resulting formula goes up to \( ek \) - we would like to avoid this superfluous blow up in read. Two, a read-k formula has size \( O(kn) \), which severely hinders its power of computation - for instance, determinant and permanent cannot even be expressed in this model when \( k \) is a constant [Kal85]. This calls for the following definition.

**Definition 1.2.** An occur-k formula is a rooted tree with internal nodes labelled by + and \( \times \) (power-product gate). A \( \times \) gate, on inputs \( g_1, \ldots, g_m \) with incoming edges labelled \( e_1, \ldots, e_m \in \mathbb{N} \), computes \( g_1^{e_1} \cdots g_m^{e_m} \). At the leaves of this tree are depth-2 formulas computing sparse polynomials (leaf nodes), where every variable occurs in at most \( k \) of these sparse polynomials.

---

1 Estimates the bit complexity of the hitting-set generator; constant factors not stressed (also in higher exponents).

2 We assume a zero or large characteristic.

3 Since algebraic independence satisfies the matroid property cf. [Oxl92], transcendence degree is well-defined.
Size of a $\times \lambda$ gate is defined as the integer $(e_1 + \cdots + e_m)$ associated with its incoming edges, while size of a $+$ gate is counted as one. Size of a leaf node is the size of the corresponding depth-2 formula. With these conventions, size of an occur-$k$ formula is defined to be the total size of all its gates (and leaf nodes) plus the number of edges. Depth is defined to be the number of layers of $+$ and $\times \lambda$ gates plus 2 (the ‘plus 2’ accounts for the depth-2 formulas at the leaves). Thus, occur-$k$ is more relaxed than the traditional read-$k$ as it packs the “power of powering” (to borrow from [GKPS11]), and the leaves are sparse polynomials (at most $kn$ many) whose dependence on its variables is arbitrary. E.g. $(x_3^2 x_2 + x_1^2 x_2^3 + x_1 x_4)^e$ is not read-1 but is trivially depth-3 occur-1.

**Theorem 1.3.** A hitting-set for any depth-$D$ occur-$k$ formula of size $s$ can be constructed in time polynomial in $s^R$, where $R = (2k)^{2D-2^D}$ (assuming char$(\mathbb{F}) = 0$ or $> s^R$).

A tighter analysis for depth-4 occur-$k$ formulas yields a better time complexity. Note that a depth-4 occur-$k$ formula allows unbounded top fanin. Also, it can be easily seen to subsume $\Sigma \Pi \Sigma \Pi(k)$ multilinear circuits studied by [SV11, KMSV10].

**Theorem 1.4.** A hitting-set for any depth-$4$ occur-$k$ formula of size $s$ can be constructed in time polynomial in $s^{k^2}$ (assuming char$(\mathbb{F}) = 0$ or $> s^{4k}$).

For constant-depth, the above theorems not only remove the restriction of multilinearity (and relax read-$k$ to occur-$k$), but further improve upon the time complexity of [AvMV11] and [SV11]. The hitting-set generator of [AvMV11] works in time $n^{k^2(2k)+O(kD)}$, and hence is super-exponential when $k = \Omega(s^{\varepsilon/2D}2^D)$ for any positive $\varepsilon < 1$ and a constant $D$, whereas the generator in Theorem 1.3 runs in sub-exponential time for the same choice of parameters. The running time of [SV11] is $s^{O(k^2)}$, which is slightly worse than that of Theorem 1.3.

Since any polynomial has an exponential-sized depth-2, occur-1 formula (just the sparse representation), proving lower bounds on this model is an interesting proposition in its own right.

**Definition 1.5.** [LR34] Let $S_n$ denote the permutation group on $n$ points and $\mathbb{C}^\times$ be the nonzero complex numbers. For any map $\chi : S_n \rightarrow \mathbb{C}^\times$, the immanant of a matrix $M = (x_{ij})_{n \times n}$ with respect to $\chi$ is defined as $\text{Imm}_\chi(M) = \sum_{\sigma \in S_n} \chi(\sigma) \prod_{i=1}^{n} x_{i\sigma(i)}$.

Determinant & permanent are special cases of the immanant with $\chi$ as the alternating sign character & the identity character, respectively. Denote $\text{Imm}_\chi(M)$ by $\text{Imm}_n$ for an arbitrarily fixed $\chi$.

**Theorem 1.6.** Any depth-$4$ occur-$k$ formula that computes $\text{Imm}_n$ must have size $s = 2^{\Omega(n/k^2)}$ over any field of characteristic zero (even counting each $\times \lambda$ gate as size one).

Thus, if each variable occurs in at most $n^{1/2-\varepsilon}$ ($0 < \varepsilon < 1/2$) many underlying sparse polynomials, it takes an exponential sized depth-4 circuit to compute $\text{Imm}_n$. Our next result is an exponential lower bound on the model for which hitting-set was developed in [BMS11] (but no lower bound was shown). It is also an improvement over the result obtained in [GKPS11] which holds only for more restricted depth-4 circuits over reals.
Theorem 1.7. Let $C$ be any circuit. Let $f_1, \ldots, f_m$ be sparse polynomials (of any degree) with sparsity bounded by $s$ and their trdeg bounded by $r$. If $C(f_1, \ldots, f_m)$ computes $\text{Imm}_n$ then $s = 2^{O((n/r)\log r)}$ over any field of characteristic zero.

Which means, any circuit involving fewer than $n^{1-\varepsilon}$-polynomials at the last levels, must have exponential size to compute $\text{Imm}_n$. (The models of Theorem 1.6 & 1.7 are incomparable). The next result is on the model for which hitting-set is given by Theorem 1.1.

Theorem 1.8. Let $C$ be any circuit and $T_1, \ldots, T_m$ be products of linear polynomials. If $C(T_1, \ldots, T_m)$ computes $\text{Imm}_n$ then $\text{trdeg}_F\{T_1, \ldots, T_m\} = \Omega(n)$ over any field of characteristic zero.

Which means, any circuit involving only $o(n)$ many algebraically independent $\Pi\Sigma$-polynomials at the last levels cannot compute $\text{Imm}_n$.

1.2. Deterministic testing of algebraic independence. The construction of hitting-set generators (stated in the previous section) also implies deterministic algorithms for certain special cases of the following problem: Given a set of polynomials as arithmetic circuits, check deterministically if they are algebraically independent. In fact, for these special cases we only require a blackbox access to the input circuits.

In this respect, it can be said that our hitting-set generators (and to some extent the algorithms for certain special cases of the following problem: Given a set of polynomials (assuming this to be a transcendence basis of the field of characteristic zero.

Let $T_1, \ldots, T_r$ be products of linear polynomials at the last levels. If we could find an $r$-variate linear map $\varphi$ that keeps $\varphi \circ J(T_1, \ldots, T_r)$ nonzero, then $\varphi(T_1), \ldots, \varphi(T_r)$ are again algebraically independent and it can be shown that for any $C$: $C(T_1, \ldots, T_m) = 0$ iff $C(\varphi(T_1), \ldots, \varphi(T_m)) = 0$. Since $T_i$’s are not sparse, the Jacobian is usually a difficult polynomial to work with, and so is finding $\varphi$. However, for the special models in this paper we are able to design $\varphi$ - mainly because Jacobian (being defined via partial derivatives) has a nice ‘linearizing effect’, on the circuit product gates, that factors itself. The $\varphi$ ultimately provides a hitting-set for $C(T_1, \ldots, T_m)$, as we reduce to a situation where $r$ is constant.

The initial idea for lower bounds is similar. Suppose $\text{Imm}_n = C(T_1, \ldots, T_m)$. Then, by algebraic dependence, $J(\text{Imm}_n, T_1, \ldots, T_r) = 0$. Our proofs then exploit

\footnote{The classical (Jacobian-based) efficient algorithm for testing the algebraic independence of general circuits, over large or zero characteristic, is randomized and whitebox.}
the nature of this identity for the special models. This part requires proving certain combinatorial properties of the immanant.

**Remark** - The dependence of our results on the field characteristic is owing to the fact that the Jacobian criterion, which involves taking derivatives, is used to characterize algebraic independence. We believe that the condition on the field characteristic in our results is probably not a fundamental requirement - rather, lifting this condition is perhaps a technical hurdle due to the lack of a suitable criterion that captures algebraic independence for low characteristic fields. (Intuitively, a low characteristic enables more 'cancellations', polynomial identities and configurations.) Recently, this direction of research is investigated in [MSS12], where a new criterion enables more 'cancellations', polynomial identities and configurations. Re- algebraic independence for low characteristic fields. (Intuitively, a low characteristic is perhaps a technical hurdle due to the lack of a suitable criterion that captures in our results is probably not a fundamental requirement - rather, lifting this condition from our results.

2. Preliminaries: Jacobian and faithful homomorphisms. Our contribution, in this section, is an elementary proof of Theorem 2.4 which was originally proved in [BMS11] using Krull’s *Hauptidealsatz*. Here, we state the main properties of the Jacobian and faithful homomorphisms without proofs - for details, refer to [BMS13].

**Definition 2.1.** The Jacobian of a set of polynomials \( f = \{f_1, \ldots, f_m\} \) in \( \mathbb{F}[x_1, \ldots, x_n] \) is defined to be the matrix \( J_x(f) = (\partial_{x_j} f_i)_{m \times n} \), where \( \partial_{x_j} f_i = \partial f_i / \partial x_j \).

Let \( S \subseteq x = \{x_1, \ldots, x_n\} \) and \( |S| = m \). Then \( J_S(f) \) denotes the minor (i.e. determinant of the submatrix) of \( J_x(f) \) formed by the columns corresponding to the variables in \( S \).

**Fact 2.2** (Jacobian criterion). Let \( f \subset \mathbb{F}[x] \) be a finite set of polynomials of degree at most \( d \), and \( \text{trdeg}_F f \leq r \). If \( \text{char}(\mathbb{F}) = 0 \) or \( \text{char}(\mathbb{F}) > d^r \), then \( \text{trdeg}_F f = \text{rank}_{\mathbb{F}(x)} J_x(f) \).

The proof of this fact may be found in [BMS13].

**Definition 2.3.** A homomorphism \( \Phi : \mathbb{F}[x] \to \mathbb{F}[y] \) (\( y \) is another set of variables) is said to be faithful to a finite set of polynomials \( f \subset \mathbb{F}[x] \) if \( \text{trdeg}_F f = \text{trdeg}_F \Phi(f) \).

**Theorem 2.4** (Faithful is useful). Let \( \Phi \) be a homomorphism faithful to \( f = \{f_1, \ldots, f_m\} \subset \mathbb{F}[x] \). Then for any \( C \in \mathbb{F}[y_1, \ldots, y_m] \), \( C(f) = 0 \iff C(\Phi(f)) = 0 \).

**Proof.** Since \( \Phi \) is faithful to \( f \), there is a transcendence basis (say, \( f_1, \ldots, f_s \)) of \( f \) such that \( \Phi(f_1), \ldots, \Phi(f_s) \) is a transcendence basis of \( \Phi(f) \). Elements of the field \( \mathbb{K} = \mathbb{F}(f) \) can be written as polynomials in \( f_{s+1}, \ldots, f_m \) with coefficients from \( \mathbb{F}(f_1, \ldots, f_s) \). Suppose \( C(f) \) is a nonzero element of \( \mathbb{K} \), then there is an inverse \( Q \in \mathbb{K} \) such that \( Q \cdot C(f) = 1 \). Since \( Q \) is a polynomial in \( f_{s+1}, \ldots, f_m \), by clearing off the denominators of these coefficients in \( Q \), we get an equation \( Q \cdot C(f) = P(f_1, \ldots, f_s) \), where \( Q \) is a nonzero polynomial in \( f \) and \( P \) is a nonzero polynomial in \( f_1, \ldots, f_s \). Applying \( \Phi \) to both sides of the equation, we conclude that \( C(\Phi(f)) = \Phi(\Phi(f)) \neq 0 \), otherwise \( P(\Phi(f_1), \ldots, \Phi(f_s)) = \Phi(\Phi(f_1, \ldots, f_s)) = 0 \) which is not possible as \( \Phi(f_1), \ldots, \Phi(f_s) \) are algebraically independent and \( P \) is a nontrivial polynomial.
Recipe for faithful homomorphisms. All the PITs in this paper would proceed by constructing faithful homomorphisms for a certain set of polynomials. The following fact describes the changes in the Jacobian after a “change of variables”.

**FACT 2.5 (Chain rule).** For any finite set of polynomials \( f \subset \mathbb{F}[x] \) and a homomorphism \( \Phi : \mathbb{F}[x] \to \mathbb{F}[y] \), we have \( \mathcal{J}_y(\Phi(f)) = \Phi (\mathcal{J}_x(f)) \cdot \mathcal{J}_y(\Phi(x)) \) (where \( \Phi \) applied to a matrix/set refers to the matrix obtained by applying \( \Phi \) to every entry).

Proof. Follows directly from the chain-rule of differentiation. \( \square \)

The recipe for faithful homomorphisms use the following “rank preserving linear maps” studied by Gabizon and Raz [GR05].

**LEMMA 2.6.** [GR05] Let \( A \) be a \( r \times n \) matrix with entries in a field \( \mathbb{F} \), and let \( t \) be an indeterminate. Then, \( \text{rank}_{\mathbb{F}(t)}(A \cdot (t^i))_{i \in [n], j \in [r]} = \text{rank}_{\mathbb{F}} A \).

**LEMMA 2.7 (Recipe for faithful maps).** Let \( f \subset \mathbb{F}[x] \) be a finite set of polynomials of degree at most \( d \), \( \text{trdeg}_x f \leq r \), and \( \text{char}(\mathbb{F}) = 0 \) or \( d' > r \). Let \( \Psi : \mathbb{F}[x] \to \mathbb{F}[z] \) be a homomorphism such that \( \text{rank}_{\mathbb{F}[x]} \mathcal{J}_x(f) = \text{rank}_{\mathbb{F}(z)} \Psi(\mathcal{J}_x(f)) \).

Then, the map \( \Phi : \mathbb{F}[x] \to \mathbb{F}[z, t, y_1, \ldots, y_r] \) that maps, for all \( i, x_i \mapsto \left( \sum_{j=1}^r y_j t^i \right) + \Psi(x_i) \) is a homomorphism faithful to \( f \).

Proof. Without loss of generality, let \( \text{trdeg}_x f = r \), which then (by Jacobian criterion) is the rank of \( \mathcal{J}_x(f) \). We intend to show that the matrix \( \mathcal{J}_y(\Phi(f)) \) is of rank \( r \), which would imply (by Jacobian criterion) that \( \text{trdeg}_x(\Phi(f)) = r \).

Consider the projection \( J' \) of \( \mathcal{J}_y(\Phi(f)) \) obtained by setting \( y_1 = \cdots = y_r = 0 \).

\[
J' = [\mathcal{J}_y(\Phi(f))]_{y=0} = [\mathcal{J}_x(f) \cdot \mathcal{J}_y(\Phi(x))]_{y=0} = \Psi(\mathcal{J}_x(f) \cdot \mathcal{J}_y(\Phi(x)))
\]

where the first line follows simply by chain rule of differentiation.

Observe that the matrix \( \mathcal{J}_y(\Phi(x)) \) is exactly the Vandermonde matrix that is present in Lemma 2.6. Also, \( \Psi(\mathcal{J}_x(f)) \) has entries in \( \mathbb{F}(z) \), and by assumption has the same rank as \( \mathcal{J}_x(f) \). Hence, by Lemma 2.6

\[
\text{rank}_{\mathbb{F}(t,z)} J' = \text{rank}_{\mathbb{F}(t,z)} (\Psi(\mathcal{J}_x(f) \cdot \mathcal{J}_y(\Phi(x)))) = \text{rank}_{\mathbb{F}(z)} \Psi(\mathcal{J}_x(f)) = r.
\]

And since \( J' \) is just a projection of \( \mathcal{J}_y(\Phi(f)) \), the rank of the latter must also be \( r \). Hence, \( \Phi \) is indeed faithful. \( \square \)

3. Hitting-set for depth-3 circuits of constant transcendence degree.

Let \( C \) be any circuit and \( D \) be the circuit \( C(T_1, \ldots, T_m) \), where each \( T_i \) is of the form \( \prod_{\ell=1}^{d} \ell_{ij}, \) every \( \ell_{ij} \) is a linear polynomial in \( \mathbb{F}[x_1, \ldots, x_n] \). Denote by \( T \) the set \( \{T_1, \ldots, T_m\} \) and by \( L(T_i) \) the multiset of linear polynomials that constitute \( T_i \). Suppose \( \text{trdeg}_x T = k \leq r \) and \( T_k = \{T_1, \ldots, T_k\} \) be a transcendence basis of \( T \).

Since \( \mathcal{J}_x(T_k) \) has full rank (\( \text{char}(\mathbb{F}) = 0 \) or \( \text{char}(\mathbb{F}) > d' \)), without loss of generality assume that the columns corresponding to \( x_k = \{x_1, \ldots, x_k\} \) form a nonzero \( k \times k \) minor of \( \mathcal{J}_x(T_k) \). By Lemma 2.7, if we construct a \( \Psi : \mathbb{F}[x] \to \mathbb{F}[z] \) that keeps \( \mathcal{J}_x(T_k) \) nonzero then \( \Psi \) can easily be extended to a homomorphism \( \Phi : \mathbb{F}[x] \to \mathbb{F}[z, y_1, y_2, \ldots, y_r] \) that is faithful to \( T \). And hence, by Theorem 2.4 it would follow that \( \Phi(D) = 0 \) if and only if \( D = 0 \).

If \( T_i = \prod_{\ell=1}^{d} \ell_{ij} \) then \( \partial_x T_i = T_i \cdot \left( \sum_{j=1}^{d} \partial_j \ell_{ij} / \ell_{ij} \right) \). By expanding, using this additive structure of \( \partial_x T_i \) and the linearity of determinant with respect to rows, the determinant \( \mathcal{J}_x(T_k) \) takes the following form,
Call a set of linear polynomials independent if the corresponding homogeneous linear parts (i.e. the constant-free parts) are \( \mathbb{F} \)-linearly independent. The term \( J_{x_k}(\ell_1, \cdots, \ell_k) \) ensures that the above sum is only over those \( \ell_1, \cdots, \ell_k \) that are independent linear polynomials (otherwise the Jacobian vanishes). The sum has the form of a depth-3 circuit, call it \( H_0 \), and we intend to construct a low-variate \( \Psi \) such that \( \Psi(H_0) \neq 0 \).

We show that this is achieved by a \( \Psi \) that preserves the independence of a 'small' set of linear polynomials - which we call a certificate of \( H_0 \). (The certifying path technique is from \cite{SS11}.)

**Certificate of \( H_0 \):** We can assume that each of the terms \( J_{x_k}(\ell_1, \cdots, \ell_k) \) in equation \((3.1)\) are nonzero field constants. Let \( \mathcal{L}(H_0) \) be the set of all linear polynomials occurring in the denominator terms “\( \ell_1 \cdots \ell_k \)” of all the summands in sum \((3.1)\). By adjusting the field constants at the numerators, we can assume that no two linear polynomials in \( \mathcal{L}(H_0) \) are constant multiple of each other. This means, the depth-3 circuit \( H_0 \) has the form \( H_0 = T \cdot \sum_{L} \alpha_L/\ell_1 \cdots \ell_k \), where \( T := \prod_{i=1}^{k} T_i \), \( \alpha_L \) is a nonzero field constant and the sum runs over some sets \( L = \{ \ell_1, \cdots, \ell_k \} \) of \( k \) independent linear polynomials in \( \mathcal{L}(H_0) \).

Define, content of a depth-3 circuit \( G = \sum_i P_i \), where \( P_i \) is a product of linear polynomials, as \( \text{cont}(G) := \gcd_i \{P_i\} \), and let the simple part \( \text{sim}(G) := G/\text{cont}(G) \).

Hence \( \text{cont}(H_0) = \gcd_{L} \{T/\ell_1 \cdots \ell_k\} \) and

\[
\text{sim}(H_0) = F_0 \cdot \sum_{L} \alpha_L/\ell_1 \cdots \ell_k, \quad \text{where } F_0 = \frac{T}{\text{cont}(H_0)}.
\]  

(3.2)

Note that \( F_0 \) is simply the product of the linear polynomials in \( \mathcal{L}(H_0) \) and so \( \deg(F_0) = |\mathcal{L}(H_0)| \). For any \( \ell \in \mathcal{L}(H_0) \), the terms in \( \text{sim}(H_0) \) that survive modulo \( \ell \) are those with \( \ell \) in the denominator “\( \ell_1 \cdots \ell_k \)” of the above expression. Hence,

\[
H_1 := \text{sim}(H_0) \mod \ell_1 = \frac{F_0}{\ell_1} \cdot \sum_{L = \{\ell_2, \cdots, \ell_k\}} \alpha_L/\ell_2 \cdots \ell_k.
\]

We can treat \( H_1 \) as a depth-3 circuit in one less variable: Suppose that \( \ell_1 = c_1 x_1 + \sum_{i=2}^{n} c_i x_i \) where \( c_i \)'s \( \in \mathbb{F} \) and \( c_1 \neq 0 \), then we can replace \( x_1 \) by \(- \sum_{i=2}^{n} c_i x_i / c_1 \) in \( \text{sim}(H_0) \), particularly in \( F_0/\ell_1 \) (of course, after dividing \( F_0 \) by \( \ell_1 \)) as well as in each of \( \ell_2, \cdots, \ell_k \) in the denominators, so that \( H_1 \) becomes a depth-3 circuit in \( \mathbb{F}[x_2, \ldots, x_n] \).

Therefore, it makes perfect sense to talk about \( \text{cont}(H_1) \) and \( \text{sim}(H_1) \). Observe that \( \ell_2, \cdots, \ell_k \) remain independent linear polynomials modulo \( \ell_1 \), and so \( H_1 \) is a depth-3 circuit of the ‘same nature’ as \( H_0 \) but with one less linear polynomials in the denominators. Also, the linear polynomials in \( \mathcal{L}(H_1) \) is a subset of the linear polynomials in \( \mathcal{L}(H_0) \) modulo \( \ell_1 \). Extending the above argument, it is possible to define a sequence of circuits: \( H_i := \text{sim}(H_{i-1}) \mod \ell_i \), (1 \( \leq \) \( i \) \( \leq \) \( k \)) where \( \ell_i \in \mathcal{L}(H_{i-1}) \). Further, \( \mathcal{L}(H_i) \) is a subset of \( \mathcal{L}(H_{i-1}) \) modulo \( \ell_i \), which implies that essentially there are independent linear polynomials, say \( \ell_1, \cdots, \ell_k \), in \( \mathcal{L}(H_0) \) such that \( \ell_i = \ell_i \mod (\ell_1, \ldots, \ell_{i-1}) \) and therefore \( H_i = \text{sim}(H_{i-1}) \mod (\ell_1, \ldots, \ell_i) \).
Lemma 3.1 (Certifying path). There exist independent linear polynomials \( \{\ell_1, \ldots, \ell_k\} \subseteq L(H_0) \) such that \( H_i \neq 0 \mod (\ell_1, \ldots, \ell_i) \), \( \forall i \in [k] \), and \( H_k \) is a nonzero product of linear polynomials in \( L(H_0) \) modulo \( (\ell_1, \ldots, \ell_k) \).

Proof. The proof is by induction on \( k \) and follows the sketch given while defining \( \sim(\cdot) \). The degree of the nonzero polynomial \( \sim(H_0) \) is \( |L(H_0)| - k \). By Chinese remaindering, there exists an \( \ell_1 \in L(H_0) \) such that \( H_1 := \sim(H_0) \mod \ell_1 \neq 0 \). In the base case \( (k = 1) \), it is easy to see that \( H_1 \) is a nonzero product of linear polynomials modulo \( \ell_1 \). For any larger \( k \), the depth-3 polynomial \( H_1 \) has exactly the same form as \( H_0 \) but with \( k - 1 \) independent linear polynomials in the denominators. Inducting on this smaller value \( k - 1 \), keeping in mind that \( L(H_1) \subseteq L(H_0) \) modulo \( (\ell_1, \ldots, \ell_i) \), completes the proof. \( \square \)

A set \( \{\ell_1, \ldots, \ell_k\} \), satisfying Lemma 3.1, is called a certifying path of \( H_0 \). Fix a certifying path \( \{\ell_1, \ldots, \ell_k\} \). Let \( \Psi : \mathbb{F}[x] \rightarrow \mathbb{F}[z_1, z_{k+1}] \) be such that \( \Psi(\ell_1), \ldots, \Psi(\ell_k) \) are independent linear polynomials in \( \mathbb{F}[z] \) and for every \( \ell \in \cup_{i=1}^k L(T_i) \), \( \ell \neq 0 \mod (\ell_1, \ldots, \ell_k) \) if \( \Psi(\ell) \neq 0 \mod (\Psi(\ell_1), \ldots, \Psi(\ell_k)) \). We call such a \( \Psi \) a rank-(\( k+1 \)) preserving map for \( L(H_0) \). It can be shown that one of the maps \( \Psi : x_i \mapsto \sum_{j=1}^{k+1} z_j b_j \), where \( b \) runs over \( d_{kn}(k+1)^2 \) distinct elements of \( \mathbb{F} \), is a rank-(\( k+1 \)) preserving map for \( H_0 \). (It is a simple corollary Lemma 2.6).

Theorem 3.2. If \( \Psi : \mathbb{F}[x] \rightarrow \mathbb{F}[z_1, z_{k+1}] \) is a rank-(\( k+1 \)) preserving map for \( L(H_0) \), then \( \Psi(H_0) \neq 0 \).

Proof. Let \( \{\ell_1, \ldots, \ell_k\} \) be the certifying path of \( H_0 \) fixed above. The proof is by reverse induction on \( k \): Assuming \( \Psi(H_i) \neq 0 \mod \Psi(I) \), we show that \( \Psi(H_{i-1}) \neq 0 \mod \Psi(I_{i-1}) \) for \( k \geq i \geq 2 \). The base case is easy, as by Lemma 3.1, \( H_k \) is a nonzero product of linear polynomials in \( L(H_0) \) modulo \( I_k \). Hence, by the definition of a rank-(\( k+1 \)) preserving map, \( \Psi(H_k) \neq 0 \mod \Psi(I_k) \) (ideal generated by independent linear polynomials is an integral domain).

The proof of the inductive step proceeds as follows. By construction, \( H_{i-1} = \text{cont}(H_{i-1}) \cdot \sim(H_{i-1}) = \text{cont}(H_{i-1}) \cdot [q_i \ell_i + H_i] \mod I_{i-1} \), for some polynomial \( q_i \). Which means, \( \Psi(H_{i-1}) = \Psi(\text{cont}(H_{i-1})) \cdot \Psi(q_i) \Psi(\ell_i) + \Psi(H_i) \mod \Psi(I_{i-1}) \). If \( \Psi(q_i) \Psi(\ell_i) + \Psi(H_i) = 0 \mod \Psi(I_{i-1}) \), then \( \Psi(H_i) = 0 \mod \Psi(I_i) \) which contradicts the induction hypothesis. Also, by Lemma 3.1, \( H_{i-1} \neq 0 \mod I_{i-1} \) implies that \( \text{cont}(H_{i-1}) \neq 0 \mod I_{i-1} \). Since, \( i \geq 2 \), the linear polynomials in the term \( \text{cont}(H_{i-1}) \) belong to \( L(H_0) \) modulo the ideal \( I_{i-1} \), once again by using the rank-(\( k+1 \)) preserving property of \( \Psi \), we infer that \( \Psi(\text{cont}(H_{i-1})) \neq 0 \mod \Psi(I_{i-1}) \). Therefore, \( \Psi(H_{i-1}) \neq 0 \mod \Psi(I_{i-1}) \). Finally, to obtain \( \Psi(H_0) \neq 0 \) from \( \Psi(H_1) \neq 0 \mod \Psi(I_1) \), use the same argument as above and that \( \Psi(\ell) \neq 0 \) for every \( \ell \in \cup_{i=1}^k L(T_i) \). \( \square \)

Since \( H_0 \) was just a maximal non-zero minor of \( J(T_1, \ldots, T_r) \), we have the following corollary (using Fact 2.2).

Corollary 3.3. Let \( \Psi : \mathbb{F}[x] \rightarrow \mathbb{F}[z_1, z_{r+1}] \) be a rank-(\( r+1 \)) preserving map for \( L(T_1) \cup \ldots \cup L(T_m) \) (where \( \{T_1, \ldots, T_m\} \) are products of linear functions with transcendence degree bounded by \( r \)). Then, \( \Phi : \mathbb{F}[x] \rightarrow \mathbb{F}[z_1, z_{r+1}, t, y_1, \ldots, y_{r+1}] \) defined as \( \Phi(x_i) \mapsto \Psi(x_i) + \sum_{j=1}^{r+1} t^j y_j \) is a faithful homomorphism for \( \{T_1, \ldots, T_m\} \).

Proof. Let \( \text{trdeg} \{T_1, \ldots, T_m\} = k \leq r \). Since any rank-(\( r+1 \)) preserving map \( \Psi \) is also a rank-(\( k+1 \)) preserving map, Theorem 3.2 states that \( \Psi \) preserves the rank of \( J(T_1, \ldots, T_m) \). Hence by Lemma 2.7, we have that \( \Phi \) is a faithful homomorphism for \( \{T_1, \ldots, T_r\} \). \( \square \)

With this, we can prove both Theorem 1.1 and Theorem 1.9.
Theorem 1.1 (restated). Let \( C \) be a poly-degree circuit of size \( s \) and each of \( T_1, \ldots, T_m \) be a product of \( d \) linear polynomials in \( \mathbb{F}[x_1, \ldots, x_n] \) such that \( \text{trdeg}_{\mathbb{F}} \{ T_1, \ldots, T_m \} \leq r \). A hitting-set for such \( C(T_1, \ldots, T_m) \) can be constructed in time polynomial in \( n \) and \( (sd)^{r} \), assuming \( \text{char}(\mathbb{F}) = 0 \) or \( > d^r \).

Proof of Theorem 1.1. Corollary 3.3 gives a homomorphism \( \Phi : \mathbb{F}[x] \to \mathbb{F}[z_1, \ldots, z_{r+1}, t, y_1, \ldots, y_{r+1}] \) that is faithful to \( \{ T_1, \ldots, T_m \} \). By Theorem 2.4 we have that \( C(T_1, \ldots, T_m) = 0 \) if and only of \( \Phi(C(T_1, \ldots, T_m)) = 0 \). Since \( C \) is a poly-degree circuit of size \( s \), \( \Phi(C(T_1, \ldots, T_m)) \) is a polynomial of degree at most \( dsO(1) \) resp. \( nrdO(1) \) in the variables \( y, z \) resp. \( t \). Using \[ \text{Sch80, Zip79, DL78} \], we can construct a hitting-set for \( \Phi(D) \) in time polynomial in \( n(sd)^{r} \). Since construction of \( \Psi \) takes time poly\( (ndr) \), the total time taken is poly\( (n,(sd)^{r}) \). \( \square \)

Theorem 1.9 (restated). Given blackbox access to polynomials \( T_1, \ldots, T_r \) that are products of \( d \) linear polynomials in \( \mathbb{F}[x_1, \ldots, x_n] \), there is a poly\( ((nd)^{r}) \) time algorithm to test whether they are algebraically independent, assuming \( \text{char}(\mathbb{F}) = 0 \) or \( > d^r \).

Proof of Theorem 1.9. Corollary 3.3 gives a homomorphism \( \Phi : \mathbb{F}[x] \to \mathbb{F}[z_1, \ldots, z_{r+1}, t, y_1, \ldots, y_{r+1}] \) that is faithful to \( \{ T_1, \ldots, T_m \} \). Hence it suffices to check if \( \{ \Phi(T_1), \ldots, \Phi(T_r) \} \) are algebraically independent or not. Since each \( \Phi(T_i) \) is a \( O(r) \)-variate polynomial of degree poly\( ((nd)^{r}) \), they have at most poly\( ((nd)^{r}) \) monomials. It is well-known \[ \text{KS01, BOT88} \] that an \( n \)-variate degree \( d \) multivariate polynomial with \( R \) monomials can be reconstructed from poly\( (n,d,R) \) evaluations. Hence, using poly\( ((nd)^{r}) \) evaluations, each \( \Phi(T_i) \) can be explicitly written down as a sum of monomials. Hence, the Jacobian \( \mathcal{J}(\Phi(T_1), \ldots, \Phi(T_r)) \) can be written down explicitly, and an application of the lemma in \[ \text{Sch80, Zip79, DL78} \] allows us to check if \( \mathcal{J}(\Phi(T_1), \ldots, \Phi(T_r)) \) has full rank in deterministic poly\( ((nd)^{r}) \) time. \( \square \)

4. Hitting-set for constant-depth constant-occur formulas.

Bounding the top fanin - Let \( C \) belong to the class \( \mathcal{C} \) of depth-\( D \), occur-\( k \) formulas of size \( s \), with potentially large top fan-in. The following easy observation would allow us to modify \( C \) slightly to work with a bounded top fan-in formula (without increasing the other parameters too much).

Observation 4.1. If \( C(x_1, \ldots, x_n) \) is non-constant and nonzero, then there is an \( i \) such that \( \overline{C} := C(x_1, \ldots, x_{i-1}, x_i + 1, x_{i+1}, \ldots, x_n) - C(x_1, \ldots, x_n) \neq 0 \), assuming \( \text{char}(\mathbb{F}) > s^D \) (i.e. the bound on the degree of \( C \)).

If \( C \) has a \( + \) gate on top then \( C(x) = \sum_{i=1}^{m} T_i \), where \( T_i \)'s are computed by \( \times \) gates at the next level. Since \( x_i \) occurs in at most \( k \) of the \( T_i \)'s, \( \overline{C} \) has top fanin at most \( 2k \). If \( C \) has a \( \times \) gate on top then \( \overline{C} \) has a + gate on top with fanin \( 2 \) and \( \text{depth}(\overline{C}) = D + 1 \). Therefore, \( \overline{C} \) belongs to the class \( \mathcal{C} \) of depth-(\( D + 1 \)) occur-\( 2k \) formulas of size at most \( (s^2 + s) \), and a + gate on top with fanin bounded by \( 2k \). (The size of \( C(x_1, \ldots, x_{i-1}, x_i + 1, x_{i+1}, \ldots, x_n) \) is bounded by \( s^2 \) as the total sparsity of the polynomials corresponding to the leaves of \( C(x_1, \ldots, x_n) \) can grow at most quadratically as \( x_i \) is replaced by \( x_i + 1 \).) Suppose \( \mathcal{H} \) is a hitting-set for the class \( \mathcal{C} \). Form a new set \( \mathcal{H} \supseteq \mathcal{H} \) by including points \( (\alpha_1 + 1, \alpha_2, \ldots, \alpha_n), (\alpha_1, \alpha_2 + 1, \ldots, \alpha_n), \ldots, (\alpha_1, \ldots, \alpha_{n-1}, \alpha_n + 1) \) in \( \mathcal{H} \) for every point \( (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathcal{H} \). Observe that \( \mathcal{H} \) is a hitting-set for \( \mathcal{C} \) and size\( (\mathcal{H}) = n \cdot \text{size}(\mathcal{H}) \). Therefore, it is sufficient
if we construct $\tilde{H}$. By reusing symbols, assume that $C$ is a depth-$D$ occur-$k$ formula of size $s$ with a $+$ gate on top having top fanin at most $k$.

In order to construct a hitting set for $C(x) = \sum_{i=1}^{k} T_i$, we shall solve the following slightly more generalized goal:

**Goal:** Let $T_1, \ldots, T_k$ be polynomials computed by size occur-$k$ depth-$D$ circuits of size $s$ each. Construct a map $\Phi$ that is faithful to $\{T_1, \ldots, T_k\}$.

Theorem 2.4 asserts that any such faithful map would preserve the nonzeroness of $C$. Let $T_r = \{T_1, \ldots, T_r\}$ be a transcendence basis of $T$. Since $J_x(T_r)$ has full rank (char($F$) = 0 or $> s^D_r$), by Lemma 2.7, assume that the columns corresponding to $x_r = \{x_1, \ldots, x_r\}$ form a nonzero minor of $J_x(T_r)$. By Lemma 2.7 it suffices to construct a $\Psi$ that keeps $J_x(T_r) \neq 0$. 

**Proof idea** - Identify a gate with the polynomial it computes, and count level of a gate from the top - the gates $T_i$'s are at level 1. Suppose each $T_i$ is a $\times \land$ gate and $T_i = \prod_{\ell=1}^{d} P_{i,\ell}$, where $P_{i,\ell}$'s are gates at level 2. Since $T_i$ is also an occur-$k$ formula, $x_1, \ldots, x_r$ appear in at most $kr$ of the $P_{i,\ell}$'s, say $P_{i,1}, \ldots, P_{i,kr}$. Hence,

\[
\partial_j T_i = \left( \prod_{\ell=kr+1}^{d} P_{i,\ell} \right) \cdot \partial_j \left( \prod_{\ell=1}^{kr} P_{i,\ell}^{e_{i,\ell}} \right) \text{ for every } 1 \leq i, j \leq r
\]

\[
\implies J_{x_r}(T_r) = \left( \prod_{i=1}^{r} \prod_{\ell=kr+1}^{d} P_{i,\ell} \right) \cdot J_{x_r} \left( \prod_{\ell=1}^{kr} P_{i,\ell}^{e_{i,\ell}} \right)
\]

Now notice that the jacobian term on the RHS of the last equation is a polynomial in $P_{i,\ell}$ and $\partial_j P_{i,\ell}$, for $1 \leq i, j \leq r$ and $1 \leq \ell \leq kr$. (Note the irrelevance of the exponents $e_{i,\ell}$'s.) So, if $\Psi$ is faithful to the set $P := \{P_{i,\ell}, \partial_j P_{i,\ell} : 1 \leq i, j \leq r, 1 \leq \ell \leq kr\}$ and the singleton sets $\{P_{i,\ell}\}$ for $1 \leq i \leq r, kr + 1 \leq \ell \leq d$, then $\Psi(J_{x_r}(T_r)) \neq 0$. Observe that the polynomials in $P$ and the singleton sets are (zeroth and first order) derivatives of the gates at level 2, and further these sets involve (the derivatives of) disjoint groups of level 2 gates. This disjointness feature ensures that the number of such sets is at most $s$.

Thus, we have reduced the problem of constructing a faithful map $\Phi$ for $T$ (gates at level 1) to the problem of constructing a map $\Psi$ that is faithful to at most $s$ many sets each containing derivatives of gates at the second level. Now, the idea is to carry forward this argument recursively to deeper levels: In the next level of the recursion we reduce the problem to constructing a map that is faithful to at most $s$ sets containing (zeroth, first and second order) derivatives of disjoint groups of gates at level 3, and so on. Eventually, the recursion reaches the level of the sparse polynomials (the leaf nodes) where a faithful map can be constructed using ideas from [KS01].

Let us formalize this proof idea. For any multiset of variables $S$, let $\Delta_S f$ denote the partial derivative of $f$ with respect to the variables in $S$ (including repetitions, as $S$ is a multiset). Let var($S$) denote the set of distinct variables in $S$.

**Lemma 4.2** (Gcd trick). Let $G$ be any gate in $C$ and $S_1, \ldots, S_w$ be multisets of variables. Then there exists another occur-$k$ formula $G'$ for which, the vector of polynomials $(\Delta_{S_1} G, \ldots, \Delta_{S_w} G) = V_{G'} \cdot (\Delta_{S_1} G', \ldots, \Delta_{S_w} G')$ such that
1. If $G$ is a $+$ gate then $G'$ is also a $+$ gate whose children consist of at most $k \cdot |\cup_{i=1}^m \text{var}(S_i)|$ of the children of $G$, and $V_G = 1$.

2. If $G$ is a $\times \land$ gate, then $G'$ is also a $\times \land$ gate whose children consist of at most $k \cdot |\cup_{i=1}^m \text{var}(S_i)|$ of the children of $G$, and $V_G = G/G'$.

Further, the gates constituting $G'$ and $V_G$ are disjoint.

Proof.

1. Suppose $G = H_1 + \cdots + H_m$. Then at most $k \cdot |\cup \text{var}(S_i)|$ of its children depend on the variables present in $\cup \text{var}(S_i)$; let $G'$ be the sum of these children. Then, $\Delta_{S_i} G = \Delta_{S_i} G'$ as the other gates are independent of the variables in $\cup S_i$.

2. Suppose $G = H_1^{e_1} \cdots H_m^{e_m}$. Since $G$ is a gate in an occur-$k$ formula, at most $k \cdot |\cup \text{var}(S_i)|$ of the $H_i$'s depend on the variables in $\cup S_i$; call these $H_1, \ldots, H_t$. Let $G' := H_1^{e_1} \cdots H_t^{e_t}$ and $V_G := G/G'$. Then, $\Delta_{S_i} G = V_G \cdot \Delta_{S_i} G'$ as claimed.

We say that a map is faithful to a collection of sets if it is faithful to every set in the collection. Going by the ‘proof idea’, suppose at the $\ell$-th level of the recursion we want to construct a $\Psi_\ell$ that is faithful to a collection of (at most) $s$ sets of polynomials, each set containing at most $r_\ell$ partial derivatives (of order up to $c_\ell$) of the gates at level $\ell$. Moreover, the sets involve derivatives of disjoint groups of gates. To begin with: $\ell = 1$ and we wish to construct a $\Psi_1$ that is faithful to just one set $T$, so $r_1 \leq k$ and $c_1 = 0$. The next lemma captures the evolution of the recursion, and is essentially a careful analysis of the growth of the sets as we descend levels.

Lemma 4.3 (Evolution via factoring). Let $\mathcal{U}$ be a set of $r_\ell$ derivatives (of orders up to $c_\ell$) of gates $G_\mathcal{U}$ at level $\ell$, and $\mathcal{U}'$ be a transcendence basis of $\mathcal{U}$. Any $|\mathcal{U}'| \times |\mathcal{U}'|$ minor of $\mathcal{J}_k(\mathcal{U}')$ is of the form $\prod_{i} V_{i}^{e_i}$, where $V_i$'s are polynomials in at most $r_\ell+1 := (c_\ell+1) \cdot 2^{c_\ell+1} k \cdot r_\ell^2$ many derivatives (of order up to $c_\ell+1 := c_\ell+1$) of disjoint groups of children of $G_\mathcal{U}$.

Proof. Let $r' = |\mathcal{U}'|$ and $M$ be an arbitrary $r' \times r'$ sub-matrix of $\mathcal{J}_k(\mathcal{U}')$. Let $(\Delta_{S_1} G, \ldots, \Delta_{S_s} G)$ be an arbitrary row of $M$ that corresponds to certain derivatives of a node $G$ at level $\ell$. Since each of the derivatives are of order at most $c_\ell+1$, we have that $|\cup_{i=1}^s \text{var}(S_i)| \leq r'(c_\ell)+r' \leq r_\ell(c_\ell+1)$. By applying Lemma 4.2 to this row, we can express this row as $V_G \cdot (\Delta_{S_1} G', \ldots, \Delta_{S_s} G')$. So, in $\det(M) = \prod_{G} V_G \cdot \det M'$ where $M'$ is the submatrix obtained after removing the $V_G$’s common from each of the rows. Thus, the elements present inside $M'$ are of the form $\Delta_{S_i} G'$, where $G'$ has at most $kr_\ell(c_\ell+1)$ children.

Since each $|S_i| \leq c_\ell+1$ and $G$ is a node in an occur-$k$ formula, at most $k(c_\ell+1)$ children of $G'$ depend on $\text{var}(S_i)$.

If $G'$ is a $+$ gate, then $\Delta_{S_i} G'$ is the sum of the derivatives of at most $k(c_\ell+1)$ of its children (that depend on $\text{var}(S_i)$).

If $G'$ is a $\times \land$ gate computing $H_1^{e_1} \cdots H_t^{e_t}$ (where $t \leq kr_\ell(c_\ell+1)$), then $\Delta_{S_i} G'$ is a polynomial combination of the $H_i$'s and $\{\Delta_T H_j\}_{T \subseteq S_i}$ for each $H_j$ depending on $\text{var}(S_i)$.

Hence in either case, $\Delta_{S_i} G'$ is a polynomial in at most $kr_\ell(c_\ell+1)+k(c_\ell+1)2^{c_\ell+1}$ many derivatives (of order atmost $(c_\ell+1)$) of the children of $G'$. Summing across the $r'$ elements in that row, the number of derivatives used are at most $kr_\ell(c_\ell+1)+k(c_\ell+1)2^{c_\ell+1}$ many derivatives used are at most $kr_\ell^2(c_\ell+1)+2^{c_\ell+1}$. Further, each $V_G$ that was removed as a common factor in a row is just a product of gates in level $(\ell+1)$, and are disjoint from the gates whose derivatives constitute $M'$ (by Lemma 4.2).
Let $C_\ell$ denote the collection of sets for which we want to construct a faithful map $\Psi_\ell$ at the $\ell$-th level of the recursion. To begin with $C_1 = \{(T_1, \ldots, T_k)\}$ and the collection $C_{\ell+1}$ is formed from $C_\ell$ in the following fashion:

For any $S \in C_\ell$, consider a maximal non-singular minor of $J(S)$, and by Lemma 4.3, this minor can be written as a product of $V_i^{(S)}$s where each $V_i^{(S)}$ is a function of at most $r_{\ell+1}$ derivatives of polynomials computed in level $\ell + 1$. Denote the set of derivatives that $V_i^{(S)}$ depends on as $\text{Elem}(V_i^{(S)})$. Then, $C_{\ell+1} = \{\text{Elem}(V_i^{(S)}) : S \in C_\ell\}$.

It follows from the above lemma that the groups of gates whose derivatives form the different $\text{Elem}(V_i)$s are disjoint and therefore $|C_{\ell+1}| \leq s$. Using Lemma 2.7 & 4.3, we can lift a map $\Psi_{\ell+1}$ to construct $\Psi_\ell$.

**Corollary 4.4.** If $\Psi_{\ell+1}$ is faithful to $C_{\ell+1}$ then $\Psi_\ell : x_i \mapsto \left(\sum_{j=1}^{r} y_j t^j + \Psi_{\ell+1}(x_i)\right)$ is faithful to $C_\ell$, where $\{y_1, \ldots, y_r, t\}$ is a fresh set of variables.

Unfolding the recursion until we reach the level of the sparse polynomials (at level $(D-2)$), if $\Psi_{D-2}$ is a faithful homomorphism for $C_{D-2}$, then $\Psi_1$ defined by

$$\Psi_1(x_i) \mapsto \sum_{\ell=1}^{D-3} \sum_{j=1}^{r} y_j t^j + \Psi_{D-2}(x_i)$$

(4.1)

is faithful to $\{(T_1, \ldots, T_k)\}$. Hence, we are reduced to the task of constructing a faithful map for the collection $C_{D-2}$ which consists of at most $s$ sets of derivatives of $s$-sparse polynomials (i.e. consisting of at most $s$ monomials), and each set being of size at most $r_{D-2}$. Constructing a faithful homomorphism for a collection of sets of sparse polynomials can be achieved using standard techniques used in sparse-polynomial identity testing.

**Lemma 4.5.** Let $C = \{W_1, \ldots, W_s\}$ where each $W_s$ is a set of $r$ polynomials that are $n$-variate, degree $d$ and $s$-sparse. For each $1 \leq p \leq (s'nd)^4$, one of the maps $\Psi^{(p)} : \mathbb{F}[x] \to \mathbb{F}[y_1, \ldots, y_r, t, u]$ defined by

$$\Psi^{(p)} : x_i \mapsto \left(\sum_{j=1}^{r} y_j t^j + u^{(dr+1) \mod p}\right)$$

is faithful to the collection $C$.

**Proof.** For each $W \in C$. Any maximal non-zero minor of $J_\ell(W)$ is a sparse polynomial with sparsity bounded by $s'$ and degree bounded by $dr$. Using [KS01], the nonzeroness of this determinant is maintained by one of the maps $\Phi^{(p)} : x_i \mapsto u^{(dr+1) \mod p}$ as $p$ varies from 1 to $O((s'nd)^4)$. For such a map $\Phi_p$, Lemma 2.7 asserts that $\Psi^{(p)} : x_i \mapsto \sum_{j=1}^{r} y_j t^j + \Phi^{(p)}(x_i)$ is faithful to $C$. \(\square\)

With the above lemma and equation (4.1), we can achieve our goal of constructing a faithful homomorphism for $\{T_1, \ldots, T_k\}$.

**Theorem 4.6.** Let $T_1, \ldots, T_k$ be polynomials computed by depth-$D$ occur-$k$ formulas of size at most $s$ each. Then, a homomorphism that is faithful to $\{T_1, \ldots, T_k\}$ can be constructed in time polynomial in $s^R$, where $R = (2k)^{2D^2}D^2$ (assuming char($\mathbb{F}$) = 0 or > $s^R$).
Proof. Unfolding the recursion in Corollary 4.4 and equation (4.1), it suffices to construct a map $\Psi^{(p)}$ that is faithful to $C_{D-2}$ (as defined earlier). Recall that $C_{D-2}$ is a collection of $s$ sets of at most $r_{D-2}$ derivatives of $s$-sparse polynomials of degree at most $d$. Hence, we can apply Lemma 4.3 to this collection $C_{D-2}$ to get that

$$\Psi^{(p)}(x_i) \mapsto \sum_{\ell=1}^{D-2} \left( \sum_{j=1}^{r_{\ell}} y_{ij} x^j + u^{(sr_{\ell}+1) \bmod p} \right)$$

Using the relation between $r_{\ell+1}$ and $r_{\ell}$ from Lemma 4.3 it is easy to show that $r_{D-2} \leq R = (2k)^{2D-2D^2D}$ and $\sum_{\ell=1}^{D-2} r_{\ell} = O(R)$. 

With the above construction for faithful homomorphism, we can prove both Theorem 1.3 and Theorem 1.10.

**Theorem 1.3 (restated).** A hitting-set for any depth-$D$ occur-$k$ formula $C(x) = T_1 + \cdots + T_k$ of size $s$ can be constructed in time polynomial in $s^R$, where $R = (2k)^{2D-2D^2D}$ (assuming char$(\mathbb{F}) = 0$ or $> s^R$).

**Proof.** Theorem 4.6 gives a homomorphism $\Psi^{(p)}$ that is faithful to $\{T_1, \ldots, T_k\}$ and reduces the number of variables to $O(\sum_{\ell=1}^{D-2} r_{\ell}) = O(R)$. By Theorem 2.4 we have that $C = T_1 + \cdots + T_k = 0$ if and only if $\Psi^{(p)}(T_1) + \cdots + \Psi^{(p)}(T_k) = 0$. Since $C$ is a poly-degree circuit of size $s$, $\Psi^{(p)}(C)$ is a polynomial of degree at most poly$(s)$ in the $O(R)$ variables. Using the lemma in [Sch80 Zip79 DL78], we can construct a hitting-set for $\Phi(D)$ in time polynomial in $s^R$. 

**Theorem 1.10 (restated).** Let $T_1, \ldots, T_r$ be $n$-variate degree $d$ polynomials computed by depth-$D$ occur-$k$ formulas of size $s$ and presented as blackboxes. There is an $(sdn)^R$-time algorithm, where $R = r \cdot (2k)^{2D-2D^2D}$, to test whether $\{T_1, \ldots, T_r\}$ are algebraically independent, assuming char$(\mathbb{F}) = 0$ or $> s^R$.

**Proof.** Theorem 4.6 gives a homomorphism $\Psi^{(p)}$ that is faithful to $\{T_1, \ldots, T_r\}$. Hence it suffices to check whether $\{\Psi^{(p)}(T_1)\}, \ldots, \{\Psi^{(p)}(T_r)\}$ are algebraically independent or not. Since each $\Psi^{(p)}(T_i)$ is a $O(R)$-variable polynomial of degree $poly(sdn)$, they have at most $poly((sdn)^R)$ monomials. Using the interpolation algorithms from [KS01 BOT88] each $\Psi^{(p)}(T_i)$ can be explicitly written down as a sum of monomials from $poly((sdn)^R)$ evaluations. Hence, the Jacobian $J(\Psi^{(p)}(T_1), \ldots, \Psi^{(p)}(T_r))$ can be written down explicitly, and an application of the lemma in [Sch80 Zip79 DL78] allows us to check if $J(\Psi^{(p)}(T_1), \ldots, \Psi^{(p)}(T_r))$ has full rank in deterministic $poly((sdn)^R)$ time. 

4.1. Restriction to the case of depth-4.

**Theorem 1.4 (restated).** A hitting-set for any depth-4 occur-$k$ formula of size $s$ can be constructed in time polynomial in $s^{k^2}$ (assuming char$(\mathbb{F}) = 0$ or $> s^{4k}$).

**Proof.** Let $C = \sum_{i=1}^k T_i$ be a depth-4 occur-$k$ formula, where $T_i = \prod_{j=1}^d P_{ij}^{c_{ij}}$, $P_{ij}$'s are sparse polynomials. The discussion at the beginning of this section justifies the assumption that top fanin is $k$. Once again, assuming $T_r$ to be a transcendence basis of $T$, we need to design a $\Psi$ such that $\Psi(J_{x_r}(T_r)) \neq 0$. Let us count the number of $P_{ij}$'s that depend on the variables $x_r$, the remaining $P_{ij}$'s can be taken out common from every row of $J_{x_r}(T_r)$ while computing its determinant - this is the first ‘taking common’ step. Let $c_{i\ell}$ be the number of $P_{ij}$’s present in $T_i$ that depend on $x_\ell$. The total number of sparse polynomials depending on $x_r$ is therefore $\sum_{1 \leq i, \ell \leq r} c_{i\ell}$. From the condition of occur-$k$, $\sum_{i,\ell} c_{i\ell} \leq k$ and hence $\sum_{i,\ell} c_{i\ell} \leq rk \leq k^2$. Let $c_i := \sum_j c_{ij}$,
be the number of \( x_r \)-dependent \( P_{ij} \)'s present in \( T_i \). For an \( x_r \)-dependent \( P_{ij} \), we can also take \( P_{ij}^{c_{ij} - 1} \) common from the \( i \)-th row of \( J_{x_r}(T_r) \) - call this the second ‘taking common’ step. The sparsity of every entry of the \( i \)-th row of the residual matrix \( M \) - after the two ‘taking common’ steps - is bounded by \( c_i s^{c_i} \), where \( s \) is the size of \( C \). Thus, \( \det(M) \) has sparsity at most \( r! \cdot \prod_{i=1}^{c_i} c_i s^{c_i} = s^{O(k^2)} \), which implies that \( J_{x_r}(T_r) \) is a product of at most \( s + 1 \) powers of sparse polynomials, each of whose sparsity is bounded by \( s^{O(k^2)} \) and degree bounded by \( sk \). As argued before, use \([KS01]\) along with Lemma \ref{hit} to construct a hitting-set for \( C \) in time \( s^{O(k^2)} \) (assuming \( \text{char}(F) = 0 \) or \( s^{4k} \)). \( \square \)

5. Lower bounds for the immanent. For sake of simplicity, we prove the lower bounds for \( \det_n \) - determinant of an \( n \times n \) matrix \( M = (x_{ij}) \) - assuming zero characteristic. All our arguments apply to \( \text{Imm}_n(M) \) for any character \( \chi \) and this is elaborated in Section \ref{imm}. The following two lemmas are at the heart of our approach to proving lower bounds. Let \( x := \{ x_{ij} : 1 \leq i, j \leq n \} \) and \( T := \{ T_1, \ldots, T_m \} \), where \( T_i \)’s are polynomials in \( F[x] \). We shall defer the proofs of these lemmas to the end of the section.

**Lemma 5.1.** Suppose \( \det_n = C(T_1, \ldots, T_n) \), where \( C \) is any circuit and let \( T_r = \{ T_1, \ldots, T_r \} \) be a transcendence basis of \( T \) with \( r < n \). Then, there exist a set of \( r + 1 \) variables \( x_{r+1} \subset x \) and an equation \( \sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0 \) such that \( M_i \)'s are distinct first order principal minors of \( M \), \( f_i \)'s are distinct \( r \times r \) minors of \( J_{x_{r+1}}(T_r) \), not all \( f_i \)'s are zero, and \( c_i \in \mathbb{F}^* \).

**Lemma 5.2.** If \( M_1, \ldots, M_t \) are distinct first order principal minors of \( M \) and \( \sum_{i=1}^{t} f_i \cdot M_i = 0 \) (not all \( f_i \)'s are zero) then the total sparsity of the \( f_i \)'s is at least \( \frac{s^{n/2 - t}}{2^{n/2 - t}} \).

5.1. Lower bound on depth-4 occur-k formulas.

**Proof of Theorem \ref{detlb}** Let \( C \) be a depth-4 occur-k formula of size \( s \) that computes \( \det_n \). Since \( \det_n \) is irreducible we can assume a top + gate in \( C \). Then \( \tilde{C} := C(x_{11} + 1, x_{12}, \ldots, x_{nm}) - C(x) \) is a depth-4 occur-2k formula of size at most \( s^2 + s \leq 2s^2 \) and top fanin bounded by \( 2k \) (similar argument as at the beginning of Section \ref{section4}). Moreover, \( \tilde{C} \) computes the minor of \( M \) with respect to \( x_{11} \) which is essentially \( \det_{n-1} \). By reusing symbols, assume that \( C \) is a depth-4 occur-k formula with top fanin bounded by \( k \), and \( C \) computes \( \det_n \).

Let \( C = \sum_{i=1}^{k} T_i = \det_n \), where \( T_i = \prod_{j=1}^{d} P_{ij}^{c_{ij}} \), \( P_{ij} \)'s are sparse polynomials. Let \( T_r \) be a transcendence basis of \( T = \{ T_1, \ldots, T_k \} \). By Lemma \ref{lem1} we have an equation \( \sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0 \) such that \( f_i \)'s are distinct \( r \times r \) minors of \( J_{x_{r+1}}(T_r) \) for some set of \( r + 1 \) variables \( x_{r+1} \). Arguing in the same way as in the proof of Theorem \ref{detlb} (in Section \ref{section4}), we can throw away certain common terms from the minors \( f_i \)'s and get another equation \( \sum_{i=1}^{r+1} g_i M_i = 0 \), where the sparsity of each \( g_i \) is \( s^{O(k^2)} \). If we apply Lemma \ref{lem2} on this equation, we get our desired result. \( \square \)

5.2. Lower bound on circuits generated by \( \Sigma \Pi \) polynomials.

**Proof of Theorem \ref{lb}** In Lemma \ref{lem1} take the \( T_i \)'s to be sparse polynomials with sparsity bounded by \( s \). Then, in the equation \( \sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0 \), each \( f_i \) has sparsity at most \( r! \cdot s^r \leq (rs)^r \). Finally, apply Lemma \ref{lem2} to obtain the desired lower bound. \( \square \)
5.3. Lower bound on circuits generated by $\Pi \Sigma$ polynomials.

**Proof of Theorem 1.8** Let $T = \{T_1, \ldots, T_m\}$ be products of linear polynomials such that $C(T_1, \ldots, T_n) = \text{Det}_n$ with $T_k = \{T_1, \ldots, T_k\}$ being a transcendence basis (we choose to denote the transcendence degree by $k$ to be consistent with Section 3). By Lemma 5.1, we get $\sum_{i=1}^{k+1} c_i f_i M_i = 0$ where the $f_i$’s are $k \times k$ minors of $J_{x \cdot t+1} (T_k)$ and $wlog f_1 \neq 0$. Like in Section 3, by Lemma 5.1 (restated), we can rewrite the above equation in the form $H_0 := T \sum_L \alpha_L (M_{\ell +1})/\ell_1 \cdots \ell_k = 0$ where each $\alpha_L (M_{\ell +1}) := \sum_{i=1}^{k+1} \alpha_{L,i} M_i$ is an $F$-linear combination of the distinct minors $M_{\ell +1} := \{M_1, \ldots, M_{\ell +1}\}$. Observe that $H_0$ is a sum of products of linear polynomials, with ‘coefficients’ being $F$-linear combinations of $M_{\ell +1}$. And since $f_1 \neq 0$, the ‘coefficient’ of $M_1$ in $H_0$ is a nonzero depth-3 circuit.

The invariant that shall be maintained is that the coefficient of $M_1$ (modulo some linear polynomials), which is a depth-3 circuit, would stay nonzero. This would finally yield a non-trivial linear combination $\alpha_L (M_{\ell +1}) = 0$ mod $\ell_k$ (where $\ell_k$ is a set of $k$ independent linear polynomials) which we can apply the following lemma (where the proof shall be deferred to the end of the section as well).

**Lemma 5.3.** If $M_1, \ldots, M_t$ are distinct first order principal minors of $M$ and $\sum_{i=1}^t \alpha_i M_i = 0$ mod $\ell_k$ (not all $\alpha_i = 0$) for independent linear polynomials $\ell_k$, then $t + k \geq n$.

Formally, define the content of $H = T \sum_L \alpha_L (M_{\ell +1})/\ell_1 \cdots \ell_k$ as $\text{cont}(H) := \gcd_L \{T/\ell_1 \cdots \ell_k\}$, and also define $\text{sim}(H) := H/\text{cont}(H)$. Therefore $\text{sim}(H_0)$ can be written in the form $F_0 \sum_L \alpha_L (M_{\ell +1})/\ell_1 \cdots \ell_k$. The coefficient of $M_1$ in the above expression is $F_0 \sum_L \alpha_{L,1} /\ell_1 \cdots \ell_k$, which by assumption is a nonzero depth-3 circuit of degree at most $|\mathcal{L}(H_0)| - k$. Therefore by Chinese remaindering, $\exists \ell \in \mathcal{L}(H_0)$ such that this coefficient remains nonzero modulo $\ell_1$. Hence, we can define $H_1 := \text{sim}(H_0) \mod \ell_1$ which has the form $H_1 = F_0 /\ell_1 \cdot \sum_{L \in \mathcal{L}} \alpha_L (M_{\ell +1})/\ell_2 \cdots \ell_k = 0$ mod $\ell_1$. And, like in Section 3, reducing modulo $\ell = c_1 x_1 + \sum_{i > 1} c_i x_i$ (where $c_1 \neq 0$) is equivalent to replacing $x_1$ by $\sum_{i > 1} (c_i/c_1) x_i$ in the expression. Thus, we may write $H_1 = F_1 /\ell_1 \cdot \sum_{L \in \mathcal{L}} \alpha_L (M_{\ell +1} \mod \ell_1)/\ell_2 \cdots \ell_k = 0$, and the choice of $\ell_1$ maintains the invariant that the coefficient of $(M_1 \mod \ell_1)$ is nonzero.

The above argument can be repeated inductively. In general, we have $H_i = \text{sim}(H_{i-1}) \mod \ell_i$ for a similar choice of $\ell_i$ via Chinese remaindering, and $H_i$ has the form $H_i = F_i /\ell_i \cdot \sum_{L \in \mathcal{L}} \alpha_L (M_{\ell +1} \mod \ell_i)/\ell_{i+1} \cdots \ell_k = 0$. Eventually, we obtain $H_{n+1} := F_{n+1} \cdot \alpha_L (M_{\ell +1} \mod \ell_{n+1}) = 0$ while the coefficient of $M_1 \mod \ell_k$ is nonzero. This implies that $\alpha_L (M_{\ell +1}) = 0$ mod $\ell_k$ is a non-trivial equation. And Lemma 5.3 asserts that this is not possible unless $2k + 1 \geq n$ or $k \geq (n-1)/2$. $\square$

5.4. Proofs of the technical lemmas.

**Lemma 5.1 (restated).** Suppose $\text{Det}_n = C(T_1, \ldots, T_m)$, where $C$ is any circuit and let $T_r = \{T_1, \ldots, T_r\}$ be a transcendence basis of $T$ with $r < n$. Then, there exist a set of $r+1$ variables $x_{r+1} \subset x$ and an equation $\sum_{i=1}^{r+1} c_i f_i M_i = 0$ such that $M_i$’s are distinct first order principal minors of $M$, $f_i$’s are distinct $r \times r$ minors of $J_{x_{r+1}} (T_r)$, not all $f_i$’s are zero, and $c_i \in F^*$. 

**Proof.** In a column of a Jacobian matrix $J_x (\cdot)$, all the entries are differentiated with respect to a variable $x$, we will say that the column is indexed by $x$. Let $T_r = \{T_1, \cdots, T_r\}$ be a transcendence basis of $T$. Amongst the nonzero $r \times r$ minors of
Lemma 5.2 (restated). If $M_1, \ldots, M_t$ are distinct first order principal minors of $M$ and $\sum_{i=1}^t f_i \cdot M_i = 0$ (not all $f_i$’s are zero) then the total sparsity of the $f_i$’s is at least $2^{n/2-t}$.

Proof. The proof is by contradiction. The idea is to start with the equation $\sum_{i=1}^t f_i \cdot M_i = 0$ and apply two steps - sparsity reduction and fanin reduction - alternately, till we arrive at a contradiction in the form of an equation $f_j \cdot M_j = 0$, where neither $f_j$ nor $M_j$ is zero if the total sparsity of the $f_j$’s is less than $2^{n/2-t}$. With an equation of the form $\sum_{i=1}^t g_i \cdot N_i = 0$, we associate four parameters $\tau$, $s$, $\eta$, and $c$. These parameters are as follows: $\tau$ is called the fanin of the equation, $s$ is the total sparsity of the $g_i$’s (we always assume that not all the $g_i$’s are zero), every $N_i$ is a distinct first order principal minor of a symbolic $\eta \times \eta$ matrix $N = (x_{ij})$, and $c$ is the maximum number of entries of $N$ that are set as constants. To begin with, we choose $g_i = f_i$ and $N_i = M_i$ for all $1 \leq i \leq t$, so $\tau = t$, $s = s$ (the total sparsity of the $f_i$’s), $\eta = n$, $N = M$ and $c = 0$. In the ‘sparsity reduction’ step, we start with an equation $\sum_{i=1}^r g_i \cdot N_i = 0$, with parameters $\tau$, $s$, $\eta$, $c$ and arrive at an equation $\sum_{i=1}^{r'} g_i' \cdot N_i' = 0$ with parameters $\tau'$, $s'$, $\eta'$, $c'$ such that $\tau' \leq \tau$, $s' \leq s'/2$, $\eta - 1 \leq \eta' \leq \eta$, and $c' \leq c + 1$. In the ‘fanin reduction’ step, we start with an equation $\sum_{i=1}^s g_i \cdot N_i = 0$, with parameters $\tau$, $s$, $\eta$, $c$ and arrive at an equation $\sum_{i=1}^{s'} g_i' \cdot N_i' = 0$ with parameters $\tau'$, $s'$, $\eta'$, $c'$ such that one of the two cases happens - Case 1: $\tau' \leq \tau - 1$, $s' \leq s$, $\eta' = \eta - 1$, and $c' = c$; Case 2: $\tau' = 1$, $s' \leq s$, $\eta' = \eta$, and $c' \leq c + \tau$.

Naturally, starting with $\sum_{i=1}^t f_i \cdot M_i = 0$, the ‘sparsity reduction’ step can only be performed at most $\log s$ many times (since the total sparsity of the $g_i$’s reduces by at least a factor of half every time this step is executed), whereas the ‘fanin reduction’ step can be performed at most $t - 1$ times (as the fanin goes down by at least one for each such step). Finally, when this process of alternating steps ends, we have an equation of the form $g_i \cdot N_i = 0$ (Case 2 of the fanin reduction step), where $g_i \neq 0$ and $N_i$ is a principal minor of a symbolic matrix $N$ of dimension at least $n - (\log s + t - 1)$ such that at most $n - (\log s + t)$ entries of $N$ are set as constants. Now, if $\log s + t \leq n - (\log s + t)$ the $N_i$ can never be zero (by Fact 5.6) and hence we arrive at a contradiction. Therefore, $s > 2^{n/2-t}$. Now, the details of the sparsity reduction and the fanin reduction steps.

Suppose, we have an equation $\sum_{i=1}^t g_i \cdot N_i = 0$ as mentioned above. Without loss
of generality, assume that the minor \( N_i \) is the minor of \( N \) with respect to the \( i \)th diagonal element of \( N \). Call all the variables \( x_{ij} \) in \( N \) with both \( i, j > \tau \) as the white variables. These are the variables that are present in every minor \( N_i \) in the sum \( \sum_{i=1}^{\tau} g_i N_i \). The variables \( x_{ij} \) where both \( i, j \leq \tau \) are called the black variables, and the remaining are the grey variables. By assumption, \( c \) of the variables in \( N \) are set as constants.

**Sparsity reduction step** - Say \( x \) is a white variable that one of the \( g_i \)'s depends on. Writing each \( g_i \) as a polynomial in \( x \), there will be two distinct powers of \( x \) amongst the \( g_i \)'s (for if not, then \( x \) can be taken common across all \( g_i \)'s). Let \( x^\ell \) be the lowest degree and \( x^h \) be the highest. Dividing the entire equation \( \sum_{i=1}^{\tau} g_i N_i = 0 \) by \( x^\ell \), we can further assume that \( \ell = 0 \). Each of the \( g_i \)'s and \( N_i \)'s can be expressed as constants. Looking at the coefficients of \( x^0 \) and \( x^{h+1} \) in the equation yields \( \sum_{i=1}^{\tau} g_i \cdot N_i = 0 \) and \( \sum_{i=1}^{\tau} g_i \cdot N_{i,1} = 0 \). Note that \( N_i,0 \)'s can be thought of as principal minors of the \( \eta \times \eta \) matrix \( N' \) obtained by setting \( x = 0 \) in \( N \). And each of the \( N_i,1 \)'s can be thought of as minors of the \( (\eta - 1) \times (\eta - 1) \) matrix \( N' \) which is the matrix associated with the minor of \( N \) with respect to \( x \). Since the monomials in \( g_i,0 \) and \( x^h g_i,1 \) are disjoint, either the total sparsity of the \( g_i,0 \)'s or the total sparsity of the \( g_i,1 \)'s is \( \leq \eta / 2 \). Thus, one of the equations \( \sum_{i=1}^{\tau} g_i,0 \cdot N_i,0 = 0 \) or \( \sum_{i=1}^{\tau} g_i,1 \cdot N_i,1 = 0 \) yields an equation of the form \( \sum_{i=1}^{\tau} g_i N'_i = 0 \) with parameters \( \tau', s', \eta' \), \( c' \) as claimed before. (In case, we choose \( \sum_{i=1}^{\tau} g_i,1 \cdot N_i,1 = 0 \) as our next equation, we also set the variables in the same columns and rows of \( x \) to constants in such a way that a \( g_i,1 \) stays nonzero. This is certainly possible over a characteristic zero field \([Sch80, Zip79]\).) The sparsity reduction step is performed whenever the starting equation \( \sum_{i=1}^{\tau} g_i N_i = 0 \) has a white variable among the \( g_i \)'s. When all the \( g_i \)'s are free of white variables, we perform the **fanin reduction step**.

**Fanin reduction step** - When we perform this step, all the \( g_i \)'s consist of black and grey variables. Pick a row \( R \) of \( N \) barring the first \( \tau \) rows. Let \( y_1, \ldots, y_\tau \) be the grey variables occurring in \( R \) (these are, respectively, the variables in the first \( \tau \) columns of \( R \)). Starting with \( y_2 \), divide the equation \( \sum_{i=1}^{\tau} g_i N_i = 0 \) by the largest power of \( y_2 \) common across all monomials in the \( g_i \)'s, and then set \( y_2 = 0 \). This process lets us assume that there exists at least one \( g_i \) which is not zero at \( y_2 = 0 \). On the residual equation, repeat the same process with \( y_3 \) and then with \( y_4 \) and so on till \( y_\tau \). Thus, we can assume without loss of generality that in the equation \( \sum_{i=1}^{\tau} g_i N_i = 0 \) there is at least one \( g_i \) that is not zero when \( y_2, \ldots, y_\tau \) are set to zero. Observe that if \( g_1 \) is the only \( g_i \) that stays nonzero under the projection \( y_2 = \ldots = y_\tau = 0 \) then \( (g_1 N_1)(y_2 = \ldots = y_\tau = 0) = 0 \), implying that \( N_1 = 0 \) under the same projection - this is Case 2 of the fanin reduction step mentioned earlier. Now, assume that there is a \( g_i \) other than \( g_1 \) (say, \( g_2 \)) that is nonzero under the projection \( y_2 = \ldots = y_\tau = 0 \). Set all the remaining variables of row \( R \) to zero except \( y_1 \) - these are the white variables in \( R \). Since the \( g_i \)'s are free of white variables (or else, we would have performed the ‘sparsity reduction’ step), none of the \( g_i \)'s is effected by this projection. However, \( N_1 \) being a minor with respect to the first diagonal element of \( N \), vanishes completely after the projection. Any other \( N_i \) takes the form \( y_1 \cdot N'_i \), where \( N'_i \) is a principal minor of a \( (\eta - 1) \times (\eta - 1) \) matrix \( N' \) which is the matrix associated with the minor of \( N \) with respect to \( y_1 \). Therefore, after the projection, the equation \( \sum_{i=1}^{\tau} g_i N_i = 0 \) becomes \( \sum_{i=2}^{\tau} \tilde{g}_i \cdot y_1 N'_i = 0 \Rightarrow \sum_{i=2}^{\tau} \tilde{g}_i \cdot N'_i = 0 \), where \( \tilde{g}_i \) is the image of \( g_i \) under the above mentioned projection and further \( \tilde{g}_2 \neq 0 \). The \( \tilde{g}_i \)'s might still contain variables from the first column of \( N \). So, as a final step, set these variables to values so that
a nonzero \( \tilde{g}_i \) remains nonzero after this projection (the \[Sch80\] [Zip79] [DL78] lemma asserts that such values exist in plenty). This gives us the desired form \( \sum_{i=1}^{t'} g'_i \theta'_i = 0 \) with parameters \( \tau', s', \eta', c' \) as claimed before (Case 1 of the fanin reduction step mentioned earlier). \( \square \)

**Lemma 5.3** (revised). If \( M_1, \ldots, M_t \) are distinct first order principal minors of \( M \) and \( \sum_{i=1}^t \alpha_i M_i = 0 \mod \ell_k \) (not all \( \alpha_i = 0 \)) for independent linear polynomials \( \ell_k \), then \( t + k \geq n \).

**Proof.** Assume that \( t + k < n \) (with \( t \geq 1 \) it means \( k \leq n - 2 \)) and \( \alpha_i M_i = 0 \mod \ell_k \). Recall that reducing an expression modulo \( \ell = c_1 x_1 + \sum_{i>1} c_i x_i \) (with \( c_1 \neq 0 \)) is equivalent to replacing \( x_1 \) by \( \sum_{i>1} (-c_i/c_1) x_i \). Hence, as \( \ell_1, \ldots, \ell_k \) are independent linear polynomials, the equation may be rewritten as \( \sum_{i=1}^t \alpha_i M'_i = 0 \) where \( (M'_i) \)'s are minors of the matrix \( M' \) obtained by replacing \( r \) entries of \( M \) by linear polynomials in other variables. We shall call these entries as corrupted entries. Without loss of generality, we shall assume that \( M'_1 \) is the minor corresponding to the \( i \)-th diagonal variable and that all the \( \alpha_i \)'s are nonzero.

**Claim 5.4.** Each of the first \( t \) rows and columns of \( M \) must have a corrupted entry.

**Proof.** Suppose the first row (without loss of generality) is free of any corrupted entry. Then, setting the entire row to zero would make all \( M'_i = 0 \) for \( i \neq 1 \). But since \( \sum \alpha_i M'_i = 0 \), this forces \( M'_1 \) to become zero under the projection as well. This leads to a contradiction as \( M'_1 \) is a determinant of an \( (n-1) \times (n-1) \) symbolic matrix under a projection, and this cannot be zero unless \( k \geq n - 1 \) (by Fact 5.6). \( \square \)

Since \( n - k > t \), there must exist a set of \( t - 1 \) rows, say \( \{R_1, \ldots, R_{t-1}\} \), of \( M \) that are free of any corrupted entries. For each of these rows, set the \( i \)-th variable of row \( R_i \) to 1, and every other variable in \( R_1, \ldots, R_{t-1} \) to zero. These projections make \( M'_i = 0 \) for all \( i \neq t \) (as in these minors an entire row vanishes). And since \( \sum_{i=1}^{t} \alpha_i M'_i = 0 \), this forces \( M'_t \) to become zero under this projection as well. But under this projection, \( M'_t \) just reduces (up to a sign) to the minor obtained from \( M' \) by removing the columns \( \{1, \ldots, t\} \) and rows \( \{R_1, \ldots, R_{t-1}\} \cup \{t\} \). This is a determinant of an \( (n-t) \times (n-t) \) symbolic matrix, containing at most \( k - t \) corrupted entries, thus \( k - t \geq n - t \) (by Fact 5.6). But then \( k \geq n \), which contradicts our initial assumption. \( \square \)

**5.5. Extensions to immanants.** All the lower bound proofs use some very basic properties of \( \text{Det}_n \). These properties are general enough that they apply to any immanant. For any map \( \chi : S_n \to \mathbb{C}^\times \), recall the definition of the immanant of an \( n \times n \) matrix \( M = (x_{ij}) \):

\[
\text{Imm}_\chi(M) = \sum_{\sigma \in S_n} \chi(\sigma) \prod_{i=1}^{n} x_{i, \sigma(i)}
\]

From the image of the map \( \chi \) it is obvious that \( \chi(\sigma) \neq 0 \) for any \( \sigma \in S_n \).

**Definition 5.5.** The minor of \( \text{Imm}_\chi(M) \) with respect to the \((i, j)\)-th entry is defined as

\[
(\text{Imm}_\chi(M))_{i,j} = \sum_{\sigma \in S_n \atop \sigma(i) = j} \chi(\sigma) \prod_{k \neq i} x_{k, \sigma(k)}
\]
This may also be rewritten as Imm$\chi'(M_{ij})$ for a suitable map $\chi' : S_{n-1} \to \mathbb{C}^\times$, where $M_{ij}$ is the submatrix of $M$ after removing the $i$-th row and $j$-th column. From the definition, it follows directly that the partial derivative of Imm$\chi(M)$ with respect to $x_{ij}$ is precisely the minor with respect to $(i,j)$.

The only crucial fact of determinants that is used in all the proofs is that a symbolic $n \times n$ determinant cannot be zero when less than $n$ of its entries are altered.

**Fact 5.6.** Let $M'$ be the matrix obtained by setting $c < n$ entries of $M$ to arbitrary polynomials in $\mathbb{F}[x]$. Then for any map $\chi : S_n \to \mathbb{C}^\times$, we have Imm$\chi(M') \neq 0$.

**Proof.** We shall say an entry of $M'$ is corrupted if it is one of the $c$ entries of $M$ that has been replaced by a polynomial. We shall prove this by carefully rearranging the rows and columns so that all the corrupted entries are above the diagonal. Then, since all entries below the diagonal are free, we may set all of them to zero and the immanant reduces to a single nonzero monomial.

Since less than $n$ entries of $M'$ have been altered, there exists a column that is free of any corrupted entries. By relabelling the columns if necessary, let the first column be free of any corrupted entry. Pick any row $R$ that contains a corruption and relabel the rows to make this the first row. This ensures that the first column is free of any corrupted entry, and the $(n-1) \times (n-1)$ matrix defined by rows and columns, 2 through $n$, contain less than $c-1$ corruptions. By induction, the $c-1$ corruptions may be moved above the diagonal by suitable row/column relabelling. And since the first column is untouched during the process, we now have all $c$ corruptions above the diagonal. Now setting all entries below the diagonal to zeroes reduces the immanant to a single nonzero monomial. □

With this fact, all our lower bound proofs of the determinant can be rewritten for any immanant.

### 6. Conditional lower bounds for depth-$D$ occur-$k$ formulas

In this section, we present a lower bound for depth-$D$ occur-$k$ formulas similar in spirit to Theorem 1.6 by assuming the following conjecture about determinant minors.

**Conjecture 6.1.** Let $M = (x_{ij})$ be an $n \times n$ matrix, and let $x_i$ denote the $i$-th diagonal variable $x_{ii}$. Let $M'$ be a projection of $M$ by setting $c = o(n)$ of the variables in $M$ to constants. Suppose the elements $x_1, \cdots, x_k$, where $k$ is a constant independent of $n$, are partitioned into non-empty sets $S_1, \cdots, S_t$. Consider $M(S_t)$, the set of $t$-th order principal minors of $M'$, each by choosing a $t$-tuple $B \in S_1 \times \cdots \times S_t$ as pivots. Over all possible choices of $B$, we get $m := |S_1| \cdots |S_t|$ many minors. Then for any set of diagonal variables $y_m$ disjoint from $x_k$, $J_{y_m}(M(S_t)) \neq 0$.

The conjecture roughly states that the different $t$-th-order principal minors are algebraically independent. We will need a generalization of Lemma 5.1 for the purposes of this section.

**Lemma 6.2.** Let $\{f_1, \cdots, f_s, g_1, \cdots, g_t\}$ be algebraically dependent polynomials such that $\text{trdeg} \{g_t\} = t$. Let $S \subseteq \mathbb{x}$ be a fixed set of variables of size at least $s + t$. Then there exists a set of $s + t$ variables $x_{s+t} \subset \mathbb{x}$ and an equation of the form

$$\sum_{i=1}^{r} c_i \cdot F_i \cdot G_i = 0 \quad \text{where } r \leq \binom{s + t}{t}$$

such that each $c_i \in \mathbb{F}^*$, each $F_i$ is a distinct $s \times s$ minor of $J_{x_{s+t} \cap S}(f_s)$, each $G_i$ is a distinct $t \times t$ minor of $J_{x_{s+t}}(g_t)$, and not all $G_i$’s are zero.
Note that we are not asserting the nonzeroness of $F_i$'s. Also, Lemma 5.1 may be obtained from the above lemma by taking $f_1 = \text{Det}_n$, $s = 1$ and $S$ to be the set of diagonal variables.

**Proof.** The proof is along the lines of Lemma 5.1. Amongst the nonzero $t \times t$ minors of $J(x, g_t)$, pick one (call the matrix associated with the minor, $N$) that maximizes the number of variables in $S$ indexing the columns of $N$. Without loss of generality, let $x_i$ be the set of variables indexing the columns of $N$. Since $|S| \geq s + t$, there exists $s$ other variables in $S$, say $\{x_{s+1}, \ldots, x_n\}$. Consider the $(s + t) \times (s + t)$ minor of $J_K(f_i \cup g_t)$ corresponding to the columns indexed by $x_{s+t}$ - call the associated $(s + t) \times (s + t)$ matrix $\tilde{N}$.

Since $f_i, g_t$ are algebraically dependent, $\det(\tilde{N}) = 0$. Expanding $\det(\tilde{N})$ over all possible $s \times s$ minors in the first $s$ rows, we have an equation

$$\sum_{U \subseteq x_{s+t}, |U| = s} c_i \cdot F_U \cdot G_U = 0$$

where each $F_U$ is a distinct $s \times s$ minor of $J_{x_{s+t}}(f_i)$, each $G_U$ is a distinct $t \times t$ minor of $J_{x_{s+t}}(g_t)$, and $c_i \in \mathbb{F}^*$. If $G_U$ is the minor with respect to variables $x_i$, then $G_U = \det(N) \neq 0$ (by construction). It suffices to show that if $F_U$ is a minor indexed by variables outside $S$, then $G_U = 0$. This follows, just like in Lemma 5.1 by the maximality assumption on choice of $x_i$. □

The rest of this section shall be devoted to the proof of the following theorem.

**Theorem 6.3.** Assuming Conjecture 6.1, any depth-D occur-k formula that computes $\text{Det}_n$ must have size $s = 2^{\Omega(n)}$ over any field of characteristic zero.

**Proof idea:** The proof proceeds on the same lines as Theorem 1.6. If $T_1, \ldots, T_k$ is a transcendence basis of gates at level 2 computing the determinant, then $J_K(\text{Det}_n, T_1, \ldots, T_k)$ is a matrix of rank $k$. This yields a non-trivial equation of the form $\sum N_i^{(1)} G_i^{(1)} = 0$ where each of the $N_i^{(1)}$'s are principal minors of $M = (x_{ij})$ and $G_i^{(1)}$'s are $k \times k$ minors of $J_K(T_1, \ldots, T_k)$. Here is where we may use Lemma 4.3 to remove common factors and obtain an equation of the form $\sum N_i^{(1)} \tilde{G}_i^{(1)} = 0$ where $\tilde{G}_i^{(1)}$ is a polynomial of constantly many derivatives of polynomials computed at the next level. The above equation may be thought of as a polynomial relation amongst $\{N_i^{(1)}\} \cup \{\text{Elem}(\tilde{G}_i^{(1)})\}$. Applying Lemma 6.2 (with a suitable choice of $S_2$), we get an equation of the form $\sum N_i^{(2)} G_i^{(2)} = 0$ where each $N_i^{(2)}$ is a minor of $J_{S_2} \left( \{N_i^{(1)}\} \right)$, and $G_i^{(2)}$'s are Jacobians minors of $\bigcup \text{Elem}(\tilde{G}_i^{(1)})$. Again after removing common factors, this equation may be interpreted as a polynomial relation amongst the entries of $N_i^{(2)}$ (which are minors of order 2) and $\text{Elem}(\tilde{G}_i^{(2)})$.

Repeating this argument, we finally reach the level of sparse polynomials and thus obtain a non-trivial equation $\sum N_i^{(D-2)} \cdot \tilde{G}_i^{(D-2)} = 0$, where each $N_i^{(D-2)}$ is a Jacobian minor of $(D-3)$-order minors, and each $\tilde{G}_i^{(D-2)}$ is a sparse polynomial. With a slightly more careful choice of the sets $S_i$ in Lemma 6.2, each of the minors $N_i^{(D-2)}$ would be a minor of $J_{S_{D-2}}(\mathcal{M}(S_1, \ldots, S_{D-3}))$. Assuming Conjecture 6.1 we can show that such an equation is not possible unless the sparsity of the $f_i$'s is large, using a similar argument as in Lemma 5.2.
Lemma 6.4. Suppose $\text{Det}_n$ is computed by a depth-$D$ occur-$k$ formula of size $s$. Then there exist variables $x_1, \ldots, x_R$ where $R = R(k, D)$, a partition of $x_R$ into non-empty sets $S_1, \ldots, S_{D'}$, $(D' \leq (D - 2))$ polynomials $f_1, \ldots, f_m$ (not all zero) where $m = |\mathcal{M}(S_D)|^{O(R)}$ and each $f_i$ has sparsity at most $s^R$, such that

$$\sum_{i=1}^{m} f_i \cdot N_i = 0$$

where each $N_i$ is a minor of $\mathcal{J}_k(\mathcal{M}(S_{D'}))$ indexed by diagonal variables.

Proof. To begin with, suppose $\text{Det}_n = C(T_1, \ldots, T_m)$ where $T_1, \ldots, T_m$ are polynomials computed at the first level. So Lemma 5.1 gives a starting equation, though we do not really have a sparsity bound on the $f_i$’s. The proof shall proceed by transforming this equation into another, involving lower level polynomials, till we get a sparsity bound.

In a general step, we would have an equation of the form $C_\ell(\mathcal{M}(S_1, \ldots, S_{\ell-1}), T_1^{(\ell)}, \ldots, T_r^{(\ell)}) = 0$, where each $T_i^{(\ell)}$ is a derivative (of order at most $\ell$) of a polynomial computed at level $\ell$ of the circuit. Without loss of generality, we may assume that $\{T_1^{(\ell)}, \ldots, T_r^{(\ell)}\}$ are algebraically independent. Let $m_\ell := |S_1| \cdots |S_{\ell-1}|$. Choose a set of diagonal elements $S_\ell$ of size $|\mathcal{M}(S_1, \ldots, S_{\ell-1})| + r_\ell$ that is disjoint from $S_1, \ldots, S_{\ell-1}$. Applying Lemma 6.2 with $S_\ell$, we get an equation of the form

$$\sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot G_i^{(\ell)} = 0$$

where $N_i^{(\ell)}$ is an $m_\ell \times m_\ell$ minor of $\mathcal{J}_{S_\ell}(\mathcal{M}(S_{\ell-1}))$ indexed by diagonal variables, each $G_i^{(\ell)}$ is an $r_\ell \times r_\ell$ minor of $\mathcal{J}_{S_\ell}(T_r^{(\ell)})$. Since $C$ is an occur-$k$ circuit, using an argument similar to Lemma 4.3, the above equation may be rewritten as

$$V_\ell \cdot \sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot \tilde{G}_i^{(\ell)} = 0$$

where each $\tilde{G}_i^{(\ell)}$’s is a polynomial function of at most $r_{\ell+1} := (\ell+1)2^{\ell+1} \cdot k(r_\ell + m_\ell)r_\ell$ many derivatives of polynomials computed at level $\ell + 1$. Note that $V_\ell$ cannot be zero as at least one $G_i^{(\ell)}$ was guaranteed to be nonzero by Lemma 6.2. Therefore, $\sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot \tilde{G}_i^{(\ell)} = 0$. Since each $\tilde{G}_i^{(\ell)}$ is a polynomial function of $r_{\ell+1}$ derivatives at the next level, we now have $C_{\ell+1}(\mathcal{M}(S_1, \ldots, S_\ell), T_1^{(\ell+1)}, \ldots, T_r^{(\ell+1)}) = 0$.

Unfolding this recursion, we finally reach the level of sparse polynomials, at which point we have an equation of the form

$$\sum_i c_i^{(D-2)} N_i^{(D-2)} \cdot \tilde{G}_i^{(D-2)} = 0$$

and each $\tilde{G}_i^{(D-2)}$ is a $r_{D-2} \times r_{D-2}$ Jacobian minor of sparse polynomials. Hence, each $\tilde{G}_i^{(D-2)}$ is itself a polynomial of sparsity bounded by $s^{r_{D-2}}$ as claimed. □

We now have to show that an equation of the form $\sum f_i \cdot N_i = 0$ is not possible unless one of the $f_i$’s has exponential sparsity. The above lemma guarantees that at least one of the $f_i$’s are nonzero in this equation, but it could be the case that some of the $N_i$’s are zero. This was not the case in the depth-4 lower bound as each $N_i$
was just a determinant minor. However, in this case they are jacobians of minors. Conjecture 6.1 asserts that the $N_i$'s are nonzero, even if few variables are set to zero. This assumption would be enough to get the required lower bound.

**Lemma 6.5.** Let $|M(S_1,\cdots,S_D)| := m$ be a constant and let $\{N_i\}_{1 \leq t}$ be distinct $m \times m$ minors of $\mathcal{J}_x(M(S_D))$ where the columns of $N_i$ are indexed by a set $T_i$ of diagonal variables of $M$ disjoint from $\bigcup_{j=1}^DS_j$. Suppose $f_1,\cdots,f_t$ are polynomials such that $\sum_{i=1}^tf_i \cdot N_i = 0$ (not all $f_i$'s are zero). Then, assuming Conjecture 6.1 is true, the total sparsity of the $f_i$'s is $2^{|D|n}$.

**Proof.** The proof is along the similar lines like the proof of Lemma 5.2 and shall proceed by a similar series of sparsity reduction and fanin reduction steps to arrive at a contradiction. Throughout the proof, Conjecture 6.1 shall assert that $N_i$'s stay nonzero (even when few variables are set to constants). We briefly describe the sparsity reduction and the fanin reduction steps and the rest of the proof would follow in essentially an identical fashion as the proof of Lemma 5.2.

Without loss of generality, assume that $\{x_1,\cdots,x_r\}$ is the union of the sets $S_i$'s and $T_i$'s. Let $N$ refer to the matrix of indeterminates that the $N_i$'s are derived from. In our case, $N$ would be obtained by (possibly) setting few variables to constants in $M = (x_{ij})$. We'll refer to all the variables $x_{ij}$ where both $i,j > r$ as white variables; these are present in every entry of each $N_i$. The variables $x_{ij}$ where both $i,j \leq r$ shall be called black variables, and the rest called grey variables. Here again, the sparsity reduction step shall be applied whenever one of the $f_i$'s depends on a white variable, otherwise the fanin reduction steps shall be applied.

**Sparsity-reduction step** - Suppose one of the $f_i$'s depend on a white variable $x$. Then each $N_i$ can be written as $N_i = N_{i,0} + \cdots + x^mN_{i,m}$, and $f_i = f_{i,0} + \cdots + f_{i,h}x^h$. One of the two equations corresponding to the coefficient of $x^0$ and $x^{h+m}$ yields a similar equation with sparsity reduced by a factor of $1/2$. Observe that $N_{i,0}$ is just $N_i |_{x=0}$, and hence the polynomials $\{N_{i,0}\}$ may be thought of as corresponding Jacobian minors of $N'$ obtained by setting $x = 0$ in $N'$. Also, $N_{i,m}$ is obtained by replacing every entry of the matrix corresponding to $N_i$ by its minor with respect $x$. And hence, $N_{i,m}$ can be thought of as a corresponding Jacobian minor of $N_x$ obtained by taking the minor of $N$ with respect to $x$. Thus the two equations corresponding to the coefficient of $x^0$ and $x^{h+m}$ are indeed of the same form as $\sum f_iN_i = 0$. (In the case of the coefficient of $x^{h+m}$, we need to set other variables in the row/column containing $x$ as in the proof of Lemma 5.2)

**Fanin reduction step** - Without loss of generality, let $x_1 \in T_1 \setminus T_2$. Pick a row $R$ of $N$ barring the first $r$ rows, and let $y_1,\cdots,y_r$ be the grey variables in $R$ (where $y_1$ is in the same column as $x_1$). By a similar process as in the proof of Lemma 5.2 we can assume that at least one $f_i$ is nonzero when $y_2,\cdots,y_r$ are set to zero.

If one of the $f_i$'s become zero when $y_2,\cdots,y_r = 0$, then pick any white variable $y$ in row $R$ and set any variable in row $R$ to zero besides $y$. This would ensure that the fanin of the equation reduces and each $N_i$ is now $y^m \cdot N_i'$. Each $N_i'$ may be thought of as being obtained from $N_y$, the minor of $N$ with respect to $y$. The other variables in the column of $y$ can be set to values to ensure that the $f_i$'s stay nonzero to obtain an equation of the form $\sum f_iN_i' = 0$ of reduced fanin.

If none of the $f_i$'s become zero when $y_2,\cdots,y_r = 0$, then set every variable in row $R$ other than $y_1$ to zero. This ensures an entire column of the matrix corresponding to $N_1$ becomes zero (as $x_1$ indexes one of the columns of $N_1$), and hence $N_1$ becomes zero. On the other hand, $N_2$ remains nonzero and each surviving $N_i$ can be written
as \( y_1^{(m)} \cdot N_1' \), where \( N_1' \) is the corresponding Jacobian minor of \( N_{y_1} \). Again, the other variables in the column of \( y_1 \) can be set to values to ensure that \( f_i \)'s stay nonzero and we obtain an equation \( \sum f_i' N_1' = 0 \) of reduced fanin.

As in the proof of Lemma 5.2 we eventually obtain an equation of the form \( f_1 N_1 = 0 \) where \( f_1 \neq 0 \) thus implying that \( N_1 = 0 \). The number of variables that have been set to constants is bounded by \( t + \log S \) where \( S \) is the initial total sparsity of the \( f_i \)'s, and \( N_1 \) is a Jacobian minor of a symbolic matrix of dimension \( n - (\log S + t - 1) \).

Conjecture 6.1 asserts that \( N_1 \) would be nonzero unless \( \log S + t = \Omega(n - (\log S + t - 1)) \), or \( S = 2^{2\Omega(t)} \).

That concludes the proof of Theorem 6.3 as well.

7. Conclusion. Spurred by the success of Jacobian in solving the hitting-set problem for constant-trdeg depth-3 circuits and constant-occur constant-depth formulas, one is naturally inspired to investigate the strength of this approach against other ‘constant parameter’ models - the foremost of which is constant top fanin depth-4 circuits (PIT even for fanin 2?).

Another problem, which is closely related to hitting-sets and lower bounds, is reconstruction of arithmetic circuits [SY10, Chapter 5]. There is a quasi-polynomial time reconstruction algorithm [KS09a], for a polynomial computed by a depth-3 constant top fanin circuit, that outputs a depth-3 circuit with quasi-polynomial top fanin. Could Jacobian be used as an effective tool to solve reconstruction problems? If yes, then it would further reinforce the versatility of this tool.

Acknowledgments. We would like to thank the anonymous reviewers for several useful suggestions that have helped us improve the presentation of the paper.

This work was done when MA, RS and NS visited Max-Planck Institute for Informatics (Saarbrücken) and would like to thank the institute for its generous hospitality. MA was supported by Humboldt Forschungspreis and J C Bose Fellowship. RS was supported by IMPECS and MSR (India) Ph.D. Fellowship. CS was supported by the IMPECS Fellowship.

REFERENCES


[LFKN90] Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic Methods


