Electronic Colloquium on Computational Complexity, Report No. 185 (2013)

Generating Matrix Identities and Proof Complexity Lower Bounds

Fu Li* Iddo Tzameret†

December 24, 2013

Abstract

Motivated by the fundamental lower bounds questions in proof complexity, we investigate the complexity of generating identities of matrix rings, and related problems. Specifically, for a field $\mathbb{F}$ let A be a non-commutative (associative) $\mathbb{F}$-algebra (e.g., the algebra $\mathrm{Mat}_d(\mathbb{F})$ of $d \times d$ matrices over $\mathbb{F}$). We say that a non-commutative polynomial $f(x_1, \dots, x_n)$ over $\mathbb{F}$ is *an identity of* A, if for all $\overline{c} \in A^n$, $f(\overline{c}) = 0$. Let B be a set of non-commutative polynomials that forms a *basis* for the identities of A, in the following sense: for every identity f of A there exist non-commutative polynomials $g_1, \dots, g_k$, for some k, that are *substitution instances* of polynomials from B, such that f is in the (two-sided) ideal $\langle g_1, \dots, g_k \rangle$. We study the following question: *Given A, B and f as above, what is the minimal number k of such generators $g_1, ..., g_k$ for which $f \in \langle g_1, \dots, g_k \rangle$?*

In particular, we focus on the case where the algebra A is $\mathrm{Mat}_d(\mathbb{F})$, and $\mathbb{F}$ has characteristic 0. Our main technical contribution is a generalization of the lower bound presented in Hrubeš [6] (for the case $d = 1$) to *any* $d > 2$:

> For every natural number $d > 2$ and *every* finite basis B for the identities of $\mathrm{Mat}_d(\mathbb{F})$, where $\mathbb{F}$ is of characteristic 0, there exists an identity f_n with n variables, that requires $\Omega(n^{2d})$ generators (i.e., substitution instances from B) to generate.

Note that for any $d > 2$, it is an *open problem* to find a basis for the identities of $\mathrm{Mat}_d(\mathbb{F})$ (while the *existence* of a finite basis was proved by Kemer [11]). Nevertheless, using results from the theory of algebras with polynomial identities (PI-algebras) together with a generalization of the arguments in [6], we conclude the above lower bound for *every* finite basis B.

We then explore connections to lower bounds in proof complexity. We consider arithmetic proofs of polynomial identities that operate with algebraic circuits and whose axioms are the polynomial-ring axioms (which can be considered as an algebraic analogue of the Extended Frege propositional proof system). We raise the following basic question: *is it true that using the generators of the (non-commutative) polynomial identities over* $\mathrm{Mat}_d(\mathbb{F})$ *as axiom (schemes) is an optimal way to prove such identities, with respect to proof size?* Namely, is it true that 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 $X_1, \dots, X_n$? We show that a positive answer to this question may lead, under further assumptions (which are generalization of the assumptions presented in [6]), up to exponential-size lower bounds on arithmetic proofs.

*Institute for Theoretical Computer Science, The Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing. Supported in part by NSFC grant 61373002

†Institute for Theoretical Computer Science, The Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing `tzameret@tsinghua.edu.cn` Supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of P. R. China; Grants 61033001, 61061130540, 61073174, 61373002.

ISSN 1433-8092

Contents

1 Introduction

1.1 Background

Proving super-polynomial size lower bounds on strong propositional proof systems, like the Extended Frege system, is one of the central problems of proof complexity, and in general is among a handful of fundamental hardness questions in computational complexity theory. An Extended Frege proof is simply a textbook logical proof system for establishing Boolean tautologies, in which one starts from basic tautological axioms written as Boolean formulas, and derives, step by step, new tautological formulas from previous ones by using a finite set of logical sound derivation rules; including the so-called *extension axiom* enabling one to denote a possibly big formula by a *single* new variable (where the variable is used neither before in the proof nor in the last line of the proof). It is not hard to show (see [10]) that Extended Frege can equivalently be defined as a logical proof system operating with Boolean *circuits* (and without the extension axiom[1]).

[1] An additional simple technical axiom is needed to formally define this proof system ([10]).

Lower bounds on Extended Frege proofs can be viewed as lower bounds on certain nondeterministic algorithms for establishing the unsatisfiability of Boolean formulas (and thus as a progress towards separating $\boldsymbol{NP}$ from $\boldsymbol{coNP}$). It is also usually considered (somewhat informally) as related to establishing (explicit) Boolean circuit size lower bounds. In fact, it has also another highly significant consequence, that places such a lower bound as a small step towards separating $\boldsymbol{P}$ from $\boldsymbol{NP}$: showing any super-polynomial lower bound on the size of Extended Frege proofs implies that, at least with respect to "polynomial-time reasoning" (namely, reasoning in the formal theory of arithmetic denoted S_2^1), it is not possible to prove that $\boldsymbol{P} = \boldsymbol{NP}$; or in other words, it is consistent with S_2^1 that $\boldsymbol{P} \neq \boldsymbol{NP}$.

Accordingly, proving Extended Frege lower bounds is considered an extremely hard problem. In fact, even *conditional* lower bounds on strong proof systems, including Extended Frege, are not known and are considered very interesting; here, we mean a condition that is different from $\boldsymbol{NP} \neq \boldsymbol{coNP}$ (see [15]; the latter condition readily implies that any propositional proof system admits a family of tautologies with no polynomial-size proofs [3]). For suggested conditional and unconditional Extended Frege lower bounds approaches, see also the recent monograph by Krajíček [13]. The only size lower bound on Extended Frege proofs that is known to date is linear $\Omega(n)$ (where n is the size of the tautological formula proved; see [12] for a proof). Establishing *super-linear* size lower bounds on Extended Frege proofs is thus a highly interesting open problem.

Another feature of proof complexity is that, in contrast to circuit complexity, even the *existence of non-explicit* hard instances for strong propositional proof systems, including Extended Frege, are unknown. For instance, simple counting arguments cannot establish super-linear size lower bounds on Extended Frege proofs.

Due to the lack of progress on establishing lower bounds on strong propositional proof systems, it is interesting, and potentially helpful, to turn our eyes to an *algebraic analogue* of strong propositional proof systems, and try first to prove nontrivial size lower bounds in such settings. Quite recently, such algebraic analogues of Extended Frege (and Frege, which is Extended Frege without the extension axiom) were investigated by Hrubeš and the second author [7, 8]. These proof systems denoted $\mathbb{P}_c(\mathbb{F})$, called simply *arithmetic proofs*, operate with algebraic equations of the form $F = G$, where F and G are algebraic circuits over a given field $\mathbb{F}$. An arithmetic proof of a polynomial identity is a sequence of identities between algebraic circuits derived by means of simple syntactic manipulation representing the polynomial-ring axioms (e.g., associativity, distributivity, unit element, field identities, etc.). Although arithmetic proof systems are not propositional proof systems, namely they do not prove propositional tautologies, they can be regarded nevertheless as *fragments* of the propositional Extended Frege proof system when the field considered is $GF(2)$. That is, every arithmetic proof over $GF(2)$ of a polynomial identity (considered as a propositional tautology) can formally be viewed also as an Extended Frege proof.[2]

Apart from the hope that arithmetic proofs would shed light on propositional proof systems, the study of arithmetic proofs is motivated by the Polynomial Identity Testing (PIT) problem, namely the problem of deciding if a given algebraic circuit computes the zero polynomial. As a

[2]In fact, it is probably true (but was not formally verified) that arithmetic proofs serve as fragments of propositional proofs also over any other finite field, as well as over the field of rational numbers (when restricted to up to exponentially big rational numbers). That is, it is probably true that every polynomial identity proved with an arithmetic proof over the given field, can be proved with at most a polynomial increase in size in Extended Frege when we fix a certain natural translation between polynomial identities over the field and propositional tautologies. The reason for this is that one could plausibly polynomially simulate arithmetic proofs over such fields with propositional proofs in which numbers are encoded as bit-strings.

decision problem, polynomial identity testing can be solved by an efficient randomized algorithm [19, 20], but no efficient deterministic algorithm is known. In fact, it is not even known whether there is a polynomial time non-deterministic algorithm or, equivalently, whether PIT is in $\boldsymbol{NP}$. An arithmetic proof system can thus be interpreted as a specific non-deterministic algorithm for PIT: in order to verify that an arithmetic circuit C computes the zero polynomial, it is sufficient to guess an arithmetic proof of $C = 0$. Hence, if every true equality has a polynomial-size proof then PIT is in $\boldsymbol{NP}$. Conversely, the arithmetic proof system captures the common syntactic procedures used to establish equality between algebraic expressions. Thus, showing the existence of identities that require super-polynomial arithmetic proofs would imply that those syntactic procedures are not enough to solve the PIT problem efficiently.[3]

The emphasis in [7, 8] was mainly on demonstrating non-trivial *upper bounds* for arithmetic proofs (as well as lower bounds in very restricted settings). Since arithmetic proofs (at least over $GF(2)$), can also be considered as propositional proofs, arithmetic proofs were found very useful in establishing short propositional proofs for the determinant identities and other statements from linear algebra [8]. As for *lower bounds* on arithmetic proofs (operating with arithmetic circuits), the same basic linear size lower bound known for Extended Frege [12] can be shown to hold for $\mathbb{P}_c$. But any super-linear size lower bound, explicit or not, on $\mathbb{P}_c(\mathbb{F})$ proof size (for any field $\mathbb{F}$) is open. In [7] it was argued that proving lower bounds even on very restricted fragments of arithmetic proofs is a highly nontrivial open problem.

The situation we have described up to now shows how little is known about (strong) propositional (and arithmetic) proof systems, and why it is highly interesting to introduce and develop novel approaches for lower bounding proofs such as arithmetic proofs, even if these approaches yield only conditional and possibly non-explicit lower bounds.

1.2 Overview of our work

The problem of proving *quadratic* size lower bounds on arithmetic proofs $\mathbb{P}_c$ was considered by Hrubeš in [6]. The work in [6] gave several conditions and open problems, under which, quadratic size lower bounds on arithmetic 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). The current work is an attempt to extend the approach suggested in Hrubeš [6], from an approach suitable for proving up to $\Omega(n^2)$ size lower bounds on $\mathbb{P}_c$ proofs, to an approach for proving much stronger lower bounds, namely an $\Omega(n^d)$ lower bound on $\mathbb{P}_c(\mathbb{F})$ proofs, for every positive $d > 2$ and for every zero characteristic field $\mathbb{F}$; and under stronger assumptions, exponential $2^{\Omega(n)}$ lower bounds on $\mathbb{P}_c(\mathbb{F})$ proofs.

In the rest of this section we discuss in more details the main algebraic problem we investigate, our result, how this work generalizes the previous work by Hrubeš' [6], under which assumptions this work can be applied to obtain strong proof complexity lower bounds and the questions left open. For a more formal treatment, see subsequent sections.

[3] It is worth emphasizing again that arithmetic proofs are different than algebraic *propositional* proof systems like the Polynomial Calculus [2] and related systems. The latter prove propositional tautologies (a **coNP** language) while the former prove formal polynomial identities.

1.2.1 The algebraic problem

For a field $\mathbb{F}$ let A be a non-commutative (associative) $\mathbb{F}$-algebra; e.g., the algebra $\mathrm{Mat}_d(\mathbb{F})$ of $d \times d$ matrices over $\mathbb{F}$. We shall always assume, unless explicitly stated otherwise, that the field $\mathbb{F}$ has characteristic 0. A non-commutative polynomial over the field $\mathbb{F}$ and with the variables $X := \{x_1, x_2, \ldots\}$ is a formal sum of monomials where the product of variables is non-commuting. Since most polynomials in this work are non-commutative **when we talk about *polynomials* we shall mean *non-commutative polynomials,*** unless otherwise stated. The set of (non-commutative) polynomials with variables X and over the field $\mathbb{F}$ is denoted $\mathbb{F}\langle X \rangle$.

We say that a polynomial $f(x_1, \ldots, x_n)$ over $\mathbb{F}$ is *an identity of* A, if for all $\bar{c} \in A^n$, $f(\bar{c}) = 0$. Let $\mathcal{B}$ be a set of non-commutative polynomials that forms a *basis* for the identities of A, in the following sense: for every identity f of A there exist non-commutative polynomials $g_1, ..., g_k$, for some k, that are *substitution instances* of polynomials from $\mathcal{B}$, such that f is in the two-sided ideal $\langle g_1, ..., g_k \rangle$ (a *substitution instance* of a polynomial $g(x_1, \ldots, x_n) \in \mathbb{F}\langle X \rangle$ is a polynomial $g(h_1, \ldots, h_n)$, for some $h_i \in \mathbb{F}\langle X \rangle$, $i \in [n]$).

Given an $\mathbb{F}$-algebra A and an identity f of A, define $Q_{\mathcal{B}}(f)$ as the minimal number k such that there exist $g_1, \ldots, g_k \in \mathbb{F}\langle X \rangle$ for which $f \in \langle g_1, ..., g_k \rangle$, and every g_i is a substitution instance of some polynomial from $\mathcal{B}$.

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. Namely, these are the (non-commutative) polynomials that are identically zero polynomials *when considered as commutative polynomials.* For instance, $x_1x_2 - x_2x_1$ is a non-zero polynomial from $\mathbb{F}\langle X \rangle$ which is an identity over $\mathscr{A}$.

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 $(x_1 + x_2)x_3 - x_3(x_1 + x_2) = x_1x_3 - x_3x_1 + x_2x_3 - x_3x_2$.

We can now present Hrubeš [6] lower bound, with the aid of our notations:

Theorem 1 ([6]). *For any field and every n, 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.

The generalization. For the sake of describing our generalization, let us treat (the $\mathbb{F}$-algebra) $\mathbb{F}$ as the matrix algebra $\mathrm{Mat}_1(\mathbb{F})$ of 1×1 matrices with entries from $\mathbb{F}$. An *algebra with polynomial identities,* or in short a ***PI-algebra*** (PI stands for Polynomial Identities), is simply 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.

In this work, we exploit known results about the structure of the identities of matrix algebras and the general theory of PI-algebras to completely generalize Hrubeš [6] lower bound above, 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 0:

Theorem 5. *Let $\mathbb{F}$ be any field of characteristic 0. For every 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 $Q_{\mathcal{B}}(f) = \Omega(n^{2d})$.*

Notice that similar to [6], the lower bound in this theorem is *non-explicit*. We do not know of an upper bound that holds on $Q_{\mathcal{B}}(f)$, for every identity f with n variables.

We now discuss shortly the proof of Theorem 5.

The study of algebras with polynomial identities is a fairly developed subject in algebra (see the monographs by Drensky [5] and Rowen [18] on this topic). Within it, perhaps the most well known works are about the identities of matrix algebras. In particular, the well-known theorem of Amitsur and Levitzky from 1950 [1] is the following:

Amitsur-Levitzki Theorem ([1])**.** *Let $\mathcal{S}_d$ 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:*

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

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 $\mathrm{Mat}_d(\mathbb{F})$*.*

The first step in proving Theorem 5 is to use the Amitsur-Levitzki Theorem: we show that when $\mathcal{E} = \{S_{2d}(x_1, \ldots, x_{2d})\}$ then there exists an $f \in \mathbb{F}\langle X \rangle$ with $2n$ variables and degree $2d + 1$, such that $Q_{\mathcal{E}}(f) = \Omega(n^{2d})$. To this end, we use a counting argument which is a generalization of the counting argument in [6]. In several places the argument is slightly more involved than in [6]; e.g., when we need to show the existence of a *single* polynomial that achieves the lower bound (see Lemma 12).

Note that $\mathcal{E} = \{S_{2d}(x_1, \ldots, x_{2d})\}$ is *not* a basis of $\mathrm{Mat}_d(\mathbb{F})$, namely there are identities of $\mathrm{Mat}_d(\mathbb{F})$ that are not generated by substitution instances of S_{2d} (also notice that $Q_{\mathcal{B}}(f)$ can be defined for any $\mathcal{B} \subseteq \mathbb{F}\langle X \rangle$). Thus, the second step in the proof of Theorem 5 is dedicated to showing that for *all finite bases $\mathcal{B}$ of the identities of* $\mathrm{Mat}_d(\mathbb{F})$ the following holds for the hard identity f considered in the theorem: $Q_{\mathcal{B}}(f) = \Theta(Q_{\mathcal{E}}(f))$. For this purpose, we use several structural results from the theory of algebras with polynomial identities (see Section 4.1.3).

One interesting feature of our proof (and theorem), is that it is in fact 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 a result of Drensky [4] (see Section 6.3). However, a highly nontrivial result of Kemer [11], shows that for any natural *d there exists* a finite basis for $\mathrm{Mat}_d(\mathbb{F})$. Our proof shows roughly that for the hard instances f in Theorem 5 no generators different from the S_{2d} generators can contribute to the generation of f.

We also demonstrate that turning the non-explicit hard identities f *from Theorem 5* into explicit ones, means finding explicit tensors with high tensor-rank:

Theorem (informal)**.** *For any $d \geq 1$, if the hard identity f of* $\mathrm{Mat}_d(\mathbb{F})$ *in Theorem 5 is explicit, then there exists an explicit tensor $A : [n]^{2d+1} \to \{0, 1\}$ with tensor-rank $\Omega(n^{2d})$.*

This is a generalization (to any order), of a similar observation made in [6] for order 3 tensors. This corollary can be interpreted as an evidence that the *specific* hard instances we provide in Theorem 5 are not good candidates for proof complexity hardness (see about connections to proof complexity below), because we expect these instances not to have small circuits. Nevertheless, this does not rule out that other hard instances (namely, hard for the $Q_{\mathcal{B}}$) *are* suitable to achieve hardness results in proof complexity.

1.2.2 Potential applications to proof complexity

The lower bound described above, though interesting in itself, is motivated by possible applications to proof complexity lower bounds. As observed in [6], for the case of $d = 1$, it is relatively immediate to prove that the minimal number of generators needed to generate an identity f over $\mathrm{Mat}_d(\mathbb{F})$ is a lower bound on the number of distinct instances of commutativity axioms $h \cdot g = g \cdot h$ needed in any arithmetic proof of f. Thus, one can hope to get *up to quadratic* lower bounds on the number of lines (and equivalently the size) in $\mathbb{P}_c$ proofs this way (due to the quadratic upper bound $Q_{\{[x_1,x_2]\}}(f) = O(n^2)$).

Conditional lower bounds on *fragments* of arithmetic proofs. We can show that for each $d > 1$, there is a connection between the measure $Q_{\mathcal{B}}(\cdot)$ and fragments of arithmetic proofs, as explained in what follows.

For each $d \geq 1$, denote by $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$ the following arithmetic proof system operating with equations between algebraic circuits: consider the proof systems $\mathbb{P}_c(\mathbb{F})$ and *replace* the commutativity axiom $h \cdot g = g \cdot h$ by a finite basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$ (namely, add a new axiom $H = 0$ for each polynomial h in the basis, where H is a (non-commutative algebraic) circuit computing h).[4] It is not hard to show the following:

Theorem. *For every identity $F = 0$, where F is a non-commutative circuit that computes a non-commutative polynomial f which is an identity of $\mathrm{Mat}_d(\mathbb{F})$, the number of lines of a $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$-proof of $F = 0$ is lower bounded up to a constant factor (depending on the choice of finite basis $\mathcal{B}$) by $Q_{\mathcal{B}}(f)$.*

Thus, as a corollary of this theorem, we get the following conditional lower bound:

Assumption: for any fixed $d \geq 1$, assume that

1. 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 $Q_{\mathcal{B}}(f) = \Omega(n^d)$, for some basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$; *and*

2. the size of f_n is $O(n^r)$, for some constant $r < d$.

Conclusion: A lower bound of $\Omega(n^{d-r})$ (in terms of the circuit equation proved) on the size of $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$ proofs.

Note that we know by Theorem 5 that Assumption 1 is true for a specific f. But we do not know whether this f conforms to Assumption 2 (it seems plausible to assume that this specific f does not, because of the connection to tensor-rank mentioned above).

Apart from formulating the systems $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$, which constitute a hierarchy (for increasing d's) of weaker and weaker fragments of $\mathbb{P}_c(\mathbb{F})$ [5], we also formulate proof systems for the ***free-trace algebra*** [17] (see Section 6.3).

[4] Formally, we should fix a specific finite basis $\mathcal{B}$ for the sake of definiteness of $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$. However, different choices of bases can only increase the number of lines in a $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$-proof by a constant factor.

[5] Though not necessarily a proper hierarchy: we do not know if $\mathbb{P}_{\mathrm{Mat}_{d-1}}(\mathbb{F})$ has any speed-up over $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$ for identities of $\mathrm{Mat}_d(\mathbb{F})$.

Conditional polynomial-size lower bounds on the size of arithmetic proofs and the main open problem. Here we investigate the possibility to achieve arbitrary polynomial-size lower bounds on $\mathbb{P}_c(\mathbb{F})$ proofs using the measure $Q_{\mathcal{B}}(\cdot)$. We introduce a seemingly very natural way to achieve this goal, namely to demonstrate a connection between $Q_{\mathcal{B}}(\cdot)$ (for arbitrary $d > 1$ and a basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$) and the size of $\mathbb{P}_c(\mathbb{F})$ proofs. However, we fall short of actually proving the connection and we leave it as an intriguing open problem, that we show may lead, under further assumptions, up to exponential-size lower bounds on arithmetic proofs.

Informally, the question we raise is this: can it be shown that proving matrix identities by reasoning with polynomials whose variables $X_1, \ldots, X_n$ range over matrices is as efficient as proving matrix identities using polynomials whose variables range over the *entries* of the matrices $X_1, \ldots, X_n$? We explain our approach in what follows.

First, note that although, for any fixed $d > 1$ and every identity f of $\mathrm{Mat}_d(\mathbb{F})$, it is possible to use $Q_{\mathcal{B}}(f)$ to lower bound the size of $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$ proofs (as described above), it is not clear if and how we can use $Q_{\mathcal{B}}(f)$ to bound $\mathbb{P}_c(\mathbb{F})$ proof sizes (observe that any identity of $\mathrm{Mat}_d(\mathbb{F})$, for any d, can be generated with $O(n^2)$ instances of the *commutator*; though this is provably not true by Theorem 5, for "higher-order commutators", namely, when considering generators for the identities of $\mathrm{Mat}_d(\mathbb{F})$, for $d > 1$).

Thus, it seems that the most plausible way to connect $Q_{\mathcal{B}}(f)$ for $d > 1$ with the size of $\mathbb{P}_c$ proofs is via the following translation: consider a nonzero identity f of $\mathrm{Mat}_d(\mathbb{F})$, for some $d > 1$. Then f is a nonzero non-commutative polynomial in $\mathbb{F}\langle X\rangle$. If we substitute each (matrix) variable x_i in f by a $d \times d$ matrix of *entry-variables* $\{x_{ijk}\}_{j,k\in[d]}$, then now f corresponds to d^2 *commutative* zero polynomials: $f = 0$ says that for every (i, j) and for every possible assignment of field $\mathbb{F}$ elements to the (i, j)-entry of each of the matrix variables in f (when the product and addition of matrices are done in the standard way) the (i, j)-entry evaluates to 0. Accordingly, 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 commutative polynomials* (see Definition 15). We denote the set of d^2 circuits corresponding to the identity F by $[\![F]\!]_d$ (and we can extend it naturally to equations between circuits: $[\![F = G]\!]_d$)).

Example: let $d = 2$ and let $f = xy - yx$ (it is obviously not an identity of $\mathrm{Mat}_2(\mathbb{F})$, but we use it only for the sake of example). And let $F = xy - yx$ be the corresponding circuit (in fact, formula) computing f. Then we substitute matrices for x, y to get:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \cdot \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} - \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \cdot \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}.$$

And the $(1, 1)$-entry non-commutative circuit (in fact formula) corresponding to this is: $(x_{11}y_{11} + x_{12}y_{21}) - (y_{11}x_{11} + y_{12}x_{21})$.

It is not hard to show that $\left|[\![F]\!]_d\right| = O\left(d^3|F|\right)$, for every non-commutative circuit F (where $\left|[\![F]\!]_d\right|$ is the total sizes of all circuits in $[\![F]\!]_d$).

The main question we raise in this work is the following:

Main Open Problem. *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, the minimal number of lines in an arithmetic* *proof of the collection of d^2 (entry-wise) equations $\llbracket F = 0 \rrbracket_d$ corresponding to F is lower bounded (up to a constant factor) by $Q_{\mathcal{B}}(f)$. And in symbols:*

$$\left| \vdash_{\mathbb{P}_c(\mathbb{F})} \llbracket F = 0 \rrbracket_d \right| = \Omega(Q_{\mathcal{B}}(f)). \tag{1}$$

(Where $\left| \vdash_{\mathbb{P}_c(\mathbb{F})} \llbracket F = 0 \rrbracket_d \right|$ is the minimal size of a $\mathbb{P}_c(\mathbb{F})$ proof containing all the circuit-equations in $\llbracket F = 0 \rrbracket_d$.)

The conditional lower bound we get is now similar to that in the previous sub-section, except that it holds for $\mathbb{P}_c$ and not only for fragments of $\mathbb{P}_c$:

Assumptions: for any fixed $d \geq 1$, assume that

1. there exists a family of identities of $f_n \in \mathbb{F}\langle X \rangle$ of $\mathrm{Mat}_d(\mathbb{F})$, with n variables, such that $Q_{\mathcal{B}}(f) = \Omega(n^d)$, for some basis $\mathcal{B}$ of the identities of $\mathrm{Mat}_d(\mathbb{F})$; and
2. the size of f_n is $O(n^r)$, for some constant $r < d$; and
3. Equation 12 holds.

Conclusion: A lower bound of $\Omega(n^{d-r})$ (in terms of the circuit equation proved) on the size of $\mathbb{P}_c(\mathbb{F})$ proofs.

We also present a straight forward *propositional version* of the Main Open Problem, by simply considering $\mathbb{F}$ to be $GF(2)$, adding to $\mathbb{P}_c(\mathbb{F})$ the Boolean axioms $x_i^2 + x_i$ and considering matrix identities over $\mathrm{Mat}_d(\mathbb{F})$ (see Section 6.2).

Conditional *exponential-size* lower bounds on arithmetic proofs. Assuming the answer to the Main Open Question above is positive (i.e., Equation 12 holds), we show under which further conditions we get *exponential-size* lower bounds on arithmetic proofs $\mathbb{P}_c(\mathbb{F})$.

The idea is simply to take the dimension d of the matrix algebras as a parameter by itself.

Assumptions:

1. Assume that for any d and any basis B_d of the identities of $\mathrm{Mat}_d(\mathbb{F})$ the number of lines in any $\mathbb{P}_c(\mathbb{F})$ proof of $\llbracket F = 0 \rrbracket_d$ is at least $\mathcal{C}_{B_d} \cdot Q_{B_d}(f)$, where $\mathcal{C}_{B_d}$ is a number depending on B_d (this is the Main Open Problem; where there $\mathcal{C}_{B_d}$ is a constant).
2. Assume that for any d and any basis B_d of the identities of $\mathrm{Mat}_d(\mathbb{F})$, there exists a number c_{B_d} such that for all sufficiently large n there exists an identity $f_{n,d}$ such that $Q_{B_d}(f_{n,d}) \geq c_{B_d} \cdot n^{2d}$. (The existence of such identities are known from our lower bound.)
3. Assume that $c_{B_d} \cdot \mathcal{C}_{B_d} = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right)$.
4. Assume that the algebraic circuit size of $f_{n,d}$ is is at most $\mathrm{poly}(n)$.

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

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

which (by assumption 4) is *exponential* in the size of the identities $f_{n,d}$ proved.

We wish to justify to a certain extent the new Assumptions 3 above (which lets us obtain the exponential lower bound). We shall use the special hard polynomials f that we proved exist in Theorem 5 for this purpose. First, note that Assumption 2 holds for the case of these f, by Theorem 5. We can also show (see Section 6.1) that the function c_{B_d} does not decrease too fast. That is, for the polynomial f we can prove the following:

$$Q_{\mathcal{B}_{n/4}}(f) = \Omega\left(\frac{2^n}{n^{5/2}\ln n}\right).$$

We can use the fact that c_{B_d} does not decrease too fast to get the following (conditional exponential lower bound):

Proposition 2. *Assume the Main Open Problem as written in Assumption 1 above and suppose that $\mathcal{C}_{B_{n/4}} = \Omega(1/\mathrm{poly}(n))$. Then, there exists a family of non-commutative circuits $\{F_n\}_{n=1}^{\infty}$ (computing the family of polynomials $\{f_n\}_{n=1}^{\infty}$) such that the number of lines in any $\mathbb{P}_c(\mathbb{F})$ proof of $[\![F_n = 0]\!]_{n/4}$ is at least $\mathcal{C}_{B_{n/4}}\Omega\left(\frac{2^n}{n^{5/2}}\ln n\right) = \Omega\left(\frac{2^n}{poly(n)}\right) = 2^{\Omega(n)}$.*

Note that this will give us a (conditional) exponential-size lower bound on $\mathbb{P}_c(\mathbb{F})$ proofs only if moreover the algebraic circuit size of $\{F_n\}_{n=1}^{\infty}$ is small enough (e.g., if Assumption 4 above holds). Though we do not believe that the algebraic circuit size of $\{F_n\}_{n=1}^{\infty}$ is small, the proposition above shows that at least potentially the parameters of our suggested framework can be accommodated to yield exponential lower bounds.

1.3 Summary

We summarize shortly several of the novel parts of our work:

1. The generalization of Hrubeš [6] work to "higher order commutativity axioms"; Obtaining a possible stronger lower bounds on proof systems;
2. The novel technical feature: the use of results from PI-theory to conclude the lower bound for any finite basis of the identities of $\mathrm{Mat}_d(\mathbb{F})$, for any d;
3. Putting forth, and formulating in a precise manner, the Main Open Problem: what is the relative efficiency between (i) proof systems establishing matrix identities by proving (non-commutative) identities whose variables range over matrices, and (ii) proof systems establishing matrix identities as entry-wise (commutative) polynomials.
4. Suggesting a new hierarchy of weaker and weaker (but not necessarily strictly) proof systems, that are fragments of arithmetic proofs; namely, the proof systems $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$, for increasing d's.

1.4 Connection to previous works

Apart from the connection to [6], we may consider the relation of the current work to the work of Hrubeš and Tzameret [8] that obtained polynomial-size (arithmetic and propositional) proofs for certain identities concerning matrices. As far as we see, there are no direct relations between these two works: in the current work we are studying matrix identities whose number of matrices (i.e., variables) grows with the number of variables n (if the number of matrices in the matrix identities over $\mathrm{Mat}_d(\mathbb{F})$ is m then the number of variables in the translation of the identities to a set of d^2 identities is $d^2 \cdot n$). Whereas in [8] the number of matrices was fixed and only the dimension of the matrices grows.

Note also that the matrix identities studied in [8] are not even translation (via $[\![\cdot]\!]$) of matrix identities over $\mathrm{Mat}_d(\mathbb{F})$: for instance

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \det\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \det\begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

is equal to $(ad-bc)\cdot(eh-fg) = (ae+bg)(cf+dh) - (af+bh)(ce+dg)$. But note that, e.g., in a translation of a matrix identity over $\mathrm{Mat}_d(\mathbb{F})$, variables of the same matrix cannot product each other.

2 Preliminaries

2.1 Algebras with polynomial identities

For a natural number n, put $[n] := \{1, 2, ..., n\}$. We use lower case letters a, b, c for constants from the underlying field, x, y, z for variables and $\overline{x}, \overline{y}, \overline{z}$ 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).

A polynomial is a formal sum of monomials, where a monomial is a product of (possibly non-commuting) variables and a constant from the underlying field. For two polynomials $f(x_1, \ldots, x_n)$ and g we say that g *is a substitution instance of* f if $g = f(h_1, \ldots, h_n)$ for some polynomials $h_1, \ldots, h_n$; and we sometimes denote $f(h_1, \ldots, h_n)$ 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 positive integer, we also use the notation $\overline{H}|_{H_j\leftarrow f}$, to denote the vector of polynomials that replace the j^{th} coordinate H_j in $\overline{H}$ by a polynomial $f \in \mathbb{F}\langle X\rangle$, where $j \in [k]$.

Definition 1. *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]

Perhaps the most prominent example of an $\mathbb{F}$-algebra is the algebra of $d \times d$ matrices, for some positive natural number d, with entries from $\mathbb{F}$ (with the usual addition and multiplication of

[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.

matrices). We denote this algebra by $\mathrm{Mat}_d(\mathbb{F})$. Note indeed that $\mathrm{Mat}_d(\mathbb{F})$ is an associative algebra but not a commutative one (i.e., the multiplication of matrices is non-commutative because AB does not necessarily equal BA, for two $d \times d$ matrices A, B).

Definition 2. *Let $\mathbb{F}\langle X\rangle$ denote the associative algebra of all polynomials such that the variables $X := \{x_1, x_2, \ldots\}$ are non-commutative with respect to multiplication. We call $\mathbb{F}\langle X\rangle$ the **free algebra (over X)**.*

For example, $x_1x_2 - x_2x_1 + x_3x_2x_3^2 - x_2x_3^3$, $x_1x_2 - x_2x_1$ and 0 are three distinct polynomials in $\mathbb{F}\langle X\rangle$.

Note that the set $\mathbb{F}\langle X\rangle$ forms a non-commutative ring. We sometimes call $\mathbb{F}\langle X\rangle$ *the ring of non-commutative polynomials* and call the polynomials from $\mathbb{F}\langle X\rangle$ *non-commutative polynomials.* Throughout this paper, unless otherwise stated, a polynomial is meant to be a non-commutative polynomial, namely a polynomial from the free algebra $\mathbb{F}\langle X\rangle$.

We now introduce the concept of a *polynomial identity algebra*, PI-algebra for short:

Definition 3. *Let A be an $\mathbb{F}$-algebra. An **identity of** A is a polynomial $f(x_1, ..., x_n) \in \mathbb{F}\langle X\rangle$ such that:*

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

*A **PI-algebra** is simply 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 $a, b \in A$, it holds that $ab - ba = 0$, and so $x_ix_j - x_jx_i$ 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 usually $\mathbb{F}[X]$.

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 $f_1, ..., f_n$ from I we have $\sum_{i=1}^{n} g_i \cdot f_i \cdot h_i \in I$, for all $g_1, ..., g_n, h_1, ..., h_n \in A$.

Definition 4. *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 $f(x_1, ..., x_n) \in \mathcal{T}$ then $f(g_1, ..., g_n) \in \mathcal{T}$, for any $g_1, ..., g_n \in \mathbb{F}\langle X\rangle$.

It is easy to see the following:

Fact 3. *The set of identities of an (associative) algebra is a T-ideal.*

A basis of a T-ideal $\mathcal{T}$ is a set of polynomials whose substitution instances generate $\mathcal{T}$ *as an ideal*:

Definition 5. *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:*

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

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

for $h_i, \ell_i, g_{i1}, ..., g_{in_i} \in \mathbb{F}\langle X\rangle$ 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 $B \subseteq \mathbb{F}\langle X\rangle$ if $\mathcal{T} = T(B)$.

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.

Note that the concept of a T-ideal is already somewhat 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 *homogenous* if all its monomials have the same total degree. Given a polynomial f, the *homogenous 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 the vector $(\overline{C})^{(j)}$ denotes the vector consisting of the jth-homogeneous parts of the circuits $C_1, C_2, \ldots, C_{2d}$.

Definition 6. *$S_d(x_1, x_2, \ldots, x_d)$ denotes the* ***standard identity*** *of degree d as follows:*

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

where $\mathcal{S}_d$ denotes the symmetric group on d elements and $sgn(\sigma)$ is the sign of the permutation σ.

For n polynomials $f_1, \ldots, f_n$ where $n \geq 2, n \in \mathbb{Z}$, we define the ***generalized-commutator*** $[f_1, \ldots, f_n]$ as follows:

$$[f_1, f_2] := f_1f_2 - f_2f_1, \qquad \text{(in case } n = 2)$$

$$\text{and} \qquad [f_1, \ldots, f_{n-1}, f_n] := [[f_1, \ldots, f_{n-1}], f_n], \qquad \text{for } n > 2.$$

A polynomial $f \in \mathbb{F}\langle X\rangle$ with n variables is homogenous *with degrees* $(1, \ldots, 1)$ (n times) if in every monomial the power of every variable $x_1, \ldots, x_n$ is precisely 1. In other words, every monomial is of the form $\alpha \cdot \prod_{i=1}^{n} x_{\sigma(i)}$, for some permutation σ of order n and some scalar α. For the sake of simplicity, we shall talk in the sequel about ***polynomial of degree*** n, when referring to polynomial with degrees $(1, \ldots, 1)$ (n times). Thus, any polynomial with n variables is homogenous of total-degree n.

2.2 Algebraic circuit

Definition 7. *Let $\mathbb{F}$ be a field, and let $X = \{x_1, \ldots, x_n\}$ be a set of input variables. An* ***arithmetic (or algebraic) circuit*** *is a directed acyclic graph, where the in-degree of nodes is at most 2. Every leaf of the graph (namely, a node of in-degree 0) is labelled with either an input variable or a field element. Every other node of the graph is labelled with either + or ×(in the first case the node is a sum-gate and in the second case a product-gate). Every edge in the graph is labelled with an arbitrary field element. A node of out-degree 0 is called an output-gate of the circuit.*

Every node and every edge in an *arithmetic circuit* computes a polynomial in the commutative polynomial-ring $\mathbb{F}[X]$ in the following way. A leaf just computes the input variable or field element that labels it. the sum of the polynomials computed by the two edges that reach it. A product-gate computes the product of the polynomials computed by the two edges that reach it. We say that a polynomial $g \in \mathbb{F}[X]$ is computed by the circuit if it is computed by one of the circuit's output-gates.

The size of a circuit Φ is defined to be the number of edges in Φ, and is denoted by $|\Phi|$.

Definition 8. *Let $\mathbb{F}$ be a field, and let $X = \{x_1, \ldots, x_n\}$ be a set of input variables. A **non-commutative arithmetic circuits** is similarly to the arithmetic circuits defined above, with the following additional feature: given any $\times$-gate of fanin 2, its children are labeled by a fixed order.*

Every node and every edge in a *non-commutative arithmetic circuit* computes a noncommutative polynomial in the free algebra $\mathbb{F}\langle X \rangle$ in exactly the same way as the arithmetic circuit does, except that at each $\times - gate$, the ordering among the children is taken into account in defining the polynomial computed at the gate.

The size of a noncommutative circuit Φ is also defined to be the number of vertices in Φ, and is denoted by $|\Phi|$.

3 The Complexity Measure

Let A be a PI-algebra (Definition 3) and let $\mathcal{T}$ be the T-ideal (Definition 4) consisting of all identities of A (see Fact 3). Assume that B is a basis for the T-ideal $\mathcal{T}$, that is, $T(B) = \mathcal{T}$. Then every $f \in \mathcal{T}$ is a consequence of B, namely, can be written as a linear combination of substitution instance of polynomials from B as follows:

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

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

A very natural question, from the complexity point of view, is the following: *What is the minimal number of distinct substitution instances $B_i(g_{i1}, \ldots, g_{in_i})$ of generators from B that must occur in (2)?* Or in other words, *how many distinct substitution instances of generators are needed to generate f above?*

Formally, we have the following:

Definition 9 ($Q_B(f)$)**.** *For a set of polynomials $B \subseteq \mathbb{F}\langle X \rangle$, define $Q_B(f)$ as the smallest (finite) k such that there exist substitution instances $g_1, g_2, \ldots, g_k$ of polynomials from B with*

$$f \in \langle g_1, g_2, \ldots, g_k \rangle,$$

where $\langle g_1, g_2, \ldots, g_k \rangle$ is the two-sided ideal generated by $g_1, g_2, \ldots, g_k$.

If the set B is a singleton $B = \{h\}$, we shall sometimes write $Q_h(\cdot)$ instead of $Q_{\{h\}}(\cdot)$.

Accordingly, we extend Definition 9 to a *sequence* of polynomials and let $Q_B(f_1, \ldots, f_n)$ be the smallest k such that there exist some substitution instances $g_1, g_2, \ldots, g_k$ of polynomials from B with

$$f_i \in \langle g_1, g_2, \ldots, g_k \rangle, \quad \text{for all } i \in [k].$$

Note that $Q_B(f)$ 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 the 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 problem, that goes by the name *The Specht Problem.* Fortunately, for matrix algebras we can use the solution of the Specht problem given by Kemer [11]. 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 find such a finite basis for most matrix algebras (namely for most 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 proof systems polynomially simulate each other; see e.g. [12]):

Proposition 4. *Let A be some $\mathbb{F}$-algebra and let B_0 and B_1 be two* finite *bases for the identities of A. Then, there exists a constant c (that depends only on B_0, B_1) such that for any identity f of A:*

$$Q_{B_0}(f) \le c \cdot Q_{B_1}(f).$$

Proof. Assume that $B_0 = \{A_1, A_2, \ldots, A_k\}$ and $B_1 = \{B_1, B_2, \ldots, B_\ell\}$. And suppose that $Q_{B_1}(f) = q$ 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 B_0 and B_1 are bases for A, there exists a constant r such that $B_{i_j} \in \left\langle A_{j_1}(\overline{h_{j_1}}), ..., A_{j_r}(\overline{h_{j_r}})\right\rangle$, 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 $l \in [r]$ (formally, $r = \max\{Q_{B_0}(B_i) \ : \ i \in [\ell]\}$).

Note that if $B_{i_j} \in \left\langle A_{j_1}(\overline{h_{j_1}}), \ldots, A_{j_r}(\overline{h_{j_r}})\right\rangle$, then for any substitution $\overline{g}_j$ (of polynomials to the variables X) we have $B_{i_j}(\overline{g_j}) \in \left\langle \left(A_{j_1}(\overline{h_{j_1}})\right)(\overline{g_j}), \ldots, \left(A_{j_r}(\overline{h_{j_r}})\right)(\overline{g_j})\right\rangle$. Thus, every $B_{i_j}(\overline{g_j})$ is generated by r substitution instances of polynomials from B_0, for any $j \in [q]$. Therefore, f can be generated with at most $r \cdot q$ substitution instances of generators from B_0, that is,

$$Q_{B_0}(f) \le r \cdot Q_{B_1}(f) \qquad \text{where } r = \max\{Q_{B_0}(B_i) \ : \ i \in [\ell]\}. \tag{3}$$

QED

4 Matrix Algebras

Hrubeš' work. For an identity f in a commutative algebra, we define the notation $Q_{\{[x,y]\}}(f)$ as the minimal number of substitution instances of the commutativity axioms $[x,y] = 0$ we need to generate f in the two-sided ideal.

For example, $Q_{[x,y]}(x_1x_2 - x_2x_1)$ is 1. And $Q_{[x,y]}(x_1x_2 - x_2x_1 + x_1x_3 - x_3x_1)$ is also 1 since the formula $x_1x_2 - x_2x_1 + x_1x_3 - x_3x_1$ equals $[x_2 + x_3, x_1]$. In [6] it was concluded that there is an identity f with n variables, such that:

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

We wish to extend this result to matrix algebras. Let $\mathrm{Mat}_d(\mathbb{F})$ denote the $d \times d$ matrix algebra over $\mathbb{F}$, that is, the set of all $n \times n$ matrices with entries from $\mathbb{F}$, with the usual operations of matrices. First of all, we extend the notation $Q_{[x,y]}(f)$, which only count the instances of one axiom, to the notation $Q_{A_1,A_2,\ldots,A_n}$ which count the instances of n axioms $A_1 = 0, A_2 = 0, \ldots, A_n = 0$.

Concerning matrix algebras, the following is the famous Amitsur-Levitzky Theorem:

Amitsur-Levitzki Theorem ([1]). *For any natural number d and any field $\mathbb{F}$ (in fact, any commutative ring) the standard identity $S_{2d}(x_1, x_2, \dots, x_{2d})$ of degree $2d$ is an identity of* $\mathrm{Mat}_d(\mathbb{F})$.

Further, it can be shown that $\mathrm{Mat}_d(\mathbb{F})$ does not have identities of degree smaller than $2d$. And that the identities of $\mathrm{Mat}_d(\mathbb{F})$ can be *finitely* generated [11]. That is, there must be a finite generating set for $\mathrm{Mat}_d(\mathbb{F})$. By Lemma 4 no matter which finite generating set $\{A_1, A_2, ..., A_k\}$ for $\mathrm{Mat}_d(\mathbb{F})$ we choose, the value $Q_{A_1,A_2,\dots A_k}$ is the same up to a constant factor.

Our main theorem is the following:

Theorem 5. *Let $\mathbb{F}$ be any field of characteristic 0. 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 $Q_{\mathcal{B}}(P) = \Omega(n^{2d})$.*

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.

4.1 The lower bound

We start by proving a lower bound on $Q_{S_{2d}}$, that is, we prove a lower bound on the number of substitution instances of S_{2d} identities needed to generate a certain identity (though S_{2d} is *not* known to be the basis of the T-ideal of the identities over $\mathrm{Mat}_d(\mathbb{F})$) .

Lemma 6. *For any natural $d \geq 1$ and any field $\mathbb{F}$ of characteristic 0 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: It can be shown that the lemma also holds for any finite *field* $\mathbb{F}$. Since in Section 4.1.3 we need to assume that the field is of characteristic 0, we prove the lemma only for fields of characteristic 0 .

For proving the lemma, we introduce the following definition:

Definition 10. *A polynomial $P \in \mathbb{F}\langle X\rangle$ with n variables $x_1, \dots, x_n$ is called an **s-polynomial** if:*

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

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

Lemma 7. *For any $P_1, P_2, \dots, P_{2d} \in \mathbb{F}\langle X\rangle$ where d is a positive integer, $S_{2d}(P_1, P_2, \dots, P_{2d})$ is the zero polynomial if there exists $i \in [2d]$ such that P_i is a constant.*

Proof. For a fixed $\mathcal{I} \in [2d]$, we have $P_{\mathcal{I}} = c \in \mathbb{F}$.

For convenience, write the set $\{x \in [2d] | x \neq \mathcal{I}\}$ as $[2d]/\mathcal{I}$, the permutation $\begin{pmatrix} 1 & 2 & \dots & m-1 & m & m+1 & \dots & 2d \\ i_1 & i_2 & \dots & i_{m-1} & \mathcal{I} & i_m & \dots & i_{2d-1} \end{pmatrix}$ as σ_m where $\{i_1, \dots, i_{2d-1}\} = [2d]/\mathcal{I}$.

Then

$$S_{2d}(P_1, P_2, \dots, P_{2d}) = \sum_{\sigma \in \mathcal{S}_{2d}} sgn(\sigma) \prod_{i=1}^{2d} P_{\sigma(i)}$$

$$
\begin{aligned}
&= \prod_{\{i_1,i_2,\dots,i_{2d-1}\}=[2d]/\mathcal{I}} \sum_{m=1}^{2d} sgn(\sigma_m) \prod_{j=1}^{m-1} P_{i_j} P_{\mathcal{I}} \prod_{j=m}^{2d-1} P_{i_j} \\
&= \prod_{\{i_1,i_2,\dots,i_{2d-1}\}=[2d]/\mathcal{I}} \sum_{m=1}^{2d} sgn(\sigma_m) c \prod_{j=1}^{2d-1} P_{i_j} \\
&= c \prod_{\{i_1,i_2,\dots,i_{2d-1}\}=[2d]/\mathcal{I}} \left(\sum_{m=1}^{2d} sgn(\sigma_m) \right) \prod_{j=1}^{2d-1} P_{i_j} \\
&= c \prod_{\{i_1,i_2,\dots,i_{2d-1}\}=[2d]/\mathcal{I}} \left(\sum_{m=1}^{d} (sgn(\sigma_{2m-1}) + sgn(\sigma_{2m})) \right) \prod_{j=1}^{2d-1} P_{i_j} \\
&= c \prod_{\{i_1,i_2,\dots,i_{2d-1}\}=[2d]/\mathcal{I}} \left(\sum_{m=1}^{d} 0 \right) \prod_{j=1}^{2d-1} P_{i_j} \\
&= 0.
\end{aligned}
$$

QED

Any s-polynomial has the following property:

Lemma 8. *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

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

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

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

That is

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

By Lemma 7, for some $j \in [r], i \in [2d]$, the polynomial $S_{2d}(\overline{P}_j)$ equals to the zero polynomial if some $\overline{P}_{ji}$ is a constant. Namely $S_{2d}(\overline{P_j})^{(2d)} = S_{2d} \left(\left(\overline{P_j}\right)^{(1)} \right)$, for all $j \in [r]$. Then,

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

That is,

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

Moreover,

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

Thus

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

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

4.1.1 The counting argument

Notation. *If $B \subseteq \mathbb{F}\langle X\rangle$ contains only one 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 9. *For any field $\mathbb{F}$ of characteristic 0, there exist s-polynomials $P_1,\ldots,P_n$ which are identities of $\mathrm{Mat}_d(\mathbb{F})$ in n variables, such that $Q_{S_{2d}}(P_1,\ldots,P_n) = \Omega(n^{2d})$ (and $Q_{S_{2d}}(P_1,\ldots,P_n)$ is finite).*

In Section 4.1.3 we show that, if $\mathbb{F}$ is of characteristic 0 then this lower bound holds for *any finite basis 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 [6] that there exists a sequence of polynomials $P_1, P_2, \ldots, P_n$ that require $\Omega\left(n^{2d}\right)$ substitution instances of the $S_{2d}(x_1,\ldots,x_{2d})$ identities to generate (all of the polynomials in the sequence) in a two-sided ideal.

Recall that an s-polynomial (Definition 10) is of the following form:

$$\sum_{j_1<j_2<\ldots<j_{2d}\in[n]} c_{i_{j_1j_2\cdots j_{2d}}} S_{2d}(x_{j_1},x_{j_2},\ldots,x_{j_{2d}}), \quad \text{where } c_{i_{j_1j_2\cdots j_{2d}}} \in \{0,1\}. \tag{4}$$

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 any choice of n s-polynomials $P_1,\ldots,P_n$ there are ℓ vectors of polynomials $\overline{Q_1},\ldots,\overline{Q_\ell}$ 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 8, for any choice of $P_1,\ldots,P_n$ and $\overline{Q}_1,\ldots,\overline{Q}_\ell$, for every $i\in[n]$:

$$P_i = \sum_{j=1}^{\ell} c_{i_j} S_{2d}\left(\overline{Q_j}^{(1)}\right) = \sum_{j=1}^{\ell} c_{i_j} S_{2d}\left(\sum_{m=1}^{n} a_{mj_1}x_m, \sum_{m=1}^{n} a_{mj_2}x_m, \ldots, \sum_{m=1}^{n} a_{mj_{2d}}x_m\right)$$

(for some $c_{i_j}, a_{mj_k} \in \mathbb{F}$).

Consider a vector $(c_{1_j},\ldots,c_{n_j},a_{k1m},\ldots,a_{k(2d)m})$ $(m\in[n], k\in[\ell])$. By linearity of S_{2d}:

$$\sum_{k=1}^{\ell} c_{i_k} S_{2d}\left(\sum_{m=1}^{n} a_{k1m}x_m, \sum_{m=1}^{n} a_{k2m}x_m, \ldots, \sum_{m=1}^{n} a_{k(2d)m}x_m\right) = \tag{5}$$

$$\sum_{j_1<j_2<\ldots<j_{2d}\in[n]} c_{i_{j_1j_2\cdots j_{2d}}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}) \qquad (\text{where } c_{i_{j_1j_2\cdots j_{2d}}} \in \mathbb{F}). \tag{6}$$

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

Claim. *Consider the coefficients $c_{1_j}, \ldots, c_{n_j}, a_{k1m}, \ldots, a_{k(2d)m}$ and the coefficients $c_{i_{j_1j_2\cdots j_{2d}}}$ in Equation 5* as variables. *Then, Equation 5 defines a degree-$(2d+1)$ polynomial map $\phi : \mathbb{F}^{(2d+1)nl} \to \mathbb{F}^{n\binom{n}{2d}}$ that maps each vector*

$$(c_{1_j}, \ldots, c_{n_j}, a_{k1m}, \ldots, a_{k(2d)m}), \quad \textit{for } m \in [n], k \in [\ell],$$

to

$$(c_{1_{j_1j_2\cdots j_{2d}}}, \ldots, c_{n_{j_1j_2\cdots j_{2d}}}), \quad \textit{for } j_1 < j_2 < \ldots < j_{2d} \in [n].$$

We omit the details of the proof of this claim. We have the following lemma:

Lemma 10 ([9], Lemma 5)**.** *For any field $\mathbb{F}$, if $\mu : \mathbb{F}^n \to \mathbb{F}^m$ is a polynomial map of degree $d > 0$, then $|\mu(\mathbb{F}^n) \bigcap \{0,1\}^m| \le (2d)^n$.*

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

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

Recall that for any choice of n s-polynomials $P_1, \ldots, P_n$ there are ℓ vectors of polynomials $\overline{Q_1}, \ldots, \overline{Q_\ell}$ from $\mathbb{F}\langle X\rangle$, such that

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

For convenience, we use $\overline{\mathcal{C}}$ for the $0-1$ vector $(c_{1_{j_1j_2\cdots j_{2d}}}, \ldots, c_{n_{j_1j_2\cdots j_{2d}}})$, where $c_{i_{j_1j_2\cdots j_{2d}}} \in \{0,1\}, i \in [n], j_1 < j_2 < \ldots < j_{2d} \in [n]$. Since for every possible $\overline{\mathcal{C}}$, the following polynomials are s-polynomials:

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

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

$$\sum_{j_1<j_2<\ldots<j_{2d}\in[n]} \mathcal{C}_{i_{j_1j_2\cdots j_{2d}}} S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}) \in \left\langle S_{2d}(\overline{Q_1}), \ldots, S_{2d}(\overline{Q_\ell})\right\rangle, i \in [n].$$

That is, there exists a vector $(c_{1_j}, \ldots, c_{n_j}, a_{k1m}, \ldots, a_{k(2d)m})$ $(m \in [n], k \in [\ell])$, such that $\phi(c_{1_j}, \ldots, c_{n_j}, a_{k1m}, \ldots, a_{k(2d)m}) = \overline{\mathcal{C}}$.

Therefore, every possible $\overline{\mathcal{C}}$ 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 $\overline{\mathcal{C}} = (c_{1_{j_1j_2\cdots j_{2d}}}, \ldots, c_{n_{j_1j_2\cdots j_{2d}}})$, where $c_{i_{j_1j_2\cdots j_{2d}}} \in \{0,1\}, i \in [n], j_1 <, \ldots, < j_{2d} \in [n]$. Hence,

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

This implies that

$$(2(2d+1))^{(2d+1)nl} \geq 2^{n\binom{n}{2d}}. \tag{7}$$

Using the ln function on both sides:

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

Hence,

$$l > \frac{\binom{n}{2d}\ln 2}{(2d+1)\ln(4d+2)}. \tag{8}$$

Namely

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

$$\text{(where } c \in \mathbb{F}\text{), hence}$$

$$l = \Omega\left(n^{2d}\right).$$

QED

4.1.2 Combining the polynomials into one

Here we show that there exists already a *single* polynomial, denoted $P^\star$ such that $Q_{S_{2d}}(P^\star) = \Omega(n^{2d})$. This is done in a manner which is similar to the work of Hrubeš [6]; however, there is a further complication here, which is dealt via the technical Lemma 12.

Lemma 11. *Let $P_1, \dots, P_n$ be s-polynomials in n variables $x_1, \dots, x_n$, and let $z_1, \dots, z_n$ be new variables, different from $x_1, \dots, x_n$. Let $P^\star := \sum_{i=1}^n z_i P_i$. Then*

$$Q_{S_{2d}}(P^\star) \geq \frac{1}{2d+1} Q_{S_{2d}}(P_1, \dots, P_n). \tag{9}$$

Specifically, for any field $\mathbb{F}$ of characteristic 0 and every $d \geq 1$, there exists a polynomial with n variables such that $Q_{S_{2d}}(P^\star) = \Omega(n^{2d})$.

Proof. For convenience, call the new variables $z_1, \dots, z_n$ the Z-variables. Given a polynomial f, the ***Z-homogenous 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-independent*.

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

Claim. *For any ℓ Z-independent polynomials $\overline{G}_1, \overline{G}_2, \dots, \overline{G}_\ell \in \mathbb{F}\langle X\rangle$, if*

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

then

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

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

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

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

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

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

■ Claim

In the following, 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

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

Namely

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

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

$$P^\star = \sum_{j=1}^{\ell}\sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{G}_j) g_{ji} \in \left\langle S_{2d}(\overline{G}_1), \ldots, S_{2d}(\overline{G}_{(2d+1)\cdot\ell})\right\rangle.$$

then we can, by the above claim, show that

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

which is the conclusion we want to prove:

$$Q_{S_{2d}}(P_1, \ldots, P_n) \le (2d+1)\cdot\ell.$$

Now, to find the $(2d+1)\cdot\ell$ Z-independent vectors of polynomials $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_{(2d+1)\cdot\ell}$ which generate $P^\star$, let $[\cdot]$ be a map that maps a polynomial $P \in \mathbb{F}\langle X\rangle$ to a polynomial $[P]$ that is defined by the following three properties:

1. The map $[\cdot]$ is linear, namely $[\alpha G + \beta H] = \alpha\,[G] + \beta\,[H]$ for any polynomials G, H and $\alpha, \beta \in \mathbb{F}$; and

2. Let M be a monomial whose Z-homogenous 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-independent. Then

$$[M] = [M_1 z M_2] = z M_2 M_1\,; \quad \text{and}$$

3. For a monomial M whose Z-homogenous part is not of degree 1, $[M] = 0$.

For convenience, in what follows, given the polynomials f, g and the vector of polynomials $\overline{H}$, we denote $(f)^0_Z, (\overline{H})^0_Z, (g)^0_Z$ by $\mathcal{F}, \overline{\mathcal{H}}, \mathcal{G}$, respectively.

Claim. *For any polynomials $f_1, g_1, \ldots, f_k, g_k$ and vector of polynomials $\overline{H}$ with variables $X_1, \ldots, X_n, z_1, \ldots, z_n$:*

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

Proof of claim: Consider the following:

$$\begin{aligned}
\left[\sum_{i=1}^{k} f_i S_{2d}(\overline{H}) g_i\right] &= \left[(\sum_{i=1}^{k} f_i S_{2d}(\overline{H}) g_i)^1_Z\right] \quad \text{by Property 3 of } [\cdot] \\
&= \left[\sum_{i=1}^{k} (f_i)^1_Z S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i + \sum_{i=1}^{k}\sum_{j=1}^{2d} \mathcal{F}_i S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow (H_j)^1_Z}\right) \mathcal{G}_i + \sum_{i=1}^{k} \mathcal{F}_i S_{2d}(\overline{\mathcal{H}})(g_i)^1_Z\right] \\
\text{(by linearity of } [\cdot]) \quad &= \sum_{i=1}^{k} \left[(f_i)^1_Z S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i\right] + \sum_{j=1}^{2d} \left[\sum_{i=1}^{k} \mathcal{F}_i S_{2d}\left(\overline{\mathcal{H}}|_{\mathcal{H}_j \leftarrow (H_j)^1_Z}\right) \mathcal{G}_i\right] + \sum_{i=1}^{k} \left[\mathcal{F}_i S_{2d}(\overline{\mathcal{H}})(g_i)^1_Z\right].
\end{aligned}$$

For every $i \in [n]$, assume $(f_i)^1_Z = h_1 z h_2$ where h_1, h_2 are Z-independent polynomials and z is a Z-variable, then

$$\left[(f_i)^1_Z S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i\right] = \left[h_1 z h_2 S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i\right] = z h_2 S_{2d}(\overline{\mathcal{H}}) \mathcal{G}_i h_1 \in \left\langle S_{2d}(\overline{\mathcal{H}})\right\rangle$$

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

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

By Lemma 12, which is proved below, we have

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

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

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

$$P^\star = [P^\star]$$

$$= \left[\sum_{j=1}^{\ell} \sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{H}_j) g_{ji} \right]$$
$$= \sum_{j=1}^{\ell} \left[\sum_{i=1}^{t_j} f_{ji} S_{2d}(\overline{H}_j) g_{ji} \right]$$
$$\in \left\langle S_{2d}(\overline{H}_j), S_{2d}(\overline{H}_j|_{H_{jq} \leftarrow \sum_{m=1}^{t_j} \mathcal{G}_{jm}\mathcal{F}_{jm}}) \right\rangle \quad \text{for any } j \in [\ell], q \in [2d].$$

Namely 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-independent polynomials that generate $P^\star$, concluding the theorem. QED

Lemma 12. *Let $X = \{x_1, x_2, \ldots, x_n\}$ and $f_1, g_1, \ldots, f_k, g_k \in \mathbb{F}\langle X \rangle$. Let $Z = \{z, z_1, z_2, \ldots, z_n\}$ and assume that n is an even positive integer, and let $\overline{P}$ be a vector of polynomials $(P_1, P_2, \ldots, P_n)$ with variable set $X \cup Z$. We denote $(\overline{P})_Z^0$, $(f_i)_Z^0, (g_i)_Z^0$ by $\overline{\mathcal{P}}, \mathcal{F}_i, \mathcal{G}_i, i \in [k]$, respectively. Then, for any $j \in [n]$, it holds that*

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

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

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

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

Proof. For a fixed $\mathcal{I} \in [n]$, we have $(P_{\mathcal{I}})_Z^1 = \mathcal{U} z \mathcal{V}$, where $z \in Z, \mathcal{U}, \mathcal{V} \in \mathbb{F}\langle X \rangle$ and $\mathcal{U}, \mathcal{V}$ are Z-independent.

For a permutation $\sigma \in \mathcal{S}_n$ and the polynomial vector $\overline{P} = (P_1, \ldots, P_n)$, we let

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

We write $\mathcal{S}_n/m$ to denote the set $\{\sigma \in \mathcal{S}_n \mid \sigma(m) = \mathcal{I}\}$.

And define

$$\pi_m = \begin{pmatrix} 1 & 2 & \ldots & n-m & n-m+1 & n-m+2 & \ldots & n \\ m+1 & m+2 & \ldots & n & m & 1 & \ldots & m-1 \end{pmatrix} \forall m \in [n].$$

Fact 13. $sgn(\pi_m) = (-1)^{m(n-m)+m-1} = (-1)^{nm-m(m-1)-1} = -1.$

Fact 14. $\overline{P}_{\sigma[m+1,n]} \cdot \overline{P}_{\sigma[1,m-1]} = \overline{P}_{\sigma\pi_m[1,n-m]} \cdot \overline{P}_{\sigma\pi_m[n-m+2,n]}$*, for all $\sigma \in \mathcal{S}_n/m$.*

Fact 15. $(\mathcal{S}_n/m)\pi_m = \mathcal{S}_n/(n-m+1)$.

So we have the following:

$$
\begin{aligned}
&\left[\sum_{i=1}^{k}\mathcal{F}_i s_n(\overline{\mathcal{P}}|_{\mathcal{P}_\mathcal{I}\leftarrow\mathcal{U}z\mathcal{V}})\mathcal{G}_i\right]\\
=&\left[\sum_{i=1}^{k}\mathcal{F}_i\sum_{\sigma\in\mathcal{S}_n} sgn(\sigma)(\overline{\mathcal{P}}_{\sigma[1,n]})|_{\mathcal{P}_\mathcal{I}\leftarrow\mathcal{U}z\mathcal{V}}\mathcal{G}_i\right]\\
=&\left[\sum_{i=1}^{k}\mathcal{F}_i\sum_{m=1}^{n}\sum_{\substack{\sigma\in\mathcal{S}_n\\ \sigma^{-1}(i)=m}} sgn(\sigma)(-1)^m(\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{P}_{\sigma(m)}\overline{\mathcal{P}}_{\sigma[m+1,n]})|_{\mathcal{P}_\mathcal{I}\leftarrow\mathcal{U}z\mathcal{V}}\mathcal{G}_i\right]\\
=&\left[\sum_{i=1}^{k}\mathcal{F}_i\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n/m} sgn(\sigma)(-1)^m(\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{P}_\mathcal{I}\overline{\mathcal{P}}_{\sigma[m+1,n]})|_{\overline{\mathcal{P}}_\mathcal{I}\leftarrow\mathcal{U}z\mathcal{V}}\mathcal{G}_i\right]\\
=&\left[\sum_{i=1}^{k}\mathcal{F}_i\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n/m} sgn(\sigma)(-1)^m(\overline{\mathcal{P}}_{\sigma[1,m-1]}\mathcal{U}z\mathcal{V}\overline{\mathcal{P}}_{\sigma[m+1,n]})\mathcal{G}_i\right]\\
=&z\mathcal{V}\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n/m} sgn(\sigma)(-1)^m\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}\\
=&z\mathcal{V}\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n/m} sgn(\sigma)(-1)^m\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]}\mathcal{U}\qquad\text{by }\textbf{Fact }14\\
=&z\mathcal{V}\sum_{m=1}^{n}\sum_{\sigma\in\mathcal{S}_n/m} sgn(\sigma\pi_m)sgn(\pi_m)(-1)^m\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]}\mathcal{U}.\\
&\text{let }\pi=\sigma\pi_m\text{, then }\pi\pi_m^{-1}=\sigma,\\
=&z\mathcal{V}\sum_{m=1}^{n}\sum_{\pi\pi_m^{-1}\in\mathcal{S}_n/m} sgn(\pi)(-1)(-1)^m\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]}\mathcal{U}\qquad\text{by }\textbf{Fact }13\\
=&-z\mathcal{V}\sum_{m=1}^{n}\sum_{\pi\in\mathcal{S}_n/(n-m+1)} sgn(\pi)(-1)^m\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]}\mathcal{U}\qquad\text{by }\textbf{Fact }15\\
&\text{let }m'=n-m+1\text{, then }m=n-m'-1,\\
=&-z\mathcal{V}\sum_{m'=1}^{n}\sum_{\pi\in\mathcal{S}_n/m'} sgn(\pi)(-1)^{n-m'+1}\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]}\mathcal{U}\\
=&-(-1)^{n+1}z\mathcal{V}\sum_{m'=1}^{n}\sum_{\pi\in\mathcal{S}_n/m'} sgn(\pi)(-1)^{m'}\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]}\mathcal{U}\\
=&z\mathcal{V}S_n(\overline{\mathcal{P}}|_{\mathcal{P}_\mathcal{I}\leftarrow\sum_{i=1}^{k}\mathcal{G}_i\mathcal{F}_i})\mathcal{U}
\end{aligned}
$$

$$\in \left\langle S_n(\overline{\mathcal{P}}|_{\mathcal{P}_{\mathcal{I}} \leftarrow \sum_{i=1}^{k} \mathcal{G}_i \mathcal{F}_i}) \right\rangle .$$

QED

4.1.3 Concluding the lower bound for every basis of the identities of $\mathrm{Mat}_d(\mathbb{F})$

Here we show that the $\Omega(n^{2d})$ lower bound proved in previous sections holds (for every $d > 2$ and) *every finite basis of the identities of* $\mathrm{Mat}_d(\mathbb{F})$, when $\mathbb{F}$ is of characteristic 0. To this end, we use several results from the theory of PI-algebras (for more on PI-theory see the monographs [18, 5]).

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

We need the following definition:

Definition 11. *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 commutators.)*

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

We need the following proposition:

Proposition 16 (Proposition 4.3.3 in [5])**.** *If R is a unitary PI-algebra over a field $\mathbb{F}$ of characteristic* 0*, then every identity of R can be generated by multi-homogenous commutator polynomials.*

Remark. Multi-homogenous *and* commutator polynomials*, in the current paper, are called* multilinear *and* proper polynomials *in [5], respectively.*

Lemma 17. *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 in which every polynomial is a multi-homogenous commutator polynomial.*

Proof. By Kemer [11], the identities of any $\mathbb{F}$-algebra, for any $\mathbb{F}$, can be generated by a finite set of identities. Namely $\mathcal{T}$ has a finite basis $\{A_1, , \dots, A_k\}$, for some positive integer k.

By Proposition 16, for a fixed identity of R, we can find finite many multi-homogenous commutator polynomials to generate. Thus, each A_i, $i \in [k]$, can be generated by finite many multi-homogenous commutator polynomials. Then there are finite many multi-homogenous commutator polynomials generating the basis $\{A_1, , \dots, A_k\}$ of $\mathcal{T}$, and hence, also finite many multi-homogenous commutator identities generating $\mathcal{T}$.

QED

Lemma 18. *Let $f \in \mathbb{F}\langle X\rangle$ be a multi-homogenous commutator polynomial with n variables. If x_i is a constant for some $i \in [n]$, then $f(x_1, \dots, x_n) \equiv 0$ (that is, f is the zero polynomial).*

Proof. In the proof, when we talk about the commutator, we mean the non-zero polynomial $[x_{t_1}, \dots, x_{t_s}]$ for all possible $t_1, \dots, t_s \in [n]$ and some natural number $s \geq 2$. It is easy to check that if we replace a variable by a constant $c \in \mathbb{F}$ in the commutator $[x_{t_1}, \dots, x_{t_s}]$, then the commutator equals 0.

By the definition of commutator polynomial, we know

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

where $0 \neq c_i \in \mathbb{F}$ and $m, n \in \mathbb{N}$, and B_{ij} is some commutator $[x_{i_1}, \ldots, x_{i_s}]$.

For a fixed $\mathcal{I} \in [n]$, by the definition of multi-homogenous polynomial, f must be linear in $x_{\mathcal{I}}$, namely $c_i \prod_{j=1}^{k_i} B_{ij}$ must be linear in $x_{\mathcal{I}}$ for every $i \in [m]$. Then there must be a $j_0 \in [k]$ such that B_{ij_0} is linear in $x_{\mathcal{I}}$. That is, $B_{ij_0}|_{x_{\mathcal{I}} \leftarrow c} = 0$. Furthermore, $\prod_{j=1}^{k_i} B_{ij}|_{x_{\mathcal{I}} \leftarrow c} = 0$ for all $i \in [m]$. Namely $f|_{x_{\mathcal{I}} \leftarrow c} = 0$. QED

By lemma 9 and lemma 11, we know that there exist s-polynomials $P_1, \ldots, P_n$ in n variables $x_1, \ldots, x_n$ that are identities over $\mathrm{Mat}_d(\mathbb{F})$, such that putting $P^\star := \sum_{i=1}^{n} z_i P_i$, where $z_1, \ldots, z_n$ are new variables, we have:

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

The following is the main lemma of this section:

Lemma 19. *Let $d > 2$, and let $\mathcal{B}$ be some basis for the T-ideals of the identities of $\mathrm{Mat}_d(\mathbb{F})$. Then, there are constants c, c' such that for any 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 theorem we need the following two lemmata.

Lemma 20. *For any natural number $d > 2$, every multi-homogenous identity (with any number of variables) over $\mathrm{Mat}_d(\mathbb{F})$ of degree at most $2d+1$ is a consequence of the standard identity S_{2d}.*

Proof. By Leron [14], we know that for any $d > 2$ every multi-homogenous identity of $\mathrm{Mat}_d(\mathbb{F})$ with degree $2d+1$ is a consequence of the standard identity S_{2d}. By Exercise 7.1.2 in [5], there are no identities of degree less than $2d$ in $\mathrm{Mat}_d(\mathbb{F})$ and every multi-homogenous polynomial identity of degree $2d$ in $\mathrm{Mat}_d(\mathbb{F})$ is also a consequence of the standard identity S_{2d}. QED

By Lemma 17, there is a basis $\{A_1, A_2, \ldots, A_m\}$ of $\mathrm{Mat}_d(\mathbb{F})$, where $A_1, \ldots, A_m$ are all multi-homogenous commutator identities (Definition 11).

Lemma 21. *Let P be an identity of $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$ and let G be a basis $\{A_1, A_2, \ldots, A_m\}$ of $\mathrm{Mat}_d(\mathbb{F})$, where $A_1, \ldots, A_m$ are all multi-homogenous commutator identities of $\mathrm{Mat}_d(\mathbb{F})$. And assume $Q_G(P) = k$, that is, k is the minimal number such that exist k substitution instances $B_1, B_2, \ldots, B_k$ of $A_1, A_2, \ldots, A_m$, for which:*

$$P \in \langle B_1, B_2, \ldots, B_k \rangle.$$

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

Proof. Assume there is A_j (for $j \in [m]$) in the basis G such that the degree of $A_j(\overline{x})$ is greater than $2d+1$. In the following, we show that none of B_ℓ ($\ell \in [k]$) is a substitution instance of A_j.

Assume otherwise. Hence, there is a $B_{\mathcal{I}}$, $\mathcal{I} \in [k]$, such that $B_{\mathcal{I}}$ is the substitution instance of A_j. Since $A_j(\overline{x})$ is homogeneous, every term in $A_j(\overline{x})$ is of degree greater than $2d+1$.

We consider the following two cases:

Case 1: Every term in the $A_j(\overline{Q})$, which is a substitution instances of $A_j(\overline{x})$, 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 homogenous parts of f of degree at most j. We consider the $2d+1$ homogenous part, that is:

$$\begin{aligned} P &= (P)^{(2d+1)} \\ &\in \left\{(h)^{(2d+1)} \mid h \in \langle B_1, B_2, \ldots, B_k\rangle\right\} \subset \left\langle (B_1)^{(\leq 2d+1)}, \ldots, (B_k)^{(\leq 2d+1)}\right\rangle. \end{aligned}$$

But $(B_{\mathcal{I}})^{(\leq 2d+1)} = \left(A_j(\overline{Q})\right)^{(\leq 2d+1)} = 0$, because, in this case, every term in $A_j(\overline{Q})$ is of degree greater than $2d+1$. So P can also belong to the ideal generated by $\left\{(B_1)^{(\leq 2d+1)}, (B_2)^{(\leq 2d+1)}, \ldots, (B_k)^{(\leq 2d+1)}\right\} \setminus (B_{\mathcal{I}})^{(\leq 2d+1)}$. This means $Q_G(P) = k-1$ which contradicts $Q_G(P) = k$. Thus the assumption is false.

Case 2: There is a term of degree at most $2d+1$ in $A_j(\overline{Q})$, which is a substitution instance of $A_j(\overline{x})$.

But we assumed that every term in $A_j(\overline{x})$ must be of degree greater than $2d+1$. This means one of the coordinates of $\overline{Q}$ must be a constant. That is, $A_j(\overline{Q}) = 0$ (by Lemma 18). So P can be generated by $\{B_1, B_2, \ldots, B_k\} \setminus B_i$. Hence, $Q_G(P) = k-1$, which contradicts $Q_G(P) = k$. Thus the assumption is false.

Now we can conclude that the assumption that there is a $B_{\mathcal{I}}$, $\mathcal{I} \in [k]$, such that $B_{\mathcal{I}}$ is a substitution instance of A_j is false. So none of B_ℓ ($\ell \in [k]$) is a substitution instance of A_j.

QED

We are now back to the proof of Lemma 19:

Proof. Let $\mathcal{B}$ be a basis $\{A_1, A_2, \ldots, A_m\}$ of $\mathrm{Mat}_d(\mathbb{F})$, where $A_1, \ldots, A_m$ are all multi-homogenous commutator identities of $\mathrm{Mat}_d(\mathbb{F})$. Let

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

For any identity P of $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$, by Lemma 21,

$$Q_{(\mathcal{B})^{(\leq 2d+1)}}(P) = Q_{\mathcal{B}}(P).$$

This also means that every identity of $\mathrm{Mat}_d(\mathbb{F})$ of degree at most $2d+1$ can be generated by $(\mathcal{B})^{(\leq 2d+1)}$. Thus, S_{2d} can be generated by $(\mathcal{B})^{(\leq 2d+1)}$. Write $(\mathcal{B})^{(\leq 2d+1)}$ as the set $\{A'_1, A'_2, \ldots, A'_{m'}\}$, $m' \leq m$, where the degree of A'_i ($\forall i \in [m']$) is less than $2d+1$. By Lemma 20, $A'_1, \ldots, A_{m'}$ is generated by S_{2d}. Then, by Equation 3 in Proposition 4, for any identity P over $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$:

$$\frac{1}{Q_{(\mathcal{B})^{(\leq 2d+1)}}(S_{2d})} Q_{S_{2d}}(P) \leq Q_{(\mathcal{B})^{(\leq 2d+1)}}(P) \leq \left(\max_{B \in \mathcal{B}'} Q_{S_{2d}}(B)\right) Q_{S_{2d}}(P) \quad d > 2. \tag{10}$$

Namely, for every identity P of $\mathrm{Mat}_d(\mathbb{F})$ of degree $2d+1$,there are constants c, c' such that:

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

QED

We can now conclude the main theorem of this section, Theorem 4.1.3, which we restate for convenience:

Theorem. *Let $\mathbb{F}$ be any field of characteristic 0. 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 $Q_{\mathcal{B}}(P) = \Omega(n^{2d})$.*

Note on the case of $d = 2$. When $d = 2$, Lemma 19 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 over $\mathrm{Mat}_2(\mathbb{F})$, but it is possible to show (though it is not entirely trivial) that f cannot be generated by S_4.

5 Relations to tensor-rank

Here we show that in order to make the hard (non-explicit) instances f from Theorem 5 into explicit ones, means finding explicit tensors with high tensor-rank. This generalizes (to any order) a similar observation made in [6] for order 3 tensors. This means that the *specific* hard instances we provide in Theorem 5 are not good candidates for proof complexity hardness, because it is reasonable to assume they do not have small size circuits.

Definition 12. *A tensor $A : [n]^r \to \mathbb{F}$ is a **simple tensor** if there exist r vectors $a_1, \ldots, a_r : [n] \to \mathbb{F}$ such that $A = a_1 \otimes \cdots \otimes a_r$, where $\otimes$ denotes tensor product, that is, A is defined by $A(i_1, i_2, \ldots, i_r) = a_1(i_1) \cdots a_r(i_r)$.*

Definition 13. *For a tensor A, the **tensor rank** $rank(A)$ is the minimal k such that there exist k simple tensors $A_1, A_2, \ldots, A_k : [n]^r \to \mathbb{F}$ such that $A = \sum_{i=1}^{k} A_i$.*

Definition 14. *For a natural number n, let A be a tensor $[n]^{r+1} \to \mathbb{F}$. We define the **corresponding polynomials** (from $\mathbb{F}\langle X\rangle$) **of the tensor** A as follows:*

$$f_{j_0} := \sum_{j_1, j_2, \ldots, j_r \in [n]} A(j_0, j_1, \ldots, j_r) S_r(x_{j_1}, x_{j_2}, \ldots, x_{j_r}), \quad \forall j_0 \in [n].$$

By the following theorem, if we find an collection of *explicit*[8] s-polynomials $f_1, \ldots, f_n$ over $\mathrm{Mat}_d(\mathbb{F})$ such that $Q_{S_{2d}}(f_1, \ldots, f_n)$ is $\Omega(n^{2d})$, then we can find an *explicit*[9] tensor $A : [n]^{2d+1} \to \{0, 1\}$ with rank $\Omega(n^{2d})$, where the corresponding polynomials of A are the s-polynomials $f_1, \ldots, f_n$.

Theorem 22. *For a natural number n, let $A_{f_1,\ldots,f_n}$ be a tensor $[n]^{r+1} \to \mathbb{F}$ and let $f_1, \ldots, f_n \in \mathbb{F}\langle X\rangle$ be the corresponding polynomials of $A_{f_1,\ldots,f_n}$, then:*

$$Q_{S_{2d}}(f_1, \ldots, f_n) \leq rank(A_{f_1,\ldots,f_n}).$$

[8] A polynomial is said to be *explicit* if the coefficient of a monomial of degree d is computable by algebraic circuits of size at most $\mathrm{poly}(d)$, where d is a natural number.

[9] A tensor $T : [n]^r \to \mathbb{F}$ is called explicit if $T(i_1, \ldots, i_r)$ can be computed by algebraic circuits of size at most polynomial in $\mathrm{poly}(r \lg n)$, that is, at most polynomial in the size of the input $(i_1, \ldots, i_r)$.}

Proof. Assume $rank(A_{f_1,\ldots,f_n}) = R$. Namely we can find R simple tensors $A_1, A_2, \ldots, A_R$ such that

$$A_{f_1,\ldots,f_n} = \sum_{i=1}^{R} A_i. \tag{11}$$

For every $i \in [R]$, by simple tensor's definition, there exist $2d+1$ vectors $a_0^{(i)}, a_1^{(i)}, \ldots, a_{2d}^{(i)} : [n] \to F$ such that $A_i = a_0^{(i)} \otimes a_1^{(i)} \otimes \cdots \otimes a_{2d}^{(i)}$. Namely $A_i(i_0, i_1, i_2, \ldots, i_{2d}) = a_0^{(i)}(i_0)a_1^{(i)}(i_1)\cdots a_{2d}^{(i)}(i_{2d})$, where $i_0, \ldots, i_{2d} \in [n]$.

Concerning the corresponding polynomials $f_1, \ldots, f_n$ of $A_{f_1,\ldots,f_n}$, for every $j_0 \in [n]$,

$$\begin{aligned}
f_{j_0} &= \sum_{j_1,j_2,\ldots,j_r \in [n]} A_{f_1,\ldots,f_n}(j_0, \ldots, j_{2d}) S_{2d}(x_{j_1}, \ldots, x_{j_{2d}}) \\
&= \sum_{j_1,j_2,\ldots,j_r \in [n]} \sum_{i=1}^{R} A_i(j_0, \ldots, j_{2d}) S_{2d}(x_{j_1}, \ldots, x_{j_{2d}}) \quad \text{(by 11)} \\
&= \sum_{i=1}^{R} \sum_{j_1,j_2,\ldots,j_r \in [n]} A_i(j_0, \ldots, j_{2d}) S_{2d}(x_{j_1}, \ldots, x_{j_{2d}}) \\
&= \sum_{i=1}^{R} a_0^{(i)}(j_0) \sum_{j_1,j_2,\ldots,j_r \in [n]} a_1^{(i)}(j_1) \cdots a_{2d}^{(i)}(j_{2d}) S_{2d}(x_{j_1}, x_{j_2}, \ldots, x_{j_{2d}}) \\
&= \sum_{i=1}^{R} a_0^{(i)}(j_0) S_{2d}\left(\sum_{1 \le j \le n} a_1^{(i)}(j)x_j, \sum_{1 \le j \le n} a_2^{(i)}(j)x_j, \ldots, \sum_{1 \le j \le n} a_{2d}^{(i)}(j)x_j \right) \\
&= \sum_{i=1}^{R} a_0^{(i)}(j_0) S_{2d}(\overline{P}_i)
\end{aligned}$$

(For convenience, write $\left(\sum_{1 \le j \le n} a_1^{(i)}(j)x_j, \sum_{1 \le j \le n} a_2^{(i)}(j)x_j, \ldots, \sum_{1 \le j \le n} a_{2d}^{(i)}(j)x_j\right)$ as $\overline{P}_i$, for any $i \in [R]$).

Namely

$$f_1, \ldots, f_n \in \left\langle S_{2d}\left(\overline{P}_1\right), \ldots, S_{2d}\left(\overline{P}_R\right) \right\rangle.$$

Thus $Q_{S_{2d}}(f_1, \ldots, f_n) \le R$, namely $Q_{S_{2d}}(f_1, \ldots, f_n) \le \text{rank}(A_{f_1,\ldots,f_n})$. QED

By the above theorem, we have the following:

Corollary 23. *If there exists a n explicit collection of s-polynomials $f_1, \ldots, f_n$ (that are all identities of) $\mathrm{Mat}_d(\mathbb{F})$, such that $Q_{S_{2d}}(f_1, \ldots, f_n) = \Omega(n^{2d})$, then there exists an explicit tensor $A : [n]^{2d+1} \to \{0, 1\}$ with tensor-rank $\Omega(n^{2d})$.*

6 Relations to proof complexity and the main open problem

Here we seek to find connections between the work we have done above to the problem of proving lower bounds in proof complexity.

Consider a matrix identity f over $\mathrm{Mat}_d(\mathbb{F})$. It is a non-commutative polynomial. Let f be a nonzero polynomial identity over $\mathrm{Mat}_d(\mathbb{F})$. Then f is a nonzero non-commutative polynomial from $\mathbb{F}\langle X\rangle$. If we substitute each (matrix) variable x_i in f by a $d\times d$ matrix of *entry-variables* $\{x_{ijk}\}_{j,k\in[n]}$, then now f corresponds to d^2 commutative zero polynomials, one for each entry computed by f. Accordingly, 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 commutative polynomials*. Formally, we define the set of d^2 non-commutative circuits corresponding to the non-commutative circuit F as follows:

Definition 15 ($[\![F]\!]_d$, $[\![F=0]\!]_d$)**.** *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 circuits which are generated from bottom to top in the circuit of F according to the following rules:*

1. *every variable x in F corresponds to d^2 new variables $x_{ij}, i,j\in[d]$;*
2. *every plus gate $X\oplus Y$, where X,Y represent two circuits, in F 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} which have been generated before;*
3. *every multiplication gate $X\otimes Y$ in F corresponds to d^2 plus gates $\oplus_{ij}, i,j\in[d]$ where each plus gate $\oplus_{ij}$ is connected to d multiplication gates $\otimes_k, k\in[d]$ which represent the multiplication of two corresponding circuit X_{ik} and Y_{kj} that have been generated before. (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 $\otimes_k$, $k\in[d]$.)*

We define $[\![F=0]\!]_d$ to be the set of equations between circuits, where each circuit in $[\![F]\!]_d$ equals the circuit 0.

Fact 24. *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, that is the number of nodes in F and $\left|[\![F]\!]_d\right|$ denotes the sum of size of all circuits in $[\![F]\!]_d$). Thus, if we fix the dimension of a matrix as a constant, then we can claim that $|[\![f]\!]_d| = \Theta(|f|)$.

First, we recall the arithmetic proof system $\mathbb{P}_c(\mathbb{F})$ (introduced in [8], and almost similarly in [7]) for deriving (commutative) polynomial identities over a field $\mathbb{F}$. The system manipulate arithmetic equations, that is, expressions of the form $F=G$ where F,G are circuits. Let $\mathbb{F}$ be a field. The system $\mathbb{P}_c(\mathbb{F})$ proves equations of the form $F=G$, where F,G are circuits over $\mathbb{F}$. The inference rules are:

$$\frac{F=G}{G=F} \qquad\qquad \frac{F=G \qquad G=H}{F=H}$$
$$\frac{F_1=G_1 \qquad F_2=G_2}{F_1+F_2=G_1+G_2} \qquad\qquad \frac{F_1=G_1 \qquad F_2=G2}{F_1\times F_2=G_1\times G_2}\,.$$

The axioms are equations of the following form, with F,G,H ranging over circuits:

$$Identity: \qquad F=F$$

$$\begin{array}{lll} \textit{Multiplication commutativity}: & F \cdot G = G \cdot F & \\ \textit{Addition commutativity}: & F + G = G + F & \\ \textit{Associativity}: & F + (G + H) = (F + G) + H & F \cdot (G \cdot H) = (F \cdot G) \cdot H \\ \textit{Distributivity}: & F \cdot (G + H) = F \cdot G + F \cdot H & \\ \textit{Zero element}: & F + 0 = F \quad F \cdot 0 = 0 & \\ \textit{Unit element}: & F \cdot 1 = F & \\ \textit{Field identities}: & c = a + b \quad d = a' \cdot b' & \end{array}$$

where in the Field identities $a, a', b, b', c, d \in \mathbb{F}$, such that the equations hold in $\mathbb{F}$.

$$\textit{Circuit axiom}: \quad F = F' \quad \text{if } F \text{ and } F' \text{ are (syntactically) identical when both are un-winded into } \textit{formulas}.$$

Note that the Circuit axiom can be verified in polynomial time (see e.g., [10]).

A proof π in $\mathbb{P}_c(\mathbb{F})$ is a sequence of equations $F_1 = G_1, F_2 = G_2, \ldots, F_k = G_k$, with F_i, G_i circuits, such that every equation is either an axiom, or was obtained from previous equations by one of the derivation rules. An equation $F_i = G_i$ appearing in a proof is also called a *proof-line*. Denote by $|\vdash_{\mathbb{P}_c(\mathbb{F})} F|$ the minimum number of lines in a $\mathbb{P}_c$ proof of $F = 0$. We say that π is a $\mathbb{P}_c$ proof of *a set* of equations if π is a $\mathbb{P}_c$ and it contains all the equations in the set as proof-lines).

For $\mathbb{F}$ an infinite field, f is an identity in $\mathrm{Mat}_d(\mathbb{F})$ iff $[\![F = 0]\!]_d$ has a $\mathbb{P}_c(\mathbb{F})$ proof. This is easy to prove as follows: assume by contradiction otherwise, then there must be an assignment A that makes $g \neq 0$. This follows since the field is infinite (and so every non zero polynomial has an assignment that does not nullifies the polynomial). But this assignment A (extended in any way to all entries) makes the matrix identity nonzero, in contradiction to the assumption that it is a matrix identity.

The main open question we raise in this work is the following:

Main Open Problem. *Let d be a positive natural number and let $\mathcal{B}$ be a (finite) basis of the T-ideal of the identities of $\mathrm{Mat}_d(\mathbb{F})$. Assume that $f \in \mathbb{F}\langle X\rangle$ is an identity over $\mathrm{Mat}_d(\mathbb{F})$, and let F be a non-commutative algebraic circuit computing f. Then, the minimal number of lines in an arithmetic proof of the collection of d^2 (entry-wise) equations $[\![F = 0]\!]_d$ corresponding to F is lower bounded (up to a constant factor) in $Q_{\mathcal{B}}(f)$. And in symbols:*

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

6.1 Conditions for exponential lower bounds

Can we, even potentially, obtain exponential lower bounds on $\mathbb{P}_c(\mathbb{F})$ proof size using the measure $Q_B(\cdot)$ and assuming the Main Open Question has a positive answer? The answer is yes, under certain assumptions. We write the assumptions formally:

Assumptions:

1. Assume that for any d and any basis B_d of the identities of $\mathrm{Mat}_d(\mathbb{F})$ the number of lines in any $\mathbb{P}_c(\mathbb{F})$ proof of $[\![F = 0]\!]_d$ is at least $\mathcal{C}_{B_d} \cdot Q_{B_d}(f)$, where $\mathcal{C}_{B_d}$ is a number depending on B_d (this is the Main Open Problem; where there $\mathcal{C}_{B_d}$ is a constant).

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

3. Assume that $c_{B_d} \cdot \mathcal{C}_{B_d} = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right)$.

4. Assume that the algebraic circuit size of $f_{n,d}$ is is at most $\mathrm{poly}(n)$.

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

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

which (by assumption 4) is *exponential* in the size of the identities $f_{n,d}$ proved.

We wish to justify to a certain extent the new Assumptions 3 above (which lets us obtain the exponential lower bound). We shall use the s-polynomials for this. First, note that Assumption 2 holds for the case of the s-polynomials, by Theorem 5.

We now show that the function c_{B_d} does not decrease too fast. By Equations 8,9 and 10 in Section 4.1, we know that for any natural number d, there is an s-polynomial f, such that:

$$Q_{B_d}(f) \geq \frac{1}{Q_{(B_d)^{(\leq 2d+1)}}(S_{2d})} \frac{1}{2d+1} \frac{\binom{n}{2d} \ln 2}{(2d+1)\ln(4d+2)}.$$

Let $\mathcal{B}_d$ be a set of identities of $\mathrm{Mat}_d(\mathbb{F})$ that contains the S_{2d} identities. Hence,

$$Q_{(\mathcal{B}_d)^{(\leq 2d+1)}}(S_{2d}) = 1.$$

Thus

$$Q_{\mathcal{B}_d}(f) \geq \frac{1}{2d+1} \frac{\binom{n}{2d} \ln 2}{(2d+1)\ln(4d+2)}.$$

If we let $d = n/4$, then

$$Q_{\mathcal{B}_{n/4}}(f) \geq \frac{1}{n/2+1} \frac{\binom{n}{n/2} \ln 2}{(n/2+1)\ln(n+2)}.$$

By Stirling's formula, we get that $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$. Hence, $\binom{n}{n/2} \sim \frac{2^{n+1/2}}{\sqrt{n\pi}}$. Then

$$Q_{\mathcal{B}_{n/4}}(f) = \Omega\left(\frac{2^n}{n^{5/2}\ln n}\right).$$

This shows that the function c_{B_d} does not decrease too fast.

We can use the fact that c_{B_d} does not decrease too fast to get the following (conditional exponential lower bound):

Proposition 25. *Assume the Main Open Problem as written in Assumption 1 above and suppose that* $\mathcal{C}_{B_{n/4}} = \Omega(1/\mathrm{poly}(n))$*. Then, there exists a family* $\{f_n\}_{n=1}^{\infty}$ *of s-polynomials such that the number of lines in any* $\mathbb{P}_c(\mathbb{F})$ *proof of* $[\![F_n = 0]\!]_{n/4}$ *is at least* $\mathcal{C}_{B_{n/4}}\Omega\left(\frac{2^n}{n^{5/2}}\ln n\right) = \Omega\left(\frac{2^n}{poly(n)}\right) = 2^{\Omega(n)}$*.*

(Note that we get a exponential lower bound for the lines of proofs in $\mathbb{P}_c$ in the above consequence without considering about the size of the s-polynomials.)

6.2 A propositional version of the Main Open Problem

We wish to comment on the applicability of our suggested framework, for achieving propositional Extended Frege lower bounds.

It seems that the most natural way to connect the complexity, measure $Q_{\mathcal{B}}(\cdot)$ to the number of lines in an Extended Frege (see, e.g., [12] or [10] for a formal definition of Extended Frege) proof is to require that the Main Open Problem states an *even stronger* statement. Admittedly, this makes the new assumption, shown below, quite speculative at the moment.

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, we can consider arithmetic proofs over $GF(2)$ augmented with the Boolean axioms $x_i^2 + x_i$, for each variables x_i, to obtain a propositional proof system which formally *is* an Extended Frege proof system (see [8]). Denote this system $\mathbb{P}_c(\mathbb{F}) + \{x_i^2 + x_i \ : \ x_i \in X\}$.

Then, there is no clear reason to rule out the following:

Main Open Problem—the Propositional Case over $GF(2)$**.** *Let* $\mathbb{F} = GF(2)$*, 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, the minimal number of lines in a* $\mathbb{P}_c(\mathbb{F}) + \{x_i^2 + x_i \ : \ x_i \in X\}$ *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* $Q_{\mathcal{B}}(f)$*. And in symbols:*

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

(Where, as before, $\left| \vdash_{\mathbb{P}_c(\mathbb{F})+\{x_i^2+x_i \,:\, x_i\in X\}} [\![F = 0]\!]_d \right|$ is the minimal size of a $\mathbb{P}_c(\mathbb{F}) + \{x_i^2 + x_i \ : \ x_i \in X\}$ proof containing all the circuit-equations in $[\![F = 0]\!]_d$.)

Comment: One can plausibly consider the same propositional version of the main open problem, with $\mathbb{F}$ being the rational numbers, and hence of characteristic 0 (for we which we have more knowledge about $Q_{\mathcal{B}}(\cdot)$, as obtained in our work). However, the way to translate arithmetic proofs $\mathbb{P}_c$ over the rationals is less immediate than the same translation for the case of $GF(2)$, and we have not verified formally the details of such a translation.

6.3 Proof systems for matrix identities

The proof system $\mathbb{P}_c(\mathbb{F})$ works for proving identities over commutative fields. Here we formulate a fragment of $\mathbb{P}_c(\mathbb{F})$ that proves matrix $\mathrm{Mat}_d(\mathbb{F})$ identities, for every given d. In what follows, $\mathbb{F}$ always denotes the field of characteristic 0.

We can define a new proof system $\mathbb{P}_{\mathrm{Mat}_d}(\mathbb{F})$ for proving identities over $\mathrm{Mat}_d(\mathbb{F})$. Since, for $d > 2$, the set of generators for the identities over $\mathrm{Mat}_d(\mathbb{F})$ are still not well understood, we shall formulate a system only for $\mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F})$, since for the identities of $\mathrm{Mat}_2(\mathbb{F})$ Drensky [4] has found a basis.

Definition 16 (The system $\mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F})$: proofs of identities over $\mathrm{Mat}_2(\mathbb{F})$)**.** $\mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F})$ *is the equational circuit proof system whose set of proper axioms consists of the following equations:*

$$
\begin{array}{lll}
\textit{Addition commutativity}: & f+g=g+f & \\
\textit{Associativity}: & f+(g+h)=(f+g)+f & f\cdot(g\cdot h)=(f\cdot g)\cdot h \\
\textit{Distributivity}: & f\cdot(g+h)=f\cdot g+f\cdot h & \\
\textit{Zero element}: & f+0=f & f\cdot 0=0 \\
\textit{Unit element}: & f\cdot 1=f & \\
\textit{Genertators}: & S_4(x,y,z,w)=0 & [[x,y]^2,z]=0 \\
\textit{Field identities}: & c=a+b & d=a'\cdot b'
\end{array}
$$

where in the Field identities $a, a', b, b', c, d \in \mathbb{F}$, such that the equations hold in $\mathbb{F}$.

Circuit axiom : $F = F'$ *if F and F' are (syntactically) identical when both are un-winded into* formulas.

Denote $\pi_{Mat_2}(f) : \mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F}) \vdash f = 0$ by the shortest proof for the equation $f = 0$ in system $\mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F})$ and denote $|\pi_{Mat_2}(f)|$ by the number of lines in $\pi_{Mat_2}(f)$.

7 Proof systems for identities of different algebras

We can consider proof systems for identities of algebras different than matrix algebras. Specifically, we can enlarge the language of polynomial identities from $\mathbb{F}\langle X\rangle$ to the *free trace polynomial algebra* $Tr\mathbb{F}\langle X\rangle$, as we now describe (see Razmyslov [16]).

First, define the trace function $Tr(\cdot) : \mathbb{F}\langle X\rangle \to \mathbb{F}$ as a function with the following congruence:

$$
\begin{aligned}
&[Tr(f), g] = 0,\\
&[Tr(f), Tr(g)] = 0,\\
&Tr(fg) = Tr(gf),\\
&Tr(\alpha f + \beta g) = \alpha Tr(f) + \beta Tr(g),
\end{aligned}
$$

where f and g range over $\mathbb{F}\langle X\rangle$ and α and β range over the field $\mathbb{F}$. For any $k \in \mathbb{N}$ and any $B_1, \ldots, B_k \in \mathbb{F}\langle X\rangle$, define the ***trace monomial*** as the following product:

$$B_1 Tr(B_2) Tr(B_3) \cdots Tr(B_k).$$

A ***trace polynomial*** is defined to be a sum of trace monomials.

Definition 17. *Let $Tr\mathbb{F}\langle X\rangle$ denote the associative algebra of trace polynomials such that the variables $X := \{x_1, x_2, \ldots\}$ are non-commutative with respect to multiplication. We call $Tr\mathbb{F}\langle X\rangle$ the* ***free trace polynomial algebra (over X)****. (More precisely, we have distributivity and associativity of product; and commutativity and associativity of additions in the free trace polynomial algebra, which are defined in the same way as for the free algebra $\mathbb{F}\langle X\rangle$.)*

From now on, we only talk about the free trace polynomial over a matrix algebra, and we will only consider the function $Tr(\cdot)$ *as the ordinary trace function.* Namely for a matrix M, the value $Tr(M)$ is the sum of diagonal elements of M.

Then a trace polynomial $f(x_1,\ldots,x_n) \in Tr\mathbb{F}\langle X\rangle$ is called a *trace identity* of $\mathrm{Mat}_d(\mathbb{F})$ if $f(B_1,\ldots,B_n)=0$ for any matrices $B_1,\ldots,B_n \in \mathrm{Mat}_d(\mathbb{F})$.

It is easy to see that the free algebra $\mathbb{F}\langle X\rangle$ is contained in $Tr\mathbb{F}\langle X\rangle$. Actually we construct the free trace polynomial algebra $Tr\mathbb{F}\langle X\rangle$ as a generalized free algebra, which will play the same role for trace identities as the free algebra plays for ordinary identities.

Furthermore, a two-sided ideal $\mathscr{T}$ of the free trace polynomial algebra $Tr\mathbb{F}\langle X\rangle$ is called a *trace T-ideal* if for any trace polynomial $f(x_1,\ldots,x_n)$ contained in $\mathscr{T}$ and any $g_1,\ldots,g_n \in \mathbb{F}\langle X\rangle$, the trace polynomial $f(g_1,\ldots,g_n)$ is contained in $\mathscr{T}$. Let $B \in Tr\mathbb{F}\langle X\rangle$ be a set of trace polynomials and let $\mathscr{T}$ be a trace T-ideal. We say that $\mathscr{T}$ is generated by B, if every $f \in \mathscr{T}$ can be written as:

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

for $h_i,\ell_i \in Tr\mathbb{F}\langle X\rangle, g_{i1},...,g_{in_i} \in \mathbb{F}\langle X\rangle$ and $B_i \in B$ (for all $i \in I$). A trace identity $f=0$ is called *a consequence of (or generated by)* the trace identities $g=0$, where g ranges over B, if $f \in \mathscr{T}$.

We have the following theorem:

Theorem 26 (Razmyslov [16], Theorem 2)**.** *All trace identities of the* $\mathrm{Mat}_d(\mathbb{F})$ *are consequences of the Cayley-Hamilton identity* f_d *which can be computed by the following recurrence:*

$$f_1 = y - Tr(y), \qquad f_n = f_{n-1}y - \frac{1}{n}Tr(f_{n-1}y), \quad n \geq 2.$$

We can thus construct a proof system $\mathbb{P}_{\mathrm{Tr}_d}(\mathbb{F})$ for proving the trace identities over $\mathrm{Mat}_d(\mathbb{F})$. For the sake of comparison with $\mathbb{P}_{\mathrm{Mat}_2}(\mathbb{F})$, we shall give only the definition of $\mathbb{P}_{\mathrm{Tr}_2}(\mathbb{F})$, as follows:

Definition 18 (The system $\mathbb{P}_{\mathrm{Tr}_2}(\mathbb{F})$: proofs of trace identities over $\mathrm{Mat}_2(\mathbb{F})$)**.** $\mathbb{P}_{\mathrm{Tr}_2}(\mathbb{F})$ *is the arithmetic proof system operating with equations between circuits whose set of proper axioms consists of the following equations:*

$$\begin{array}{ll}
\textit{Addition Commutativity}: & f+g=g+f \\
\textit{Associativity}: & f+(g+h)=(f+g)+h \qquad f\cdot(g\cdot h)=(f\cdot g)\cdot h \\
\textit{Distributivity}: & f\cdot(g+h)=f\cdot g+f\cdot h \\
\textit{Zero element}: & f+0=f \qquad f\cdot 0=0 \\
\textit{Unit element}: & f\cdot 1=f \\
\textit{Genertator}: & x^2-Tr(x)x+\frac{1}{2}(Tr^2(x)-Tr(x^2))I=0 \\
\textit{Trace Commutativity}: & [Tr(f),Tr(g)]=0 \qquad [Tr(f),g]=0 \\
 & Tr(f\cdot g)=Tr(g\cdot f) \\
\textit{Trace Linearity}: & Tr(\alpha f+\beta g)=\alpha Tr(f)+\beta Tr(g) \\
 & \textit{where } f \textit{ and } g \textit{ range over } \mathbb{F}\langle X\rangle \textit{and } \alpha \textit{ and } \beta \textit{ range over the field } \mathbb{F} \\
\textit{Field identities}: & c=a+b \qquad d=a'\cdot b'
\end{array}$$

where in the Field identities $a,a',b,b',c,d \in \mathbb{F}$*, such that the equations hold in* $\mathbb{F}$*.*

$$\textit{Circuit axiom}: \qquad F = F' \quad \textit{if } F \textit{ and } F' \textit{ are (syntactically) identical when both are un-winded into } \text{formulas}.$$

Denote by $\pi_{Tr_2}(f) : \mathbb{P}_{\mathrm{Tr}_2}(\mathbb{F}) \vdash f = 0$ the smallest proof of the equation $f = 0$ in the system $\mathbb{P}_{\mathrm{Tr}_2}(\mathbb{F})$ and denote by $|\pi_{Tr_2}(f)|$ the number of lines in $\pi_{Tr_2}(f)$. Also, denote by $|\pi_C([\![f]\!]_d)|$ the minimal number of lines in a $\mathbb{P}_c(\mathbb{F})$ proof of all the equations in $[\![f]\!]_d$. We have the following:

Proposition 27.

$$|\pi_{Mat_2}(f)| = \Omega(|\pi_{Tr_2}(f)|).$$
$$|\pi_{Mat_2}(f)| = \Omega(|\pi_C([\![f]\!]_d)|).$$

Observation: we can obtain the standard identity S_4 and the *Hall identity* $[[x,y]^2, z]$ from the Cayley-Hamilton Theorem by constant many steps.

We now prove this observation. For the standard identity: since for any matrix variable x which belongs to $Mat_2(\mathbb{F})$, we already know the following polynomial is zero polynomial

$$P(x) := x^2 - Tr(x)x + \frac{1}{2}(Tr^2(x) - Tr(x^2))I.$$

And we linearize the identity and get the following

$$P(x+y)-P(x)-P(y) = (xy+yx)-(Tr(x)y+Tr(y)x)+\frac{1}{2}((Tr(x)Tr(y)+Tr(y)Tr(x))-Tr(xy+yx))I.$$

Since

$$Tr(x)Tr(y) = Tr(y)Tr(x), Tr(xy) = Tr(yx),$$

we see that $M_2(\mathbb{F})$ also satisfies the trace identity:

$$f(x,y) = (xy + yx) - (Tr(x)y + Tr(y)x) + (Tr(x)Tr(y) - Tr(xy))I.$$

Now we replace x, y by $z_{\sigma_1}z_{\sigma_2}, z_{\sigma_3}z_{\sigma_4}$ where σ is a permutation from the symmeTric group $\mathcal{S}_4$:

$$\begin{aligned} 0 &= \sum_{\sigma\in\mathcal{S}_4} sgn(\sigma)f(z_{\sigma_1}z_{\sigma_2}, z_{\sigma_3}z_{\sigma_4}) \\ &= \sum_{\sigma\in\mathcal{S}_4} sgn(\sigma)[z_{\sigma_1}z_{\sigma_2}z_{\sigma_3}z_{\sigma_4} + z_{\sigma_3}z_{\sigma_4}z_{\sigma_1}z_{\sigma_2} - \\ &\quad Tr(z_{\sigma_1}z_{\sigma_2})z_{\sigma_3}z_{\sigma_4} - Tr(z_{\sigma_3}z_{\sigma_4})z_{\sigma_1}z_{\sigma_2} + \\ &\quad Tr(z_{\sigma_1}z_{\sigma_2})Tr(z_{\sigma_3}z_{\sigma_4})I - Tr(z_{\sigma_1}z_{\sigma_2}z_{\sigma_3}z_{\sigma_4})I] \\ &= 2\sum_{\sigma\in\mathcal{S}_4} sgn(\sigma)z_{\sigma_1}z_{\sigma_2}z_{\sigma_3}z_{\sigma_4}. \end{aligned}$$

For the Hall identity:

$$P([x,y]) = [x,y]^2 - Tr([x,y])[x,y] + \frac{1}{2}(Tr^2([x,y]) - Tr([x,y]^2))I.$$

Since

$$Tr([x,y]) = Tr(xy) - Tr(yx) = 0,$$
$$P([x,y]) = [x,y]^2 - \frac{1}{2}Tr([x,y]^2)I.$$

Namely $[x,y]^2 = \frac{1}{2}Tr([x,y]^2)I$, that is $[[x,y]^2, z] = [\frac{1}{2}Tr([x,y]^2)I, z] = 0$.

Acknowledgements

We wish to thank V. Arvind, Michael Forbes, Kristoffer Arnsfelt Hansen, Jan Krajíček, Periklis Papakonstantinou, Youming Qiao, Ran Raz 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.

References

[1] S. A. Amitsur and J. Levitzki. Minimal identities for algebras. In *Proc. Amer. Math. Soc. (2)*, pages 449–463, 1950. 1.2.1, 4

[2] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996)*, pages 174–183, New York, 1996. ACM. 3

[3] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44(1):36–50, 1979. 1.1

[4] Vesselin Drensky. A minimal basis of identities for a second-ordermatrix algebra over a field of characteristic 0. *Algebra i Logika*, 20(3):291–299, May–June 1981. Translation. 1.2.1, 6.3

[5] Vesselin Drensky. *Free Algebras and PI-Algebras.* Springer-Verlag, Singapore, 1999. 1.2.1, 4.1.3, 16, 4.1.3, 4.1.3

[6] Pavel Hrubeš. How much commutativity is needed to prove polynomial identities? *Electronic Colloquium on Computational Complexity, ECCC*, (Report no.: TR11-088), June 2011. (document), 1.2, 1.2.1, 1, 1.2.1, 1.2.2, 1, 1.4, 4, 4.1.1, 4.1.2, 5

[7] Pavel Hrubeš and Iddo Tzameret. The proof complexity of polynomial identities. In *Proceedings of the 24th IEEE Conference on Computational Complexity (CCC)*, pages 41–51, 2009. 1.1, 6

[8] Pavel Hrubeš and Iddo Tzameret. Short proof for the determinant identities. In *Proceedings of the 44th Annual ACM Symposium on the Theory of Computing (STOC)*, New York, 2012. ACM. 1.1, 1.4, 6, 6.2

[9] Pavel Hrubeš and Amir Yehudayoff. Arithmetic complexity in ring extensions. *Theory of Computing*, 7:119–129, 2011. 10

[10] Emil Jeřábek. Dual weak pigeonhole principle, Boolean complexity, and derandomization. *Ann. Pure Appl. Logic*, 129(1-3):1–37, 2004. 1.1, 1, 6, 6.2

[11] Alexander Kemer. Finite basability of identities of associative algebras. *Algebra i Logika*, 26(5):597–641, 650, 1987. (document), 1.2.1, 3, 4, 4.1.3

[12] Jan Krajíček. *Bounded arithmetic, propositional logic, and complexity theory*, volume 60 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1995. 1.1, 3, 6.2

[13] Jan Krajíček. *Forcing with random variables and proof complexity*, volume 382 of *London Mathematical Society Lecture Notes Series*. Cambridge Press, 2010. 1.1

[14] Uri Leron. Multilinear identities of the matrix ring. *Transactions of the American Mathematical Society*, 183:175–202, Sep. 1973. 4.1.3

[15] Pavel Pudlák. Twelve problems in proof complexity. In *Proceedings of CSR*, 2008. 1.1

[16] Ju. P. Razmyslov. Identities with trace in full matrix algebras over a field of characteristic. *zero. Izv. Akad. Nauk SSSR Ser. Mat.*, 38:723–756, 1974. 7, 26

[17] Yu P Razmyslov. Trace identities and central polynomials in the matrix superalgebras $m_{n,k}$. *Sbornik: Mathematics*, 56(1):187–206, 1987. 1.2.2

[18] Louis Halle Rowen. *Polynomial identities in ring theory*. Pure and Applied Mathematics. Academic Press, 1980. 1.2.1, 4.1.3

[19] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, 1980. 1.1

[20] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation*, pages 216–226. Springer-Verlag, 1979. 1.1

ECCC ISSN 1433-8092

http://eccc.hpi-web.de