From Sylvester-Gallai Configurations to Rank Bounds: Improved Black-box Identity Test for Depth-3 Circuits

Nitin Saxena
Hausdorff Center for Mathematics
Bonn - 53115, Germany
ns@hcm.uni-bonn.de

C. Seshadhri
IBM Almaden
San Jose - 95126, USA
csesha@us.ibm.com

Abstract—We study the problem of identity testing for depth-3 circuits of top fanin $k$ and degree $d$. We give a new structure theorem for such identities. A direct application of our theorem improves the known deterministic $d^{\Omega(k^2)}$-time black-box identity test over rationals (Kayal & Saraf, FOCS 2009) to one that takes $d^{O(k^2)}$-time. Our structure theorem essentially says that the number of independent variables in a real depth-3 identity is very small. This theorem affirmatively settles the strong rank conjecture posed by Dvir & Shpilka (STOC 2005).

We devise a powerful algebraic framework and develop tools to study depth-3 identities. We use these tools to show that any depth-3 identity contains a much smaller nucleus identity that contains most of the “complexity” of the main identity. The special properties of this nucleus allow us to get almost optimal rank bounds for depth-3 identities.

Keywords—depth-3 circuit; identities; Sylvester-Gallai; incidence configuration; Chinese remaindering; ideal theory

I. INTRODUCTION

Polynomial identity testing (PIT) ranks as one of the most important open problems in the intersection of algebra and computer science. We are provided an arithmetic circuit that computes a polynomial $p(x_1, x_2, \ldots, x_n)$ over a field $F$, and we wish to test if $p$ is identically zero (in other words, if $p$ is the zero polynomial). In the black-box setting, we do not have access to the circuit. We are only allowed to evaluate the polynomial $p$ at various domain points. The main goal is to devise a deterministic (preferably black-box) polynomial time algorithm for PIT. Heintz & Schnorr [HS80], Kabanets & Impagliazzo [KI04] and Agrawal [Agr05], [Agr06] have shown connections between deterministic algorithms for identity testing and circuit lower bounds, emphasizing the importance of this problem. For a detailed exposition, see surveys [Sax09], [AS09].

Even for the special case of depth-3 circuits, this question is still open. This may seem quite depressing. It is. Nonetheless, there exist concrete results that justify both our ignorance and the acceptance of results on depth-3 PIT in major publishing venues. Agrawal and Vinay [AV08] showed that an efficient black-box identity test for depth-4 essentially leads to subexponential lower bounds.

A depth-3 circuit $C$ over a field $F$ is of the form $C(x_1, \ldots, x_n) = \sum_{i=1}^{k} T_i$, where $T_i$ (a multiplication term) is a product of at most $d$ linear polynomials with coefficients in $F$. We are especially interested in the case $F = \mathbb{Q}$. In this section, we will just assume this unless explicitly mentioned otherwise. The size of the circuit $C$ can be expressed in three parameters: the number of variables $n$, the degree $d$, and the top fanin (or the number of terms) $k$. Such a circuit is referred to as a $\Sigma \Pi \Sigma(k, d, n)$ circuit. PIT algorithms for depth-3 circuits were first studied by Dvir & Shpilka [DS06]. There have been many recent results in this area by Kayal & Saxena [KS07] (in the non-black-box setting), Karnin & Shpilka [KS08], Saxena & Seshadhri [SS09], and Kayal & Saraf [KS09b]. Our main result is a better black-box tester for $\Sigma \Pi \Sigma$ circuits over $\mathbb{Q}$. We get a running time of $nd^{k^2}$, an exponential improvement (in $k$) over the previous best of $ndk^8$ [KS09b]. Table I details the time complexities of previous algorithms. These time complexities are actually bounds on the total number of bit operations. Also, the running times are technically polynomial in the stated times.

**Theorem 1:** Consider circuits over $\mathbb{Q}$. There exists a deterministic black-box algorithm for PIT on $\Sigma \Pi \Sigma(k, d, n)$ circuits, whose time complexity is $\text{poly}(nd^{k^2})$.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Time complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[KS08]</td>
<td>$nd^{(2k^2 + \log k + 2d)}$</td>
</tr>
<tr>
<td>[SS09]</td>
<td>$ndk^4 \log d$</td>
</tr>
<tr>
<td>[KS09b]</td>
<td>$ndk^8$</td>
</tr>
<tr>
<td>This paper</td>
<td>$nd^{k^2}$</td>
</tr>
</tbody>
</table>

This is the first result that gives a time complexity both polynomial in $d$ and singly-exponential in $k$ for $\mathbb{Q}$. This is not too far from the best non-black-box algorithm for $\Sigma \Pi \Sigma$ circuits, which runs in $\text{poly}(nd^k)$ time [KS07]. This result closes the gap (almost) between black-box and non-black-box algorithms.

All these results go via rank bounds for depth-3 identities,
introduced by Dvir & Shpilka [DS06]. This is a very interesting quantity associated with these circuits, and roughly speaking, bounds the maximum number of “free variables” that can be present in a depth-3 identity. If a $\Sigma\Pi\Sigma(k, d, n)$ circuit has rank $r$, then there exists a linear transformation that converts this to an equivalent $\Sigma\Pi\Sigma(k, d, r)$ circuit. (This linear transformation is very easy to determine.) The remarkable insight of [DS06] was that the rank of every $\Sigma\Pi\Sigma(k, d, n)$ identity is very low. Any $\Sigma\Pi\Sigma(k, d, r)$-circuit can be completely expanded out in poly$(kd^r)$ time. Hence, low rank bounds for identities imply efficient non-black-box PIT algorithms.

Karnin & Shpilka [KS08] showed how small rank bounds for identities imply efficient black-box PIT algorithms. This opened the door for black-box algorithms for depth-3 PIT. Indeed, all known algorithms for this problem come as a consequence of their result. Rank bounds have also found applications in learning $\Sigma\Pi\Sigma$ circuits [Shp09], [KS09a]. Hence, the rank and file of researchers studying this problem are interested in proving small rank bounds. As mentioned earlier, we focus on the field $\mathbb{Q}$. Dvir & Shpilka [DS06] initiated this line of work by showing that the rank of a simple, minimal $\Sigma\Pi\Sigma(k, d)$ identity is $2^{O(k^2)}(\log d)^{k-2}$. There are basic constructions of rank $\Omega(k)$ identities over $\mathbb{Q}$ [DS06]. Dvir & Shpilka [DS06] conjectured that the rank should be bounded by poly$(k)$. This rank bound was improved to $O(k^3 \log d)$ by Saxena & Seshadhri [SS09]. Kayal & Saraf [KS09b] achieved a breakthrough by proving a rank bound independent of $d$. Their bound was $k^{O(k)}$. We finally settle the Dvir-Shpilka conjecture and show a rank bound of $O(k^2)$.

The advances of Kayal & Saraf were obtained through the use of incidence geometry theorems, like the famous Sylvester-Gallai theorem. This theorem states that for any set $S$ of points in the Euclidean plane, not all collinear, there exists a line passing through exactly two points in $S$. Generalizations of this to higher dimensions are called Sylvester-Gallai theorems (see survey [BM90]). This theorem and its generalizations have connections to rank bounds for depth-3 circuits. The result of [KS09b] gave an intricate combinatorial construction that converts depth-3 identities to sets of colored points in Euclidean space. This allowed the use of Sylvester-Gallai theorems to bound the rank.

Our contribution comes through a new algebraic framework for studying depth-3 identities. This has many benefits. Firstly, it allows for a much more “efficient” use of Sylvester-Gallai theorems to bound the rank. This leads to nearly optimal rank bounds. Secondly, the connection between Sylvester-Gallai theorems and rank bounds is far more transparent, at the loss of some color from the theorems. Theorem 4 gives a simple formula that relates the depth-3 rank to Sylvester-Gallai bounds. A nice byproduct of this connection is the improvement of rank bounds over arbitrary fields. Thirdly, we get a deep structural theorem about depth-3 identities over any field. Every such identity contains a nucleus identity expressed on few variables. This nucleus, in some sense, captures all the complexity of the original identity, and has some very special properties. A better understanding of this nucleus may lead us to the goal of a truly polynomial time algorithm.

A. Definitions and results

We recall that a depth-3 circuit $C$ over a field $\mathcal{F}$ is: $C(x_1, \ldots, x_n) = \sum_{i=1}^{k} T_i$, where $T_i$ is a product of $d_i$ linear polynomials $\ell_{i,j}$ over $\mathcal{F}$. For the purposes of studying identities we can assume, by homogenization, that $\ell_{i,j}$‘s are linear forms (i.e. linear polynomials with a zero constant coefficient) and $\forall i, d_i = d$. It will be convenient to state our results in terms of arbitrary fields.

Definition 2: [DS06]

- **Simple Circuit:** $C$ is a simple circuit if there is no nonzero linear form dividing all the $T_i$‘s.
- **Minimal Circuit:** $C$ is a minimal circuit if for every proper subset $S \subset [k]$, $\sum_{i \in S} T_i$ is nonzero.
- **Rank of a circuit:** The coefficients of $\ell_{i,j}$ form an $n$-dimensional vector over $\mathbb{F}$. The rank of the circuit, rk$(C)$, is defined as the rank of the set of all linear forms $\ell_{i,j}$ viewed as vectors.

The rank of a circuit can be interpreted as the minimum number of independent variables required to express $C$. The definition of simple and minimal circuits are used to remove certain pathological cases. The rank question is: for a simple and minimal $\Sigma\Pi\Sigma(k, d, n)$ identity over field $\mathcal{F}$, what is the maximal possible rank? A trivial upper bound on the rank (for any $\Sigma\Pi\Sigma$-circuit) is $kd$, since that is the total number of linear forms involved in $C$. A substantially smaller rank bound than $kd$ shows that identities do not have as many “degrees of freedom” as general circuits.

Before we state our results, it will be helpful to explain Sylvester-Gallai configurations. A set of points $S$ with the property that every line through two points of $S$ passes through a third point in $S$ is called a Sylvester-Gallai configuration. The famous Sylvester-Gallai theorem states that the only Sylvester-Gallai configuration in $\mathbb{R}^2$ is a set of collinear points. This basic theorem about point-line incidences was extended to higher dimensions [Han65], [BE67]. We define the notion of Sylvester-Gallai rank bounds. This is a clean and convenient way of expressing these theorems.

**Definition 3:** Let $S$ be a finite subset of the projective space $\mathbb{P}^n$. Alternately, $S$ is a set of non-zero vectors in $\mathbb{P}^{n+1}$ without multiples: no two vectors in $S$ are scalar multiples of each other. Suppose, for every set $V \subset S$ of $k$ linearly independent vectors, the linear span of $V$ contains at least $k + 1$ vectors of $S$. Then, the set $S$ is said to be $SG_k$-closed.
The largest possible rank of an \(SG_k\)-closed set of at most \(m\) vectors in \(\mathbb{F}^n\) (for any \(n\)) is denoted by \(SG_k(\mathbb{F}, m)\).

The Sylvester-Gallai theorem states higher dimensional analogues [Han65], [BE67] can be interpreted to say \(SG_k(\mathbb{R}, m) \leq 2(k - 1)\). Our main theorem is a simple, clean expression of how Sylvester-Gallai influences identities. We state this for general fields.

**Theorem 4 (From \(SG_k\) to Rank):** Let \(|\mathbb{F}| > d\). The rank of a simple and minimal \(\Sigma\Pi\Sigma(k, d)\) identity over \(\mathbb{F}\) is at most \(2k^2 + k \cdot SG_k(\mathbb{F}, d)\).

A direct application of the \(SG_k(\mathbb{R}, m)\) bound yields an almost optimal rank bound for real depth-3 identities. For completeness, we state the exact rank bound obtained. We have a slightly stronger version (Theorem 18) of the above theorem that gives better constants.

**Theorem 5 (Depth-3 Rank Bounds):** Let \(C\) be a \(\Sigma\Pi\Sigma(k, d)\) circuit, over field \(\mathbb{R}\), that is simple, minimal and zero. Then, \(rk(C) < 3k^2\).

As discussed before, an application of this result to Lemma 4.10 of [KS08] gives a deterministic black-box identity test for \(\Sigma\Pi\Sigma(k, d, n)\) circuits over \(\mathbb{Q}\). Formally, we get the following **hitting set generator** for \(\Sigma\Pi\Sigma\) circuits with real coefficients.

**Corollary 6 (Black-box PIT over \(\mathbb{Q}\)):** There is a deterministic algorithm that takes as input a triple \((k, d, n)\) of natural numbers and in time \(\text{poly}(ndk^2)\), outputs a hitting set \(H \subset \mathbb{Z}^n\) with the following properties:

1. Any \(\Sigma\Pi\Sigma(k, d, n)\) circuit \(C\) over \(\mathbb{R}\) computes the zero polynomial iff \(\forall a \in H, C(a) = 0\).
2. \(H\) has at most \(\text{poly}(ndk^2)\) points.
3. The total bit-length of each point in \(H\) is \(\text{poly}(kn \log d)\).

**1) Other fields:** What about other fields? The rank bounds of [DS06] and [SS09] hold for arbitrary fields, whereas the rank bound of [KS09b] holds only for \(\mathbb{R}\). It has been observed that for finite fields, the rank of an \(\Sigma\Pi\Sigma\) identity can be as large as \(\Omega(k \log d)\) [KS07], [SS09]. Hence, the \(O(k^3 \log d)\) bound proved by [SS09] is almost optimal. As a small bonus, we give a slight improvement upon this bound using our approach. This requires Sylvester-Gallai theorems over arbitrary fields, an interesting question in itself. It was shown that \(SG_2(\mathbb{C}, m) \leq 3\) [EPS06], and certain lower bounds for locally decodable codes implied \(SG_2(\mathbb{F}, m) = O(\log m)\). (Concretely, Corollary 2.9 of [DS06] can be used to prove that \(SG_2(\mathbb{F}, m) = O(\log m)\). This is an extension of theorems in [GKST02] that prove this for \(\mathbb{F}_2\).) Otherwise than this, nothing was previously known. One of our auxiliary theorems, of independent interest, gives a high-dimensional Sylvester-Gallai bound for all fields. Applying the stronger version of Theorem 4, we get our rank bound.

**Theorem 7 (\(SG_k\) for all fields):** For any field \(\mathbb{F}\) and \(k, m \in \mathbb{N}^{>1}, \ SG_k(\mathbb{F}, m) \leq 9k \lg m\). (There is a construction that shows that \(SG_k(\mathbb{F}_p, m) = \Omega(k \cdot \log_p m)\).)

Let \(C\) be a \(\Sigma\Pi\Sigma(k, d)\) circuit, over an arbitrary field \(\mathbb{F}\), that is simple, minimal and zero. Then, \(rk(C) < 3k^2(\lg 2d)\).

**B. History**

And now, for a brief history of PIT algorithms. The first randomized polynomial time PIT algorithm, which was a black-box algorithm, was given (independently) by Schwartz [Sch80] and Zippel [Zip79]. Randomized algorithms that use less randomness were given by Chen & Kao [CK00], Lewin & Vadhan [LV98], and Agrawal & Biswas [AB03]. Klivans & Spielman [KS01] observed that even for depth-3 circuits for bounded top fanin, deterministic identity testing was open. Progress towards this was first made by the quasi-polynomial time algorithm of Dvir & Shpilka [DS06]. The problem was resolved by a polynomial time algorithm given by Kayal and Saxena [KS07], with a running time exponential in the top fanin. Both these algorithms were non-black-box. As for black-box algorithms, the authors are quite sure that the reader has heard enough history. Identity tests are known only for very special depth-4 circuits [AM07], [Sax08], [SV09], [KMSV09]. Agrawal and Vinay [AV08] showed that an efficient black-box identity test for depth-4 circuits will actually give a quasi-polynomial black-box test, and subexponential lower bounds, for circuits of all depths that compute low degree polynomials. Thus, understanding depth-3 identities seems to be a natural first step towards the goal of PIT.

**II. Proof Outline, Ideas, and Organization**

Our proof of the rank bound comprises of several new ideas, both at the conceptual and the technical levels. Instead of giving proofs in this extended abstract, we will only provide the intuition and the overall argument. We recommend the interested reader to see the full version of this paper [SS10]. The full proof of Theorem 4 is extremely technical, requires many definitions, and involve many algebraic arguments. Our attempt is to convey with main ideas without getting into too much formalism or mathematical details. We describe all the major milestones, many of which are interesting in their own right. Indeed, it is the authors’ opinion that the reader has little to gain from simply reading the detailed proofs without getting the essence of the ideas.

The intuition portion is divided into three subsections, each dealing with a separate component of the final proof.
Each portion proves an interesting structural theorem. The three notions that are crucially used and developed are: ideal Chinese remaindering, matchings and Sylvester-Gallai rank bounds. Related notions have appeared (in some form) in the works of Kayal & Saxena [KS07], Saxena & Seshadhri [SS09] and Kayal & Saraf [KS09b] respectively, to prove different kinds of results. The first two steps set up the algebraic framework and prove theorems that hold for all fields. The third step is where the Sylvester-Gallai theorems are brought in.

A. Notation and definitions

We will denote the set \{1, \ldots, n\} by \[n\]. We fix the base field to be \( \mathbb{F} \), so the circuits compute multivariate polynomials in the polynomial ring \( R := \mathbb{F}[x_1, \ldots, x_n] \). We use \( \mathbb{F}^* \) to denote \( \mathbb{F} \setminus \{0\} \).

A linear form is a linear polynomial in \( R \) with zero constant term. We will denote the set of all linear forms by \( L(R) := \{ \sum_{i=1}^n a_ix_i \mid a_1, \ldots, a_n \in \mathbb{F} \} \). Clearly, \( L(R) \) is a vector (or linear) space over \( \mathbb{F} \) and will be quite useful. Much of what we do shall deal with multi-sets of linear forms (sometimes polynomials in \( R \)), equivalence classes inside them, and various maps across them. A list of linear forms is a multi-set of forms with an arbitrary order associated with them. The actual ordering is unimportant: we will heavily use maps between lists, and the ordering allows us to define these maps unambiguously. The usual set operations between lists can be naturally defined.

**Definition 8:** We collect some important definitions from [SS09]:

### [Multiplication term, \( L(\cdot) \) & \( M(\cdot) \)]

A multiplication term \( f \) is an expression in \( R \) given as (the product may have repeated \( \ell^\prime \)'s), \( f := c \cdot \prod_{\ell \in \mathbb{F}} L(\ell) \), where \( c \in \mathbb{F}^* \) and \( S \) is a list of nonzero linear forms. The list of linear forms in \( f, L(f) \), is just the list \( S \) of forms occurring in the product above. For a list \( S \) of linear forms we define the multiplication term of \( S, M(S) \), as \( \prod_{\ell \in \mathbb{F}} L(\ell) \) or 1 if \( S = \emptyset \).

### [Forms in a Circuit]

We will represent a \( \Sigma\Pi\Sigma(k,d) \) circuit \( C \) as a sum of \( k \) multiplication terms of degree \( d \), \( C = \sum_{i=1}^k T_i \). The list of linear forms occurring in \( C \) is \( L(C) := \bigcup_{\ell \in \mathbb{F}} L(T_i) \). Note that \( L(C) \) is a list of size exactly \( kd \). The rank of \( C, \text{rk}(C) \), is just the number of linearly independent linear forms in \( L(C) \). (Remark: for the purposes of this paper \( T_i \)'s are given in circuit representation and thus the list \( L(T_i) \) is unambiguously defined from \( C \)).

### [Similar forms]

For any two polynomials \( f,g \in R \) we call \( f \) similar to \( g \) if there exists \( c \in \mathbb{F}^* \) such that \( f = cg \). We say \( f \) is similar to \( g \) mod \( I \), for some ideal \( I \) of \( R \), if there is some \( c \in \mathbb{F}^* \) such that \( f \equiv cg \) (mod \( I \)). Note that “similarity mod \( I \)” is an equivalence relation (reflexive, symmetric and transitive) and partitions any list of polynomials into equivalence classes.

### [Span \( \text{sp}(\cdot) \)]

For any \( S \subseteq L(R) \) we let \( \text{sp}(S) \subseteq L(R) \) be the linear span of the linear forms in \( S \) over the field \( \mathbb{F} \). (Conventionally, \( \text{sp}(\emptyset) = \{0\} \)).

### [Matchings]

Let \( U,V \) be lists of linear forms and \( I \) be a subspace of \( L(R) \). An \( I \)-matching \( \pi \) between \( U,V \) is a bijection \( \pi \) between lists \( U,V \) such that: for all \( \ell \in U \), \( \pi(\ell) \in F^* \ell + I \).

When \( f,g \) are multiplication terms, an \( I \)-matching between \( f,g \) would mean an \( I \)-matching between \( L(f),L(g) \).

B. Step 1: Matching the Gates in an Identity

We will show that all the multiplication terms of a minimal \( \Sigma\Pi\Sigma \) identity can be matched by a low rank space \( K \), spanned by “few” linear forms in \( L(R) \).

**Theorem 9 (Matching-Nucleus):** Let \( C = T_1 + \cdots + T_k \) be a \( \Sigma\Pi\Sigma(k,d) \) circuit that is minimal and zero. Then there exists a linear subspace \( K \) of \( L(R) \) such that:

1. \( \text{rk}(K) < k^2 \).
2. \( \forall i \in [k] \), there is a \( K \)-matching \( \pi_i \) between \( T_1,T_i \).

The idea of matchings within identities was first introduced in [SS09], but nothing as powerful as this theorem has been proven. This theorem gives us a space of small rank, independent of \( d \), that contains most of the “complexity” of \( C \). All forms in \( C \) outside \( K \) are just mirrored in the various terms. This starts connecting the algebra of depth-3 identities to a combinatorial structure. Indeed, the graphical picture (explained in detail below) that this theorem provides, really gives an intuitive grasp on these identities. The proof of this involves some interesting generalizations of the Chinese Remainder Theorem to some special ideals.

**Definition 10 (mat-nucleus):** Let \( C \) be a minimal \( \Sigma\Pi\Sigma(k,d) \) identity. The linear subspace \( K \) given by Theorem 9 is called mat-nucleus of \( C \).

The notion of mat-nucleus is easier to see in the representation of the \( \Sigma\Pi\Sigma(4,d) \) circuit \( C = \sum_{i \in [4]} T_i \) given in Figure 1a. The four bubbles refer to the four multiplication terms of \( C \) and the points inside the bubbles refer to the linear forms in the terms. The proof of Theorem 9 gives mat-nucleus as the space generated by the linear forms in the dotted box. The linear forms that are not in mat-nucleus lie “above” the mat-nucleus and are all (mat-nucleus)-matched, i.e. \( \forall \ell \in (L(T_i) \setminus \text{mat-nucleus}) \), there is a form similar to \( \ell \) modulo mat-nucleus in each \( (L(T_i) \setminus \text{mat-nucleus}) \). Thus the essence of Theorem 9 is: the mat-nucleus part of the terms of \( C \) has low rank \( k^2 \), while the part of the terms above mat-nucleus all look “similar”.

**Proof Idea for Theorem 9:** The key insight in the construction of mat-nucleus is a reinterpretation of the non-black-box identity test of Kayal & Saxena [KS07] as a structural result for \( \Sigma\Pi\Sigma \) identities. Roughly speaking,
KS07] showed that $C = 0$ iff for every path $(v_1, v_2, v_3)$ (where $v_i \in L(T_i)$): $T_4 \equiv 0 \pmod{v_1, v_2, v_3}$ or in ideal terms, $T_4 \in \langle v_1, v_2, v_3 \rangle$. (This is technically false, but it portrays the right idea.) Paths are depicted in Figure 1b. Thus, it is enough to go through all the $d^k$ paths to certify the zeroness of $C$. This is why the time complexity of the identity test of [KS07] is dominated by $d^k$. Now if we are given a $\Sigma\Pi\Sigma(4, d)$ identity $C$ which is minimal, then we know that $T_1 + T_2 + T_3 \neq 0$. Thus, by applying the above interpretation of [KS07] to $T_1 + T_2 + T_3$ we will get a path $(v_1, v_2)$ such that $T_3 \not\equiv \langle v_1, v_2 \rangle$. Since $C = 0$ this means that $T_3 + T_4 \equiv 0 \pmod{v_1, v_2}$ but $T_3, T_4 \neq 0 \pmod{v_1, v_2}$ (if $T_4$ is in $\langle v_1, v_2 \rangle$ then so will be $T_3$). Thus, $T_3 \equiv -T_4 \pmod{v_1, v_2}$ is a nontrivial congruence and it immediately gives us a $\langle v_1, v_2 \rangle$-matching between $T_3, T_4$. By repeating this argument with a different permutation of the terms we could match different terms (by a different ideal), and finally we expect to match all the terms (by the union of the various ideals).

This argument has numerous technical problems, the most important one being that it does not really work. But all issues can be taken care of by suitable algebraic generalizations. A major stumbling block is the presence of repeating forms. It could happen that $(\mod v_1), v_2$ occurs in many terms, or in the same term with a higher power. The most important tool developed is an ideal version of Chinese remaindering that forces us to consider not just linear forms $v_1, v_2$, but multiplication terms $v_1, v_2$ dividing $T_1, T_2$ respectively.

C. Step 2: Certificate for Linear Independence of Gates

Theorem 9 gives us a space $K$, of rank $< k^2$, that matches $T_1$ to each term $T_i$. In particular, this means that the list $L_K(T_i) := L(T_i) \cap K$ has the same cardinality $d'$ for each $i \in [k]$. In fact, if we look at the corresponding multiplication terms $K_i := M(L_K(T_i)), i \in [k]$, then they again form a $\Sigma\Pi\Sigma(k, d')$ identity!

Precisely, $C' = \sum_{i \in [k]} \alpha_i K_i$ for some $\alpha_i$’s in $\mathbb{F}^*$ is an identity. We would like $C'$ to somehow mimic the structure of $C$. Of course $C'$ is simple but is it again minimal? Unfortunately, it may not be. As we will see in Step 3, when $C'$ somewhat “mirrors” the structure of $C$, then bounding the rank of the forms “outside” $K$ becomes possible. Step 2 involves increasing the space $K$ (but not by too much) that gives us a $C'$ with the right behavior. Specifically, if $T_1, \ldots, T_k'$ are linearly independent (i.e. $\beta_i \not\equiv 0$) s.t. $\sum_{i \in [k']} \beta_i T_i = 0$, then so are $K_1, \ldots, K_k'$. The following can be seen as an important structural theorem of depth-3 identities.

**Theorem 11 (Nucleus):** Let $C = \sum_{i \in [k]} T_i$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity and let $\{T_i | i \in I\}$ be a maximal set of linearly independent terms ($1 \leq k' := |I| < k$). Then there exists a linear subspace $K$ of $\mathbb{F}(R)$ such that:

1. $\text{rk}(K) < 2k^2$.
2. $\forall i \in [k], \, \exists \pi_i$ between $T_1, T_i$.
3. (Define $\forall i \in I, \, K_i := M(L_K(T_i))$. The terms $\{K_i | i \in I\}$ are linearly independent.

**Definition 12 (nucleus):** Let $C$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity. The linear subspace $K$ given by Theorem 11 is called the nucleus of $C$. The subspace $K$ induces an identity $C' = \sum_{i \in [k]} \alpha_i K_i$ which we call the nucleus identity.

The notion of the nucleus is easier to grasp when $C$ is a $\Sigma\Pi\Sigma(k, d)$ identity that is strongly minimal, i.e. $T_1, \ldots, T_{k-1}$ are linearly independent. Clearly, such a $C$ is also minimal. For such a $C$, Theorem 11 gives a nucleus $K$ such that the corresponding nucleus identity is strongly minimal. The structure of $C$ is very strongly represented by $C'$. As a bonus, we actually end up greatly simplifying the polynomial-time PIT algorithm of Kayal & Saxena [KS07] (although we will not discuss this point in detail in this paper).

**Proof Idea for Theorem 11:** The first two properties in the theorem statement are already satisfied by mat-nucleus of $C$. So we incrementally add linear forms to the space mat-nucleus till it satisfies property (3) and becomes the nucleus. The addition of linear forms is guided by the ideal version of Chinese remaindering. For convenience assume $T_1, T_2, T_3$ to be linearly independent. Then, by homogeneity and equal degree, we have an equivalent ideal statement:
Let $T_2 \notin (T_1)$ and $T_3 \notin (T_1, T_2)$. Even in this general setting the path analogy (used in the last subsection) works and we essentially get linear forms $v_1 \in L(T_1)$ and $v_2 \in L(T_2)$ such that: $T_2 \notin (v_1)$ and $T_3 \notin (v_1, v_2)$. We now add these forms $v_1, v_2$ to the space mat-nucleus, and call the new space $K$. It is expected that the new $K_1, K_2, K_3$ are now linearly independent.

Not surprisingly, the above argument has numerous technical problems. But it can be made to work by careful applications of the ideal version of Chinese remaining.

**D. Step 3: Invoking Sylvester-Gallai Theorems**

As explained in Section I-A, we rephrase the standard Sylvester-Gallai theorems in terms of Sylvester-Gallai closure and rank bounds (Definition 3). Using some linear algebra and combinatorial tricks, we prove the first ever general Sylvester-Gallai bound for all fields.

**Theorem 13 (General Sylvester-Gallai):** For any field $\mathbb{F}$ and $k, m \in \mathbb{N}^{>1}$, $\mathrm{SG}_k(\mathbb{F}, m) \leq 9k \lg m$.

The following definition is very helpful in applying Sylvester-Gallai rank bounds to our scenario.

**Definition 14 (SG operator):** $[\mathrm{SG}_k(\cdot)]$ Let $k, m \in \mathbb{N}^{>1}$. Suppose a set $S \subseteq \mathbb{F}^n$ has rank greater than $\mathrm{SG}_k(\mathbb{F}, m)$ (were $\# S \leq m$). Then, by definition, $S$ is not $\mathrm{SG}_k$-closed. In this situation we say the $k$-dimensional Sylvester-Gallai operator $\mathrm{SG}_k(S)$ (applied on $S$) returns a set of $k$ linearly independent vectors $V$ in $S$ whose span has no point in $S \setminus V$.

Let $C$ be a simple and strongly minimal $\Sigma \Pi \Sigma(k, d)$ identity. Theorem 11 gives us a nucleus $K$, of rank $< 2k^2$, that matches $T_i$ to each term $T_i$. As seen in Step 2, if we look at the corresponding multiplication terms $K_i := M(L_K(T_i))$, $i \in [k]$, then they again form a $\Sigma \Pi \Sigma(k, d')$ “nucleus identity” $C' = \sum_{i \in [k]} \alpha_i K_i$, for some $\alpha_i$'s in $\mathbb{F}^*$, which is simple and strongly minimal. Define the non-nucleus part of $T_i$ as $L^c_K(T_i) := L(T_i) \setminus K$, for all $i \in [k]$ (c in the exponent annotates “complement”, since $L(T_i) = L_K(T_i) \cup L^c_K(T_i)$).

What can we say about the rank of $L^c_K(T_i)$? Define the non-nucleus part of $C$ as $L^c_K(C) := \cup_{i \in [k]} L^c_K(T_i)$. Our goal in Step 3 is to bound $\mathrm{rk}(L^c_K(C) \bmod K)$ by $2k$ when the field is $\mathbb{R}$. This will give us a rank bound of $\mathrm{rk}(K) + \mathrm{rk}(L^c_K(C) \bmod K) < (2k^2 + 2k)$ for simple and strongly minimal $\Sigma \Pi \Sigma(k, d)$ identities over $\mathbb{R}$. The proof is mainly combinatorial, based on higher dimensional Sylvester-Gallai theorems and a property of set partitions, with a sprinkling of algebra.

We apply the $\mathrm{SG}_k$ operator not directly on the forms in $L(C)$ but on a suitable truncation of those forms. So we need another definition.

**Definition 15 (Non-$K$ rank):** Let $K$ be a linear subspace of $L(R)$. Then $L(R) / K$ is again a linear space (the quotient space). Let $S$ be a list of forms in $L(R)$. The non-$K$ rank of $S$ is defined to be $\mathrm{rk}(S \bmod K)$ (i.e. the rank of $S$ when viewed as a subset of $L(R) / K$).

Let $C$ be a $\Sigma \Pi \Sigma(k, d)$ identity with nucleus $K$. The non-$K$ rank of the non-nucleus part $L^c_K(T_i)$ is called the non-nucleus rank of $T_i$. Similarly, the non-$K$ rank of the non-nucleus part $L^c_K(C) := \cup_{i \in [k]} L^c_K(T_i)$ is called the non-nucleus rank of $C$.

We give an example to explain the non-$K$ rank. Let $R = \mathbb{F}[z_1, \ldots, z_n, y_1, \ldots, y_m]$. Suppose $K = \langle z_1, \ldots, z_n \rangle$ and $S \subseteq L(R)$. We can take any element $\ell$ in $S$ and simply drop all the $z_i$ terms, i.e. ‘truncate’ the $z$-part of $\ell$. This gives a set of linear forms over the $y$ variables. The rank of these is the non-$K$ rank of $S$.

We are now ready to state the theorem that is proved in Step 3. It basically shows a neat relationship between the non-nucleus part and Sylvester-Gallai.

**Theorem 16 (Bound for simple, strongly minimal identities):** Let $|\mathbb{F}| > d$. The non-nucleus rank of a simple and strongly minimal $\Sigma \Pi \Sigma(k, d)$ identity over $\mathbb{F}$ is at most $\mathrm{SG}_k(\mathbb{F}, d)$. More specifically, (for nucleus $K$) the vectors in $L(C) \setminus K$ form an $\mathrm{SG}_k-1$-closed set.

Observe that this theorem together with Theorem 11 gives a complete structure theorem for strongly minimal depth-3 identities. One can make suitable claims for identities that are not strictly minimal. Essentially, we just take a subset of linearly independent terms, say $T_1, \ldots, T_k$, that form a basis for $\{T_i|i \in [k]\}$. We can now construct strongly minimal identities using these terms and apply the above theorem. Specifically, we get the following.

**Definition 17 (Independent-fanin):** Let $C = \sum_{i \in [k]} T_i$ be a $\Sigma \Pi \Sigma(k, d)$ circuit. The independent-fanin of $C$, $\mathrm{ind-fanin}(C)$, is defined to be the size of the maximal $I \subseteq [k]$ such that $\{T_i|i \in I\}$ are linearly independent polynomials.

We now state the following stronger version of the main theorem.

**Theorem 18 (Final bound):** Let $|\mathbb{F}| > d$. The rank of a simple, minimal $\Sigma \Pi \Sigma(k, d)$ identity over reals is at most $2k^2 + (k - k') \cdot \mathrm{SG}_k(\mathbb{F}, d)$ where $k'$ is at most $3k^2$, proving the main theorem over reals. Likewise, for any $\mathbb{F}$, we get the rank bound of $2k^2 + (k - k') \cdot \mathrm{SG}_k(\mathbb{F}, d)$.
\[ \leq 2k^2 + (k - k')^2 g' k' \leq 2k^2 + \frac{g^2}{2} \leq 3k^2 \log 2d, \]
proving the main theorem.

**Proof Idea for Theorem 16:** Basically, we apply the $SG_k(\cdot)$ operator on the non-nucleus part of the term $T_1$, i.e. we treat a linear form $\sum a_i x_i$ as the point $(1, \frac{a_2}{a_1}, \ldots, \frac{a_n}{a_1}) \in \mathbb{F}^n$ for the purposes of Sylvester-Gallai and then we consider $SG_k(L_{g_k}(T_1))$ assuming that the non-nucleus rank of $T_1$ is more than $SG_k(\mathbb{F}, d)$. This application of Sylvester-Gallai is much more direct compared to the methods used in [KS09b]. There, they effectively needed to prove versions of Sylvester-Gallai that dealt with colored points and needed a **hyperplane decomposition** property after applying a $SG_{\alpha \circ \psi}(\cdot)$ operator on $L(C)$. Since, modulo the nucleus, all multiplication terms look essentially the same, it suffices to focus attention on just one of them. Hence, we apply the $SG_k$-operator on a single multiplication term.

Assume $C$ is a simple, strongly minimal $\Sigma \Pi \Sigma(\ell, d)$ identity with terms \{ $T_i \mid i \in \{k\}$ \} and let $K$ be its nucleus given by Step 2. It will be convenient for us to fix a linear form $y_0 \in L(R)^*$ and a subspace $U$ of $L(R)$ such that we have the following *orthogonal* vector space decomposition $L(R) = \mathbb{F} y_0 \oplus U \oplus K$. This means for any form $\ell \in L(R)$,

there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}$, $u \in U$ and $v \in K$.

Furthermore, we will assume wlog that for every form $\ell \in L_{g_k}(T_1)$ the corresponding $\alpha$ is nonzero, i.e. each form in $L_{g_k}(T_1)$ is *monic* wrt $y_0$. Technically, we do not need the extra variable $y_0$ and can work in a projective space. Nonetheless, it makes the presentation easier.

**Definition 19** *(trun$(\cdot)$)*: Fix a decomposition $L(R) = \mathbb{F} y_0 \oplus U \oplus K$. For any form $\ell \in L_{g_k}(T_1)$, there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}$, $u \in U$ and $v \in K$.

The *truncated form* trun$(\ell)$ is the linear form obtained by dropping the $K$ part and normalizing, i.e. $\text{trun}(\ell) := y_0 + \alpha^{-1} u$.

Given a list of forms $S$ we define trun$(S)$ to be the corresponding set (thus no repetitions) of truncated forms.

To be precise, we fix a basis \{ $y_{1}, \ldots, y_{k}(\ell) \}$ of $U$ so that each form in $\text{trun}(L_{g_k}(T_1))$ has representation $y_0 + \sum a_i y_i$ (a’s $\in \mathbb{F}$). We view each such form as the point $(1, a_1, 1, \ldots, a_{k}(\ell))$ while applying Sylvester-Gallai on trun$(L_{g_k}(T_1))$. Assume, for the sake of contradiction, that the non-nucleus rank of $T_1$, $\text{rk}(\text{trun}(L_{g_k}(T_1))) > SG_{k-1}(\mathbb{F}, d)$. Therefore, $SG_{k-1}(\text{trun}(L_{g_k}(T_1)))$ gives $(k - 1)$ linearly independent forms $\ell_1, \ldots, \ell_{k-1} \in (y_0 + U)$ whose span contains no other linear form of trun$(L_{g_k}(T_1))$.

For simplicity of exposition, let us fix $k = 4$, $K$ spanned by $y$'s, $U$ spanned by $y$'s and $\ell_i = y_0 + y_i$ ($i \in \{3\}$). Note that (by definition) $\text{trun}(\alpha y_0 + \sum a_i z_i + \sum \beta y_i) = y_0 + \sum \frac{a_i}{\alpha} y_i$. We want to derive a contradiction using the $SG_3$-operator output $(y_0 + y_1, y_0 + y_2, y_0 + y_3)$ and the fact that $C$ is a simple, strongly minimal $\Sigma \Pi \Sigma(4, d)$ identity. Consider the setting given in Figure 2. Suppose the linear forms in $C$ that are similar to a form in \{ $y_0 + y_i + K \mid i \in [3]$ \} are exactly those depicted in the figure. All forms within a row are $K$-matched. We would like to find forms $\ell_1, \ell_2, \ell_3$ with the following properties: (1) $\ell_i \equiv c_i \ell_i (\text{mod } K)$ (for some constant $c_i$). (2) There exists some $j$ such that no $\ell_i$ divides $T_j$, but for each $T_i$ ($\ell \neq j$), some $\ell_i$ divides $T_i$. In this situation, we can choose $\ell_1 = y_0 + y_1 + z_1, \ell_2 = y_0 + y_2 + z_2$, and $\ell_3 = -y_0 - y_3 + z_2$. None of these divides $T_3$. Observe that the triple $(y_0 + y_1 + z_1, y_0 + y_2 + z_2, y_0 + y_3 + z_2)$ does not satisfy these conditions, since no appropriate $T_j$ can be found.

Take $C$ modulo the ideal $I := \langle y_0 + y_1 + z_1, y_0 + y_2 + z_2, -y_0 - y_3 + z_2 \rangle$. It is easy to see that $C \equiv T_4 (\text{mod } I)$, so $I$ “kills” the first three terms. Since $C$ is an identity, $T_4 \in I$. Thus, there is a form $\ell \in L(T_4)$ such that $\ell \in \text{sp}(\ell_1, \ell_2, \ell_3)$. Since no form from $\ell_i$ divides $T_4$, so $\ell$ must be a non-trivial combination of these forms. By the matching property, there exists some form $\ell \in L(T_1)$ such that trun$(\ell) \equiv \ell (\text{mod } I)$. In other words, trun$(\ell) \in \text{trun}(L_{g_k}(T_1))$. But that contradicts the fact that $(\ell_1, \ell_2, \ell_3)$ form an $SG_3$-tuple. This implies that the non-nucleus rank of $C$ is at most $SG_3(\mathbb{F}, d)$.

The approach above worked because we were lucky enough to find $\ell_1, \ell_2, \ell_3$ with the right properties. Can we always do this? No, because of repeating forms. Suppose, after going modulo form $\ell$, the circuit looks like $x^3 y + 2x^2 y^2 + ax^3 = 0$. This is not simple, but it does not have to be. We are only guaranteed that the original circuit is simple. Once we go modulo $\ell$, that property is lost. Now, the choice of any form kills all terms. We will use our more powerful Chinese remaining tools and the nucleus properties to deal with this. The minimality of the nucleus identity plays a crucial role here and helps us deal with such situations. We have to prove a special theorem about partitions of $[k]$ and use strong minimality (which we did not use in the above sketch).

### III. Conclusion

In this work we developed the strongest methods, to date, to study depth-3 identities. The ideal methods hinge on a classification of zerodivisors of the ideals generated by gates of a $\Sigma \Pi \Sigma$ circuit. That is useful in proving an ideal version of Chinese remaining tailor-made for $\Sigma \Pi \Sigma$ circuits, which is in turn useful to show a connection between all the gates involved in an identity. As a byproduct, it shows the existence of a low rank *nucleus identity* $C'$ inside any given $\Sigma \Pi \Sigma(k, d)$ identity $C$ (when $C$ is not minimal, $C'$ can still be defined but it might not be homogeneous). The properties of the nucleus identity are an important part of an identity and it might be useful for PIT to understand (or classify) it further. Can the rank bound for the nucleus...
identity be improved to $O(k)$? More importantly, can the rank bound for simple minimal real $\Sigma \Pi \Sigma(k, d)$ identities be improved to $O(k)$? The best constructions known, since [DS06], have rank $4(k-2)$. Over other fields, our upper bound of $O(k^2 \log d)$ still leaves some gap in understanding the exact dependence on $k$. Of course, the most important question is whether our techniques can help construct a truly polynomial time deterministic (even non-black-box) algorithm for PIT.

We generalize the notion of Sylvester-Gallai configurations to any field and define a parameter $SG_k(\mathbb{F}, m)$ associated with field $\mathbb{F}$. This number seems to be a fundamental property of a field, and as we show, is very closely related to $\Sigma \Pi \Sigma$ identities. It would be interesting to obtain bounds for $SG_k(\mathbb{F}, m)$ for different $\mathbb{F}$. For example, as also asked by [KS09b], can we nontrivially bound the number $SG_k(\mathbb{F}, m)$ for interesting fields: $C$, finite fields with large characteristic, or even $p$-adic fields? The only known $SG_k$ rank bounds are those for $\mathbb{R}$, $SG_2(\mathbb{C}, m) \leq 3$, and $SG_2(\mathbb{F}, m) \leq O(\log m)$. We shed (a little) light on $\mathbb{R}$ rank bounds by showing $SG_k(\mathbb{R}, m) = O(k \log m)$. We conjecture: $SG_k(\mathbb{F}, m)$ is $O(k)$ for zero characteristic fields, while $O(k + k \cdot \log_p m)$ for fields of characteristic $p > 1$. The latter would mean that when the characteristic is large ($p \geq m$), $SG_k(\mathbb{F}, m) = O(k)$, matching the bounds for zero characteristic fields.

**ACKNOWLEDGEMENTS**

We are grateful to Hausdorff Center for Mathematics, Bonn for the kind support, especially, hosting the second author when part of the work was done. The first author thanks Nils Frohberg for several detailed discussions that clarified the topic of incidence geometry and Sylvester-Gallai theorems. The second author is extremely grateful to Ken Clarkson and especially to T. S. Jayram for their suggestions in improving the presentation. We also thank Malte Beecken, Johannes Mittmann (for the high rank $SG_k$ construction over $\mathbb{F}_p$) and Thomas Thierauf for several interesting discussions.

**REFERENCES**


