Witnessing Matrix Identities and Proof Complexity

Fu Li[*] Iddo Tzameret[†]

January 18, 2018

Abstract

We use results from the theory of algebras with polynomial identities (PI-algebras) to study the witness complexity of matrix identities. A *matrix identity* of $d \times d$ matrices over a field $\mathbb{F}$ is a non-commutative polynomial $f(x_1, \ldots, x_n)$ over $\mathbb{F}$, such that f vanishes on every $d \times d$ matrix assignment to its variables. For every field $\mathbb{F}$ of characteristic 0, every $d > 2$ and every finite basis of $d \times d$ matrix identities over $\mathbb{F}$, we show there exists a family of matrix identities $(f_n)_{n \in \mathbb{N}}$, such that each f_n has $2n$ variables and requires at least $\Omega(n^{2d})$ many generators to generate, where the generators are substitution instances of elements from the basis. The lower bound argument uses fundamental results from PI-algebras together with a generalization of the arguments in [Hru11].

We apply this result in algebraic proof complexity, focusing on proof systems for polynomial identities (PI proofs) which operate with algebraic circuits and whose axioms are the polynomial-ring axioms [HT09, HT15], and their subsystems. We identify a decreasing in strength hierarchy of subsystems of PI proofs, in which the dth level is a sound and complete proof system for proving $d \times d$ matrix identities (over a given field). For each level $d > 2$ in the hierarchy, we establish an $\Omega(n^{2d})$ lower bound on the number of proof-steps needed to prove certain identities.

Finally, we present several concrete open problems about non-commutative algebraic circuits and speed-ups in proof complexity, whose solution would establish stronger size lower bounds on PI proofs of matrix identities, and beyond.

Keywords: Algebraic complexity, PI-algebras, Proof Complexity, Non-commutative circuits
Mathematics subject classification: 16R10, 68Q17, 03F20

1 Introduction

Proof complexity studies the computational resources required to prove different statements in different proof systems. Beginning with the seminal work of Cook and Reckhow [CR79], proof systems for propositional logic (or unsatisfiable CNF formulas) attracted most attention in proof complexity research. It is however natural and interesting to investigate the complexity of proof systems for languages different than propositional logic. One such language of interest is that of polynomial identities written as algebraic circuits. Deciding the language of polynomial identities is the Polynomial Identity Testing (PIT) problem.

An efficient probabilistic algorithm for PIT is known, due to Schwartz and Zippel [Sch80, Zip79]: when the field is sufficiently large, with high probability two different polynomials will

[*] Department of Computer Science, The University of Texas at Austin. Email: fuli.theory.research@gmail.com (Parts of this work was done while at Tsinghua University, supported in part by the NSFC Grant 61373002).

[†] Department of Computer Science, Royal Holloway, University of London, Egham Hill, Egham, TW20 0EX. Email: Iddo.Tzameret@rhul.ac.uk

differ on a randomly chosen field assignment. However, whether the PIT problem is in P, namely is solvable in deterministic polynomial-time, is a major open problem in computational complexity and derandomization theory. Moreover, even showing that there are subexponential-size *witnesses* (verifiable in polynomial-time) witnessing that two algebraic circuits compute the same polynomial, constitutes a major open problem. Formally, it is unknown whether PIT is in $\mathsf{NSUBEXP}$ (let alone in NP; cf. Kabanets-Impagliazzo [KI04]).

Hrubeš-Tzameret [HT09] raised the question whether, assuming that the PIT problem does possess short witnesses, a proof system using only symbolic manipulations (resembling a logical proof system) is enough to provide these short witnesses; Or conversely, can we prove lower bounds on such proofs? Lower bounding the size of such symbolic manipulation-based proofs would not rule out that PIT is in NP, but would at least show that certain methods and algorithms (those algorithms whose run corresponds to a symbolic proof[1]) are incapable of establishing that PIT is in NP.

To this end, natural proof systems that operate with algebraic circuits and establish polynomial identities (*PI proof systems* for short) were introduced and studied in [HT09, HT15] (see also the survey [PT16]). A PI proof starts from a set of axioms expressing properties of polynomials (e.g., distributivity and commutativity), and derives new identities between algebraic circuits, using successive additions and multiplications of identities. It turned out that these proof systems are fairly strong: PI proofs can simulate many non-trivial structural constructions from algebraic circuit complexity and admit short proofs for quite a few identities of interest (see [HT09, HT15]). Moreover, only lower bounds on very restricted fragments of PI proofs are known [HT09], and apparently it is quite hard to prove (even polynomial-size) lower bounds on PI proofs (assuming nontrivial lower bounds even exist). PI proofs over $\mathbf{GF(2)}$ were shown to constitute a subsystem of propositional (Extended Frege) proofs, and so understanding the complexity of PI proofs has important implications in propositional proof complexity, as shown in [HT15] (cf. [PT16]).

In this paper, we continue the study of polynomial identities and their associated witness and proof complexity. We focus on matrix identities; the language of matrix identities (written as non-commutative algebraic circuits) constitutes a proper sub-language of polynomial identities. We are interested in the following question: *are there short witnesses for matrix identities, and specifically, does every matrix identity have a short symbolic-proof (i.e., a proof that starts from axioms and derives the identity step by step using symbolic manipulations)?*

Matrix identities are simply non-commutative polynomials that vanish over every matrix assignment. More precisely, for a polynomial f whose variables do not commute under multiplication (hence, a *non-commutative polynomial*), we can consider f as a polynomial over the matrix ring of $d \times d$ matrices $\mathrm{Mat}_d(\mathbb{F})$, for some constant dimension d and field $\mathbb{F}$. Then, the equation $f = 0$ means that f evaluates to the zero matrix for every $\mathrm{Mat}_d(\mathbb{F})$ assignment to its variables, in which case we call f a *matrix identity* of $\mathrm{Mat}_d(\mathbb{F})$.

Similar to polynomial identities, matrix identities can be decided in probabilistic polynomial-time (over sufficiently large fields).[2] But as far as we know, it is open whether matrix identities can be decided in deterministic polynomial-time, or possess sub-exponential witnesses. Thus, it is interesting to study whether matrix identities admit short symbolic proofs and establish lower bounds on these proofs, as a way to better understand the witness-complexity of matrix identities.

[1] Like the run of a (DPLL based) SAT-solver on unsatisfiable instances corresponds to a resolution refutation [SAT09].

[2] If we randomly choose scalar matrices αI, for α a field element and I the identity matrix, then with high probability a non-identity evaluates to a nonzero matrix under the assignment (similar to the commutative case).

Furthermore, the proof complexity of matrix identities is interesting from the pure proof complexity perspective, since proof systems for matrix identities are subsystems of PI proofs, for which we lack any nontrivial lower bound. Matrix identities seem like a good step towards PI proofs lower bounds, since they possess more structure than (commutative) polynomial identities. Indeed, the languages of matrix identities, of increasing dimensions, create a fine spectrum: on the one extreme we have (commutative) polynomial identities (i.e., identities of $\text{Mat}_1(\mathbb{F})$), on the other extreme non-commutative polynomial identities, and in between we have the languages of $d \times d$ matrix identities, for increasing d's (cf. Chien and Sinclair [CS07]). (Note that the language of $d \times d$ matrix identities is contained in the language of matrix identities of lower dimensions.)

The complexity of non-commutative identities (written as algebraic formulas) is quite well understood: by Raz and Shpilka [RS05] it is decidable in P (see also the recent work of Arvind et al. [AMR16] and references therein). So, informally, the spectrum from (commutative) polynomial identities to non-commutative identities becomes apparently easier to decide as we get closer to non-commutative identities (intuitively, as we progress into "less commutative" polynomial rings we have less dependencies between variables and thus identities become easier to track).

Our first goal will be to investigate the complexity of generating matrix identities, measured by the minimal number of generator instances needed to generate a given identity. We establish unconditional lower bounds on this measure. Our second goal, is to introduce sound and complete proof systems for establishing matrix identities (of increasing dimensions). These proof systems are subsystems of PI proof systems, and form a hierarchy of subsystems within PI proofs (whose first level coincides with PI proofs). Moreover, these proof systems are robust in the sense that for each level the choice of different axioms can only cost up to a polynomial increase in size. Using our first result, we show the existence of matrix identities that require many (i.e., $\Omega(n^{2d})$) proof-steps. Our final goal is to present two natural open problems, one about algebraic circuit complexity and another about proof complexity, based on which up to exponential-size lower bounds on PI proofs (for matrix identities suitably encoded) in terms of the size of the identities proved, follow. We also discuss possible connections to *propositional* proof complexity lower bounds.

2 Overview of Results

This section provides some necessary definitions and a detailed overview of our results.

2.1 Polynomial and Matrix Identities

For a field $\mathbb{F}$ let A be a non-commutative (associative and with a unity) $\mathbb{F}$-algebra; e.g., the algebra $\text{Mat}_d(\mathbb{F})$ of $d \times d$ matrices over $\mathbb{F}$. Formally, A is an $\mathbb{F}$-algebra if A is a vector space over $\mathbb{F}$ together with a distributive multiplication operation; where multiplication in A is associative (but it need not be commutative) and there exists a multiplicative unity in A. We always assume, unless explicitly stated otherwise, that the field $\mathbb{F}$ has characteristic 0 (when we write "any field" we also include fields of finite characteristics).

Denote by $\mathbb{F}[X]$ the ring of (commutative) polynomials with coefficients from $\mathbb{F}$ and variables $X := \{x_1, x_2, \dots\}$. A *polynomial* is a formal linear combination of monomials, where a *monomial* is a product of variables. Two polynomials are identical if all their monomials have the same coefficients. A ***non-commutative polynomial*** over the field $\mathbb{F}$ is a formal linear combination of monomials, where the product of variables is *non-commuting*. Since most polynomials in this work are non-commutative, *unless otherwise stated when we talk about* polynomials *we will mean non-commutative polynomials*. Nevertheless, to avoid confusion many times we will write in brackets whether a polynomial is commutative or non-commutative. The ring of (non-commutative) polynomials with variables X and over the field $\mathbb{F}$ is denoted $\mathbb{F}\langle X\rangle$. We say that the polynomial $f(x_1,\dots,x_n)\in\mathbb{F}\langle X\rangle$ is *an identity of the algebra A*, if for all $\overline{c}\in A^n$, $f(\overline{c})=0$. In particular, when A is $\mathrm{Mat}_d(\mathbb{F})$ we say that f is a ***matrix identity of*** $\mathrm{Mat}_d(\mathbb{F})$. A ***substitution instance*** of a polynomial $g(x_1,\dots,x_n)\in\mathbb{F}\langle X\rangle$ is a polynomial $g(h_1,\dots,h_n)$, for some $h_i\in\mathbb{F}\langle X\rangle$, $i\in[n]$.

2.2 Stratification

A matrix identity is a non-commutative polynomial vanishing over all assignments of matrices to variables. Consider the algebra of 1×1 "matrices" $\mathrm{Mat}_1(\mathbb{F})$, for $\mathbb{F}$ a field of characteristic 0. Its set of identities consists of all the non-commutative polynomials that vanish over field elements. Since, by definition, the field is commutative, the identities of $\mathrm{Mat}_1(\mathbb{F})$ can be considered as the set of all (commutative) polynomial identities (written as non-commutative polynomials); in other words, these are the non-commutative polynomials such that for every multiset of variables $\left\{x_{i_j}\ :\ j\in J\right\}$ the sum of coefficients of all monomials that are products of the variables in the multiset (with any product orders) is zero. For example, $x_1x_2x_{141}-\frac{1}{2}x_2x_{141}x_1-\frac{1}{2}x_2x_1x_{141}$ is a nonzero polynomial in $\mathbb{F}\langle X\rangle$ that is an identity of $\mathrm{Mat}_1(\mathbb{F})$.[3] Equivalently, the identities of $\mathrm{Mat}_1(\mathbb{F})$ are all non-commutative polynomials in the two-sided ideal generated by the *commutators* $x_ix_j-x_jx_i$, for every pair of variables x_i,x_j.

Using matrix identities of increasing dimensions d we obtain a *stratification* of the language of (commutative) polynomial identities, i.e., of the matrix identities of $\mathrm{Mat}_1(\mathbb{F})$ (see Figure 1). Namely, we obtain the following strictly decreasing (with respect to containment) chain of languages:

$$\begin{aligned}\text{(commutative) polynomial identities} = \mathrm{Mat}_1(\mathbb{F})\text{-identities} &\supsetneq \mathrm{Mat}_2(\mathbb{F})\text{-identities} \supsetneq \dots\\ &\supsetneq \mathrm{Mat}_d(\mathbb{F})\text{-identities} \supsetneq \mathrm{Mat}_{d+1}(\mathbb{F})\text{-identities} \supsetneq \dots\end{aligned}$$

The fact that the identities of $\mathrm{Mat}_{d+1}(\mathbb{F})$ are also identities of $\mathrm{Mat}_d(\mathbb{F})$ is easy to show. The fact that the chain above is *strictly* decreasing can be proved either by elementary methods [Jeř14] or as a corollary of [AL50].

2.3 Algebraic Circuits

Let $\mathbb{F}$ be a field. Algebraic circuits and formulas over $\mathbb{F}$ compute (commutative) polynomials in $\mathbb{F}[X]$ via addition and multiplication gates, starting from the input variables and constants from the field. More precisely, an *algebraic circuit F* is a finite directed acyclic graph (DAG) with *input nodes* (i.e., nodes of in-degree zero) and a single *output node* (i.e., a node of out-degree zero). Input nodes are labeled with either a variable or a field element in $\mathbb{F}$. All the other nodes have in-degree two (unless otherwise stated) and are labeled by either an addition gate $+$ or a product gate $\times$. An input node is said to *compute* the variable or scalar that labels itself. A $+$ (or $\times$) gate is said to compute the addition (product, resp.) of the (commutative) polynomials computed by its incoming nodes. An algebraic circuit is called a *formula*, if the underlying directed acyclic graph is a tree (that is, every node has at most one outgoing edge). The *size*

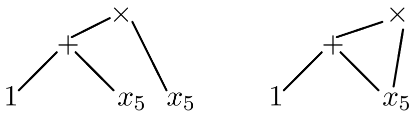
[3] Note that the problem of deciding the language of (commutative) polynomial identities (the PIT problem) written as algebraic circuits is identical to the problem of deciding the language of $\mathrm{Mat}_1(\mathbb{F})$ identities written as non-commutative algebraic circuits.

of a circuit F is the number of nodes in it, denoted $|F|$, and the *depth* of a circuit is the length of the longest directed path in it.

A ***non-commutative circuit*** is an algebraic circuit in which the children of product gates have *order*, so that a product gate is said to compute the non-commutative polynomial obtained by multiplying the (non-commutative) polynomial computed by the left child with the (non-commutative) polynomial computed by the right child (in this order). A *non-commutative formula* is a non-commutative circuit whose underlying directed acyclic graph is a tree.

For a (commutative or non-commutative) algebraic circuit F we denote by $\hat{F}$ the (commutative or non-commutative, resp.) polynomial computed by F.

We say that two algebraic circuits F, F' are *similar* if F and F' are syntactically identical when both are un-winded into *formulas* (a circuit is un-winded into a formula by duplicating every node in the directed acyclic graph that has a fan-out bigger than one, obtaining a tree instead of a DAG). The similarity relation can be decided in polynomial time (cf. [Jeř04]). For example, the following two circuits are similar, since the formula to the left is obtained by un-winding the circuit to the right into a formula (cf. [HT15]):



2.4 Proofs of Matrix Identities

We now introduce a hierarchy of proof systems for matrix identities. Each level d of the hierarchy proves $d \times d$ matrix identities over a given field. We begin with polynomial identities (PI) proofs.

2.4.1 Polynomial Identities Proofs

PI proofs as initially introduced in [HT09], denoted $\mathbf{PI}_c$ (and $\mathbf{PI}_c(\mathbb{F})$ when we wish to be explicit about the field $\mathbb{F}$), are sound and complete proof systems for the set of (commutative) polynomial identities of $\mathbb{F}$, written as equations between algebraic circuits. A PI proof starts from axioms like associativity, commutativity of addition and product, distributivity of product over addition, unit element axioms, etc., and derives new equations between algebraic circuits $F = G$ using rules for adding and multiplying two previous identities. The axioms of $\mathbf{PI}_c$ express reflexivity of equality, commutativity and associativity of addition and product, distributivity, zero element, unit element, and true identities in the field.

Algebraic circuits in PI proofs are treated as purely syntactic objects (similar to the way a propositional formula is a syntactic object in propositional proofs). Thus, simple computations such as multiplying out brackets, are done explicitly, step by step.

Definition 1 (System $\mathbf{PI}_c(\mathbb{F})$, [HT09, HT15])**.** *The system $\mathbf{PI}_c(\mathbb{F})$ proves equations of the form $F = G$, where F, G are algebraic circuits over $\mathbb{F}$. The inference rules of $\mathbf{PI}_c$ are (with F, G, H ranging over all algebraic circuits, and where an equation below a line can be inferred from the one above the line):*

$$\frac{F = G}{G = F} \qquad\qquad \frac{F = G \qquad G = H}{F = H} \qquad\qquad \frac{F_1 = G_1 \qquad F_2 = G_2}{F_1 \circ F_2 = G_1 \circ G_2} \quad \textit{for } \circ \in \{+, \cdot\}\,.$$

The axioms of $\mathbf{PI}_c$ are the following (again, F, G, H range over algebraic circuits):

$$F = F \qquad\qquad\qquad F + (G + H) = (F + G) + H$$

$$
\begin{aligned}
&F + G = G + F && F \cdot (G \cdot H) = (F \cdot G) \cdot H \\
&F \cdot G = G \cdot F && F \cdot (G + H) = F \cdot G + F \cdot H \\
&F + 0 = F && F \cdot 0 = 0 \\
&F \cdot 1 = F && \\
&a = b + c\,,\ \ a' = b' \cdot c'\,, && \text{when } a, b, c, a', b', c' \in \mathbb{F}, \text{ and the equations hold in } \mathbb{F}\,; \\
&F = F'\,, && \text{when } F, F' \text{ are } \textit{similar} \text{ circuits.}
\end{aligned}
$$

A **PI**$_c$ proof *is a sequence of equations (called* proof-lines*)* $F_1 = G_1$, $F_2 = G_2, \dots, F_k = G_k$*, with* F_i, G_i *circuits, such that every equation is either an axiom or was obtained from previous equations by one of the inference rules. The* **size** *of a proof is the total size of all circuits appearing in the proof. The* number of steps *in a proof is the number of proof-lines in it.*

A PI proof can be verified for correctness in polynomial-time (assuming the field has efficient representation; e.g., the field of rational numbers).

2.4.2 Matrix Identities Proofs

To define proof systems for matrix identities we need the concept of a *basis* of a set of identities of a given $\mathbb{F}$-algebra A (e.g., the matrix algebra $\mathrm{Mat}_d(\mathbb{F})$).

Definition 2 (Basis)**.** *We say that a set of non-commutative polynomials* $\mathcal{B}$ *forms a* **basis** *for the identities of an* $\mathbb{F}$*-algebra* A*, if the following holds: for every identity* f *of* A *there exist non-commutative polynomials* $g_1, \dots, g_k$*, for some* k*, that are substitution instances (see Sec. 2.1) of polynomials from* $\mathcal{B}$*, and such that* f *is in the two-sided ideal* $\langle g_1, \dots, g_k \rangle$ *.*

Notice that if we take out the "commutativity axiom"

$$F \cdot G = G \cdot F$$

from **PI**$_c$ proofs, we get a proof system that establishes *non-commutative* polynomial identities written as non-commutative algebraic circuits. The reason why we can consider this proof system as operating with *non-commutative* algebraic circuits is that, as mentioned above, circuits in **PI**$_c$ proofs are treated as syntactic objects and so product gates have order on their children and thus can be considered as either computing commutative or non-commutative polynomials.

Accordingly, to define proof systems for matrix identities we replace the commutativity axiom with polynomials from a basis of the matrix identities of $\mathrm{Mat}_d(\mathbb{F})$, as shown below. Intuitively, the basis of $\mathrm{Mat}_d(\mathbb{F})$-identities can be thought of as *higher-order commutativity axioms.*

For any field $\mathbb{F}$ of characteristic 0, any $d \geq 1$, and any basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$, we define the following proof system $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$, which is sound and complete for the identities of $\mathrm{Mat}_d(\mathbb{F})$ written as equations between non-commutative circuits:

Definition 3 (Proof system $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$)**.** *Let* $\mathcal{B} = \{B_1, \dots, B_k\} \subset \mathbb{F}\langle X \rangle$ *be a finite basis of* $\mathrm{Mat}_d(\mathbb{F})$*-identities, and let* $H_1, \dots, H_k$ *be non-commutative algebraic circuits such that* $\hat{H}_i = B_i$*, for all* $i \in [k]$*. The proof system* $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ *is defined by taking* $\mathbf{PI}_c(\mathbb{F})$ *(Definition 1) and replacing the commutativity axiom* $F \cdot G = G \cdot F$ *by the set of axioms* $H_1 = 0, \dots, H_k = 0$*. Additionally,* $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ *has the axioms of distributivity of product over addition from* both *left and right:* $F \cdot (G + H) = F \cdot G + F \cdot H$ *and* $(G + H) \cdot F = G \cdot F + H \cdot F$*.*[4]

[4] This is needed because we do not have anymore the commutativity axiom in our system to simulate both of these two distributivity axioms.

Note that $\mathbf{PI}_c(\mathbb{F})$ is equivalent to $\mathbf{PI}_{\mathrm{Mat}_1}(\mathbb{F})$, since the commutator $[g,h]$ is an axiom of $\mathbf{PI}_c(\mathbb{F})$ and the commutator is a basis of the identities of $\mathrm{Mat}_1(\mathbb{F})$ (and the two distributivity axioms polynomially simulate each other using the commutator axiom, and so they do not add more power to the system $\mathbf{PI}_{\mathrm{Mat}_1}(\mathbb{F})$).

Figure 1 illustrates the languages of matrix identities written as non-commutative circuits and their corresponding proof systems.

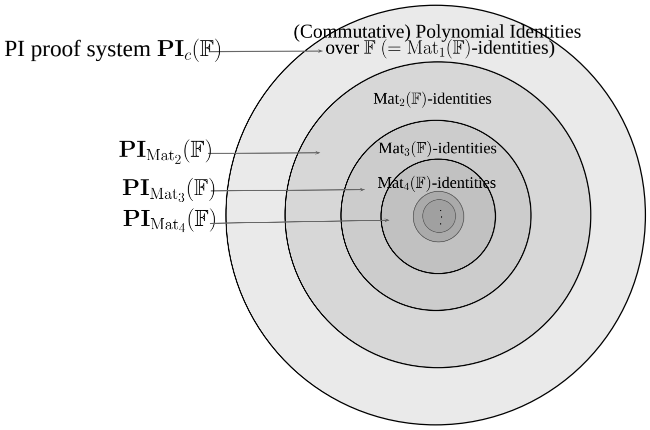

PI proof system $\mathbf{PI}_c(\mathbb{F})$
(Commutative) Polynomial Identities
over $\mathbb{F}$ (= $\mathrm{Mat}_1(\mathbb{F})$-identities)
$\mathrm{Mat}_2(\mathbb{F})$-identities
$\mathbf{PI}_{\mathrm{Mat}_2}(\mathbb{F})$
$\mathrm{Mat}_3(\mathbb{F})$-identities
$\mathbf{PI}_{\mathrm{Mat}_3}(\mathbb{F})$
$\mathrm{Mat}_4(\mathbb{F})$-identities
$\mathbf{PI}_{\mathrm{Mat}_4}(\mathbb{F})$
⋮

Figure 1: A schematic illustration of the languages of polynomial identities and their corresponding proof systems. The largest language is that of commutative polynomial identities written as non-commutative circuits (see Section 2.2).

$\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proofs are *robust* proof systems in the sense that different choices of finite bases $\mathcal{B}$ can only increase the number of lines in a $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$-proof by a constant factor. That is, for any fixed field $\mathbb{F}$ and fixed $d \geq 1$, replacing the axioms in $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ with any other finite set of axioms that are complete for $\mathrm{Mat}_d(\mathbb{F})$-identities will amount to a proof system that polynomially simulates $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ (when we use the gates algebraic gates $\cdot$, $+$, and field elements).

2.5 Main Lower Bound

Our main result is an unconditional lower bound on the size (in fact the number of proof-lines) of $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proofs, for any d, *in terms of the number of variables* n in the matrix identity proved:

Theorem 5 (Main lower bound)**.** *Let* $\mathbb{F}$ *be any field of characteristic zero, let* $d > 2$ *be any natural number and* $\mathcal{B}$ *be any finite basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*. Then, there exists a family of identities* $(f_n)_{n\in\mathbb{N}}$ *of* $\mathrm{Mat}_d(\mathbb{F})$ *each with degree* $2d+1$ *and* $2n$ *variables, such that any* $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ *proof of* f_n *requires* $\Omega(n^{2d})$ *proof-lines.*

The proof of the main lower bound is explained in the following subsection, and is based on a complexity measure defined on matrix identities and their generation in a (two-sided) ideal. The complexity measure is interesting by itself, and can be applied to identities of any algebra with polynomial identities (PI-algebras; see [Row80, Dre99] for the theory of PI-algebras), and not only matrix identities.

Comments. (i) When $d = 2$, our proof, showing the lower bound for *every* basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_2(\mathbb{F})$, does *not* hold. ***We explain this in the final paragraph of Section 5.2.***

(ii) The hard instance in the main lower bound theorem is *non-explicit*. Thus, we do not know if there are small non-commutative circuits computing the hard instances. This is the reason the lower bound holds only with respect to the number of variables n in the hard-instances and not with respect to its circuit size—the latter is the more desired result in proof complexity. Section 6 sets out an approach to achieve this latter result. However, we emphasize that in proof complexity non-explicit lower bounds are almost as interesting as explicit ones, and that for strong enough proof systems no non-explicit lower bounds are known to date (in contrast to Boolean circuit complexity in which explicitness plays a crucial role in lower bound results).

(iii) The proof-systems $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ are defined using a finite basis of the identities of $\mathrm{Mat}_d(\mathbb{F})$. An interesting feature of our proof (and theorem), is that it is an open problem to describe bases of the identities of $\mathrm{Mat}_d(\mathbb{F})$, for any $d > 2$. (For the case $d = 2$ the basis is known by Drensky [Dre81]). However, a highly nontrivial result of Kemer [Kem87], shows that for any natural d there *exists* a finite basis for $\mathrm{Mat}_d(\mathbb{F})$ (cf. [AKBK16], for a simpler proof for the zero characteristic case).

(iv) We do not know if the hierarchy of proof systems $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ for increasing d's is a *strictly* decreasing hierarchy (since we do not know if $\mathbf{PI}_{\mathrm{Mat}_{d-1}}(\mathbb{F})$ has any speed-up [namely, has smaller size proofs for some instances] over $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ for identities of $\mathrm{Mat}_d(\mathbb{F})$).

In the following section we give a detailed overview of the lower bound argument.

2.6 Proof Overview

Here we explain in detail the complexity measure we define and how to obtain the lower bound on this measure. This complexity measure is a lower bound on the minimal number of proof-lines in a corresponding $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$-proof (for the case $d = 1$ this was observed in [Hru11]), from which we conclude Theorem 5.

2.6.1 Generative Complexity of Identities

Let $\mathcal{B} \in \mathbb{F}\langle X\rangle$, and assume that A is an $\mathbb{F}$-algebra and f is an identity of A. Define

$$Q_{\mathcal{B}}(f)$$

as the minimal number k such that there exist $g_1, \dots, g_k \in \mathbb{F}\langle X\rangle$ that are all substitution instances of polynomials in $\mathcal{B}$, and such that $f \in \langle g_1, \dots, g_k\rangle$. (Note that different substitution instances of the same polynomials from $\mathcal{B}$ are counted twice.) We call $Q_{\mathcal{B}}(f)$ ***the generative complexity of f with respect to $\mathcal{B}$.***

We extend this definition by defining $Q_{\mathcal{B}}(f_1, \dots, f_m)$ as the minimal number k such that there exist $g_1, \dots, g_k \in \mathbb{F}\langle X\rangle$ that are all substitution instances of polynomials in $\mathcal{B}$, and $f_i \in \langle g_1, \dots, g_k\rangle$, for all $i \in [m]$. See Section 3.1 for more formal definitions.

Example: Let $\mathbb{F}$ be an infinite field and consider the field $\mathbb{F}$ itself as an $\mathbb{F}$-*algebra*, denoted $\mathscr{A}$. Then the identities of $\mathscr{A}$ are all the polynomials from $\mathbb{F}\langle X\rangle$ that evaluate to 0 under every assignment from $\mathbb{F}$ to the variables X. The identities of $\mathscr{A}$ are precisely the identities of $\mathrm{Mat}_1(\mathbb{F})$ discussed in Section 2.2. That is, these are the (non-commutative) polynomials that are identically zero polynomials *when considered as commutative polynomials.*

It is not hard to show that the *basis* of the algebra $\mathscr{A}$ is the *commutator* $x_1x_2-x_2x_1$, denoted $[x_1, x_2]$. In other words, every identity of $\mathscr{A}$ is generated (in the two-sided ideal) by substitution instances of the commutator. Considering $Q_{\{[x_1,x_2]\}}$, we can now ask what is $Q_{\{[x_1,x_2]\}}(x_1x_3 - x_3x_1 + x_2x_3 - x_3x_2)$? The answer is 1, since we need only *one* substitution instance of the commutator to generate the polynomial: $(x_1+x_2)x_3 - x_3(x_1+x_2) = x_1x_3 - x_3x_1 + x_2x_3 - x_3x_2$.

Hrubeš [Hru11] showed the following lower bound (using a slightly different terminology):

Theorem 1 (Hrubeš [Hru11])**.** *Let $\mathbb{F}$ be a field and let n be a positive natural number. There exists an identity $f \in \mathbb{F}\langle X \rangle$ of $\mathscr{A}$ with n variables, such that*

$$Q_{\{[x_1,x_2]\}}(f) = \Omega(n^2)\,.$$

It is also not hard to show that $Q_{\{[x_1,x_2]\}}(f) = O(n^2)$ for any identity f.

2.6.2 Lower Bounds on Generative Complexity

An *algebra with polynomial identities*, a *PI-algebra* for short, is an $\mathbb{F}$-algebra that has a non-trivial identity, that is, there is a *nonzero* $f \in \mathbb{F}\langle X \rangle$ that is an identity of the algebra.

We completely generalize Hrubeš [Hru11] lower bound above (excluding the case $d = 2$), from a lower bound of $\Omega(n^2)$ for generating identities of $\mathrm{Mat}_1(\mathbb{F})$ to a lower bound of $\Omega(n^{2d})$ for generating identities of $\mathrm{Mat}_d(\mathbb{F})$, for any $d > 2$ and any field $\mathbb{F}$ of characteristic zero. We exploit results about the structure of the identities of matrix algebras and the general theory of PI-algebras.

Theorem 4 (Lower bound on generative complexity)**.** *Let $\mathbb{F}$ be any field of characteristic zero. For every natural number $d > 2$ and every finite basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$, there exists a family of identities f_n over $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$ and $2n$ variables, such that $Q_{\mathcal{B}}(f) = \Omega(n^{2d})$.*

Similar to [Hru11], the lower bound in Theorem 4 is *non-explicit.*

Also, note that we do not know of an upper bound (in terms of n) that holds on $Q_{\mathcal{B}}(g)$, for every identity g with n variables.

The main lower bound (Theorem 5) is a corollary of Theorem 4 and the following proposition:

Proposition 6. *Let $\mathbb{F}$ be any field and let $\mathcal{B}$ be a finite basis of the identities of $\mathrm{Mat}_d(\mathbb{F})$. For every identity f of $\mathrm{Mat}_d(\mathbb{F})$, if F is a non-commutative circuit that computes f, the number of proof-lines in any $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proof of $F = 0$ is lower bounded up to a constant factor (depending on the choice of finite basis $\mathcal{B}$) by $Q_{\mathcal{B}}(f)$.*

Overview of the proof of Theorem 4. The study of algebras with polynomial identities is a fairly developed subject (see for instance the monographs by Drensky [Dre99] and Rowen [Row80]). Within this field, perhaps the most well studied topic is about the identities of matrix algebras. In particular, the well-known theorem of Amitsur and Levitzki from 1950 [AL50] is the following:

Amitsur-Levitzki Theorem ([AL50])**.** *Let $\mathfrak{S}_d$ be the permutation group on d elements and let $S_d(x_1, x_2, \ldots, x_d)$ denote the* ***standard identity*** *of degree d as follows:*

$$S_d(x_1, x_2, \ldots, x_d) := \sum_{\sigma \in \mathfrak{S}_d} sgn(\sigma) \prod_{i=1}^{d} x_{\sigma(i)}.$$

Then, for any natural number d and any field $\mathbb{F}$ (in fact, any commutative ring) the standard identity $S_{2d}(x_1, x_2, \ldots, x_{2d})$ of degree $2d$ is an identity of $\mathrm{Mat}_d(\mathbb{F})$.

Theorem 4 is proved in several steps. The main argument can be divided into two main parts, described as follows:

Part 1: We use the Amitsur-Levitzki Theorem to show that when $\mathcal{E} = \{S_{2d}(x_1, \dots, x_{2d})\}$ there exists an $f_n \in \mathbb{F}\langle X\rangle$ with $2n$ variables and degree $2d+1$, such that $Q_{\mathcal{E}}(f) = \Omega(n^{2d})$. To this end, we generalize the method in [Hru11] to "higher order commutativity axioms": using a counting argument we show the existence of n special polynomials (that we call *s-polynomials*; see Definition 8) $P_1, P_2, \dots, P_n$ over n variables each of degree $2d$ such that $Q_{\mathcal{E}}(P_1, \dots, P_n) = \Omega(n^{2d})$ (see Lemma 11). Then, we combine the n s-polynomials into a single polynomial $P^\star$ with degree $2d+1$, by adding n new variables, such that $Q_{\mathcal{E}}(P^\star) = \Omega(Q_{\mathcal{E}}(P_1, \dots, P_n))$. (The polynomial $P^\star$ will constitute the hard instance f_n.)

See the proof of Lemma 11 for a concise overview of the counting argument we use.

Part 2: In contrast to the case $d = 1$ in [Hru11], $\mathcal{E} = \{S_{2d}(x_1, \dots, x_{2d})\}$ for $d > 1$, is known *not* to be a basis of the identities of $\text{Mat}_d(\mathbb{F})$, namely there are identities of $\text{Mat}_d(\mathbb{F})$ that are not generated by substitution instances of S_{2d} (see [BDDK03, Sec. 2] and [Dre99]) (also notice that $Q_{\mathcal{B}}(f)$ can be defined for any set $\mathcal{B} \subseteq \mathbb{F}\langle X\rangle$). In this part we show roughly that for the hard instances f_n in Theorem 4 no generators different from the S_{2d} generators can contribute to its generation. More precisely, we show that when $d > 2$, for *all finite bases* $\mathcal{B}$ *of the identities of* $\text{Mat}_d(\mathbb{F})$, the following holds for f_n: $Q_{\mathcal{B}}(f_n) \geq c \cdot Q_{\mathcal{E}}(f_n)$ for some constant c that depends on $\mathcal{B}$ and d but not on n.

For this purpose, we find a special set $\mathcal{B}' \subseteq \mathbb{F}\langle X\rangle$ that serves as an "intermediate" set between $\mathcal{B}$ and $\mathcal{E}$, such that $\mathcal{B}$ is generated by $\mathcal{B}'$, and all the polynomials in $\mathcal{B}'$ that contribute to the generation of the hard instance f_n can be generated already by $\mathcal{E}$. We then show (Corollary 20) that for any basis $\mathcal{B}$, there is a specific set $\mathcal{B}'$ of polynomials of a special form, namely, *multi-homogeneous commutator polynomials* (Definition 9), that can generate $\mathcal{B}$. Based on the properties of multi-homogeneous commutator polynomials, we show that, for the hard instance f_n, only the generators of degree at most $2d+1$ in $\mathcal{B}'$ can contribute to the generation of f_n (Lemma 24). We then prove that when $d > 2$, all the generators of degree at most $2d+1$ in $\mathcal{B}'$ can be generated by $\mathcal{E}$ (this is where we use the assumption that $d > 2$ (see Lemma 23)). We thus get the conclusion $Q_{\mathcal{B}'}(f) \geq c \cdot Q_{\mathcal{E}}(f)$, when $d > 2$.

2.7 Relation to Previous Work

As mentioned above, our work generalizes Hrubeš' work [Hru11]. That work also considered proving *quadratic* size lower bounds on PI proofs $\mathbf{PI}_c$. It gave several conditions and open problems, under which, quadratic size lower bounds on PI proofs would follow, and further, showed that the general framework suggested may have potential, at least in theory, to yield Extended Frege quadratic-size lower bounds; however, we note that Extended Frege quadratic-size lower bounds are in fact *already known*, since the same lower bound on Frege from [Kra95] holds for Extended Frege[5].

Hrubeš and Tzameret [HT15] obtained polynomial-size (algebraic and propositional) proofs for certain (suitably encoded) identities concerning matrices. However, in the current work we are studying matrix identities in which the number of matrices grows with the number of variables n in the identity, whereas in [HT15] the number of matrices was fixed and only the dimension of the matrices grows.

Other results connecting non-commutative polynomials and proof complexity is the recent work of Li et at. [LTW15] (and its precursor in [Tza11]) showing that a non-commutative formula-based proof system (formally, an *Ideal Proof System* certificate in the sense of Grochow

[5] We thank Emil Jeřabek for drawing our attention to this fact.

and Pitassi [GP14], which is written as a non-commutative formula and uses the commutators as additional axioms) is sufficient to polynomially simulate Frege proofs (and over $\mathbf{GF(2)}$ is *equivalent* to Frege proofs up to quasi-polynomial size factors).

3 More Formal Preliminaries

3.1 Algebras with Polynomial Identities

For a natural number n, put $[n] := \{1, 2, \ldots, n\}$. We use lower case letters a, b, c for constants from the underlying field, x, y, z for variables, $\overline{x}, \overline{y}, \overline{z}$ for vectors of variables, f, g, h, ℓ or upper case letters such as A, B, P, Q for polynomials and $\overline{f}, \overline{g}, \overline{h}, \overline{\ell}, \overline{A}, \overline{B}, \overline{P}, \overline{Q}$, for vectors of polynomials (when the arity of the vector is clear from the context).

Recall the definition of commutative and non-commutative polynomials from Section 2.1. For two polynomials $f(x_1, \ldots, x_n)$ and g we sometimes denote the substitution instance $f(h_1, \ldots, h_n)$ by $f(\overline{h})$. For a polynomial $f(x_1, \ldots, x_n) \in \mathbb{F}\langle X\rangle$, $f\big|_{x_{i_1}\leftarrow g_{i_1}, \ldots, x_{i_k}\leftarrow g_{i_k}}$ denotes the polynomial that replaces $x_{i_1}, \ldots, x_{i_k}$ by $g_{i_1}, \ldots, g_{i_k}$ in f, respectively, where $g_{i_1}, \ldots, g_{i_k} \in \mathbb{F}\langle X\rangle, i_1, \ldots, i_k$ are distinct numbers from $[n]$ and $k \in [n]$. For a vector $\overline{H}$ of polynomials $H_1, \ldots, H_k \in \mathbb{F}\langle X\rangle$ where k is a positive integer, we use the notation $\overline{H}|_{H_j \leftarrow f}$, to denote the vector of polynomials that replaces the jth coordinate H_j in $\overline{H}$ by a polynomial $f \in \mathbb{F}\langle X\rangle$, where $j \in [k]$.

Let A be a vector space over a field $\mathbb{F}$ and $\cdot : A \times A \to A$ be a distributive multiplication operation. If $\cdot$ is associative, that is, $a_1 \cdot (a_2 \cdot a_3) = (a_1 \cdot a_2) \cdot a_3$ for all a_1, a_2, a_3 in A, then the pair $(A, \cdot)$ is called an ***associative algebra over*** $\mathbb{F}$, or an $\mathbb{F}$***-algebra***, for short.[6]

The algebra of $d \times d$ matrices $\mathrm{Mat}_d(\mathbb{F})$, for some positive natural number d, with entries from $\mathbb{F}$ (and with the usual addition and multiplication of matrices) is an example of an $\mathbb{F}$-algebra. Note that $\mathrm{Mat}_d(\mathbb{F})$ is an associative algebra but not a commutative one.

We can consider the ring of non-commutative polynomials $\mathbb{F}\langle X\rangle$ as the associative algebra of all polynomials such that the variables $X = \{x_1, x_2, \ldots\}$ are non-commutative with respect to multiplication. The ring $\mathbb{F}\langle X\rangle$ is also called the *free algebra (over X)*.

We now define formally the concept of a *polynomial identity algebra* (mentioned before):

Definition 4. *Let A be an $\mathbb{F}$-algebra. An **identity of** A is a polynomial $f(x_1, \ldots, x_n) \in \mathbb{F}\langle X\rangle$ such that:*

$$f(a_1, \ldots, a_n) = 0, \text{ for all } a_1, \ldots, a_n \in A.$$

*A **PI-algebra** is an algebra that has a non-trivial identity, that is, there is a nonzero $f \in \mathbb{F}\langle X\rangle$ that is an identity of the algebra.*

For example, every *commutative* $\mathbb{F}$-algebra A is also a PI-algebra: for any $u, v \in A$, it holds that $uv - vu = 0$, and so $x_i x_j - x_j x_i$ is a nonzero polynomial identity of A, for any positive $i \neq j \in \mathbb{N}$. A concrete example of a commutative algebra is the usual ring of (*commutative*) polynomials with coefficients from a field $\mathbb{F}$ and variables $X = \{x_1, x_2, \ldots\}$, denoted $\mathbb{F}[X]$.

An example of an algebra that is *not* a PI-algebra is the free algebra $\mathbb{F}\langle X\rangle$ itself. This is because a nonzero polynomial $f \in \mathbb{F}\langle X\rangle$ cannot be an identity of $\mathbb{F}\langle X\rangle$ (since the assignment that maps each variable to itself does not nullify f).

A ***two-sided ideal*** I of an $\mathbb{F}$-algebra A is a subset of A such that for any (not necessarily distinct) elements $f_1, \ldots, f_n$ from I we have $\sum_{i=1}^{n} g_i \cdot f_i \cdot h_i \in I$, for all $g_1, \ldots, g_n, h_1, \ldots, h_n \in A$.

[6] In general an $\mathbb{F}$-algebra can be non-associative, but since we only talk about associative algebras in this paper we use the notion of $\mathbb{F}$-algebra to imply that the algebra is associative.

Definition 5. *A* ***T-ideal*** $\mathcal{T}$ *is a two-sided ideal of* $\mathbb{F}\langle X\rangle$ *that is closed under all endomorphisms*[7]*, namely, is closed under all substitutions of variables by polynomials.*

In other words, a T-ideal is a two-sided ideal $\mathcal{T}$, such that if $f(x_1,\dots,x_n) \in \mathcal{T}$ then $f(g_1,\dots,g_n) \in \mathcal{T}$, for any $g_1,\dots,g_n \in \mathbb{F}\langle X\rangle$.

It is easy to see the following:

Fact 2. *The set of identities of an (associative) algebra is a T-ideal.*

Recall the definition of a basis of a set of identities over an algebra (Definition 2). We repeat here the definition of a basis, using the notion of a T-ideal. The basis of a T-ideal $\mathcal{T}$ is a set of polynomials whose substitution instances generate $\mathcal{T}$ *as an ideal*:

Definition 6. *Let* $B \subseteq \mathbb{F}\langle X\rangle$ *be a set of polynomials and let* $\mathcal{T}$ *be a T-ideal in* $\mathbb{F}\langle X\rangle$*. We say that* B ***is a basis for*** $\mathcal{T}$ *or that* $\mathcal{T}$ ***is generated as a T-ideal by*** B*, if every* $f \in \mathcal{T}$ *can be written as:*

$$f = \sum_{i\in I} h_i \cdot B_i(g_{i1},\dots,g_{in_i}) \cdot \ell_i\,, \tag{1}$$

for $h_i, \ell_i, g_{i1},\dots,g_{in_i} \in \mathbb{F}\langle X\rangle$ *and* $B_i \in B$ *(for all* $i \in I$*).*

Given $B \subseteq \mathbb{F}\langle X\rangle$, we write $T(B)$ to denote the T-ideal generated by B. Thus, a T-ideal $\mathcal{T}$ is generated by $B \subseteq \mathbb{F}\langle X\rangle$ iff $\mathcal{T} = T(B)$.

Examples: $T(x_1)$ is simply the set of all polynomials from $\mathbb{F}\langle X\rangle$. $T(x_1x_2 - x_2x_1)$ is the set of all non-commutative polynomials that are zero if considered as commutative polynomials.

We say that a polynomial $f \in \mathbb{F}\langle X\rangle$ is a ***consequence*** of the polynomials $\{B_i\}_{i\in I}$, if f can be written as in (1).

Note that the concept of a T-ideal is already reminiscent of logical proof systems, where generators of the T-ideal $\mathcal{T}$ are like axioms schemes and generators of a two-sided ideal containing f are like substitution instances of the axioms.

A polynomial is ***homogeneous*** if all its monomials have the same total degree. Given a polynomial f, the ***homogeneous part of degree*** j of f, denoted $f^{(j)}$ is the sum of all monomials with total degree j. We write $(C)^{(j)}$ to denote the jth-homogeneous part of the circuit C, and given the vector of circuits $\overline{C} = (C_1,\dots,C_k)$ the vector $\left(\overline{C}\right)^{(j)}$ denotes the vector $(C_1^{(j)},\dots,C_k^{(j)})$.

4 Complexity of Generating Matrix Identities

Here we formally define the complexity measure for generating a matrix identity. We repeat some of the concepts introduced already in Section 2.6.

Let A be a PI-algebra (Definition 4) and let $\mathcal{T}$ be the T-ideal (Definition 5) consisting of all identities of A (see Fact 2). Assume that $\mathcal{B}$ is a basis for the T-ideal $\mathcal{T}$ (Definition 6), that is, $T(B) = \mathcal{T}$. Then every $f \in \mathcal{T}$ is a consequence of $\mathcal{B}$, that is, can be written as a combination of substitution instances of polynomials from $\mathcal{B}$, as follows:

$$f = \sum_{i\in I} h_i \cdot B_i(g_{i1},\dots,g_{in_i}) \cdot \ell_i\,, \tag{2}$$

[7] An algebra endomorphism of A is an (algebra) homomorphism $A \to A$.

for $h_i, \ell_i, g_{i1}, \ldots, g_{in_i} \in \mathbb{F}\langle X\rangle$ and $B_i \in \mathcal{B}$ (for all $i \in I$). A very natural question, from the complexity point of view, is the following: *How many distinct substitution instances of generators are needed to generate f above?*

Formally, we have the following:

Definition 7 ($Q_{\mathcal{B}}(f)$)**.** *For any set of polynomials $\mathcal{B} \subseteq \mathbb{F}\langle X\rangle$, define $Q_{\mathcal{B}}(f)$ as the smallest (finite) k such that there exist substitution instances $g_1, \ldots, g_k$ of polynomials from $\mathcal{B}$ with*

$$f \in \langle g_1, \ldots, g_k\rangle,$$

where $\langle g_1, \ldots, g_k\rangle$ is the two-sided ideal generated by $g_1, \ldots, g_k$.

Note that we do not need to assume that $\mathcal{B}$ is a basis of all identities of the algebra A to make $Q_{\mathcal{B}}(F)$ definable. If the set $\mathcal{B}$ is a singleton $\mathcal{B} = \{h\}$, we can also write $Q_h(\cdot)$ instead of $Q_{\{h\}}(\cdot)$. We also extend Definition 7 to a *sequence* of polynomials and let $Q_{\mathcal{B}}(f_1, \ldots, f_n)$ be the smallest k such that there exist some substitution instances $g_1, \ldots, g_k$ of polynomials from $\mathcal{B}$ with

$$f_i \in \langle g_1, \ldots, g_k\rangle, \quad \text{for all } i \in [n].$$

Notice that $Q_{\mathcal{B}}(f)$ is interesting only if f is not already in the generating set. Hence, we need to make sure that the generating set does not contain f and the easiest way to do this (when considering asymptotic growth of measure) is by stipulating that the generating set is finite. Given an algebra, the question whether there exists a finite generating set of the T-ideal of the identities of the algebra is a highly non-trivial *Specht Problem.* Fortunately, for matrix algebras we can use the solution of the Specht problem given by Kemer [Kem87] (see also [AKBK16]). Kemer showed that for every matrix algebra A there exists a finite basis of the T-ideal of the identities of A. The problem to actually describe such a finite basis for most matrix algebras (namely for all values of d, for $\mathrm{Mat}_d(\mathbb{F})$) is open.

We have the following simple proposition, which is analogous to a certain extent to the fact that every two (Frege) propositional proof systems polynomially simulate each other (cf. [Kra95]):

Proposition 3 (Robustness of Q-measure)**.** *Let A be some $\mathbb{F}$-algebra and let $\mathcal{B}_0$ and $\mathcal{B}_1$ be two* finite *bases for the identities of A. Then, there exist constants c, c' (that depends only on $\mathcal{B}_0, \mathcal{B}_1$, and more precisely, $c = \max_{B\in\mathcal{B}_1}\{Q_{\mathcal{B}_0}(B)\}$ and $c' = \max_{B\in\mathcal{B}_0}\{Q_{\mathcal{B}_1}(B)\}$), such that for any identity f of A:*

$$c' \cdot Q_{\mathcal{B}_1}(f) \le Q_{\mathcal{B}_0}(f) \le c \cdot Q_{\mathcal{B}_1}(f).$$

Proof. Assume that $\mathcal{B}_0 = \{A_1, \ldots, A_k\}$ and $\mathcal{B}_1 = \{B_1, \ldots, B_\ell\}$ are the two bases (where the A_i, B_i's are polynomials from $\mathbb{F}\langle X\rangle$). And suppose that $Q_{\mathcal{B}_1}(f) = q$ and $f \in \left\langle B_{i_1}(\overline{g_1}), \ldots, B_{i_q}(\overline{g_q})\right\rangle$, for $i_j \in [\ell]$ and where $\overline{g_j} \in \mathbb{F}\langle X\rangle$ are the substitutions of polynomials for the variables of B_{i_j}. By assumption that both $\mathcal{B}_0$ and $\mathcal{B}_1$ are bases for A, there exists a constant c such that $B_{i_j} \in \left\langle A_{j_1}(\overline{h_{j_1}}), \ldots, A_{j_r}(\overline{h_{j_c}})\right\rangle$, for all $j \in [q]$, and where $\overline{h_{j_l}} \in \mathbb{F}\langle X\rangle$ are the substitutions of polynomials for the variables of A_{j_l}, for any $l \in [c]$ (more precisely, $c = \max\{Q_{\mathcal{B}_0}(B_i) \ : \ i \in [\ell]\}$).

Note that if $B_{i_j} \in \left\langle A_{j_1}(\overline{h_{j_1}}), \ldots, A_{j_c}(\overline{h_{j_c}})\right\rangle$, then for any substitution $\overline{g_j}$ (of polynomials to the variables X) we have $B_{i_j}(\overline{g_j}) \in \left\langle \left(A_{j_1}(\overline{h_{j_1}})\right)(\overline{g_j}), \ldots, \left(A_{j_c}(\overline{h_{j_c}})\right)(\overline{g_j})\right\rangle$. Thus, each of the $B_{i_j}(\overline{g_j})$'s that generate f (where $j \in [q]$), is generated by itself by at most c substitution instances of polynomials from $\mathcal{B}_0$. Therefore, f can be generated with at most $c \cdot q$ substitution instances of generators from $\mathcal{B}_0$, that is,

$$Q_{\mathcal{B}_0}(f) \le c \cdot Q_{\mathcal{B}_1}(f), \qquad \text{where } c = \max\{Q_{\mathcal{B}_0}(B_i) \ : \ i \in [\ell]\}.$$

Similarly, we have $c' \cdot Q_{\mathcal{B}_1}(f) \le Q_{\mathcal{B}_0}(f)$, for $c' = (\max_{B\in\mathcal{B}_0}\{Q_{\mathcal{B}_1}(B)\})^{-1}$. QED

5 Main Lower Bound

In this Section we prove our main lower bound on the generative complexity of matrix identities (restated from Section 2.6.2):

Theorem 4 (Main generative complexity lower bound)**.** *Let $\mathbb{F}$ be a field of characteristic zero. For every natural number $d > 2$ and for every finite basis $\mathcal{B}$ of the T-ideal of identities of* $\mathrm{Mat}_d(\mathbb{F})$*, there exists an identity P over* $\mathrm{Mat}_d(\mathbb{F})$ *of degree $2d+1$ with n variables, such that* $Q_{\mathcal{B}}(P) = \Omega\left(\binom{n}{2d}\right) = \Omega(n^{2d})$.

It is interesting to point out that although we do not necessarily know what is the (finite) generating set of $\mathrm{Mat}_d(\mathbb{F})$ we still can lower bound the number of generators needed to generate certain identities. This is due to the fact that we know some finite bases exist, and further we will have some information on the generating set of the hard instances considered (see Section 5.2).

As a corollary of Theorem 4 we obtain the main proof complexity lower bound (restated from Section 2.5):

Theorem 5 (Main proof complexity lower bound)**.** *Let $\mathbb{F}$ be any field of characteristic zero. For any natural number $d > 2$ and every finite basis $\mathcal{B}$ of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*, there exists an identity f over* $\mathrm{Mat}_d(\mathbb{F})$ *of degree $2d+1$ with n variables, such that any* $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$*-proof of f requires $\Omega(n^{2d})$ proof-lines.*

Proof. This follows from Theorem 4 and Lemma 6 proved below. QED

Lemma 6. *Let $\mathbb{F}$ be a field and let $\mathcal{B}$ be a finite basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*. For every identity f of* $\mathrm{Mat}_d(\mathbb{F})$*, if F is a non-commutative circuit that computes f, the number of lines in a* $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ *proof of $F = 0$ is lower bounded up to a constant factor (depending on the choice of finite basis $\mathcal{B}$) by $Q_{\mathcal{B}}(f)$.*

Proof. Let π be a $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proof of $F = 0$ and let T be the set of all the basis $\mathcal{B}$ axioms used in π, namely, T consists of all the equations $H = 0$ in π, where H is a substitution instance of some $B \in \mathcal{B}$. Recall that $\hat{H}$ denotes the *polynomial* computed by H. It suffices to show that $|T| \geq Q_{\mathcal{B}}(f)$, which will follow by showing that

$$f \in \left\langle h \in \mathbb{F}\langle X\rangle \;:\; h = \hat{H} \text{ and } (H = 0) \in T \right\rangle . \tag{3}$$

(3) is proved by a straightforward induction on the number of proof-lines in π (because every $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proof can be seen as computing in the ideal generated by the proof lines). QED

The rest of this section is dedicated to proving Theorem 4.

5.1 Generative Lower Bound from a Specific Set of Identities

Here we prove Lemma 7, which is a lower bound on $Q_{S_{2d}}$. That is, we prove a lower bound on the number of substitution instances of a specific set of identities S_{2d} needed to generate a certain identity. Note that S_{2d} is *not* known to be the basis of the T-ideal of the identities over $\mathrm{Mat}_d(\mathbb{F})$. More precisely, we wish to prove:

Lemma 7. *Let $d \geq 1$ be a natural number and let $\mathbb{F}$ be a field of characteristic zero, then there exists a polynomial $P \in$* $\mathrm{Mat}_d(\mathbb{F})$ *of degree $2d+1$ with n variables such that $Q_{S_{2d}}(P) = \Omega(n^{2d})$.*

Comment (on characteristic zero fields): Lemma 7 can be shown to hold for every *finite* field $\mathbb{F}$. We state and prove the lemma only for zero characteristic fields since when we apply the lemma in Section 5.2 we will need to assume that the field is of zero characteristic (see for example Proposition 19).

We introduce the following central definition:

Definition 8. *A polynomial $P \in \mathbb{F}\langle X\rangle$ with n variables $x_1, \ldots, x_n$ is called an* ***s-polynomial*** *if:*

$$P = \sum_{j_1<j_2<\ldots<j_{2d}\in[n]} c_{j_1 j_2 \ldots j_{2d}} \cdot S_{2d}\left(x_{j_1}, \ldots x_{j_{2d}}\right),$$

for some natural d and constants $c_{j_1 j_2 \ldots, j_{2d}} \in \{0,1\}$, for all $j_1 < j_2 < \ldots < j_{2d} \in [n]$.

Lemma 8. *For every $P_1, \ldots, P_{2d} \in \mathbb{F}\langle X\rangle$ where d is a positive integer, $S_{2d}(P_1, \ldots, P_{2d})$ is the zero polynomial if there exists $i \in [2d]$ such that P_i is a constant.*

Proof. Assume $P_\delta = c \in \mathbb{F}$, for some $\delta \in [2d]$. Given $i_1 \neq i_2 \neq \ldots \neq i_{2d-1} \in [n] \setminus \delta$, let σ_m denote the permutation

$$\begin{pmatrix} 1 & 2 & \ldots & m-1 & m & m+1 & \ldots & 2d \\ i_1 & i_2 & \ldots & i_{m-1} & \delta & i_m & \ldots & i_{2d-1} \end{pmatrix}.$$

Then,

$$\begin{aligned}
S_{2d}(P_1, \ldots, P_{2d}) &= \sum_{\sigma\in\mathcal{S}_{2d}} sgn(\sigma) \prod_{i=1}^{2d} P_{\sigma(i)} \qquad \text{(by definition)} \\
&= \sum_{i_1\neq i_2\neq\ldots\neq i_{2d-1}\in[2d]\setminus\delta} \sum_{m=1}^{2d} sgn(\sigma_m) \prod_{j=1}^{m-1} P_{i_j} P_\delta \prod_{j=m}^{2d-1} P_{i_j} \\
&= c \cdot \left(\sum_{i_1\neq i_2\neq\ldots\neq i_{2d-1}\in[2d]\setminus\delta} \left(\sum_{m=1}^{2d} sgn(\sigma_m) \right) \prod_{j=1}^{2d-1} P_{i_j} \right) \\
&= c \cdot \left(\sum_{i_1\neq i_2\neq\ldots\neq i_{2d-1}\in[2d]\setminus\delta} \left(\sum_{m=1}^{d} \left(sgn\left(\sigma_{2m-1}\right) + sgn(\sigma_{2m})\right) \right) \prod_{j=1}^{2d-1} P_{i_j} \right) \\
&= c \cdot \left(\sum_{i_1\neq i_2\neq\ldots\neq i_{2d-1}\in[2d]\setminus\delta} \left(\sum_{m=1}^{d} 0 \right) \prod_{j=1}^{2d-1} P_{i_j} \right) = 0\,.
\end{aligned}$$

QED

Recall that for a polynomial g, $g^{(i)}$ stands for the homogeneous component of degree i of g.

Lemma 9. *For every sequence $\overline{P}$ of $2d$ polynomials, $S_{2d}(\overline{P})^{(2d)} = S_{2d}\left(\left(\overline{P}\right)^{(1)}\right)$.*

Proof. Note that

$$S_{2d}(\overline{P})^{(2d)} = S_{2d}\left(\left(\overline{P}\right)^{(1)}\right) + \sum_{j_1+\ldots+j_{2d}=2d \text{ and } \exists i\in[2d], j_i\neq 1} S_{2d}\left((P)^{(j_1)}, \ldots, (P)^{(j_{2d})}\right).$$

But every summand in the rightmost term must have $j_r = 0$ for some $r \in [2d]$ (since otherwise $j_1 + \ldots + j_{2d} > 2d$). Thus, by Lemma 8, every summand in the rightmost term is zero. QED

We can now show that every s-polynomial has the following property:

Lemma 10. *Let f be an s-polynomial. If there exist vectors of polynomials $\overline{P_1},\dots,\overline{P_r}$ with*

$$f \in \left\langle S_{2d}(\overline{P_1}),\dots,S_{2d}(\overline{P_r})\right\rangle,$$

then there are constants c_i's such that

$$f = \sum_{i=1}^{r} c_i S_{2d}\left((\overline{P_i})^{(1)}\right).$$

Proof. Notice that the s-formula f is $2d$-homogeneous. Thus,

$$f = (f)^{(2d)} \in \left\{ (h)^{(2d)} \;\middle|\; h \in \langle S_{2d}(\overline{P_1}),\dots,S_{2d}(\overline{P_r})\rangle \right\}.$$

By Lemma 8 (and the linearity of S_{2d}), every nonzero substitution instance of S_{2d} must be of degree at least $2d$. Thus

$$f \in \left\langle S_{2d}(\overline{P_1})^{(2d)},\dots,S_{2d}(\overline{P_r})^{(2d)}\right\rangle.$$

By Lemma 9 we have

$$f \in \left\langle S_{2d}\left((\overline{P_1})^{(1)}\right),\dots,S_{2d}\left((\overline{P_r})^{(1)}\right)\right\rangle.$$

That is,

$$f = \sum_{j=1}^{r}\sum_{i=1}^{t_j} A_{ji} S_{2d}\left((\overline{P_j})^{(1)}\right) B_{ji}, \qquad \text{for some } A_{ji}, B_{ji} \in \mathbb{F}\langle X\rangle.$$

Moreover,

$$\left(A_{ji} S_{2d}\left((\overline{P_j})^{(1)}\right) B_{ji}\right)^{(2d)} = (A_{ji}B_{ji})^{(0)}\, S_{2d}\left((\overline{P_j})^{(1)}\right).$$

And thus,

$$f = \sum_{j=1}^{r} c_j S_{2d}\left((\overline{P_j})^{(1)}\right),$$

where c_j is the constant $\sum_{i=1}^{t_j}(A_{ji}B_{ji})^{(0)}$, for any $j \in [r]$. QED

5.1.1 The Counting Argument

Notation. *If $B \subseteq \mathbb{F}\langle X\rangle$ contains only a* single *polynomial g, then we write $Q_g(\cdot)$ instead of $Q_B(\cdot)$, to simplify the writing. Note that B may not be a basis for the algebra considered (e.g., we may consider identities of the* $\mathrm{Mat}_d(\mathbb{F})$ *generated by some B, where B is not a basis for (all) the identities of* $\mathrm{Mat}_d(\mathbb{F})$*).*

Lemma 11. *Let $\mathbb{F}$ be a field of characteristic zero. There exist s-polynomials $P_1,\dots,P_n$ which are identities of* $\mathrm{Mat}_d(\mathbb{F})$ *in n variables, such that $Q_{S_{2d}}(P_1,\dots,P_n) = \Omega(n^{2d})$, and where $Q_{S_{2d}}(P_1,\dots,P_n)$ is finite.*

In Section 5.2 we show that, if $\mathbb{F}$ is of characteristic zero then this lower bound holds for *all finite bases of* $\mathrm{Mat}_d(\mathbb{F})$, namely for Q_B, where B is any finite basis of $\mathrm{Mat}_d(\mathbb{F})$.

Proof. We prove, by a generalization of the counting argument from [Hru11], that there exists a sequence of polynomials $P_1,\dots,P_n$ that require $\Omega\left(n^{2d}\right)$ substitution instances of the $S_{2d}(x_1,\dots,x_{2d})$ identities to generate (all of the polynomials in the sequence) in a two-sided ideal.

Informal overview of proof. First, we show that the total number of n-tuples of s-formulas is $2^{n\binom{n}{2d}}$: each P_i (for $i = 1, \ldots, n$) is determined by the degree-$2d$ standard polynomials we choose, out of the $\binom{n}{2d}$ possibilities (the coefficients of each standard polynomial is 0-1), which amounts to $2^{\binom{n}{2d}}$ possibilities. This is powered by n because we need to choose n such P_i's. We thus get $2^{n\binom{n}{2d}}$.

Second, given a natural number ℓ, we count the total number of n-tuples of s-polynomials that can be generated with ℓ substitution instances of degree-$2d$ standard polynomials. By Lemma 10, we can assume without loss of generality that all the generators are standard polynomials of degree $2d$ in which we substitute variables by *homogeneous* linear forms with n variables. Thus, for every $i \in [n]$,

$$P_i = \sum_{j=1}^{\ell} c_{ij} s_{2d}(l_1, \ldots, l_{2d}), \qquad \text{for linear homogeneous forms } l_j\text{'s, and } c_{ij}\text{'s in } \mathbb{F}.$$

Then, the total number of different possible such n-tuples $P_1, \ldots, P_n$ is the total number of choices of scalars c_{ij}, for $i \in [n], j \in [\ell]$, and additionally the total number of choices of ℓ tuples $l_1, \ldots, l_{2d}$ of homogeneous linear forms. Each l_i is an n-variate homogeneous linear form so we have to pick n scalars for it. Altogether we have $2dn\ell + n\ell = (2d+1)n\ell$ scalar choices to make, namely we have $|\mathbb{F}|^{(2d+1)n\ell}$ possibilities. Assuming $|\mathbb{F}|$ is finite and constant, we get

$$2^{n\binom{n}{2d}} \leq |\mathbb{F}|^{(2d+1)n\ell},$$

implying that $\ell = \Omega(n^{2d})$. Using a lemma of Hrubeš-Yehudayoff [HY11] (Lemma 5 below) we show that the same argument holds for infinite fields.

Formal proof. Recall that an s-polynomial (Definition 8) is of the following form:

$$\sum_{j_1 < j_2 < \ldots < j_{2d} \in [n]} c_{j_1 j_2 \cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}), \quad \text{where } c_{j_1 j_2 \cdots j_{2d}} \in \{0, 1\}\,.$$

Assume that

$$\ell = \max\left\{Q_{S_{2d}}(P_1, \ldots, P_n) \ : \ P_i \text{ is an s-polynomial, for all } i \in [n]\right\}.$$

Then for every choice of n s-polynomials $P_1, \ldots, P_n$ there are ℓ vectors of polynomials $\overline{Q_1}, \ldots, \overline{Q_\ell}$ (defining the substitution instances of generators) from $\mathbb{F}\langle X\rangle$, such that

$$P_1, \ldots, P_n \in \left\langle S_{2d}(\overline{Q_1}), \ldots, S_{2d}(\overline{Q_\ell})\right\rangle.$$

By Lemma 10, for every $i \in [n]$,

$$P_i = \sum_{u=1}^{\ell} c_{iu} S_{2d}\left(\overline{Q_u}^{(1)}\right) = \sum_{u=1}^{\ell} c_{iu} S_{2d}\left(\sum_{j=1}^{n} a_{u1j}x_j, \sum_{j=1}^{n} a_{u2j}x_j, \ldots, \sum_{j=1}^{n} a_{u(2d)j}x_j\right),$$
$$\text{for some } c_{iu}, a_{ukj} \in \mathbb{F}, \text{ for } u \in [\ell], k \in [2d], j \in [n].$$

We will consider the scalars in the equation above (over all $i \in [n]$) as vectors of the following form:

$$\left(c_{11}, c_{12}, \ldots, c_{n\ell}, a_{111}, a_{112}, \ldots, a_{\ell(2d)(n-1)}, a_{\ell(2d)n}\right). \tag{4}$$

By linearity of S_{2d}, for all $i \in [n]$,

$$\sum_{u=1}^{\ell} c_{iu} S_{2d}\left(\sum_{j=1}^{n} a_{u1j}x_j, \sum_{j=1}^{n} a_{u2j}x_j, \ldots, \sum_{j=1}^{n} a_{u(2d)j}x_j\right) = \sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \gamma_{ij_1j_2\cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}), \qquad \text{for some } \gamma_{ij_1j_2\cdots j_{2d}}\text{'s in } \mathbb{F}. \quad (5)$$

A *polynomial map* $\mu : \mathbb{F}^s \to \mathbb{F}^m$ of degree $r > 0$ is a map $\mu = (\mu_1, \ldots, \mu_m)$, where each μ_i is a (commutative) multivariate polynomial of degree r with s variables.

Claim. *Equation* (5) *defines a degree-*$(2d+1)$ *polynomial map* $\phi : \mathbb{F}^{(2d+1)n\ell} \to \mathbb{F}^{n\binom{n}{2d}}$ *that maps each vector* (4) *to a vector*

$$(\gamma_{ij_1j_2\cdots j_{2d}} \; : \; j_1 < j_2 < \ldots < j_{2d} \in [n], \; i \in [n]) \; .$$

Proof of claim: Since the standard identity S_{2d} is a multilinear function, for each $u \in [\ell]$,

$$S_{2d}\left(\sum_{j_1\in[n]} a_{u1j_1}x_{j_1}, \ldots, \sum_{j_{2d}\in[n]} a_{u(2d)j_{2d}}x_{j_{2d}}\right) = \sum_{j_1,j_2,\ldots,j_{2d}\in[n]} a_{u1j_1}\cdots a_{u(2d)j_{2d}} S_{2d}(x_{j_1}, \ldots, x_{j_{2d}}).$$

Then, for all $i \in [n]$,

$$\begin{aligned}\sum_{u=1}^{\ell} c_{iu} S_{2d}\left(\sum_{j=1}^{n} a_{u1j}x_j, \ldots, \sum_{j=1}^{n} a_{u(2d)j}x_j\right) &= \sum_{u\in[\ell],j_1,j_2,\ldots,j_{2d}\in[n]} c_{iu}a_{u1j_1}\cdots a_{u(2d)j_{2d}} S_{2d}(x_{j_1}, \ldots, x_{j_{2d}}) \\ &= \sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \gamma_{ij_1j_2\cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}) \,.\end{aligned}$$

Therefore, for every $j_1 < j_2 < \ldots < j_{2d} \in [n], i \in [n]$, the coefficient $\gamma_{ij_1j_2\cdots j_{2d}}$ is a linear combination of the $(2d+1)$-degree terms $c_{iu}a_{u1k_1}\cdots a_{u(2d)k_{2d}}$, for all $u \in [\ell]$ and all $\{k_1, \ldots, k_{2d}\} = \{j_1, \ldots, j_{2d}\}$. ■Claim

We have the following lemma by Hrubeš and Yehudayoff [HY11]:

Lemma 12 ([HY11], Lemma 5)**.** *Let* $\mathbb{F}$ *be a field. If* $\mu : \mathbb{F}^s \to \mathbb{F}^m$ *is a polynomial map of degree* $r > 0$*, then* $|\mu(\mathbb{F}^s)\bigcap\{0,1\}^m| \le (2r)^s$*.*

Using Lemma 12, for the degree-$(2d+1)$ polynomial map $\phi : \mathbb{F}^{(2d+1)n\ell} \to \mathbb{F}^{n\binom{n}{2d}}$, we have

$$\left|\phi(\mathbb{F}^{(2d+1)n\ell})\bigcap\{0,1\}^{n\binom{n}{2d}}\right| \le (2(2d+1))^{(2d+1)n\ell} \,.$$

Denote by $\overline{\gamma}$ a 0-1 vector $(\gamma_{1j_1j_2\cdots j_{2d}}, \ldots, \gamma_{nj_1j_2\cdots j_{2d}})$, where $\gamma_{ij_1j_2\cdots j_{2d}} \in \{0,1\}, j_1 < j_2 < \ldots < j_{2d} \in [n], i \in [n]$. Since for every possible $\overline{\gamma}$, the following polynomials are s-polynomials:

$$\sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \gamma_{1j_1j_2\cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}), \quad \ldots, \quad \sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \gamma_{nj_1j_2\cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}),$$

there exist ℓ vectors of polynomials $\overline{Q_1}, \ldots, \overline{Q_\ell}$ in $\mathbb{F}\langle X\rangle$, such that

$$\sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \gamma_{ij_1j_2\cdots j_{2d}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}) \in \left\langle S_{2d}(\overline{Q_1}), \ldots, S_{2d}(\overline{Q_\ell})\right\rangle, i \in [n].$$

That is, there exists a vector $\mathbf{v} = \left(c_{11}, c_{12}, \ldots, c_{n\ell}, a_{111}, a_{112}, \ldots, a_{\ell(2d)(n-1)}, a_{\ell(2d)n}\right)$, such that $\phi(\mathbf{v}) = \overline{\gamma}$. Hence, every possible $\overline{\gamma}$ belongs to $\phi(\mathbb{F}^{(2d+1)nl}) \bigcap \{0,1\}^{n\binom{n}{2d}}$. Further, there are $2^{n\binom{n}{2d}}$ distinct vectors $\overline{\gamma}$. Therefore,

$$\left|\phi(\mathbb{F}^{(2d+1)nl}) \bigcap \{0,1\}^{n\binom{n}{2d}}\right| \geq 2^{n\binom{n}{2d}}.$$

This implies by Lemma 12, that

$$(2(2d+1))^{(2d+1)nl} \geq 2^{n\binom{n}{2d}}.$$

Using the ln function on both sides we have

$$(2d+1)nl \ln(2(2d+1)) \geq n\binom{n}{2d} \ln 2.$$

Hence,

$$l > \frac{\binom{n}{2d} \ln 2}{(2d+1)\ln(4d+2)}.$$

Namely,

$$l > c\binom{n}{2d} = c\frac{n(n-1)\cdots(n-2d+1)}{(2d)!} = \Omega\left(n^{2d}\right),$$

(for c a constant independent of n). QED

5.1.2 Combining the Polynomials into One

Here we conclude the proof of Lemma 7. That is, we show that there exists a *single* polynomial, denoted $P^\star$, such that $Q_{S_{2d}}(P^\star) = \Omega(n^{2d})$. This is done in a manner resembling [Hru11]; however, there is a further complication that is dealt with in Lemma 14 below.

Let $P_1, \ldots, P_n$ be s-polynomials in n variables $x_1, \ldots, x_n$, and let $z_1, \ldots, z_n$ be new variables, different from $x_1, \ldots, x_n$. We put

$$P^\star := \sum_{i=1}^{n} z_i P_i.$$

For convenience, we call the new variables $z_1, \ldots, z_n$ the Z-variables. Given a polynomial f, the ***Z-homogeneous part of degree j of f***, denoted $(f)_Z^{(j)}$, is the sum of all monomials where the total degree of the Z-variables is j. For example, if $f = z_1xy + z_2z_1 + z_3x + 1 + x$, then $(f)_Z^{(1)} = z_1xy + z_3x$, $(f)_Z^{(2)} = z_2z_1$, $(f)_Z^{(0)} = 1 + x$. A polynomial that does not contain any Z-variable is said to be *Z-free*.

First, we claim that $P^\star$ has the following property:

Lemma 13. *For every ℓ Z-free polynomials $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_\ell \in \mathbb{F}\langle X\rangle$, if*

$$P^\star \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\right\rangle,$$

then

$$P_1, \ldots, P_n \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\right\rangle.$$

Proof. Since $P^\star \in \langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\rangle$,

$$P^\star = \sum_{i=1}^{n} z_i P_i = \sum_{j=1}^{\ell} \sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{G}_j) g_{ji}\,,$$

for some $f_{ji}, g_{ji} \in \mathbb{F}\langle X, Z\rangle$ and some t_j's.

Note that we cannot assume that $t_j \leq 1$, because of non-commutativity: for instance, it might happen that we have two terms like $fAg + f'Ag'$ that we cannot join into a single term uAv (for some u, v).

Now, assign $z_1 = 1, z_2 = z_3 = \cdots = z_n = 0$ in $P^\star$. Since $\overline{G}_1, \ldots, \overline{G}_\ell$ do not contain $z_1, \ldots, z_n$, the $\overline{G}_1, \ldots, \overline{G}_\ell$ will remain the same. Thus,

$$P_1 = \sum_{j=1}^{\ell} \sum_{i=1}^{t_j} f'_{ji} S_{2d}(\overline{G}_j) g'_{ji}\,,$$

where $f'_{ji} = f_{ji}|_{z_1 \leftarrow 1, z_2 \leftarrow 0, \ldots, z_n \leftarrow 0}$ and $g'_{ji} = g_{ji}|_{z_1 \leftarrow 1, z_2 \leftarrow 0, \ldots, z_n \leftarrow 0}$. That is, $P_1 \in \langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\rangle$.

Similarly, we can show $P_2, \ldots, P_n \in \langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\rangle$. Therefore, $P_1, \ldots, P_n \in \langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell)\rangle$. QED

We define

$$[\![\cdot]\!] : \mathbb{F}\langle X, Z\rangle \to \mathbb{F}\langle X, Z\rangle$$

to be the map determined by the following three properties:

1. The map $[\![\cdot]\!]$ is linear, namely $[\![\alpha G + \beta H]\!] = \alpha\,[\![G]\!] + \beta\,[\![H]\!]$ for all polynomials G, H and $\alpha, \beta \in \mathbb{F}$.
2. Let M be a monomial whose Z-homogeneous part is of degree 1. Thus, M can be uniquely written as $M_1 z_i M_2$, $z_i \in Z$, where M_1, M_2 are Z-free. Then,
$$[\![M]\!] = [\![M_1 z M_2]\!] = z M_2 M_1\,.$$
3. For a monomial M whose Z-homogeneous part is not of degree 1, $[\![M]\!] = 0$.

For convenience, in what follows, given the polynomials f_i, g_i and the vector of polynomials $\overline{H}$, we denote $(f_i)^{(0)}_Z, \left(\overline{H}\right)^{(0)}_Z, (g_i)^{(0)}_Z$ by $\mathcal{F}, \overline{\mathcal{H}}, \mathcal{G}$, respectively, where $\left(\overline{H}\right)^{(0)}_Z$ is the result of applying $(\cdot)^{(0)}_Z$ on $\overline{H}$ coordinate-wise. Note that $(f_i)^{(0)}_Z, (g_i)^{(0)}_Z$ and $\left(\overline{H}\right)^{(0)}_Z$ are Z-free polynomials (vectors of polynomials, resp.).

We need to prove the following lemma before concluding Lemma 7.

Lemma 14. *Let $X = \{x_1, \ldots, x_n\}$ and $f_1, g_1, \ldots, f_k, g_k \in \mathbb{F}\langle X\rangle$. Let $Z = \{z_1, \ldots, z_n\}$ and assume that n is an even positive integer, and let $\overline{P}$ be a vector of polynomials $(P_1, \ldots, P_n)$ over the variable set $X \cup Z$. We denote $\left(\overline{P}\right)^{(0)}_Z$, $(f_i)^{(0)}_Z, (g_i)^{(0)}_Z$ by $\overline{\mathcal{P}}, \mathcal{F}_i, \mathcal{G}_i,$, respectively, for $i \in [k]$. Then, for every $\delta \in [n]$, it holds that*

$$\left[\!\!\left[\sum_{i=1}^{k} \mathcal{F}_i S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow (P_\delta)^{(1)}_Z}\right) \mathcal{G}_i\right]\!\!\right] \in \left\langle S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k} \mathcal{G}_i \mathcal{F}_i}\right)\right\rangle. \tag{6}$$

For example, when $n = 2$, this lemma shows the following:

$$\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_2\left((P_1)_Z^{(1)}, \mathcal{P}_2\right)\mathcal{G}_i \right\rrbracket \in \left\langle S_2\left(\sum_{i=1}^{k} \mathcal{G}_i\mathcal{F}_i, P_2\right)\right\rangle,$$

$$\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_2\left(\mathcal{P}_1, (P_2)_Z^{(1)}\right)\mathcal{G}_i \right\rrbracket \in \left\langle S_2\left(P_1, \sum_{i=1}^{k} \mathcal{G}_i\mathcal{F}_i\right)\right\rangle.$$

Proof. Notice that, for all $\delta \in [n]$, we have $(P_\delta)_Z^{(1)} = \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}$, where $\mathcal{U}_{tw}, \mathcal{V}_{tw} \in \mathbb{F}\langle X\rangle$ and $\mathcal{U}_{tw}, \mathcal{V}_{tw}$ are Z-free. Then, it suffices to prove that for all $\delta \in [n]$

$$\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\right)\mathcal{G}_i \right\rrbracket = -\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw} S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\mathcal{U}_{tw}. \quad (7)$$

This is because, $\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow (P_\delta)_Z^{(1)}}\right)\mathcal{G}_i \right\rrbracket = \left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\right)\mathcal{G}_i \right\rrbracket$ and $-\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw} S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\mathcal{U}_{tw} \in \left\langle S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\right\rangle$, and hence we have (6), which is the desired result.

To prove (7), it is sufficient to expand $\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_n(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}})\mathcal{G}_i \right\rrbracket$ transforming it to $-\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw} S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\mathcal{U}_{tw}$.

For the sake of convenience we let

$$\overline{P}_{\sigma[i,j]} = \begin{cases} \prod_{m=i}^{j} P_{\sigma(m)}, & i \le j; \\ 1, & i > j \end{cases},$$

where $\sigma \in \mathfrak{S}_n$, and $\mathfrak{S}_n$ is the permutation group of order n, and $\overline{P} = (P_1, \ldots, P_n)$ is a vector of polynomials. Then, we have $S_n(\overline{P}) = \sum_{\sigma\in\mathfrak{S}_n} sgn(\sigma)(\overline{P}_{\sigma[1,n]})$. Furthermore, we use $\mathfrak{S}_n/m_\delta$ to denote the set $\{\sigma \in \mathfrak{S}_n \mid \sigma(m) = \delta\}$. With the above notation, we have the following expansion

$$\begin{aligned}
&\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_n\left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\right)\mathcal{G}_i \right\rrbracket \\
=&\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i \sum_{\sigma\in\mathfrak{S}_n} sgn(\sigma)\left(\overline{\mathcal{P}}_{\sigma[1,n]}\right)|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\mathcal{G}_i \right\rrbracket \\
=&\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i \sum_{m=1}^{n} \sum_{\substack{\sigma\in\mathfrak{S}_n \\ \sigma^{-1}(\delta)=m}} sgn(\sigma)\left(\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{P}_{\sigma(m)}\overline{\mathcal{P}}_{\sigma[m+1,n]}\right)|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\mathcal{G}_i \right\rrbracket \\
=&\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i \sum_{m=1}^{n} \sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\left(\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{P}_{\delta}\overline{\mathcal{P}}_{\sigma[m+1,n]}\right)|_{\mathcal{P}_\delta \leftarrow \sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}}\mathcal{G}_i \right\rrbracket \\
=&\left\llbracket \sum_{i=1}^{k} \mathcal{F}_i \sum_{m=1}^{n} \sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\left(\overline{\mathcal{P}}_{\sigma[1,m-1]}\sum_{t=1}^{n}\sum_{w}\mathcal{U}_{tw}z_t\mathcal{V}_{tw}\overline{\mathcal{P}}_{\sigma[m+1,n]}\right)\mathcal{G}_i \right\rrbracket \\
=&\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw} \sum_{m=1}^{n} \sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\overline{\mathcal{P}}_{\sigma[m+1,n]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{U}_{tw}\,.
\end{aligned}$$

In the following, we proceed to transform the above formula to $-\sum_{t=1}^{n}\sum_j z_t\mathcal{V}_{tw}S_n(\overline{\mathcal{P}}|_{\mathcal{P}_\delta\leftarrow\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i})\mathcal{U}_{tw}$, which concludes the proof. That is, we need to prove

$$\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw}\left(\sum_{m=1}^{n}\sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\overline{\mathcal{P}}_{\sigma[m+1,n]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\sigma[1,m-1]}\right)\mathcal{U}_{tw} =$$
$$-\sum_{t=1}^{n}\sum_{w} z_t\mathcal{V}_{tw}S_n(\overline{\mathcal{P}}|_{\mathcal{P}_\delta\leftarrow\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i})\mathcal{U}_{tw}.$$

And therefore, it suffices to prove

$$\sum_{m=1}^{n}\sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\overline{\mathcal{P}}_{\sigma[m+1,n]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\sigma[1,m-1]} = -S_n(\overline{\mathcal{P}}|_{\mathcal{P}_\delta\leftarrow\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}).$$

For $m\in[n]$, consider the permutation π_m defined as:

$$\begin{pmatrix} 1 & 2 & \dots & n-m & n-m+1 & n-m+2 & \dots & n \\ m+1 & m+2 & \dots & n & m & 1 & \dots & m-1 \end{pmatrix}.$$

Note that, for π_m, we have the following facts:

Fact 15. *For every permutation $\pi\in\mathfrak{S}_n$, where n is an even integer, $sgn(\pi\pi_m^{-1}) = sgn(\pi)sgn(\pi_m) = -sgn(\pi)$.*

Fact 16. $\overline{P}_{\sigma[m+1,n]}\cdot\overline{P}_{\sigma[1,m-1]} = \overline{P}_{\sigma\pi_m[1,n-m]}\cdot\overline{P}_{\sigma\pi_m[n-m+2,n]}$, *for all $\sigma\in\mathfrak{S}_n/m_\delta$.*

Fact 17. $(\mathfrak{S}_n/m_\delta)\pi_m = \mathfrak{S}_n/(n-m+1)_\delta$.

We now have the following

$$\sum_{m=1}^{n}\sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\overline{\mathcal{P}}_{\sigma[m+1,n]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\sigma[1,m-1]}$$
$$=\sum_{m=1}^{n}\sum_{\sigma\in\mathfrak{S}_n/m_\delta} sgn(\sigma)\overline{\mathcal{P}}_{\sigma\pi_m[1,n-m]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\sigma\pi_m[n-m+2,n]} \qquad \text{by } \textbf{Fact } 16$$
$$\text{letting } \pi'=\sigma\pi_m, \text{ then } \pi'\in(\mathfrak{S}_n/m_\delta)\pi_m, \text{ and } \sigma=\pi'\pi_m^{-1},$$
$$=\sum_{m=1}^{n}\sum_{\pi'\in(\mathfrak{S}_n/m_\delta)\pi_m} sgn(\pi'\pi_m^{-1})\overline{\mathcal{P}}_{\pi'[1,n-m]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\pi'[n-m+2,n]}$$
$$=\sum_{m=1}^{n}\sum_{\pi'\in(\mathfrak{S}_n/m_\delta)\pi_m} (-sgn(\pi'))\overline{\mathcal{P}}_{\pi'[1,n-m]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\pi'[n-m+2,n]} \qquad \text{by } \textbf{Fact } 15$$
$$=-\sum_{m=1}^{n}\sum_{\pi'\in\mathfrak{S}_n/(n-m+1)_\delta} sgn(\pi')\overline{\mathcal{P}}_{\pi'[1,n-m]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\pi'[n-m+2,n]}, \qquad \text{by } \textbf{Fact } 17$$
$$\text{letting } m'=n-m+1, \text{ then } n-m=m'-1 \text{ and } n-m+2=m'+1,$$
$$=-\sum_{m'=1}^{n}\sum_{\pi'\in\mathfrak{S}_n/m'_\delta} sgn(\pi')\overline{\mathcal{P}}_{\pi'[1,m'-1]}\left(\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i\right)\overline{\mathcal{P}}_{\pi'[m'+1,n]}$$

$$= - \; S_n \left(\overline{\mathcal{P}}|_{\mathcal{P}_\delta \leftarrow \sum_{i=1}^{k} \mathcal{G}_i \mathcal{F}_i} \right).$$

QED

The following lemma suffices to conclude Lemma 7.

Lemma 18. *For every field $\mathbb{F}$ of characteristic zero and every $d \geq 1$, there exists a polynomial with n variables such that $Q_{S_{2d}}(P^\star) = \Omega(n^{2d})$. Specifically:*

$$Q_{S_{2d}}(P^\star) \geq \frac{1}{2d+1} Q_{S_{2d}}(P_1, \ldots, P_n). \tag{8}$$

Proof. Assume $Q_{S_{2d}}(P^\star) = \ell$. That is, there are k vectors of polynomials $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_\ell$ such that

$$P^\star \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_\ell) \right\rangle.$$

Or in other words

$$P^\star = \sum_{i=1}^{n} z_i P_i = \sum_{j=1}^{\ell} \sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{G}_j) g_{ji}, \qquad \text{for some } f_{ji}, g_{ji} \in \mathbb{F}\langle X, Z \rangle \text{ and some } t_j\text{'s.}$$

If we can find $(2d+1) \cdot \ell$ Z-free vectors of polynomials $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_{(2d+1)\cdot\ell}$ such that

$$P^\star \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_{(2d+1)\cdot\ell}) \right\rangle,$$

then, by Lemma 13

$$P_1, \ldots, P_n \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_{(2d+1)\cdot\ell}) \right\rangle,$$

which is the conclusion we want to prove, that is $Q_{S_{2d}}(P_1, \ldots, P_n) \leq (2d+1) \cdot \ell$.

Claim. *For every sequence of polynomials $f_1, g_1, \ldots, f_k, g_k$ and vector of polynomials $\overline{H}$, with variables $x_1, \ldots, x_n, z_1, \ldots, z_n$:*

$$\left\llbracket \sum_{i=1}^{k} f_i S_{2d}(\overline{H}) g_i \right\rrbracket \in \left\langle S_{2d}(\overline{\mathcal{H}}),\ S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_1 \leftarrow \sum_{i=1}^{k} \mathcal{G}_i \mathcal{F}_i}\right), \ldots,\ \left(\overline{\mathcal{H}}|_{\mathcal{H}_{2d} \leftarrow \sum_{i=1}^{k} \mathcal{G}_i \mathcal{F}_i}\right) \right\rangle.$$

Proof of claim: Consider the following:

$$\left\llbracket \sum_{i=1}^{k} f_i S_{2d}(\overline{H}) g_i \right\rrbracket = \left\llbracket \left(\sum_{i=1}^{k} f_i S_{2d}(\overline{H}) g_i \right)_Z^{(1)} \right\rrbracket \qquad \text{(by Property 3 of } [\cdot])$$

$$= \left\llbracket \sum_{i=1}^{k} (f_i)_Z^{(1)} S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i + \sum_{i=1}^{k} \sum_{j=1}^{2d} \mathcal{F}_i S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow (H_j)_Z^{(1)}}\right) \mathcal{G}_i + \sum_{i=1}^{k} \mathcal{F}_i S_{2d}(\overline{\mathcal{H}}) (g_i)_Z^{(1)} \right\rrbracket$$

$$\text{(by linearity of } \llbracket\cdot\rrbracket) \qquad = \sum_{i=1}^{k} \left\llbracket (f_i)_Z^{(1)} S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i \right\rrbracket + \sum_{j=1}^{2d} \left\llbracket \sum_{i=1}^{k} \mathcal{F}_i S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow (H_j)_Z^{(1)}}\right) \mathcal{G}_i \right\rrbracket +$$

$$\sum_{i=1}^{k} \left\llbracket \mathcal{F}_i S_{2d}(\overline{\mathcal{H}}) (g_i)_Z^{(1)} \right\rrbracket.$$

For every $i \in [k]$, assume $(f_i)_Z^{(1)} = \sum_{r=1}^{n} \sum_j g_{rj} z_r h_{rj}$ where g_{rj}, h_{rj} are Z-free polynomials (and $z_1, \ldots, z_n$ are the Z-variables), then

$$\left[\!\left[(f_i)_Z^{(1)} S_{2d}(\overline{\mathcal{H}})\mathcal{G}_i\right]\!\right] = \left[\!\left[\sum_{r=1}^{n}\sum_j g_{rj} z_r h_{rj} S_{2d}(\overline{\mathcal{H}})\mathcal{G}_r\right]\!\right] = \sum_{r=1}^{n}\sum_j z_r h_{rj} S_{2d}(\overline{\mathcal{H}})\mathcal{G}_r g_{rj} \in \left\langle S_{2d}(\overline{\mathcal{H}})\right\rangle ,$$

where the right most equality stems from Property 2 of $[\![\cdot]\!]$. Similarly, for every $i \in [k]$, we can show

$$\left[\!\left[\mathcal{F}_i S_{2d}(\overline{\mathcal{H}})(g_i)_Z^{(1)}\right]\!\right] \in \left\langle S_{2d}(\overline{\mathcal{H}})\right\rangle .$$

By Lemma 14,

$$\left[\!\left[\sum_{i=1}^{k} \mathcal{F}_i S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow (H_j)_Z^{(1)}}\right)\mathcal{G}_i\right]\!\right] \in \left\langle S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\right\rangle, \qquad \text{for any } j \in [2d].$$

Thus, $\left[\!\left[\sum_{i=1}^{k} f_i S_{2d}\left(\overline{H}\right) g_i\right]\!\right] \in \left\langle S_{2d}\left(\overline{\mathcal{H}}\right),\ S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_1 \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right), \ldots,\ \left(\overline{\mathcal{H}}|_{\mathcal{H}_{2d} \leftarrow \sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i}\right)\right\rangle$.
■Claim

Note that $P^\star = (P^\star)_Z^{(1)}$. By the properties of $[\![\cdot]\!]$ we have:

$$\begin{aligned} P^\star &= [\![P^\star]\!] \\ &= \left[\!\left[\sum_{j=1}^{\ell}\sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{H}_j) g_{ji}\right]\!\right] \\ &= \sum_{j=1}^{\ell}\left[\!\left[\sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{H}_j) g_{ji}\right]\!\right] \\ &\in \left\langle S_{2d}\left(\overline{\mathcal{H}}\right), S_{2d}\left(\overline{\mathcal{H}}_j|_{H_{jq} \leftarrow \sum_{m=1}^{t_j}\mathcal{G}_{jm}\mathcal{F}_{jm}}\right) \ :\ j \in [\ell],\ q \in [2d]\right\rangle . \end{aligned}$$

That is, for $P^\star = \sum_{j=1}^{\ell}\sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{H}_j) g_{ji}$, we have $(2d+1)\cdot\ell$ Z-free polynomials that generate $P^\star$, concluding the proof of Lemma 18. QED

5.2 Concluding the Lower Bound for Every Basis

Here we show that the $\Omega(n^{2d})$ lower bound proved in previous sections (Lemma 7) holds for (every $d > 2$ and) *every finite basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$, when $\mathbb{F}$ is of characteristic zero. To this end, we use several results from the theory of PI-algebras (for more on PI-theory see the monographs [Row80, Dre99]).

A polynomial $f \in \mathbb{F}\langle X\rangle$ with d variables is *multi-homogeneous with degrees* $(1, \ldots, 1)$ (d times) if in every monomial the power of every variable $x_1, \ldots, x_d$ is precisely 1. In other words, every monomial is of the form $\prod_{i=1}^{d} x_{\sigma(i)}$, for some permutation σ of order d. For the sake of simplicity, we will talk in the sequel about a ***multi-homogeneous polynomial of degree*** d, when referring to a multi-homogeneous polynomial with degrees $(1, \ldots, 1)$ (d times). Thus, every multi-homogeneous polynomial with d variables is homogeneous of total-degree d.

For $n \geq 2$ polynomials $f_1, \ldots, f_n$, define the ***generalized-commutator*** $[f_1, \ldots, f_n]$ as follows:

$$[f_1, f_2] := f_1 f_2 - f_2 f_1, \qquad \text{(in case } n = 2\text{)}$$
$$\text{and} \quad [f_1, \ldots, f_{n-1}, f_n] := [[f_1, \ldots, f_{n-1}], f_n], \qquad \text{for } n > 2.$$

Definition 9. *A polynomial $f \in \mathbb{F}\langle X\rangle$ is called a* ***commutator polynomial*** *if it is a linear combination of products of generalized-commutators. (We assume that 1 is a product of an empty set of commutator polynomials.)*

For example, $[x_1, x_2] \cdot [x_3, x_4] + [x_1, x_2, x_3]$ is a commutator polynomial.

We say that a PI-algebra is *unitary* if the product operation of the PI-algebra has a unit (e.g., the identity matrix, for matrix PI-algebras).

Proposition 19 ([Dre99, Proposition 4.3.3])**.** *If R is a unitary PI-algebra over a field $\mathbb{F}$ of characteristic zero, then every identity of R can be generated by multi-homogeneous commutator polynomials.*[8]

Corollary 20. *Let R be a unitary PI-algebra and let $\mathcal{T}$ be the T-ideal consisting of all identities of R. Then $\mathcal{T}$ has a finite basis $\mathcal{B}_0$ in which every polynomial is a multi-homogeneous commutator polynomial. Moreover, for every finite basis $\mathcal{B}_1$ of $\mathcal{T}$, there are constants c_1, c_2 such that for every identity f of R, $c_2 Q_{\mathcal{B}_1}(f) \le Q_{\mathcal{B}_0}(f) \le c_1 Q_{\mathcal{B}_1}(f)$.*

Proof. By Kemer [Kem87], for every field $\mathbb{F}$, the identities of every $\mathbb{F}$-algebra has a finite basis. Assume $\mathcal{C}$ is some finite basis guaranteed to exist. Then, by Proposition 19, each polynomial in $\mathcal{C}$ is generated by (constant) many multi-homogeneous commutator polynomials. Thus, there is a finite basis $\mathcal{B}_0$ of multi-homogeneous commutator polynomials that generate the basis of $\mathcal{T}$, meaning that $\mathcal{B}_0$ itself is a basis of $\mathcal{T}$.

By the robustness of $Q(\cdot)$ (Propositional 3), for every finite basis $\mathcal{B}_1$ of $\mathcal{T}$ and every identity f of R, there are constants c_1, c_2 depending on $\mathcal{B}_0, \mathcal{B}_1$ alone, such that $c_2 Q_{\mathcal{B}_1}(f) \le Q_{\mathcal{B}_0}(f) \le c_1 Q_{\mathcal{B}_1}(f)$.

QED

Lemma 21. *Let $f \in \mathbb{F}\langle X\rangle$ be a multi-homogeneous commutator polynomial with n variables. If x_δ is a constant for some $\delta \in [n]$, then $f(x_1, \dots, x_n) \equiv 0$ (that is, f is the zero polynomial).*

Proof. It is easy to check that if we replace a variable by a constant $c \in \mathbb{F}$ in a generalized-commutator, then the generalized-commutator becomes 0.

By the definition of a commutator polynomial,

$$f = \sum_{i=1}^{m} c_i \prod_{j=1}^{k_i} B_{ij},$$

where $c_i \in \mathbb{F}$ and $m, n \in \mathbb{N}$, and the B_{ij}'s are generalized-commutators. Since f is a multi-homogeneous polynomial, the variable x_δ occurs in every term $\prod_{j=1}^{k_i} B_{ij}$ in f (i.e., for every $i \in [m]$). Hence, for every $i \in [m]$, x_δ must occur in some B_{ij} (for some $j \in [k_i]$). But B_{ij} is a generalized-commutator, and since x_δ is constant, $B_{ij} = 0$. Therefore, every term $\prod_{j=1}^{k_i} B_{ij}$ in f is 0.

QED

By lemma 11 and lemma 18, we know that there exist s-polynomials $P_1, \dots, P_n$ in n variables $x_1, \dots, x_n$ that are identities of $\mathrm{Mat}_d(\mathbb{F})$, such that putting $P^\star {:=} \sum_{i=1}^{n} z_i P_i$, where $z_1, \dots, z_n$ are new variables, we have:

$$Q_{S_{2d}}(P^\star) \ge \frac{1}{2d+1} \cdot Q_{S_{2d}}(P_1, \dots, P_n) = \Omega(n^{2d}).$$

The following is the main lemma of this section:

[8] *Multi-homogeneous* and *commutator polynomials*, are called *multilinear* and *proper polynomials*, respectively, in [Dre99].

Lemma 22. *Let $d > 2$, and let $\mathcal{B}$ be a finite basis for the T-ideals of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*. Then, there are constants c, c' such that for every identity P over* $\mathrm{Mat}_d(\mathbb{F})$ *of degree $2d+1$:*

$$cQ_{S_{2d}}(P) \leq Q_{\mathcal{B}}(P) \leq c'Q_{S_{2d}}(P).$$

To prove this lemma we need the following two lemmas.

Lemma 23. *Let $d > 2$ be a natural number. Every multi-homogeneous identity (with any number of variables) of* $\mathrm{Mat}_d(\mathbb{F})$ *of degree at most $2d+1$ is a consequence of the standard identity S_{2d}.*

Proof. By Leron [Ler73], we know that for every $d > 2$, every multi-homogeneous identity of $\mathrm{Mat}_d(\mathbb{F})$ with degree exactly $2d+1$ is a consequence of the standard identity S_{2d}. By Drensky [Dre99, Exercise 7.1.2], there are no identities of degree less than $2d$ in $\mathrm{Mat}_d(\mathbb{F})$ and every multi-homogeneous polynomial identity of degree $2d$ in $\mathrm{Mat}_d(\mathbb{F})$ is also a consequence of the standard identity S_{2d}. QED

By Corollary 20, there is a basis $\{A_1, \dots, A_m\}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$, where $A_1, \dots, A_m$ are all multi-homogeneous commutator polynomials (Definition 9).

Lemma 24. *Let $P \in \mathbb{F}\langle X \rangle$ be an identity of* $\mathrm{Mat}_d(\mathbb{F})$ *of degree $2d+1$ and let G be a basis $\{A_1, \dots, A_m\}$ of* $\mathrm{Mat}_d(\mathbb{F})$*, where $A_1, \dots, A_m$ are all multi-homogeneous commutator identities of* $\mathrm{Mat}_d(\mathbb{F})$*. Assume that $Q_G(P) = k$, that is, k is the minimal number such that there exist k substitution instances $B_1, \dots, B_k$ of $A_1, \dots, A_m$, for which:*

$$P \in \langle B_1, \dots, B_k \rangle .$$

Then, no B_ℓ, for $\ell \in [k]$, is a substitution instance of a basis element A_j with the degree of A_j greater than $2d+1$.

Proof. Assume there exists an A_j (for $j \in [m]$) in G with degree greater than $2d+1$. We show that none of B_ℓ ($\ell \in [k]$) is a substitution instance of A_j.

Suppose otherwise, that is, suppose that there is a B_δ, $\delta \in [k]$, such that B_δ is the substitution instance $A_j(\overline{Q})$, for some $\overline{Q}$. Since A_j is homogeneous, every monomial in A_j is of degree greater than $2d+1$. We consider the following two cases:

Case 1: Every monomial in $A_j(\overline{Q})$ is of degree greater than $2d+1$.

For convenience, given a polynomial f, we denote by $f^{\leq j}$ the polynomial $\sum_{i=0}^{j} (f)^{(i)}$, namely the sum of all homogeneous parts of f of degree at most j. We consider the $2d+1$ homogeneous part, that is:

$$\begin{aligned} P &= (P)^{(2d+1)} \\ &\in \left\langle (h)^{(2d+1)} \;\middle|\; h \in \langle B_1, \dots, B_k \rangle \right\rangle \subseteq \left\langle (B_1)^{(\leq 2d+1)}, \dots, (B_k)^{(\leq 2d+1)} \right\rangle . \end{aligned}$$

But $(B_\delta)^{(\leq 2d+1)} = \left(A_j(\overline{Q})\right)^{(\leq 2d+1)} = 0$, because by assumption every monomial in $A_j(\overline{Q})$ is of degree greater than $2d+1$. So P belongs to the ideal generated by $\left\{(B_1)^{(\leq 2d+1)}, \dots, (B_k)^{(\leq 2d+1)}\right\} \setminus (B_\delta)^{(\leq 2d+1)}$. This means $Q_G(P) = k-1$, which contradicts $Q_G(P) = k$. Thus, the assumption is false.

Case 2: There is a monomial of degree at most $2d+1$ in $A_j(\overline{Q})$.

But since $A_j(\overline{x})$ is homogeneous of degree greater than $2d+1$, it contains only monomials of degrees greater than $2d+1$. This means that one of the entries in $\overline{Q}$ contains a nonzero constant term. By linearity of A_j and by Lemma 21, if we consider $\overline{Q}'$ as $\overline{Q}$ with zero instead of this nonzero constant term, we have $A_j(\overline{Q}) = A_j(\overline{Q}')$. We can continue in a similar manner, so that no polynomial has a nonzero constant term in $\overline{Q}$. And thus, eventually, every monomial in $A_j(\overline{Q})$ is of degree greater than $2d+1$, which is reduced to Case 1 above. QED

We are now ready to prove Lemma 22.

Proof of Lemma 22. Let $\mathcal{B}$ be a basis of the identities of $\mathrm{Mat}_d(\mathbb{F})$. By Corollary 20, let $\mathcal{A} := \{A_1, \ldots, A_m\}$ be a finite basis where all the A_i's are multi-homogeneous commutator identities of $\mathrm{Mat}_d(\mathbb{F})$, and there exists constants c_1, c_2 (depending only on $\mathcal{A}, \mathcal{B}$), such that for every identity f of $\mathrm{Mat}_d(\mathbb{F})$,

$$c_2 Q_{\mathcal{B}}(f) \le Q_{\mathcal{A}}(f) \le c_1 Q_{\mathcal{B}}(f). \tag{9}$$

Define

$$(\mathcal{A})^{(\le 2d+1)} := \{A_i \in \mathcal{A} \mid \text{the degree of } A_i \text{ is no more than } 2d+1\}.$$

For every identity P of $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$, by Lemma 24,

$$Q_{(\mathcal{A})^{(\le 2d+1)}}(P) = Q_{\mathcal{A}}(P). \tag{10}$$

This also means that every identity of $\mathrm{Mat}_d(\mathbb{F})$ of degree at most $2d+1$ can be generated by $(\mathcal{A})^{(\le 2d+1)}$. Thus, S_{2d} can be generated by $(\mathcal{A})^{(\le 2d+1)}$. By Lemma 23, all the polynomials in $(\mathcal{A})^{(\le 2d+1)}$ are generated by S_{2d}. Therefore, by Proposition 3, for every identity P of $\mathrm{Mat}_d(\mathbb{F})$ with degree $2d+1$:[9]

$$\frac{1}{Q_{(\mathcal{A})^{(\le 2d+1)}}(S_{2d})} \cdot Q_{S_{2d}}(P) \le Q_{(\mathcal{A})^{(\le 2d+1)}}(P) \le \left(\max_{A \in (\mathcal{A})^{(\le 2d+1)}} Q_{S_{2d}}(A) \right) \cdot Q_{S_{2d}}(P), \quad d > 2.$$

By (9) and (10) we now get that for every identity P of $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$,

$$\frac{1}{c_1 \cdot Q_{(\mathcal{A})^{(\le 2d+1)}}(S_{2d})} \cdot Q_{S_{2d}}(P) \le Q_{\mathcal{B}}(P) \le \frac{1}{c_2} \cdot \left(\max_{A \in (\mathcal{A})^{(\le 2d+1)}} Q_{S_{2d}}(A) \right) \cdot Q_{S_{2d}}(P), \quad d > 2.$$

QED

This concludes the main theorem of this section, Theorem 4.

Note on the case of $d = 2$. When $d = 2$, Lemma 22 is not true. For example, the polynomial $f = [[x_1, x_2][x_3, x_4] + [x_3, x_4][x_1, x_2], x_5]$ is an identity of $\mathrm{Mat}_2(\mathbb{F})$, but in [Ler73] it is proved that f cannot be generated by S_4. Namely the restriction $d > 2$ in Lemma 22, and also in Theorem 4, is essential for our proof.

6 Open Problems

Here we consider two open problems of independent interest, one about non-commutative algebraic circuit complexity and the other about proof complexity. Based on these open problems,

[9] Note that in Proposition 3 we can substitute the bases $\mathcal{B}_0, \mathcal{B}_1$ by every pair of sets of identities (not necessarily a pair of bases), as long as the identities in $\mathcal{B}_1$ are consequences of the identities in $\mathcal{B}_0$, and vice versa.

up to exponential-size lower bounds on PI proofs follow (that is, exponential-size in terms of the (non-commutative)[10] circuit-size of the identity proved).

Informally, the two problems are as follows:

Informal problem I. *There exist non-commutative algebraic circuits of small size that compute matrix identities of high generative complexity.*

Informal problem II. *Proving matrix identities by reasoning with polynomials whose variables $X_1, \dots, X_n$ range over* matrices *is as efficient as proving matrix identities using polynomials whose variables range over the* entries *of the matrices $X_1, \dots, X_n$?*

6.1 Matrix Proof Lower Bounds in Terms of Algebraic Circuit Size

In Theorem 5 we established polynomial $\Omega(n^{2d})$ lower bounds on the number of steps (and hence size) in matrix proofs of matrix identities with n variables. The hard instances we used in Theorem 5 were non-explicit, and so we do not know their algebraic circuit size. However, it is more interesting from the (proof) complexity perspective to have size lower bounds on $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ proofs in terms of the algebraic circuit size of the identities proved. For this purpose, we need to assume the existence of non-commutative algebraic circuits of small size that compute matrix identities of high generative complexity:

Problem I. *Prove that for some fixed $1 \le r < d$ and a fixed basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$, there exists a family of identities $f_n \in \mathbb{F}\langle X \rangle$ of $\mathrm{Mat}_d(\mathbb{F})$, with n variables, such that $Q_{\mathcal{B}}(f_n) = \Omega(n^d)$, and f_n has a non-commutative algebraic circuit of size $O(n^r)$.*

Polynomial lower bounds on $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$-proofs (assuming problem I): *There exists a family of identities f_n of $\mathrm{Mat}_d(\mathbb{F})$ whose non-commutative algebraic circuit-size is s_n, but every $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$-proof of f_n has size $\Omega((s_n)^{\frac{d}{r}})$, for some fixed $1 \le r < d$.*

Note that we *do* know by Theorem 4 that the lower bound in Problem I is true for all $d > 2$ and for some (non-explicit) family f_n. But we do not know whether f_n has small non-commutative circuits, as required in Problem I.

6.2 Polynomial-Size Lower Bounds on PI Proofs

Here we propose the possibility that every polynomial-size lower bound on matrix identities proofs $\mathbf{PI}_{\mathrm{Mat}_d}(\mathbb{F})$ (Definition 3) can be lifted to lower bounds on PI proofs $\mathbf{PI}_c(\mathbb{F})$ (Definition 1).

Consider a nonzero identity $f \in \mathbb{F}\langle X \rangle$ of $\mathrm{Mat}_d(\mathbb{F})$, for some $d > 1$. If we substitute each (matrix) variable x_ℓ in f by a $d \times d$ matrix of *entry-variables* $\{x_{\ell jk}\}_{j,k \in [d]}$ (and consider product as matrix product and addition as entry-wise addition), then f corresponds to d^2 *commutative* zero polynomials (in case $\mathbb{F}$ is not big enough, these may be nonzero commutative polynomials that compute the zero function over $\mathbb{F}$), each computing an entry of the $d \times d$ zero matrix computed by f (see the example below and Proposition 26).

Accordingly, assume that $\mathbb{F}$ is a sufficiently big field, and let F be a non-commutative circuit computing f. Then under the above substitution of d^2 entry-variables to each variable in F, we get d^2 non-commutative circuits, each computing the zero polynomial *when considered as*

[10] PI proofs operate with equations between (commutative) algebraic circuits. However, since these algebraic circuits are written as purely syntactic objects in PI proofs, implicitly we have an order on children of product gates. Hence, we can consider algebraic circuits in PI proofs as non-commutative circuits.

commutative polynomials (see Definition 10).[11] We denote the set of d^2 circuits corresponding to the identity F by $[\![F]\!]_d$ (and we extend it naturally to equations between circuits: $[\![F = G]\!]_d$).

Example: Let $d = 2$ and let $f = x_1x_2 - x_2x_1$ (it is not an identity of $\mathrm{Mat}_2(\mathbb{F})$, but we use it only for the sake of example). And let $F = x_1x_2 - x_2x_1$ be the corresponding circuit (in fact, formula) computing f. Then we substitute entry variables for x_1, x_2 to get:

$$\begin{pmatrix} x_{111} & x_{112} \\ x_{121} & x_{122} \end{pmatrix} \cdot \begin{pmatrix} x_{211} & x_{212} \\ x_{221} & x_{222} \end{pmatrix} - \begin{pmatrix} x_{211} & x_{212} \\ x_{221} & x_{222} \end{pmatrix} \cdot \begin{pmatrix} x_{111} & x_{112} \\ x_{121} & x_{122} \end{pmatrix}.$$

And the $(1, 1)$-entry non-commutative circuit (formula) in $[\![F]\!]_d$, is:

$$(x_{111}x_{211} + x_{112}x_{221}) - (x_{211}x_{111} + x_{212}x_{121}).$$

Formally, we define the set of d^2 non-commutative circuits corresponding to the non-commutative circuit F as follows:

Definition 10 ($[\![F]\!]_d$)**.** *Let F be a non-commutative circuit computing the polynomial $f \in \mathbb{F}\langle X\rangle$, such that f is an identity of $\mathrm{Mat}_d(\mathbb{F})$. We define $[\![F]\!]_d$ as the set of d^2 (commutative) circuits that are generated from bottom to top in the circuit F as follows:*

1. *Every variable x_ℓ in F corresponds to d^2 new variables $x_{\ell ij}, i, j \in [d]$;*
2. *Every plus gate $X \oplus Y$ in F, where X, Y are two circuits, corresponds to d^2 plus gates $\oplus_{ij}, i, j \in [d]$ where each plus gate $\oplus_{ij}$ connects the corresponding circuit X_{ij} and Y_{ij} (that were generated before);*
3. *Every multiplication gate $X \otimes Y$ in F corresponds to d^2 plus gates $\oplus_{ij}$, for $i, j \in [d]$, where each plus gate $\oplus_{ij}$ is connected to d multiplication gates $\otimes_k$, for $k \in [d]$, each a product of X_{ik} and Y_{kj}. (Formally, plus gates have* fan-in two*, and so $\oplus_{ij}$ is the root of a binary tree whose internal nodes are all plus gates and whose d leaves are the product gates $\otimes_k$, $k \in [d]$.)*

Denote by $[\![F = 0]\!]_d$ the set of equations between circuits, where each circuit in $[\![F]\!]_d$ equals the circuit 0*.*

Fact 25. *Since every gate in F corresponds to at most d^3 gates in $[\![F]\!]_d$, we have:*

$$\left|[\![F]\!]_d\right| = O\left(d^3|F|\right)$$

(where $|F|$ denotes the size of F and $\left|[\![F]\!]_d\right|$ denotes the sum of sizes of all circuits in $[\![F]\!]_d$). Thus, when the dimension d of a matrix is constant, we have $|[\![f]\!]_d| = O(|f|)$.

For a *set* of identities S we say that $\mathbf{PI}_c(\mathbb{F})$ *proves* S, in symbols $\vdash_{\mathbf{PI}_c(\mathbb{F})} S$, if there exists a $\mathbf{PI}_c(\mathbb{F})$ proof that contains all the identities in S. We denote by $|\vdash_{\mathbf{PI}_c(\mathbb{F})} S|$ the minimal size of a $\mathbf{PI}_c(\mathbb{F})$ proof of S.

Proposition 26. *Let F be a non-commutative algebraic circuit computing f. For large enough fields $\mathbb{F}$ (specifically, for characteristic zero fields), $f \in \mathbb{F}\langle X\rangle$ is an identity of $\mathrm{Mat}_d(\mathbb{F})$ iff $[\![F = 0]\!]_d$ has a $\mathbf{PI}_c(\mathbb{F})$ proof.*

[11] Recall that the same algebraic circuit, assuming it has order on children of product gates, can be considered as both a commutative and a non-commutative circuit.

Proof. Since $\mathbf{PI}_c(\mathbb{F})$ is a complete proof system for (commutative) polynomial identities written as equations between algebraic circuits, it suffices to show that every circuit in $[\![F]\!]_d$ computes (as a commutative circuit) the zero polynomial (i.e., the zero in $\mathbb{F}[X]$). Suppose that f is an identity of $\mathrm{Mat}_d(\mathbb{F})$ and assume by a way of contradiction that there is a nonzero polynomial $g \in \mathbb{F}[X]$ in $[\![F]\!]_d$. Then, there must be an assignment α of field elements such that $g(\alpha) \neq 0$ (this follows since the field is infinite, and so every nonzero polynomial has an assignment that does not nullify the polynomial). Extend the assignment α in any way to all the entry-variables in $[\![F]\!]_d$ and denote this extended assignment by α'. Thus, the set of $\mathrm{Mat}_d(\mathbb{F})$ matrices determined by this α' cannot nullify f, contradicting the assumption that f is an identity of $\mathrm{Mat}_d(\mathbb{F})$. The converse direction is similar. QED

Problem II. *Let d be a positive natural number and let $\mathcal{B}$ be a finite basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*. Assume that $f \in \mathbb{F}\langle X\rangle$ is an identity of* $\mathrm{Mat}_d(\mathbb{F})$*, and let F be a non-commutative algebraic circuit computing f. Prove that*

$$\left| \vdash_{\mathbf{PI}_c(\mathbb{F})} [\![F = 0]\!]_d \right| = \Omega(Q_{\mathcal{B}}(f)). \tag{11}$$

The conditional lower bound we get now is similar to that in Section 6.1, except that it holds for $\mathbf{PI}_c(\mathbb{F})$ and not only for matrix proofs:

Polynomial lower bounds on PI proofs $\mathbf{PI}_c(\mathbb{F})$ (assuming Problems I and II): *There exists a family of identities f_n of* $\mathrm{Mat}_d(\mathbb{F})$ *whose non-commutative algebraic circuit F_n has size s_n, but every $\mathbf{PI}_c(\mathbb{F})$-proof of $[\![F_n = 0]\!]_d$ has size $\Omega(s_n^{d/r})$, for some fixed $1 \le r < d$.*

6.3 The Propositional Case

We now discuss the applicability of our suggested framework to obtaining lower bounds on the size of *propositional proofs*.

Given a commutative algebraic circuit C over $GF(2)$, we can think of the circuit equation $C = 0$ as a *Boolean* circuit computing a tautology, instead of an algebraic circuit: interpreting $+$ as XOR, $\cdot$ as $\wedge$, and $=$ as logical equivalence $\equiv$ (that is, $\leftrightarrow$). Accordingly, if we augment to the $\mathbf{PI}_c(\mathbb{F})$ proof system, where $\mathbb{F} = \mathbf{GF(2)}$, the axioms $x_i^2 + x_i = 0$, for every variable x_i, we obtain a propositional proof system which formally *is* an Extended Frege proof system (see [HT15]). Denote this system by $\mathbf{PI}_c(\mathbb{F}) + \{x_i^2 + x_i = 0 \ : \ x_i \in X\}$.

Propositional version of Problem I. *Let $\mathbb{F} = \mathbf{GF(2)}$, let d be a positive natural number and let $\mathcal{B}$ be a (finite) basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$*. Assume that $f \in \mathbb{F}\langle X\rangle$ is an identity of* $\mathrm{Mat}_d(\mathbb{F})$*, and let F be a non-commutative algebraic circuit computing f. Then,*

$$\left| \vdash_{\mathbf{PI}_c(\mathbb{F})+\{x_i^2+x_i=0\,:\,x_i\in X\}} [\![F = 0]\!]_d \right| = \Omega(Q_{\mathcal{B}}(f)). \tag{12}$$

As before, $\left| \vdash_{\mathbf{PI}_c(\mathbb{F})+\{x_i^2+x_i\,=0:\,x_i\in X\}} [\![F = 0]\!]_d \right|$ is the minimal size of a $\mathbf{PI}_c(\mathbb{F}) + \{x_i^2 + x_i = 0 \ : \ x_i \in X\}$ proof of $[\![F = 0]\!]_d$ (which by the above mentioned, is the minimal Extended Frege proof size of $[\![F = 0]\!]_d$ up to polynomial factors). In other words, the minimal size in a $\mathbf{PI}_c(\mathbb{F}) + \{x_i^2 + x_i \ = 0 : \ x_i \in X\}$ proof of the collection of d^2 (entry-wise) equations $[\![F = 0]\!]_d$ corresponding to F is lower bounded (up to a constant factor) by $Q_{\mathcal{B}}(f)$.

Comment: One can consider the same propositional version of the main open problem, with $\mathbb{F}$ being the rational numbers, and hence of characteristic zero (for we which we have more knowledge about $Q_{\mathcal{B}}(\cdot)$, as obtained in our work). However, the way to translate PI proofs $\mathbf{PI}_c$ over the rationals is less immediate than the same translation for the case of $\mathbf{GF(2)}$.

6.4 Exponential-Size Lower Bounds

Assuming Problem II (Equation (11)) is settled, we show under which parameters one gets *exponential-size* lower bounds on $\mathbf{PI}_c(\mathbb{F})$ proofs. The idea is to let the dimension d of the matrix algebras grow with n (the number of variables in the hard instances). Therefore, if the growth rate of the minimal proof size of the hard instances is exponential in d (like the non-explicit hard instances in Theorem 5), while the growth rate of the algebraic circuit size of the hard instances is only polynomial d, we obtain an exponential lower bound.

For this approach we need to set up the assumptions more carefully:

Refinement of Problems I and II:

1. *Problem II*: For every d and every basis $\mathcal{B}_d$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$ the size of every $\mathbf{PI}_c(\mathbb{F})$ proof of $[\![F = 0]\!]_d$ is at least $\mathcal{C}_{\mathcal{B}_d} \cdot Q_{\mathcal{B}_d}(f)$, where $\mathcal{C}_{\mathcal{B}_d}$ is a number depending on $\mathcal{B}_d$ and F is a non-commutative algebraic circuit computing f (this is the same as Problem II except that here we explicitly show $\mathcal{C}_{\mathcal{B}_d}$).

2. Assume that for some sufficiently large d and some basis $\mathcal{B}_d$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$, there exists a number $c_{\mathcal{B}_d}$, such that for all sufficiently large n there exists an identity $f_{n,d}$ with $Q_{\mathcal{B}_d}(f_{n,d}) \geq c_{\mathcal{B}_d} \cdot n^{2d}$. (The existence of such identities are known from our unconditional lower bound in Theorem 5.)

3. Assume that for the $c_{\mathcal{B}_d}$ in item 2 above: $c_{\mathcal{B}_d} \cdot \mathcal{C}_{\mathcal{B}_d} = \Omega\left(\frac{1}{\mathrm{poly}(d)}\right)$.

4. *Refinement of Problem I*: Assume there exist non-commutative algebraic circuits $F_{n,d}$ computing $f_{n,d}$ from item 2 of size $\mathrm{poly}(n,d)$.

Corollary (assuming assumptions 1 to 4 above hold): There exists a polynomial size (in n) family of identities between algebraic circuits, for which every $\mathbf{PI}_c(\mathbb{F})$ proof requires $2^{\Omega(n)}$ number of proof-lines.

Proof. By the assumptions, every $\mathbf{PI}_c(\mathbb{F})$ proof of $[\![F_{n,d} = 0]\!]_d$ has size at least $\mathcal{C}_{\mathcal{B}_d} \cdot Q_{\mathcal{B}_d}(f_{n,d}) = \mathcal{C}_{\mathcal{B}_d} \cdot c_{\mathcal{B}_d} \cdot n^{2d}$. Consider the family $\{f_{n,d}\}_{n=1}^{\infty}$, *where d is a function of n*, and take $d = n/4$. Then, we get the following lower bound on the size of every $\mathbf{PI}_c(\mathbb{F})$ proof of the family $\{f_{n,d}\}_{n=1}^{\infty}$:

$$c_{\mathcal{B}_d} \cdot \mathcal{C}_{\mathcal{B}_d} \cdot n^{2d} = \frac{1}{\mathrm{poly}(n/4)} \cdot n^{n/2} = 2^{\Omega(n)},$$

which (by assumption 4 and Fact 25) is *exponential* in the algebraic circuit-size of the identities $[\![F_{n,d} = 0]\!]_d$ proved. QED

Acknowledgements

We wish to thank V. Arvind, Albert Atserias, Michael Forbes, Emil Jeřabek and Amir Shpilka for useful discussions related to this work. We are also greatly indebted to Vesselin Drensky for his help with the bibliography and providing us with his monograph [Dre99].

References

[AKBK16] Eli Aljadeff, Alexei Kanel-Belov, and Yaakov Karasik. Kemer's theorem for affine PI algebras over a field of characteristic zero. *J. Pure Appl. Algebra*, 220(8):2771–2808, 2016. 2.5, 4

[AL50] S. A. Amitsur and J. Levitzki. Minimal identities for algebras. In *Proc. Amer. Math. Soc. (2)*, pages 449–463, 1950. 2.2, 2.6.2

[AMR16] V. Arvind, Partha Mukhopadhyay, and S. Raja. Randomized polynomial time identity testing for noncommutative circuits. *ArXiV*, 2016. 1

[BDDK03] Francesca Benanti, James Demmel, Vesselin Drensky, and Plamen Koev. Computational approach to polynomial identities of matrices - a survey. *Ring Theory: Polynomial Identities and Combinatorial Methods, Proc. of the Conf. in Pantelleria*, 235:141–178, 2003. Lect. Notes in Pure and Appl. Math. Eds. A. Giambruno, A. Regev, and M. Zaicev. 2.6.2

[CR74a] Stephen A. Cook and Robert A. Reckhow. Corrections for "On the lengths of proofs in the propositional calculus (preliminary version)". *SIGACT News*, 6(3):15–22, July 1974. 6.4

[CR74b] Stephen A. Cook and Robert A. Reckhow. On the lengths of proofs in the propositional calculus (preliminary version). In *Proceedings of the 6th Annual ACM Symposium on Theory of Computing (STOC 1974)*, pages 135–148, 1974. For corrections see Cook-Reckhow [CR74a]. 6.4

[CR79] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *J. Symb. Log.*, 44(1):36–50, 1979. This is a journal-version of Cook-Reckhow [CR74b] and Reckhow [Rec76]. 1

[CS07] Steve Chien and Alistair Sinclair. Algebras with polynomial identities and computing the determinant. *SIAM J. Comput.*, 37(1):252–266, 2007. 1

[Dre81] Vesselin Drensky. A minimal basis of identities for a second-order matrix algebra over a field of characteristic 0. *Algebra i Logika*, 20(3):291–299, May–June 1981. Translation. 2.5

[Dre99] Vesselin Drensky. *Free Algebras and PI-Algebras*. Springer-Verlag, Singapore, 1999. 2.5, 2.6.2, 2.6.2, 5.2, 19, 8, 5.2, 6.4

[GP14] Joshua A. Grochow and Toniann Pitassi. Circuit complexity, proof complexity, and polynomial identity testing. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*, pages 110–119, 2014. Full version at `arXiv:abs/1404.3820`. 2.7

[Hru11] Pavel Hrubeš. How much commutativity is needed to prove polynomial identities? *Electronic Colloquium on Computational Complexity, ECCC*, (Report no.: TR11-088), June 2011. (document), 2.6, 2.6.1, 1, 2.6.2, 2.6.2, 2.6.2, 2.7, 5.1.1, 5.1.2

[HT09] Pavel Hrubeš and Iddo Tzameret. The proof complexity of polynomial identities. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 41–51, 2009. (document), 1, 2.4.1, 1

[HT15] Pavel Hrubeš and Iddo Tzameret. Short proofs for the determinant identities. *SIAM J. Comput.*, 44(2):340–383, 2015. (A preliminary version appeared in Proceedings of the 44th Annual ACM Symposium on the Theory of Computing (STOC'12)). (document), 1, 2.3, 1, 2.7, 6.3

[HY11] Pavel Hrubeš and Amir Yehudayoff. Arithmetic complexity in ring extensions. *Theory of Computing*, 7:119–129, 2011. 5.1.1, 5.1.1, 12

[Jeř04] Emil Jeřábek. Dual weak pigeonhole principle, Boolean complexity, and derandomization. *Ann. Pure Appl. Logic*, 129(1-3):1–37, 2004. 2.3

[Jeř14] Emil Jeřabek. Personal communication, 2014. 2.2

[Kem87] Alexander Kemer. Finite basability of identities of associative algebras. *Algebra i Logika*, 26(5):597–641, 650, 1987. 2.5, 4, 5.2

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complexity*, 13(1-2):1–46, 2004. Preliminary version in the *35th Annual ACM Symposium on Theory of Computing (STOC 2003)*. 1

[Kra95] Jan Krajíček. *Bounded arithmetic, propositional logic, and complexity theory*, volume 60 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. 2.7, 4

[Ler73] Uri Leron. Multilinear identities of the matrix ring. *Transactions of the American Mathematical Society*, 183:175–202, Sep. 1973. 5.2, 5.2

[LTW15] Fu Li, Iddo Tzameret, and Zhengyu Wang. Non-commutative formulas and frege lower bounds: a new characterization of propositional proofs. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 412–432, 2015. Full Version: `http://arxiv.org/abs/1412.8746`. 2.7

[PT16] Tonnian Pitassi and Iddo Tzameret. Algebraic proof complexity: Progress, frontiers and challenges. *ACM SIGLOG News*, 3(3), 2016. 1

[Rec76] Robert A. Reckhow. *On the lengths of proofs in the propositional calculus*. PhD thesis, University of Toronto, 1976. 6.4

[Row80] Louis Halle Rowen. *Polynomial identities in ring theory*. Pure and Applied Mathematics. Academic Press, 1980. 2.5, 2.6.2, 5.2

[RS05] Ran Raz and Amir Shpilka. Deterministic polynomial identity testing in non-commutative models. *Comput. Complex.*, 14(1):1–19, April 2005. Preliminary version in the *19th Annual IEEE Conference on Computational Complexity (CCC 2004)*. 1

[SAT09] *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*, 2009. 1

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980. Preliminary version in the *International Symposium on Symbolic and Algebraic Computation (EUROSAM 1979)*. 1

[Tza11] Iddo Tzameret. Algebraic proofs over noncommutative formulas. *Inf. Comput.*, 209(10):1269–1292, 2011. 2.7

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation (EUROSAM 1979)*, pages 216–226. Springer-Verlag, 1979. 1