

# Black-Box Identity Testing of Noncommutative Rational Formulas in Deterministic Quasipolynomial Time

V. Arvind\*    Abhranil Chatterjee<sup>†</sup>    Partha Mukhopadhyay<sup>‡</sup>

September 27, 2023

## Abstract

Rational Identity Testing (RIT) is the decision problem of determining whether or not a given noncommutative rational formula computes zero in the free skew field. It admits a deterministic polynomial-time *white-box* algorithm [GGdOW20, IQS18, HH21], and a randomized polynomial-time black-box algorithm [DM17] via singularity testing of linear matrices over the free skew field.

Designing a subexponential-time deterministic RIT algorithm in *black-box* is a major open problem in this area. Despite being open for several years, this question has seen very limited progress. In fact, the only known result in this direction is the construction of a quasipolynomial-size hitting set for rational formulas of only *inversion height* two [ACM22].

In this paper, we settle this problem and obtain a deterministic quasipolynomial-time RIT algorithm for the general case in the black-box setting. Our algorithm uses ideas from the theory of finite dimensional division algebras, algebraic complexity theory, and the theory of generalized formal power series.

---

\*Institute of Mathematical Sciences (HBNI), and Chennai Mathematical Institute, Chennai, India. Email: arvind@imsc.res.in.

<sup>†</sup>Indian Statistical Institute, Kolkata, India. Email: abhneil@gmail.com. Research Supported by INSPIRE Faculty Fellowship provided by the Department of Science and Technology, Government of India.

<sup>‡</sup>Chennai Mathematical Institute, Chennai, India. Email: partham@cmi.ac.in.

# 1 Introduction

The goal of algebraic circuit complexity is to understand the complexity of computing multivariate polynomials and rational functions using basic arithmetic operations, such as additions, multiplications, and inverses. Algebraic formulas and algebraic circuits are some of the well-studied computational models.

In the commutative setting, the role of inverses is well understood, but in noncommutative computation, it is quite subtle. To elaborate, it is known that *any* commutative rational expression can be expressed as  $fg^{-1}$  where  $f$  and  $g$  are two commutative polynomials [Str73]. However, noncommutative rational expression such as  $x^{-1} + y^{-1}$  or  $xy^{-1}x$  cannot be represented as  $fg^{-1}$  or  $f^{-1}g$  for any noncommutative polynomials  $f$  and  $g$ . Therefore, the presence of *nested inverses* makes a rational expression more complicated, for example  $(z + xy^{-1}x)^{-1} - z^{-1}$ . Another issue is that a noncommutative rational expression is not always defined on a matrix substitution. For a noncommutative rational expression  $r$ , its *domain of definition* is the set of matrix tuples (of any dimension) where  $r$  is defined. Two rational expressions  $r_1$  and  $r_2$  are *equivalent* if they agree on every matrix substitution in the intersection of their domain of definition. This induces an equivalence relation on the set of all noncommutative rational expressions (with nonempty domain of definition). Interestingly, this computational definition was used by Amitsur in the characterization of the *universal* free skew field [Ami66]. The free skew field consists of these equivalence classes, called *noncommutative rational functions*. One can think of the free skew field  $\mathbb{F}\langle x_1, \dots, x_n \rangle$  is the smallest field that contains the noncommutative polynomial ring  $\mathbb{F}\langle x_1, \dots, x_n \rangle$ . It has been extensively studied in mathematics [Ami66, Coh71, Coh95, FR04].

The complexity-theoretic study of this model was initiated by Hrubeš and Wigderson [HW15]. Computationally (and in this paper), noncommutative rational functions are represented by algebraic formulas using addition, multiplication, and inverse gates over a set of noncommuting variables, called noncommutative rational formulas. They also addressed the *rational identity testing* problem (RIT): decide efficiently whether a given noncommutative rational formula  $r$  computes the zero function in the free skew field. Equivalently, the problem is to decide whether  $r$  is zero on its domain of definition, follows from Amitsur's characterization. For example, the rational expression  $(x + xy^{-1}x)^{-1} + (x + y)^{-1} - x^{-1}$  is a rational identity, known as Hua's identity [Hua49]. Rational expressions exhibit peculiar properties which seem to make the RIT problem quite different from the noncommutative polynomial identity testing. For example, Bergman has constructed an explicit rational formula, of inversion height two, which is an identity for  $3 \times 3$  matrices but not an identity for  $2 \times 2$  matrices [Ber76]. Also, the apparent lack of *canonical representations*, like a sum of monomials representation for polynomials, and the use of nested inverses in noncommutative rational expressions complicate the problem. This motivates the definition of *inversion height* of a rational formula which is the maximum number of inverse gates in a path from an input gate to the output gate. The *inversion height* of a rational function is the minimum over the inversion heights of the formulas representing the function. For example, consider the rational expression  $(x + xy^{-1}x)^{-1}$ . Even though it has a nested inverse, it follows from Hua's identity that it represents a rational function of inversion height one. In fact, Hrubeš and Wigderson obtain the following interesting bound on the inversion height of a formula [HW15].

**Fact 1.** *For any noncommutative  $n$ -variate rational function  $r$  computed by a rational formula of size  $s$  there is a rational formula of size  $\text{poly}(n, s)$  of inversion height  $O(\log s)$ .*

Consequently, in the black-box setting we can assume w.l.o.g. that the input rational formula of some size  $s$  has inversion height  $O(\log s)$ .

Hrubeš and Wigderson have given an efficient reduction from the RIT problem to the singularity testing problem of linear matrices in noncommuting variables over the free skew field (NSINGULAR). It is the noncommutative analogue of Edmonds' problem of symbolic determinant identity testing (SINGULAR). While SINGULAR can be easily solved in randomized polynomial time using Polynomial Identity Lemma [DL78, Zip79, Sch80], finding a deterministic algorithm remains completely elusive [KI04].

Remarkably, NSINGULAR  $\in P$  thanks to two independent breakthrough results [GGdOW16, IQS18]. In particular, the algorithm of Garg, Gurvits, Oliveira, and Wigderson [GGdOW16] is analytic in nature and based on operator scaling which works over  $\mathbb{Q}$ . The algorithm of Ivanyos, Qiao, and Subrahmanyam [IQS18] is purely algebraic. Moreover, the algorithm in their paper [IQS18] works over  $\mathbb{Q}$  and fields with positive characteristics. Subsequently, a third algorithm based on convex optimization is also developed by Hamada and Hirai [HH21]. Not only are these beautiful results, but they have also enriched the field of computational invariant theory greatly [BFG<sup>+</sup>19, DM20, MW19]. As a consequence, RIT can also be solved in deterministic polynomial time in the *white-box* setting. Both the problems admit a randomized polynomial-time black-box algorithm due to Derksen and Makam [DM17].

Two central open problems in this area are to design faster deterministic algorithms for the NSINGULAR problem and RIT problem in the black-box setting, raised in [GGdOW16]. The algorithms in [GGdOW16] and [IQS18] are inherently sequential and they are unlikely to be helpful for designing a black-box algorithm. It is well-known [GGdOW16] that an efficient black-box algorithm (via a hitting set construction) for NSINGULAR would generalize the celebrated quasi-NC algorithm for bipartite matching significantly [FGT21]. Even for the RIT problem (which could be easier than the NSINGULAR problem), the progress towards designing an efficient deterministic black-box algorithm is very limited. In fact, only very recently a deterministic quasipolynomial-time black-box algorithm for identity testing of rational formulas of inversion height two has been designed [ACM22]. Another very recent result shows that certain ABP (algebraic branching program)-hardness of polynomial identities (PI) for matrix algebras will lead to a black-box derandomization of RIT in almost general setting [ACG<sup>+</sup>22]. However, such a hardness result has not established so far. It is interesting to note that in the literature of identity testing, the NSINGULAR problem and the RIT problem stand among rare examples where deterministic polynomial-time white-box algorithms are designed but for the black-box case no deterministic subexponential-time algorithm is known.

## 1.1 Our Result

In this paper, we give a deterministic quasipolynomial-time *black-box* algorithm for the RIT problem via a hitting set construction. More precisely, we show the following:

**Theorem 2** (Black-box RIT). *For the class of  $n$ -variate noncommutative rational formulas of size  $s$  and inversion height  $\theta$ , we can construct a hitting set  $\mathcal{H}_{n,s,\theta} \subseteq \text{Mat}_{\ell_\theta}(\mathbb{Q})^n$  of size  $\leq (ns)^{\theta^{O(1)} \log^2(ns)}$  in deterministic  $\leq (ns)^{\theta^{O(1)} \log^2(ns)}$  time where  $\ell_\theta \leq (ns)^{\theta^{O(1)}}$ .*

As an immediate corollary of Theorem 2 and Fact 1, we obtain a quasipolynomial-size hitting set for any noncommutative rational functions computed by a rational formula of polynomial size.

## 1.2 Proof Idea

The main idea of our proof is to construct a hitting set for rational formulas of every inversion height inductively. At a high level, this approach builds on the framework introduced in [ACM22, ACG<sup>+</sup>23]. As we have already defined, a noncommutative rational formula is nonzero in the free skew field if there exists a nonzero matrix substitution. However, the difficulty is that unlike a noncommutative polynomial, a rational formula can be undefined for a matrix substitution. It happens if there is an inverse on top of a subformula evaluated to a singular matrix. Informally speaking, it is somewhat easier to maintain that subformulas evaluate to nonzero matrices, but it is much harder to maintain that they evaluate to *non singular* matrices.

One of the possible ways to tackle this problem is to evaluate rational formulas on division algebras. Finite dimensional division algebras are associative algebras where every nonzero element is invertible. Taking this informal idea forward, one of the key developments in [ACM22] was to embed the hitting set obtained by Forbes and Shpilka [FS13] for noncommutative *polynomials* computed by ABPs in a cyclic division algebra (see Section 2.3 for the definition of a cyclic division algebra) suitably small index i.e. the dimension of its matrix representation. This idea of embedding inside a division algebra is proved to be very useful for us. Let us formally define the notion of hitting set for rational formulas (of any arbitrary inversion height) inside a division algebra.

**Definition 3** (Division algebra hitting set). For a class of rational formulas, a division algebra hitting set is a hitting set over some division algebra where every point in the hitting set is a division algebra tuple.

The advantage of such a hitting set is that, whenever a rational formula evaluates to some nonzero value (over a tuple in the hitting set), the output is invertible. If we have a division algebra hitting set for rational formulas of inversion height  $\theta - 1$ , can we efficiently construct a division algebra hitting set for rational formulas of inversion height  $\theta$ ? In that case, we could inductively build a division algebra hitting set for rational formulas of every inversion height. This is the main technical step we implement here using several conceptual and technical ideas.

At this point, we take a detour and carefully examine the connection between the RIT and NSINGULAR problems. It is known that RIT is polynomial-time reducible to NSINGULAR [HW15]. But do we need the full power of NSINGULAR to solve the RIT problem? Consider the following special case of NSINGULAR. The input is a linear matrix  $T(x_1, \dots, x_n)$  and a matrix tuple  $(p_1, \dots, p_n) \in D_1^n$  for a cyclic division algebra  $D_1$ . The promise is that there is a submatrix  $T'$  of size  $s - 1$  such that  $T'(p_1, \dots, p_n)$  is invertible. It is easier to think such a tuple  $(p_1, \dots, p_n)$  as a *witness*. The question is to check the singularity of  $T$  over the free skew field. We show that the construction of a hitting set for rational formulas of inversion height  $\theta$  reduces to this special case where the witness is some tuple in the hitting set for height  $\theta - 1$ .

We then consider the shifted matrix  $T(x_1 + p_1, \dots, x_n + p_n)$ . Using Gaussian elimination, we could convert the shifted matrix of form:

$$U \cdot T(x_1 + p_1, \dots, x_n + p_n) \cdot V = \left[ \begin{array}{c|c} I_{s-1} - L & A_j \\ \hline B_i & C_{ij} \end{array} \right], \quad (1)$$

where the entries of  $L, A_j, B_i, C_{ij}$  are homogeneous  $D_1$ -linear forms. Here  $B_i$  is a row vector  $A_j$  is a column vector. At a high level, it has a conceptual similarity with the idea used in [BBJP19] in

approximating commutative rank. It is not too difficult to prove that  $T$  is invertible, if and only if,  $C_{ij} - B_i(I_{s-1} - L)^{-1}A_j = C_{ij} - B_i(\sum_{k \geq 0} L^k)A_j$  is a nonzero series. Using a standard result of noncommutative formal series [Eil74, Corollary], this is equivalent in saying that the truncated polynomial  $C_{ij} - B_i(\sum_{k \leq \ell(s-1)\ell} L^k)A_j$  is nonzero where  $\ell$  is the index of  $D_1$ . However, the series and the polynomial will have division algebra elements interleaving in between the variables. Such series (res. polynomials) is called a generalized series (resp. generalized polynomial) and has been studied extensively in the work of Volčič (see [Vol18] for more details). We can also define a notion of generalized ABP similarly. Finally, (up to a certain scaling by scalars) the upshot is that the division algebra hitting set construction for rational formulas reduces to the division algebra hitting set construction for such generalized ABPs.

We now consider such generalized ABPs where the coefficients lie inside a cyclic division algebra  $D_1$  of index  $\ell_1$ , call it  $D_1$ -ABPs. Our goal is to construct a division algebra hitting set for such ABPs. To do so, we introduce new noncommuting indeterminates for every variable and use the following mapping:

$$x_i \mapsto \sum C_{jk} \otimes y_{ijk},$$

where  $\{C_{jk}\}$  is the basis of  $D_1$ . The idea is to overcome the problem of interleaving division algebra elements using the property of tensor products. This substitution reduces the problem to identity testing of a noncommutative ABP in the  $\{y_{ijk}\}$  variables. Luckily, a division algebra hitting set construction for noncommutative ABPs is already known [ACM22]. However, in general, the tensor product of two division algebras is not a division algebra. At this point, we use a result of [Pie82] that states that the tensor product of two cyclic division algebras of index  $\ell_1$  and  $\ell_2$  is a cyclic division algebra of index  $\ell_1\ell_2$  if  $\ell_1$  and  $\ell_2$  are relatively prime. However, the division algebra hitting set construction for noncommutative ABPs is known for division algebras whose index is only a power of two [ACM22]. To use the result of [Pie82] in several stages recursively, we need a division algebra hitting set construction whose index is a power of any *arbitrary* prime  $p$ .

We now informally describe how to find a hitting set for noncommutative formulas (more generally for noncommutative ABPs) in a division algebra of arbitrary prime power index. For simplicity, suppose the prime is  $p$  and the ABP degree is  $p^d$ . In [FS13], it is assumed that the degree of the ABP is  $2^d$  and the construction has a recursive structure. In particular, it is by a reduction to the hitting set construction for ROABPs (read-once algebraic branching programs) over the commutative variables  $u_1, u_2, \dots, u_{2^d}$ . The recursive step in their construction is by combining hitting sets (via hitting set generator  $\mathcal{G}_{d-1}$ ) for two halves of degree  $2^{d-1}$  [FS13] with a rank preserving step of matrix products to obtain the generator  $\mathcal{G}_d$  at the  $d^{\text{th}}$  step. More precisely,  $\mathcal{G}_d$  is a map from  $\mathbb{F}^{d+1} \rightarrow \mathbb{F}^{2^d}$  that stretches the seed  $(\alpha_1, \dots, \alpha_{d+1})$  to a  $2^d$  tuple for the read-once variables.

For our case, the main high-level idea is to decompose the ABP of degree  $p^d$  in  $p$  consecutive windows each of length  $p^{d-1}$ . One can adapt the rank preserving step for *two matrix products* in [FS13] even for the case of *p many matrix products*. However, the main difficulty is to ensure that the hitting set points lie inside a division algebra. For our purpose, we take a classical construction of cyclic division algebras [Lam01, Chapter 5]. The cyclic division algebra  $D = (K/F, \sigma, z)$  is defined using an indeterminate  $x$  as the  $\ell$ -dimensional vector space:

$$D = K \oplus Kx \oplus \dots \oplus Kx^{\ell-1},$$

where the (noncommutative) multiplication for  $D$  is defined by  $x^\ell = z$  and  $xb = \sigma(b)x$  for all

$b \in K$ . Here  $\sigma : K \rightarrow K$  is an automorphism of the Galois group  $\text{Gal}(K/F)$ . The field  $F = \mathbb{Q}(z)$  and  $K = F(\omega)$ , where  $z$  is an indeterminate and  $\omega$  is an  $\ell^{\text{th}}$  primitive root of unity. The matrix representation of a general element in  $D$  is of the following form:

$$\begin{bmatrix} 0 & b & 0 & \cdots & 0 \\ 0 & 0 & \sigma(b) & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma^{\ell-2}(b) \\ z\sigma^{\ell-1}(b) & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Roughly, the plan will be to (inductively) assume that the construction follows the  $\sigma$ -automorphism in each window of length  $\rho^{d-1}$ , and then we need to satisfy the  $\sigma$ -action at each  $\rho - 1$  boundaries. More technically, to embed the hitting set of [FS13], we need to choose  $\ell = \rho^L$  appropriately larger than  $\rho^d$ . As it turns out the construction of the division algebra requires a tower of extension fields of  $F$ , with a higher-order root of unity at each stage.

Specifically, let  $\omega_i = \omega^{\rho^{a_i}}$  for  $a_1 > a_2 > \cdots > a_d > a_{d+1} > 0$ , where  $a_i$  are positive integers suitably chosen. Let  $K_i = F(\omega_i)$  be the cyclic Galois extension for  $1 \leq i \leq d + 1$  giving a tower of extension fields

$$F \subset F(\omega_1) \subset F(\omega_2) \subset \cdots \subset F(\omega_d) \subset F(\omega_{d+1}) \subset F(\omega).$$

We require two properties of  $\omega_i$ ,  $1 \leq i \leq d + 1$ . Firstly, for the hitting set generator  $\mathcal{G}_i$  we will choose the root of unity as  $\omega_i$  and the variable  $\alpha_i$  will take values only in the set  $W_i = \{\omega_i^j \mid 1 \leq j \leq \rho^{L-a_i}\}$ . We also require that the  $K$ -automorphism  $\sigma$  has the property that for all  $1 \leq i \leq d + 1$  the map  $\sigma^{\rho^i}$  fixes  $\omega_i$ . In fact we will ensure that  $\sigma^{\rho^i}$  has  $F(\omega_i)$  as its fixed field. The construction of  $D$  satisfying the above properties is the main technical step in Section 3.

It turns out that implementing all these ideas leads to a quasipolynomial-size hitting set for rational formulas of *any* constant inversion height. More, precisely for rational formulas of inversion height  $\theta$ , the size of the hitting set would be  $(ns)^{2^{O(\theta)} \log ns}$  and the final division algebra index will be  $(ns)^{2^\theta}$ . But to get the quasipolynomial-size hitting set for arbitrary rational formulas (where  $\theta = O(\log s)$ ) several further technical ideas are required.

We first need to analyze the source of the blow-up of  $2^{O(\theta)}$  in the exponent of the dimension. Let the indices at the  $(\theta - 1)^{\text{th}}$  and  $\theta^{\text{th}}$  stages are  $\ell_{\theta-1}$  and  $\ell_\theta$  respectively. It turns out that  $\ell_\theta = \ell_{\theta-1} \cdot \rho_\theta^{O(\rho_\theta \log(\ell_{\theta-1} s n \rho_\theta))}$ . Therefore,  $\ell_\theta \approx \ell_{\theta-1}^2$  which recursively leads to  $\ell_\theta \approx \ell_0^{2^\theta}$ . Indeed, our proof technique reveals that the parameters appear from the hitting set construction of a  $D_{\theta-1}$ -ABP of the number of variables  $\ell_{\theta-1}^2 n$ , width  $2\ell_{\theta-1} s$ , and degree  $\ell_{\theta-1}(2s + 1)$ . The degree parameter appears from the truncation of  $D_{\theta-1}$ -series at degree  $\ell_{\theta-1}(2s + 1)$ . This can be easily managed down to  $2s + 1$ , if we use a generalization of [Eil74, Theorem 8.3] over division algebra [DK21, Example 8.2]. The dependence of the size of the hitting set on the number of variables can be improved by a log-product trick over two variables  $\{y_0, y_1\}$  that replaces  $x_i$  by  $y_{b_1} y_{b_2} \cdots y_{b_{\log n}}$  where  $b_{\log n} \cdots b_1$  is the binary representation of  $i$ . This trick will increase the degree from  $\rho^d$  to  $\rho^d \log n$  but we will be fine since the dependence on the degree is only logarithmic. The hardest part is to improve the dependence on width. Here somewhat surprisingly, we give a construction such that the index of the division algebra has no dependence on the width in the exponent. This is achieved by adjoining the base field  $F$  by a complex root of unity  $\omega_0$  of a sufficiently large order of a prime power. Moreover, this prime is different from all the primes used in the recursion.

Implementing all these steps we get a quasipolynomial-size hitting set over  $\mathbb{Q}(\omega, \omega_0, z)$ . Then we show how to transfer the hitting set over  $\mathbb{Q}$  itself by a relatively standard idea that treats the parameters  $\omega, \omega_0$  and  $z$  as *fresh indeterminates*  $t_1, t_2, t_3$  and vary them over a suitably chosen quasipolynomial-size set over  $\mathbb{Q}$ .

### 1.3 Organization

In [Section 2](#), we provide a background on algebraic complexity theory, cyclic division algebras, and noncommutative formal power series. [Section 3](#) contains the construction of a quasipolynomial-size hitting set for noncommutative ABPs over cyclic division algebras whose index is any prime power. The result of [Section 4](#) is the construction of a hitting set for generalized ABPs defined over cyclic division algebras. In [Section 5](#), we construct a hitting set for NSINGULAR problem given a witness. The main result is proved in [Section 6](#) in two parts: [Section 6.1](#) gives the proof for rational formulas of constant inversion height and [Section 6.2](#) gives the proof for arbitrary rational formulas. Finally, we conclude in [Section 7](#).

## 2 Preliminaries

### 2.1 Notation

Throughout the paper, we use  $\mathbb{F}, F, K$  to denote fields, and  $\text{Mat}_m(\mathbb{F})$  (resp.  $\text{Mat}_m(F), \text{Mat}_m(K)$ ) to denote  $m$ -dimensional matrix algebra over  $\mathbb{F}$  (resp. over  $F, K$ ). Similarly,  $\text{Mat}_m(\mathbb{F})^n$  (resp.  $\text{Mat}_m(F)^n, \text{Mat}_m(K)^n$ ) denote the set of  $n$ -tuples over  $\text{Mat}_m(\mathbb{F})$  (resp.  $\text{Mat}_m(F), \text{Mat}_m(K)$ ), respectively.  $D$  is used to denote finite-dimensional division algebras. We use  $p$  to denote an arbitrary prime number. Let  $\underline{x}$  denote the set of variables  $\{x_1, \dots, x_n\}$ . Sometimes we use  $\underline{p} = (p_1, \dots, p_n)$  and  $\underline{q} = (q_1, \dots, q_n)$  to denote the matrix tuples in suitable matrix algebras where  $n$  is clear from the context. The free noncommutative ring of polynomials over a field  $\mathbb{F}$  is denoted by  $\mathbb{F}\langle \underline{x} \rangle$ . For matrices  $A$  and  $B$ , their usual tensor product is denoted by  $A \otimes B$ . For a polynomial  $f$  and a monomial  $m$ , we use  $[m]f$  to denote the coefficient of  $m$  in  $f$ .

### 2.2 Algebraic Complexity Theory

**Definition 4** (Algebraic Branching Program). An *algebraic branching program* (ABP) is a layered directed acyclic graph. The vertex set is partitioned into layers  $0, 1, \dots, d$ , with directed edges only between adjacent layers ( $i$  to  $i + 1$ ). There is a *source* vertex of in-degree 0 in the layer 0, and one out-degree 0 *sink* vertex in layer  $d$ . Each edge is labeled by an affine  $\mathbb{F}$ -linear form in variables, say,  $x_1, x_2, \dots, x_n$ . The polynomial computed by the ABP is the sum over all source-to-sink directed paths of the ordered product of affine forms labeling the path edges.

The *size* of the ABP is defined as the total number of nodes and the *width* is the maximum number of nodes in a layer, and the *depth* is the number of layers in the ABP.<sup>1</sup> An ABP can compute a commutative or a noncommutative polynomial, depending on whether the variables  $x_1, x_2, \dots, x_n$  occurring in the  $\mathbb{F}$ -linear forms are commuting or noncommuting. ABPs of width  $w$

---

<sup>1</sup>We choose to call it depth instead of length which can be used for encoding lengths etc.

can also be defined as an iterated matrix multiplication  $\underline{u}^T \cdot M_1 M_2 \cdots M_\ell \cdot \underline{v}$ , where  $\underline{u}, \underline{v} \in \mathbb{F}^n$  and each  $M_i$  is of form  $\sum_{i=1}^n A_i x_i$  for matrices  $A_i \in \text{Mat}_w(\mathbb{F})$ , assuming without loss of generality that all matrices  $M_j$ ,  $1 \leq j \leq \ell$  are  $w \times w$ .

We say a set  $\mathcal{H} \subseteq \mathbb{F}^n$  is a hitting set for a (commutative) algebraic circuit class  $\mathcal{C}$  if for every  $n$ -variate polynomial  $f$  in  $\mathcal{C}$ ,  $f \neq 0$  if and only if  $f(\underline{a}) \neq 0$  for some  $\underline{a} \in \mathcal{H}$ .

A special class of ABPs in commuting variables are the *read-once* ABPs (in short, ROABPs). In ROABPs a different variable is used for each layer, and the edge labels are univariate polynomials over that variable. For the class of ROABPs, Forbes and Shpilka [FS13] obtained the first quasipolynomial-time black-box algorithm by constructing a hitting set of quasipolynomial size.

**Theorem 5.** [FS13] *For the class of polynomials computable by a width  $r$ , depth  $d$ , individual degree  $< n$  ROABPs of known order, if  $|\mathbb{F}| \geq (2dnr^3)^2$ , there is a  $\text{poly}(d, n, r)$ -explicit hitting set of size at most  $(2dn^2r^4)^{\lceil \log d+1 \rceil}$ .*

Indeed, they proved a more general result.

**Definition 6** (Hitting Set Generator). A polynomial map  $\mathcal{G} : \mathbb{F}^t \rightarrow \mathbb{F}^n$  is a generator for a circuit class  $\mathcal{C}$  if for every  $n$ -variate polynomial  $f$  in  $\mathcal{C}$ ,  $f \equiv 0$  if and only if  $f \circ \mathcal{G} \equiv 0$ .

**Theorem 7.** [FS13, Construction 3.13, Lemma 3.21] *For the class of polynomials computable by a width  $r$ , depth  $d$ , individual degree  $< n$  ROABPs of known order, one can construct a hitting set generator  $\mathcal{G} : \mathbb{F}^{\lceil \log d+1 \rceil} \rightarrow \mathbb{F}^d$  of degree  $dnr^4$  efficiently.*

As a consequence, Forbes and Shpilka [FS13], obtain an efficient construction of quasipolynomial-size hitting set for noncommutative ABPs as well. Consider the class of noncommutative ABPs of width  $w$ , and depth  $d$  computing polynomials in  $\mathbb{F}\langle x \rangle$ . The result of Forbes and Shpilka provide an explicit construction (in quasipolynomial-time) of a set  $\text{Mat}_{d+1}(\mathbb{F})$ , such that for any ABP (with parameters  $w$  and  $d$ ) computing a nonzero polynomial  $f$ , there always exists  $(p_1, \dots, p_n) \in \mathcal{H}_{n,w,d}$ ,  $f(p) \neq 0$ .

**Theorem 8** (Forbes and Shpilka [FS13]). *For all  $w, d, n \in \mathbb{N}$ , if  $|\mathbb{F}| \geq \text{poly}(d, n, w)$ , then there is a hitting set  $\mathcal{H}_{n,w,d} \subset \text{Mat}_{d+1}(\mathbb{F})$  for noncommutative ABPs of parameters  $|\mathcal{H}_{n,w,d}| \leq (wdn)^{O(\log d)}$  and there is a deterministic algorithm to output the set  $\mathcal{H}_{n,w,d}$  in time  $(wdn)^{O(\log d)}$ .*

## 2.3 Cyclic Division Algebras

A division algebra  $D$  is an associative algebra over a (commutative) field  $\mathbb{F}$  such that all nonzero elements in  $D$  are units (they have a multiplicative inverse). In this paper, we are interested in finite-dimensional division algebras. Specifically, we focus on cyclic division algebras and their construction [Lam01, Chapter 5]. Let  $F = \mathbb{Q}(z)$ , where  $z$  is a commuting indeterminate. Let  $\omega$  be an  $\ell^{\text{th}}$  primitive root of unity. To be specific, let  $\omega = e^{2\pi i/\ell}$ . Let  $K = F(\omega) = \mathbb{Q}(\omega, z)$  be the cyclic Galois extension of  $F$  obtained by adjoining  $\omega$ . So,  $[K : F] = \ell$  is the degree of the extension. The elements of  $K$  are polynomials in  $\omega$  (of degree at most  $\ell - 1$ ) with coefficients from  $F$ .

Define  $\sigma : K \rightarrow K$  by letting  $\sigma(\omega) = \omega^k$  for some  $k$  relatively prime to  $\ell$  and stipulating that  $\sigma(a) = a$  for all  $a \in F$ . Then  $\sigma$  is an automorphism of  $K$  with  $F$  as fixed field and it generates the Galois group  $\text{Gal}(K/F)$ .



The division algebra  $D = (K/F, \sigma, z)$  is defined using a new indeterminate  $x$  as the  $\ell$ -dimensional vector space:

$$D = K \oplus Kx \oplus \cdots \oplus Kx^{\ell-1},$$

where the (noncommutative) multiplication for  $D$  is defined by  $x^\ell = z$  and  $xb = \sigma(b)x$  for all  $b \in K$ . The parameter  $\ell$  is called the index of  $D$  [Lam01, Theorem 14.9].

The elements of  $D$  has matrix representation in  $K^{\ell \times \ell}$  from its action on the basis  $\mathcal{X} = \{1, x, \dots, x^{\ell-1}\}$ . I.e., for  $a \in D$  and  $x^j \in \mathcal{X}$ , the  $j^{\text{th}}$  row of the matrix representation is obtained by writing  $x^j a$  in the  $\mathcal{X}$ -basis.

For example, the matrix representation  $M(x)$  of  $x$  is:

$$M(x)[i, j] = \begin{cases} 1 & \text{if } j = i + 1, i \leq \ell - 1 \\ z & \text{if } i = \ell, j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$M(x) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ z & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

For each  $b \in K$  its matrix representation  $M(b)$  is:

$$M(b)[i, j] = \begin{cases} b & \text{if } i = j = 1 \\ \sigma^{i-1}(b) & \text{if } i = j, i \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

$$M(b) = \begin{bmatrix} b & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma(b) & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2(b) & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^{\ell-2}(b) & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^{\ell-1}(b) \end{bmatrix}$$

**Proposition 9.** For all  $b \in K$ ,  $M(bx) = M(b) \cdot M(x)$

Also, the matrix representation of  $xb = \sigma(b)x$  is easy to see in the basis  $\{1, x, \dots, x^{\ell-1}\}$ :

$$M(\sigma(b)x) = \begin{bmatrix} 0 & \sigma(b) & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2(b) & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^3(b) & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^{\ell-1}(b) \\ \sigma^\ell(b)z & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Define  $C_{i,j} = M(\omega^{j-1}) \cdot M(x^{i-1})$  for  $1 \leq i, j \leq \ell$ . Observe that  $\mathfrak{B} = \{C_{i,j}, i, j \in [\ell]\}$  is a  $F$ -generating set for the division algebra  $D$ .

**Fact 10.** *The  $F$ -linear span of  $\mathfrak{B}$  is the cyclic division algebra  $D$  in the matrix algebra  $\text{Mat}_\ell(K)$ .*

The following proposition is a standard fact.

**Proposition 11.** [[Lam01](#), Section 14(14.13)] *The  $K$ -linear span of  $\mathfrak{B}$  is the entire matrix algebra  $\text{Mat}_\ell(K)$ .*

The following theorem gives us a way of constructing new division algebras using tensor products. This construction plays an important role in our main result.

**Theorem 12.** [[Pie82](#), Proposition, Page 292] *Let  $K, L$  be cyclic extensions of the field  $F$  such that their extension degrees,  $[K : F]$  and  $[L : F]$ , are relatively prime. Let  $D_1 = (K/F, \sigma_1, z)$ , and  $D_2 = (L/F, \sigma_2, z)$  be the corresponding cyclic division algebras as defined above. Then their tensor product  $D_1 \otimes D_2$  is also a cyclic division algebra.*

## 2.4 Noncommutative Rational Series

Let  $D$  be a division algebra and  $P$  be a series over the noncommuting variables  $x_1, x_2, \dots, x_n$  defined as follows:

$$P = c - B \left( \sum_{k \geq 0} L^k \right) A,$$

where  $c$  is a  $D$ -linear form (over  $x_1, \dots, x_n$ ),  $B$  (resp.  $A$ ) is a  $1 \times s$  (resp.  $s \times 1$ ) dimensional vector, and  $L$  is a  $s \times s$  matrix. The entries of  $B, L, A$  are  $D$ -linear forms over  $x_1, \dots, x_n$ . Furthermore, the variables  $x_i : 1 \leq i \leq n$  commute with the elements in  $D$ . Define the truncated polynomial  $\tilde{P}$  as follows:

$$\tilde{P} = c - B \left( \sum_{k \leq s-1} L^k \right) A. \quad (2)$$

The next statement shows that the infinite series  $P \neq 0$  is equivalent in saying that  $\tilde{P}$  is nonzero. The proof of the fact is standard when  $D$  is a (commutative) field [[Eil74](#), Corollary 8.3, Page 145]. For the case of division algebras, a sketch of the proof can be found [[DK21](#), Example 8.2, Page 23]. However, we include a self-contained proof.

**Fact 13.** *The infinite series  $P \neq 0$  if and only if its truncation  $\tilde{P} \neq 0$ .*

*Proof.* If  $P = 0$ , then obviously  $\tilde{P} = 0$ , since the degrees in different homogeneous components do not match. Now, suppose  $\tilde{P} = 0$ . Notice that the terms in  $c$  are linear forms and the degree of any term in  $B \left( \sum_{k \geq 0} L^k \right) A$  is at least two. Hence,  $c$  must be zero. Write the row and column vectors  $B$  and  $A$  as  $B = \sum_{\ell} B_{\ell} x_{\ell}$ ,  $A = \sum_{\ell} A_{\ell} x_{\ell}$ . Similarly, write  $L = \sum_{\ell} L_{\ell} x_{\ell}$ .

Suppose  $BL^s A$  contributes a nonzero monomial (word)  $w = x_{i_1} x_{i_2} \dots x_{i_{s+2}}$ . Clearly the coefficient of  $w$  is  $B_{i_1} L_{i_2} \dots L_{i_{s+1}} A_{i_{s+2}}$ . Consider the vectors  $v_1 = B_{i_1}$ ,  $v_2 = B_{i_1} L_{i_2}$ ,  $\dots$ ,  $v_{s+1} = B_{i_1} L_{i_2} \dots L_{i_{s+1}}$  corresponding to the prefixes  $w_1 = x_{i_1}$ ,  $w_2 = x_{i_1} x_{i_2}$ ,  $\dots$ ,  $w_{s+1} = x_{i_1} \dots x_{i_{s+1}}$ . These vectors  $v_i, 1 \leq i \leq s+1$  all lie in the (left)  $D$ -module  $D^s$  which has rank  $s$ . As  $D$  is a division algebra, these vectors cannot all be  $D$ -linearly independent. Hence, there are elements  $\lambda_1, \dots, \lambda_{s+1}$  in  $D$ , not all zero, such that the linear combination  $\lambda_1 v_1 + \dots + \lambda_{s+1} v_{s+1} = 0$ . However,  $v_{s+1} A_{i_{s+2}} \neq 0$  by the assumption. Hence, there is at least one vector  $v_{\ell} : 1 \leq \ell \leq s$  such that  $v_{\ell} A_{i_{s+2}} \neq 0$ . This means

that the coefficient of the word  $w_\ell x_{i_s+2}$ , which is of length at most  $s + 1$ , is nonzero in  $\widetilde{P}$ , which is not possible by assumption.

Now, with  $k = s$  as the base case, we can inductively apply the above argument to show that  $BL^k A$  is zero for each  $k \geq s$ . ■

## 2.5 Generalized Formal Power Series

We now define the notion of generalized series first introduced by Volčič. For a detailed exposition, see [Vol18].

A *generalized word* or a *generalized monomial* in  $x_1, \dots, x_n$  over the matrix algebra  $\text{Mat}_m(\mathbb{F})$  allows the matrices to interleave between variables. That is to say, a generalized monomial is of the form:  $a_0 x_{k_1} a_2 \cdots a_{d-1} x_{k_d} a_d$ , where  $a_i \in \text{Mat}_m(\mathbb{F})$ , and its degree is the number of variables  $d$  occurring in it. A finite sum of generalized monomials is a *generalized polynomial* in the ring  $\text{Mat}_m(\mathbb{F})\langle \underline{x} \rangle$ . A *generalized formal power series* over  $\text{Mat}_m(\mathbb{F})$  is an infinite sum of generalized monomials such that the sum has finitely many generalized monomials of degree  $d$  for any  $d \in \mathbb{N}$ . The ring of generalized series over  $\text{Mat}_m(\mathbb{F})$  is denoted  $\text{Mat}_m(\mathbb{F})\langle\langle \underline{x} \rangle\rangle$ .

A generalized series (resp. polynomial)  $S$  over  $\text{Mat}_m(\mathbb{F})$  admits the following canonical description. Let  $E = \{e_{i,j}, 1 \leq i, j \leq m\}$  be the set of elementary matrices. Express each coefficient matrix  $a$  in  $S$  in the  $E$  basis by a  $\mathbb{F}$ -linear combination and then expand  $S$ . Naturally each monomial of degree- $d$  in the expansion looks like  $e_{i_0, j_0} x_{k_1} e_{i_1, j_1} x_{k_2} \cdots e_{i_{d-1}, j_{d-1}} x_{k_d} e_{i_d, j_d}$  where  $e_{i_l, j_l} \in E$  and  $x_{k_l} \in \underline{x}$ . We say the series  $S$  (resp. polynomial) is identically zero if and only if it is zero under such expansion i.e. the coefficient associated with each generalized monomial is zero.

The evaluation of a generalized series over  $\text{Mat}_m(\mathbb{F})$  is defined on any  $k'm \times k'm$  matrix algebra for some integer  $k' \geq 1$  [Vol18]. To match the dimension of the coefficient matrices with the matrix substitution, we use an inclusion map  $\iota : \text{Mat}_m(\mathbb{F}) \rightarrow \text{Mat}_{k'm}(\mathbb{F})$ , for example,  $\iota$  can be defined as  $\iota(a) = a \otimes I_{k'}$  or  $\iota(a) = I_{k'} \otimes a$ . Now, a generalized monomial  $a_0 x_{k_1} a_1 \cdots a_{d-1} x_{k_d} a_d$  over  $\text{Mat}_m(\mathbb{F})$  on matrix substitution  $(p_1, \dots, p_n) \in \text{Mat}_{k'm}(\mathbb{F})^n$  evaluates to

$$\iota(a_0) p_{k_1} \iota(a_1) \cdots \iota(a_{d-1}) p_{k_d} \iota(a_d)$$

under some inclusion map  $\iota : \text{Mat}_m(\mathbb{F}) \rightarrow \text{Mat}_{k'm}(\mathbb{F})$ . All such inclusion maps are known to be compatible by the Skolem-Noether theorem [Row80, Theorem 3.1.2]. Therefore, if a series  $S$  is zero with respect to some inclusion map  $\iota : \text{Mat}_m(\mathbb{F}) \rightarrow \text{Mat}_{k'm}(\mathbb{F})$ , then it is zero w.r.t. any such inclusion map.

We naturally extend the definition of usual ABPs (Definition 4) to the generalized ABPs.

**Definition 14** (Generalized Algebraic Branching Program). A *generalized algebraic branching program* is a layered directed acyclic graph. The vertex set is partitioned into layers  $0, 1, \dots, d$ , with directed edges only between adjacent layers ( $i$  to  $i + 1$ ). There is a *source* vertex of in-degree 0 in the layer 0, and one out-degree 0 *sink* vertex in layer  $d$ . Each edge is labeled by a generalized linear form of  $\sum_{i=1}^n a_i x_i b_i$  where  $a_i, b_i \in \text{Mat}_m(\mathbb{F})$  for some integer  $m$ . The generalized polynomial computed by the ABP is the sum over all source-to-sink directed paths of the ordered product of generalized linear forms labeling the path edges.

### 3 Division Algebra Hitting Set for Noncommutative ABPs

Fix a prime number  $\rho$ . In particular,  $\rho$  is independent of the input ABP. The main result of this section shows that the quasipolynomial-size hitting set construction for noncommutative ABPs by Forbes and Shpilka [FS13] can be adapted to a more general setting where the hitting set points lie in a finite-dimensional cyclic division algebra whose index is a power of  $\rho$ . We note that such construction is already known when the index is a power of 2 [ACM22].

Let  $F$  be a characteristic zero field. Let  $\{u_1, u_2, \dots, u_\rho\}$  be commuting indeterminates. The ring  $\text{Mat}_r(F[u_i])$  consists of  $r \times r$  matrices whose entries are univariate polynomials in  $u_i$  over  $F$ . Equivalently, an element  $M \in \text{Mat}_r(F[u_i])$  can be seen as a univariate polynomial with matrix coefficients in  $\text{Mat}_r(F)$ . Its degree  $\deg(M)$  is the largest integer such that the matrix coefficient of  $u_i^{\deg(M)}$  in  $M$  is nonzero. Let  $\bar{F}$  denote the algebraic closure of  $F$ . The following lemma is a generalization of [FS13, Lemma 3.5].

**Lemma 15.** *For each  $i \in [\rho]$ , let  $M_i \in \text{Mat}_r(F[u_i])$  be of degree  $< n$  and  $\omega \in \bar{F}$  be a root of unity whose (finite) order is at least  $n^\rho$ . Let  $K = F(\omega)$  be the field extension by  $\omega$ . Then for any  $\alpha \in F$  and any  $\mu \geq n$ ,*

$$\text{span}_K \left\{ [u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i) \right\} \supseteq \text{span}_K \left\{ M_1(\omega^\ell \alpha) M_2((\omega^\ell \alpha)^\mu) \cdots M_\rho((\omega^\ell \alpha)^{\mu^{\rho-1}}) \right\}.$$

Moreover, except for  $< n^\rho r^2$  many values of  $\alpha$  in  $F$ ,

$$\text{span}_K \left\{ [u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i) \right\} = \text{span}_K \left\{ \{M_1(\omega^\ell \alpha) M_2((\omega^\ell \alpha)^\mu) \cdots M_\rho((\omega^\ell \alpha)^{\mu^{\rho-1}})\} \right\}.$$

Here  $\ell$  varies from  $\{0, 1, \dots, r^2 - 1\}$ .

*Proof.* By span we will always mean the  $K$ -linear span.

$$\begin{aligned} \prod_{i=1}^{\rho} M_i(u_i) &= \sum_{j_1, j_2, \dots, j_\rho} \left( [u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i) \right) \cdot u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}. \\ \text{Therefore, } \prod_{i=1}^{\rho} M_i((\omega^\ell \alpha)^{\mu^{i-1}}) &= \sum_{j_1, j_2, \dots, j_\rho} \left( [u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i) \right) \cdot (\omega^\ell \alpha)^{j_1 + j_2 \mu + \dots + j_\rho \mu^{\rho-1}}, \quad (3) \end{aligned}$$

which proves the first part of the lemma.

We now define a rectangular matrix  $C \in \text{Mat}_{n^\rho, r^2}(F)$  as follows. Each row of  $C$  is indexed by a tuple  $(j_1, j_2, \dots, j_\rho) \in \{0, 1, \dots, n-1\}^\rho$ . For each such tuple  $(j_1, j_2, \dots, j_\rho)$ , treating the  $r \times r$  matrix  $[u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i)$  as an  $r^2$ -dimensional vector, we define it as the corresponding row  $C_{(j_1, j_2, \dots, j_\rho)}$ . By definition,

$$\text{row-span}(C) = \text{span} \left\{ [u_1^{j_1} u_2^{j_2} \cdots u_\rho^{j_\rho}] \prod_{i=1}^{\rho} M_i(u_i) \right\}.$$

Now consider the matrix  $A_\alpha \in \text{Mat}_{r^2, n^p}(K)$  whose columns are indexed by tuples  $(j_1, j_2, \dots, j_p) \in \{0, 1, \dots, n-1\}^p$  and let

$$(A_\alpha)_{\ell, (j_1, j_2, \dots, j_p)} = (\omega^\ell \alpha)^{j_1 + j_2 \mu + \dots + j_p \mu^{p-1}}.$$

We note that this is used in the rank extractor construction by Gabizon and Raz [GR08]. Applying their result it follows that, except for at most  $n^p r^2$  values of  $\alpha$ , the rank of the product matrix  $\text{rank}(A_\alpha C) = \text{rank}(C)$ . Multiplying the  $\ell^{\text{th}}$  row of  $A_\alpha$  with  $C$  we get

$$(A_\alpha)_\ell C = \sum_{j_1, j_2, \dots, j_p} \left( [u_1^{j_1} u_2^{j_2} \dots u_p^{j_p}] \prod_{i=1}^p M_i(u_i) \right) \cdot (\omega^\ell \alpha)^{j_1 + j_2 \mu + \dots + j_p \mu^{p-1}} = \prod_{i=1}^p M_i((\omega^\ell \alpha)^{\mu^{i-1}}).$$

Therefore,  $\text{row-span}(A_\alpha C) = \text{span}\{M_1(\omega^\ell \alpha) M_2((\omega^\ell \alpha)^\mu) \dots M_p((\omega^\ell \alpha)^{\mu^{p-1}})\}$ .

As  $\text{row-span}(C)$  contains  $\text{row-span}(A_\alpha C)$ , if  $\text{rank}(C) = \text{rank}(A_\alpha C)$  then we have  $\text{row-span}(C) = \text{row-span}(A_\alpha C)$ . Therefore, barring at most  $n^p r^2$  values of  $\alpha$ ,  $\text{row-span}(C) = \text{row-span}(A_\alpha C)$ . ■

Now we informally discuss how Lemma 15 is used for the hitting set construction. W.l.o.g, we can assume that the degree of the ABP is  $p^d$  for some integer  $d$ . We group the ABP layers into  $p$  sets where each set has  $p^{d-1}$  consecutive matrix products (over different variables for each of the sets). Then, roughly speaking, the next lemma gives a method to show that the span of the full matrix product can be captured by span of the matrix products over a *single* variable. A crucial component will be Lemma 15. The next lemma is a generalization of [FS13, Lemma 3.7].

**Lemma 16.** Consider  $p$  many families of  $r \times r$  matrices  $\mathcal{M}_1 = \{M_{1,1}, M_{1,2}, \dots, M_{1,p^{d-1}}\}, \dots, \mathcal{M}_p = \{M_{p,1}, M_{p,2}, \dots, M_{p,p^{d-1}}\}$  where for the  $j^{\text{th}}$  family the entries are univariate polynomials over  $F[u_j]$  of degree less than  $n$ . Let  $(f_1(u), f_2(u), \dots, f_{p^{d-1}}(u)) \in F[u]$  be polynomials of degree at most  $m$ . Let  $\omega \in \bar{F}$  be a root of unity of order more than  $(p^{d-1} n m)^p$  and  $K = F(\omega)$ . Define polynomials in indeterminate  $v$ :

$$\begin{aligned} f'_i(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_1}) q_\ell(v), \quad 1 \leq i \leq p^{d-1} \\ f'_{i+p^{d-1}}(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_2}) q_\ell(v), \quad 1 \leq i \leq p^{d-1} \\ &\vdots \\ f'_{i+(p-1)p^{d-1}}(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_p}) q_\ell(v), \quad 1 \leq i \leq p^{d-1} \end{aligned}$$

where  $\mu_j = \mu^{j-1}$ ,  $\mu = 1 + p^{d-1} n m$ , and  $q_\ell(v)$  is the corresponding Lagrange interpolation polynomial.

Then, for all but  $(p^{d-1} n m r)^p$  many values of  $\alpha_d$ , the  $K$ -linear span of the matrix coefficients of the matrix product  $\prod_{j=1}^p \prod_{i=1}^{p^{d-1}} M_{j,i}(f_i(u_j))$  is contained in the  $K$ -linear span of the matrix coefficients of the product  $\prod_{j=1}^p \prod_{i=1}^{p^{d-1}} M_{j,i}(f'_i(v))$ .

*Proof.* As before, all spans are  $K$ -linear spans. Let  $\gamma = \rho^{d-1}$  and for each  $j$ , let  $R_j(u_j) = \prod_{i=1}^{\rho^{d-1}} M_{j,i}(f_i(u_j))$ . Note that  $R_j(u_j)$  is a matrix of univariate polynomials in  $u_j$  of degree less than  $\gamma nm$ . By definition,

$$\prod_{j=1}^{\rho} \prod_{i=1}^{\rho^{d-1}} M_{j,i}(f_i(u_j)) = \prod_{j=1}^{\rho} R_j(u_j).$$

**Lemma 15** implies that the span of the coefficients of  $\prod_{j=1}^{\rho} R_j(u_j)$  is contained in the span of  $\prod_{j=1}^{\rho} R_j((\omega^\ell \alpha)^{\mu_j})$ , where  $\mu_j = \mu^{j-1}$  for  $\mu > \rho^{d-1} nm$ .

For each  $j$ , let  $T_j(v) = \prod_{i=1}^{\rho^{d-1}} M_{j,i}(f'_{i+(j-1)\rho^{d-1}}(v))$ . By the definition of the Lagrange interpolation polynomials, letting  $q_\ell(\beta_k) = \delta_{\ell k}$  where each  $\beta_k$  is distinct, we have

$$T_j(\beta_\ell) = \prod_{i=1}^{\rho^{d-1}} M_{j,i}(f_i((\omega^\ell \alpha_d)^{\mu_j})) = R_j((\omega^\ell \alpha_d)^{\mu_j}).$$

$$\text{Hence, } \text{span} \left\{ \prod_{j=1}^{\rho} R_j(u_j) \right\} \subseteq \text{span} \left\{ \prod_{j=1}^{\rho} T_j(v) \right\}_{v \in K}. \quad \blacksquare$$

Now we are ready to prove the main theorem of the section.

**Theorem 17.** *Let  $\rho$  be any prime number. For the class of  $n$ -variate degree  $\tilde{d}$  noncommutative polynomials computed by homogeneous ABPs of width  $r$ , we can construct a hitting set  $\widehat{\mathcal{H}}_{n,r,\tilde{d}} \subseteq D_2^n$  of size  $(nr\tilde{d})^{O(\rho \log \rho \log \tilde{d})}$  in  $(nr\tilde{d})^{O(\rho \log \rho \log \tilde{d})}$  time. Here  $D_2$  is a cyclic division algebra of index  $\ell_2 = \rho^L$  where  $L \leq O(\rho \log(nr\tilde{d}))$ .*

*Proof.* We will set  $\ell_2 = \rho^L$  as the index of the division algebra  $D_2$ , where  $\rho$  is the given prime and  $L$  will be determined in the analysis below. One of the necessary conditions is that  $\rho^L > \tilde{d}$ .

One of the key ideas in [FS13] is to convert the given ABP into a set-multilinear form and eventually a read-once form. More specifically, they replace the noncommutative variable  $x_i$  by the matrix  $M(x_i)$ :

$$M(x_i) = \begin{bmatrix} 0 & x_{i1} & 0 & \cdots & 0 \\ 0 & 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & x_{i\tilde{d}} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

and the variables  $x_{i1}, x_{i2}, \dots, x_{i\tilde{d}}$  will be replaced by  $u_1^i, u_2^i, \dots, u_{\tilde{d}}^i$ . Obviously, these matrices are nilpotent matrices and they are not elements of any division algebra. These variables will be finally substituted by the output of a generator  $\mathcal{G}_{\log \tilde{d}}$  that stretches a seed  $(\alpha_1, \alpha_2, \dots, \alpha_{\log \tilde{d}+1})$  to  $(f_1(\alpha), f_2(\alpha), \dots, f_{\tilde{d}}(\alpha))$ .

Here, our plan will be to replace  $x_i$  by the following matrix  $M(x_i)$  :

$$M(x_i) = \left[ \begin{array}{cccc|ccc} 0 & f_1^i(\underline{\alpha}) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & f_2^i(\underline{\alpha}) & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f_d^i(\underline{\alpha}) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & f_{\tilde{d}+1}^i(\underline{\alpha}) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & f_{\ell_2-1}^i(\underline{\alpha}) \\ z f_{\ell_2}^i(\underline{\alpha}) & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array} \right],$$

where the tuple  $(f_1(\underline{\alpha}), \dots, f_{\ell_2}(\underline{\alpha}))$  will be the output of a generator of seed length  $O(\log \ell_2)$ . Additionally, if we can maintain the property that each such matrix is a circulant matrix that represent a cyclic division algebra element of the form shown in [Proposition 9](#), we will be able to implement the construction. Now we discuss the implementation of these ideas.

Choose  $\omega = e^{\frac{2\pi i}{p^L}}$ , a primitive root of unity of order  $p^L$ . Let  $F = \mathbb{Q}(z)$  and  $K = F(\omega)$  be its (finite) extension by  $\omega$ . Using the construction described in [Section 2.3](#), we consider the cyclic division algebra  $D_2 = (K/F, \sigma, z)$ . We fix the  $K$ -automorphism  $\sigma$  as

$$\sigma(\omega) = \omega^{p^\kappa+1},$$

where the positive integer  $\kappa$  will be suitably chosen in the following analysis, fulfilling the constraints of [Lemma 16](#) and some additional requirements.

Let  $d = \log_p \tilde{d}$ , where we assume (without loss of generality) that  $\tilde{d}$  is a power of  $p$ . Let  $\omega_i = \omega^{p^{a_i}}$  for  $a_1 > a_2 > \cdots > a_d > a_{d+1} > 0$ , where  $a_i$  are positive integers to be chosen. We denote by  $K_i$  the cyclic Galois extension  $K_i = F(\omega_i)$  of  $F$  by  $\omega_i$ , for  $1 \leq i \leq d+1$ . This gives a tower of field extensions

$$F \subset F(\omega_1) \subset F(\omega_2) \subset \cdots \subset F(\omega_d) \subset F(\omega_{d+1}) \subset F(\omega) = K.$$

We require two properties of  $\omega_i, 1 \leq i \leq d+1$ .

1. For the hitting set generator  $\mathcal{G}_i$  we will choose the root of unity as  $\omega_i$  and the variable  $\alpha_i$  will take values only in the set

$$W_i = \{\omega_i^j \mid 1 \leq j \leq p^{L-a_i}\}.$$

2. We require that the  $K$ -automorphism  $\sigma$  has the property that for all  $1 \leq i \leq d+1$  the map  $\sigma^{p^i}$  fixes  $\omega_i$ . It is enough to ensure that  $\sigma^{p^i}$  has  $F(\omega_i)$  as its fixed field.

We take up the second property. As  $\sigma(\omega) = \omega^{p^\kappa+1}$ , we have  $\sigma(\omega_i) = \omega^{p^{a_i}(p^\kappa+1)}$ . Therefore,

$$\sigma^{p^i}(\omega_i) = \omega^{p^{a_i}(p^\kappa+1)^{p^i}}.$$

Now,  $(p^\kappa + 1)^{p^i} = \sum_{j=0}^{p^i} \binom{p^i}{j} p^{\kappa j}$ . Choosing  $\kappa = L/2$ , we have  $\omega^{p^{\kappa j}} = 1$  for  $j \geq 2$ . Therefore,

$$\sigma^{p^i}(\omega_i) = \omega^{p^{a_i}(p^{i+\kappa}+1)} = \omega_i \cdot \omega^{p^{a_i+i+\kappa}}.$$

We can set  $a_i + i + \kappa = L$  for  $1 \leq i \leq d + 1$  to ensure that  $\sigma^{\rho^i}$  fixes  $\omega_i$ . Putting  $L = 2\kappa$ , we obtain

$$a_i = \kappa - i \text{ for } 1 \leq i \leq d + 1. \quad (4)$$

It remains to choose  $\kappa$ . In the construction of our hitting set generator  $\mathcal{G}_i$ , the parameter  $\alpha_i$  will take values only in  $W_i$  defined above. We note that  $|W_i| = \rho^{L-a_i} = \rho^{\kappa+i}$ . By Lemma 16 there are at most  $(\rho^d nmr)^\rho$  many bad values of  $\alpha_i$  for any  $i$ . Thus, it suffices to choose  $\kappa$  such that  $\rho^\kappa > (\rho^d nmr)^\rho$ . It suffices to set

$$\kappa = \lceil \rho d \rceil + \lceil \rho \log_2(nmr) \rceil + 1.$$

The choice of  $\kappa$  determines the value of parameter  $\mu$  in Lemma 33. Since  $L = 2\kappa$ , it follows that  $\rho^L > \tilde{d}$  holds.

Coming back to the modified construction of  $\mathcal{G}_d$ , inductively, we can assume that we have already constructed hitting set generators for each *window* of length  $\rho^{d-1}$ . More precisely, let  $\mathcal{G}_{d-1} : (\alpha_1, \dots, \alpha_{d-1}, u) \mapsto (f_1(u), f_2(u), \dots, f_{\rho^{d-1}}(u))$  (where the polynomial  $f_i(u) \in K_{d-1}[u]$ , for  $1 \leq i \leq \rho^{d-1}$ ) with the above two properties has already been constructed. Namely for each window, suppose  $f_{i+1}(u) = \sigma(f_i(u))$  holds for all  $i \leq \rho^{d-1} - 1$ . Now define  $\mathcal{G}_d : (\alpha_1, \dots, \alpha_d, v) \mapsto (f'_1(v), f'_2(v), \dots, f'_{\rho^d}(v))$  using Lemma 16.

Since the Lagrange interpolation polynomial  $q_\ell(v)$  has only integer coefficients,  $\sigma(q_\ell(v)) = q_\ell(v)$ . Therefore, for every  $j^{\text{th}}$  window (where  $j \in \{1, 2, \dots, \rho\}$ ) we have that  $1 + (j-1)\rho^{d-1} \leq i \leq j\rho^{d-1} - 1$ , we have  $f'_{i+1}(v) = \sigma(f'_i(v))$ .

Now, consider each boundary condition, i.e.  $i = j\rho^{d-1}$ . We need to ensure that  $\sigma(f'_{j\rho^{d-1}}(v)) = f'_{1+j\rho^{d-1}}(v)$ . Equivalently, we need to ensure that

$$\sigma \left( \sum_{\ell=1}^{\rho^2} f_{\rho^{d-1}}((\omega_d^\ell \alpha_d)^{\mu_{j-1}}) q_\ell(v) \right) = \sum_{\ell=1}^{\rho^2} f_1((\omega_d^\ell \alpha_d)^{\mu_j}) q_\ell(v).$$

We prove it by induction on  $j$ . Inductively, we can enforce it by requiring that

$$\sigma^{(j-1)\rho^{d-1}} \left( \sum_{\ell=1}^{\rho^2} f_1(\omega_d^\ell \alpha_d) q_\ell(v) \right) = \sum_{\ell=1}^{\rho^2} f_1((\omega_d^\ell \alpha_d)^{\mu_j}) q_\ell(v).$$

Since  $\alpha_d$  will be chosen from  $W_d$  (all powers of  $\omega_d$ ), we can write  $\omega_d^\ell \alpha_d = \omega_d^{j'}$  for some  $j'$ . Now,  $\sigma^{(j-1)\rho^{d-1}}(f_1(\omega_d^{j'})) = f_1(\sigma^{(j-1)\rho^{d-1}}(\omega_d^{j'}))$  as  $\sigma^{\rho^{d-1}}$  fixes all coefficients of  $f_1$  (because  $f_1(u) \in K_{d-1}[u]$ ). Now,

$$\sigma^{(j-1)\rho^{d-1}}(\omega_d^{j'}) = \omega_d^{j' \cdot (\rho^\kappa + 1)^{(j-1)\rho^{d-1}}} = \omega_d^{j'(1+\rho^{d-1+\kappa})^{j-1}} = (\omega_d^\ell \alpha_d)^{\mu_{j-1}},$$

which verifies that the choice of  $\mu$  in Lemma 16 is  $1 + \rho^{d-1+\kappa}$ .

As already discussed, the parameter  $v$  (whose place holder is  $\alpha_{d+1}$  in the description of  $\mathcal{G}_d$ ) should vary over a set of size  $O((\rho^d nmr)^\rho)$ . This way we ensure that  $f_{i+1} = \sigma(f_i)$  for  $1 \leq i \leq \rho^d - 1$ . Now define  $f_{\rho^d+j} = \sigma(f_{\rho^d+j-1})$  for  $1 \leq j \leq \rho^d$ . The fact that  $\mathcal{G}_d$  is indeed a generator follows from the span preserving property and the proof is identical to the proof given in [FS13]. For our



case, it uses [Lemma 16](#). To see the final hitting set size, we note that the seed  $(\alpha_1, \dots, \alpha_d, \alpha_{d+1}) \in S_1 \times S_2 \times \dots \times S_{d+1}$ , where  $S_i \subseteq W_i$  and  $|S_i| = p^\kappa$ . Each seed  $(\alpha_1, \dots, \alpha_{d+1})$  defines a  $n$ -tuple over  $D_2^n$  in the hitting set. So the size of the hitting set is  $(p^d n m r)^{O(d\rho)}$ . Since  $m$  is the degree of the generators at every stage which is bounded by the degree of the Lagrange interpolation polynomial  $r^2$ , we can simplify,  $|\widehat{H}_{n,r,\vec{d}}| \leq (nr\vec{d})^{O(\rho \log \vec{d} \log p)}$ . ■

## 4 Division Algebra Hitting Set for Generalized ABPs over Cyclic Division Algebras

In this section, we will consider generalized ABPs where the coefficients are from a cyclic division algebra. We will construct hitting sets for such ABPs inside another cyclic division algebra, applying the result of [Section 3](#).

**Definition 18** ( $D_1$ -ABP). Let  $D_1$  be a cyclic division algebra of index  $\ell_1$ . We define a  $D_1$ -ABP as a generalized ABP  $\mathcal{A}$  in  $\{x_1, x_2, \dots, x_n\}$  variables (as defined in [Definition 14](#)) where each edge is labeled by  $\sum_{i=1}^n a_i x_i b_i : a_i, b_i \in D_1$ . The ABP  $\mathcal{A}$  computes a generalized polynomial over  $D_1$ .

The main result of this section is an efficient hitting set construction for such ABPs. Before we state the main result of this section, we first prove some properties of a  $D_1$ -ABP where  $D_1 = (K_1/F, \sigma_1, z)$  is a cyclic division algebra of index  $\ell_1$ .

**Claim 19.** *For any nonzero  $n$ -variate degree- $d$   $D_1$ -ABP  $\mathcal{A}$  of width  $r$ , for every  $d' \geq \ell_1 d$ , there is a  $d' \times d'$  matrix tuple such that the  $D_1$ -ABP is nonzero evaluated on that tuple.*

*Proof.* Fix an edge of  $\mathcal{A}$  and let its label be  $\sum_{i=1}^n a_i x_i b_i$ , for  $a_i, b_i \in D_1$ . Replace each  $a_i, b_i \in D_1$  by its matrix representation in  $\text{Mat}_{\ell_1}(K_1)$  and the variable  $x_i$  by  $Z_i$ , an  $\ell_1 \times \ell_1$  matrix whose  $(j, k)^{\text{th}}$  entry is a new noncommuting indeterminate  $z_{ijk}$ . Therefore, each edge is now labeled by an  $\ell_1 \times \ell_1$  matrix whose entries are  $K_1$ -linear terms in  $\{z_{ijk}\}$  variables. After the substitution,  $\mathcal{A}$  is now computing a matrix  $M$  of degree- $d$  noncommutative polynomials. Clearly, it is an identity-preserving substitution. I.e.,  $\mathcal{A}$  is nonzero if and only if  $M$  is nonzero. Therefore, if  $\mathcal{A}$  is nonzero, we can find a  $d \times d$  matrix substitution for the  $\{z_{ijk}\}$  variables such that  $M$  evaluated on that substitution is nonzero.<sup>2</sup> Hence, we obtain an  $\ell_1 d \times \ell_1 d$  matrix tuple for the  $\underline{x}$  variables such that  $\mathcal{A}$  is nonzero on that substitution. ■

**Claim 20.** *Suppose for a nonzero  $n$ -variate degree- $d$   $D_1$ -ABP  $\mathcal{A}$  of width  $r$ , there is a matrix tuple  $(p_1, \dots, p_n) \in \text{Mat}_{d'}(K_1)^n$  such that the ABP is nonzero evaluated on that tuple. Let  $\widetilde{D}_1 = (\widetilde{K}_1/F, \widetilde{\sigma}, z)$  be a cyclic division algebra of index  $d'$ , where  $K_1$  is a subfield of  $\widetilde{K}_1$ . Then there is a tuple in  $\widetilde{D}_1^n$  such that the  $D_1$ -ABP  $\mathcal{A}$  is nonzero evaluated on that tuple as well.*

*Proof.* Let  $\{\widetilde{C}_{j,k}\}_{1 \leq j,k \leq \ell_1 d'}$  be the basis of the division algebra  $\widetilde{D}_1$ . By [Proposition 11](#), we can write each matrix  $p_i = \sum_{j,k} \lambda_{ijk} \widetilde{C}_{jk}$  where each  $\lambda_{ijk} \in \widetilde{K}_1$ . Define new commuting indeterminates  $\{u_{ijk}\}$  and let  $\widetilde{p}_i = \sum u_{ijk} \widetilde{C}_{jk}$ . Evaluating  $\mathcal{A}$  on  $(\widetilde{p}_1, \dots, \widetilde{p}_n)$  then gives a nonzero matrix of commutative

<sup>2</sup>In fact,  $\lceil d/2 \rceil + 1$ -dimensional matrix substitutions will suffice [[AL50](#)].

polynomials, as it is nonzero if  $u_{ijk} \leftarrow \lambda_{ijk}$ . We can now find a substitution for each  $u_{ijk} \leftarrow \gamma_{ijk} \in \mathbb{Q}$  such that such a nonzero polynomial evaluates to nonzero. Hence, we can define a tuple  $(q_1, \dots, q_n)$  where each  $q_i = \sum \gamma_{ijk} C_{jk}$  such that  $\mathcal{A}$  is nonzero on  $(q_1, \dots, q_n)$ . Now the proof follows since each  $q_i \in \widetilde{D}_1$  as stated in [Fact 10](#). ■

**Lemma 21.** *For any nonzero  $n$ -variate degree- $d$   $D_1$ -ABP  $\mathcal{A}$  of width  $r$ , there is a cyclic division algebra  $\widetilde{D}_1$  of index  $\ell_1 \ell_2$  (where  $\ell_2 \geq d$  and  $\ell_2$  is relatively prime to  $\ell_1$ ) and a tuple in  $\widetilde{D}_1^n$  such that  $\mathcal{A}$  is nonzero evaluated on that tuple.*

*Proof.* Consider a cyclic division algebra  $D_2$  of index  $\ell_2$ . Define  $\widetilde{D}_1 = D_1 \otimes D_2$ . By assumption,  $\ell_2 (\geq d)$  is relatively prime to  $\ell_1$ . Therefore,  $\widetilde{D}_1$  is also a cyclic division algebra by [Theorem 12](#). Now the proof follows from [Claim 19](#) and [Claim 20](#). ■

We are now ready to prove the main result of this section.

**Theorem 22** (Division algebra hitting set for  $D_1$ -ABPs). *Let  $D_1$  be a cyclic division algebra of index  $\ell_1$  and  $\rho_2$  be any prime that is not a divisor of  $\ell_1$ . For the class of  $n$ -variate degree- $d$   $D_1$ -ABPs of width  $r$ , we can construct a hitting set  $\widehat{\mathcal{H}}_{n,r,d}^{D_1} \subseteq \widetilde{D}_1^n$  of size  $(\ell_1 n r d)^{O(\rho_2 \log \rho_2 \log d)}$  in deterministic  $(\ell_1 n r d)^{O(\rho_2 \log \rho_2 \log d)}$ -time where  $\widetilde{D}_1$  is a cyclic division algebra of index  $\ell_1 \ell_2$ . Here  $\ell_2 = \rho_2^{L_2}$  where  $L_2 = O(\rho_2 \log(\ell_1 n r d))$ . Moreover,  $D_1$  is a subalgebra of  $\widetilde{D}_1$ .*

*Proof.* Let  $\{C_{jk}\}_{1 \leq j,k \leq \ell_1}$  be the basis of  $D_1$ . Introduce a set of noncommuting indeterminates  $\{y_{ijk}\}_{i \in [n], j,k \in [\ell_1]}$ . Consider the following mapping:

$$x_i \mapsto \sum_{j,k} C_{jk} \otimes y_{ijk}.$$

Equivalently, each  $x_i$  is substituted by an  $\ell_1 \times \ell_1$  matrix. Fix a  $D_1$ -ABP  $\mathcal{A}$ . Consider each edge of  $\mathcal{A}$  labeled as  $\sum_{i=1}^n a_i x_i b_i$  where  $a_i, b_i \in D_1$ . Replace each  $a_i, b_i \in D_1$  by its matrix representation in  $\text{Mat}_{\ell_1}(K_1)$  and  $x_i$  by the  $\ell_1 \times \ell_1$  matrix  $\sum_{j,k} C_{jk} \otimes y_{ijk}$ . Therefore, each edge is now labeled by an  $\ell_1 \times \ell_1$  matrix whose entries are  $K_1$ -linear terms in  $\{y_{ijk}\}$  variables. After the substitution,  $\mathcal{A}$  is now computing a matrix  $M$  of degree- $d$  noncommutative polynomials in  $\{y_{ijk}\}$  variables.

**Claim 23.** *If the  $D_1$ -ABP  $\mathcal{A}(x)$  is nonzero then the matrix  $M \in \text{Mat}_{\ell_1}(\mathbb{F}\langle y \rangle)$  is nonzero.*

*Proof.* If  $\mathcal{A}(x)$  is nonzero, then it is nonzero evaluated at some  $(p_1, \dots, p_n) \in \widetilde{D}_1^n$  where  $\widetilde{D}_1 = D_1 \otimes D_2$  ([Lemma 21](#)). We can therefore expand the  $D_1$  component in the  $\{C_{jk}\}$  basis and write each  $p_i = \sum C_{jk} \otimes q_{ijk}$  for some  $q_{ijk} \in D_2$ . Therefore  $M$  is nonzero under the substitution each  $y_{ijk} \leftarrow q_{ijk}$ . ■

We now claim that each entry of  $M$  is computable by a small ABP.

**Claim 24.** *For each  $1 \leq j, k \leq \ell_1$ , the  $(j, k)^{\text{th}}$  entry of the matrix  $M \in \text{Mat}_{\ell_1}(\mathbb{F}\langle y \rangle)$  is computable by an  $\ell_1^2 n$ -variate degree- $d$  noncommutative homogeneous ABP of width  $\ell_1 r$ .*

*Proof.* For each vertex  $v$  in the  $D_1$ -ABP  $\mathcal{A}$ , make  $\ell_1$  copies of  $v$  (including the source  $S$  and sink  $T$ ), let us call it  $(v, 1), \dots, (v, \ell_1)$ . For any two vertices  $u$  and  $v$ , suppose the edge is labeled by  $\sum_{i=1}^n a_i x_i b_i$  and  $M_{u,v}$  be the corresponding  $\ell_1 \times \ell_1$  matrix after substitution. Then for each  $1 \leq \hat{j}, \hat{k} \leq \ell_1$ , we add an edge  $((u, \hat{j}), (v, \hat{k}))$  labeled by the  $(\hat{j}, \hat{k})^{th}$  entry of  $M_{u,v}$ . Note that product of the edge labels of a path exactly captures the corresponding matrix product. Therefore, if we consider the ABP with source  $(S, \hat{j})$  and sink  $(T, \hat{k})$ , it is computing the  $(\hat{j}, \hat{k})^{th}$  entry of the matrix  $M$ . Note that the width of the new ABP is  $\ell_1 r$ . ■

We now consider a nonzero entry of the matrix  $M$  and get a division algebra hitting set for this  $\ell_1^2 n$ -variate degree- $d$  noncommutative homogeneous ABP of width  $\ell_1 r$  inside a cyclic division algebra  $D_2$  of index  $\ell_2$  where  $\ell_2$  is relatively prime to  $\ell_1$  by construction. For  $\ell_2$  we need to use any prime  $\rho_2$  that is not a divisor of  $\ell_1$ . Let  $\ell_2 = \rho_2^{L_2}$ . Define  $\tilde{D}_1 = D_1 \otimes D_2$ .

$$\text{Finally, } \widehat{\mathcal{H}}_{n,r,d}^{D_1} = \left\{ (q_1, \dots, q_n) : q_i = \sum_{j,k} C_{jk} \otimes q_{ijk} \quad \text{where } (q_{111}, \dots, q_{n\ell_1\ell_1}) \in \widehat{\mathcal{H}}_{\ell_1^2 n, \ell_1 r, d} \right\}. \quad (5) \quad \blacksquare$$

By [Theorem 17](#), the size of  $\widehat{\mathcal{H}}_{\ell_1^2 n, \ell_1 r, d}$  is  $(\ell_1 n r d)^{O(\rho_2 \log \rho_2 \log d)}$  and  $L_2$  is  $O(\rho_2 \log(\ell_1 n r d))$ .

## 5 Hitting Set for NSINGULAR given a Witness

In this section, we consider the NSINGULAR problem for linear matrices of size  $s \times s$  under the promise that we already have a witness matrix tuple such that a submatrix of size  $s - 1$  is invertible on that tuple. The result of this section is crucial for the hitting set construction for rational formulas in [Section 6](#).

More precisely, we construct the hitting set for rational formulas inductively on the inversion height. To construct a hitting set for inversion height  $\theta$  from inversion height  $\theta - 1$ , we will use the promised version of the NSINGULAR problem.

**Theorem 25.** *Let  $T(\underline{x})$  be a linear matrix of size  $s$  in  $\{x_1, \dots, x_n\}$  variables invertible over  $F\langle\langle \underline{x} \rangle\rangle$  and  $D_1$  be a cyclic division algebra of index  $\ell_1$ . Let  $\rho_2$  be any prime which is not a divisor of  $\ell_1$ . Then, given a tuple  $(p_1, \dots, p_n) \in D_1^n$  such that there is a submatrix  $T'$  of  $T$  of size  $s - 1$  such that  $T'(p)$  is invertible, we can construct a hitting set  $\tilde{\mathcal{H}}_{n,s,\ell_1}^p \subseteq \tilde{D}_1^n$  of size  $(\ell_1 n s)^{O(\rho_2 \log \rho_2 \log(\ell_1 s))}$  in deterministic  $(\ell_1 n s)^{O(\rho_2 \log \rho_2 \log(\ell_1 s))}$ -time where  $\tilde{D}_1$  is a cyclic division algebra of index  $\ell_1 \rho_2^{O(\rho_2 \log(\ell_1 n s))}$ , such that if  $T(\underline{x})$  is invertible then for some  $(q_1, \dots, q_n) \in \tilde{\mathcal{H}}_{n,s,\ell_1}^p$ ,  $T(q)$  is invertible.*

*Proof.* We can find two invertible transformations  $U, V$  in  $\text{Mat}_s(D_1)$  such that

$$U \cdot T(p_1, p_2, \dots, p_n) \cdot V = \left[ \begin{array}{c|c} I_{s-1} & 0 \\ \hline 0 & 0 \end{array} \right],$$

where  $I_{s-1}$  is the identity matrix whose diagonal elements are the identity element of  $D_1$ . This is possible since one can do Gaussian elimination over division algebras.

Notice that  $T(\underline{x} + \underline{p}) = T(\underline{p}) + T(\underline{x})$ . Hence, we can write

$$T(\underline{x} + \underline{p}) = U^{-1} \cdot \left( \left[ \begin{array}{c|c} I_{s-1} & 0 \\ \hline 0 & 0 \end{array} \right] + U \cdot T(\underline{x}) \cdot V \right) \cdot V^{-1}.$$

$$\text{Equivalently, } T(\underline{x} + \underline{p}) = U^{-1} \cdot \left[ \begin{array}{c|c} I_{s-1} - L & A_i \\ \hline B_j & C_{ij} \end{array} \right] \cdot V^{-1},$$

where each entry of  $L, A_i, B_j, C_{ij}$  are  $D_1$ -linear forms in  $\underline{x}$  variables with no constant term. We can simplify it further by multiplying both sides by invertible matrices and writing,

$$T(\underline{x} + \underline{p}) = U^{-1} U' \left[ \begin{array}{c|c} I_{s-1} - L & 0 \\ \hline 0 & C_{ij} - B_i(I_{s-1} - L)^{-1} A_j \end{array} \right] V' V^{-1}. \quad (6)$$

$$\text{where, } U' = \left[ \begin{array}{c|c} I_{s-1} & 0 \\ \hline B_i(I_{s-1} - L)^{-1} & 1 \end{array} \right], \quad V' = \left[ \begin{array}{c|c} I_{s-1} & (I_{s-1} - L)^{-1} A_j \\ \hline 0 & 1 \end{array} \right].$$

$$\text{Let, } P_{ij}(\underline{x}) = C_{ij} - B_i(I_{s-1} - L)^{-1} A_j. \quad (7)$$

We can also represent  $P_{ij}$  as a series:

$$P_{ij}(\underline{x}) = C_{ij} - B_i \left( \sum_{k \geq 0} L^k \right) A_j.$$

This is a generalized series (in  $\underline{x}$  variables) over the division algebra  $D_1$  where the division algebra elements can interleave in between the variables.

**Claim 26.**

$$\text{Define, } \tilde{P}_{ij}(\underline{x}) = C_{ij} - B_i \left( \sum_{0 \leq k \leq (s-1)\ell_1} L^k \right) A_j.$$

$$\text{Then, } P_{ij}(\underline{x}) = 0 \iff \tilde{P}_{ij}(\underline{x}) = 0.$$

*Proof.* If we substitute each  $x_i$  by the generic  $\ell_1 \times \ell_1$  matrix of noncommuting indeterminates  $Z_i = (z_{ijk})_{1 \leq j, k \leq \ell_1}$ , the generalized series  $P_{ij}$  then computes a matrix of recognizable series over the variables  $\{z_{ijk}\}_{1 \leq i \leq n, 1 \leq j, k \leq \ell_1}$  (the proof is similar to [Claim 24](#)). Then [Fact 13](#) implies that if we truncate  $P_{ij}(Z)$  within degree  $(s-1)\ell_1$ , we get a nonzero matrix of polynomials computed by ABPs. Note that, substituting each  $x_i$  by the generic  $\ell_1 \times \ell_1$  matrix  $Z_i = (z_{ijk})_{j,k}$  in  $\tilde{P}_{ij}$  will have the same effect. Therefore,

$$P_{ij}(\underline{x}) = 0 \iff \tilde{P}_{ij}(\underline{x}) = 0. \quad \blacksquare$$

$$\text{We can now write, } P_{ij} = 0 \iff \left( C_{ij} = 0 \quad \text{and} \quad \text{for each } 0 \leq k \leq (s-1)\ell_1, \quad B_i L^k A_j = 0 \right),$$

where each  $B_i L^k A_j$  is a generalized polynomial over  $D_1$ , indeed it is a  $D_1$ -ABP.

The following statement now reduces the singularity testing to identity testing of a  $D_1$ -ABP.<sup>3</sup>

<sup>3</sup>In a recent work [?], a similar idea is used to show a polynomial-time reduction from NSINGULAR to identity testing of noncommutative ABPs in the white-box setting.

**Claim 27.**  $T(\underline{x})$  is invertible over  $F\langle\langle\underline{x}\rangle\rangle$  if and only if  $\widetilde{P}_{ij} \neq 0$ .

*Proof.* Let  $\widetilde{P}_{ij}$  be zero, therefore  $P_{ij}$  is also zero, and for every matrix substitution it is zero. Assume to the contrary,  $T(\underline{x})$  is invertible over  $F\langle\langle\underline{x}\rangle\rangle$ . Then, there exists a matrix tuple  $(p'_1, \dots, p'_n) \in \text{Mat}_{k\ell_1}(K)^n$  for some large enough integer  $k$  and an extension field  $K$ , such that  $T(p')$  is invertible. We now evaluate Equation (6) substituting each  $x_i \mapsto p'_i - p_i \otimes I_k$ . Clearly,  $P_{ij}$  must be nonzero on that substitution which leads to a contradiction.

For the other direction, if  $\widetilde{P}_{ij} \neq 0$ , then there exists a matrix tuple  $(q_1, \dots, q_n) \in \widetilde{D}_1^n$  where  $\widetilde{D}_1$  is a cyclic division algebra of index  $\ell_1\ell_2$  (see Lemma 21), such that  $\widetilde{P}_{ij}(q) \neq 0$ . We then evaluate  $T(\underline{x} + \underline{p})$  on  $(tq_1, \dots, tq_n)$  where  $t$  is a commutative variable. Clearly, the infinite series  $P_{ij}$  is nonzero at  $tq$ . Also,  $(I_{s-1} - L)(tq)$  is invertible.

However, this also shows that  $P_{ij}(tq)$  is a nonzero matrix of rational expressions in  $t$ , and the determinant of  $(I_{s-1} - L)(tq)$  is a nonzero polynomial. Since the degrees of the polynomials in the rational expressions and the determinant are bounded by a polynomial, we can vary the parameter  $t$  over a polynomial-size set  $\Gamma \subset \mathbb{Q}$  such that  $P_{ij}(tq)$  and  $\det(I_{s-1} - L)(tq)$  are nonzero, for some  $t \in \Gamma$ . As we need to only avoid the roots of the numerator and the denominator polynomials present in  $P_{ij}(tq)$ , and the roots of  $\det(I_{s-1} - L)(tq)$ , it suffices to choose  $\Gamma \subset \mathbb{Q}$  of size  $\text{poly}(s, \ell_1, \ell_2)$ . Therefore,  $T(tq + \underline{p} \otimes I_{\ell_2})$  is invertible for some  $t \in \Gamma$  by Equation (6). ■

Let  $k_0$  be the minimum  $k$  such that  $B_i L^k A_j \neq 0$ . Now apply Theorem 22 on  $B_i L^{k_0} A_j$  to construct a hitting set  $\widehat{\mathcal{H}}_{n,s-1,\ell_1(s-1)}^{D_1}$  of size  $\leq (ns\ell_1)^{O(\rho_2 \log \rho_2 \log(s\ell_1))}$  inside a division algebra  $\widetilde{D}_1$  of index  $\ell_1\ell_2$ , where  $\ell_2 = \rho_2^{L_2}$  for a prime  $\rho_2$  that does not divide  $\ell_1$ . Moreover,  $L_2 = O(\rho_2 \log(\ell_1 n s \rho_2))$ . Hence the set  $\Gamma$  can be chosen to be of size  $(ns\ell_1)^{O(\rho_2)}$ .

This gives the final hitting set,

$$\widehat{\mathcal{H}}_{n,s,\ell_1}^p = \left\{ (tq_1 + p_1 \otimes I_{\ell_2}, \dots, tq_n + p_n \otimes I_{\ell_2}) : q \in \widehat{\mathcal{H}}_{n,s-1,\ell_1(s-1)}^{D_1} \text{ and } t \in \Gamma \right\}. \quad (8)$$

## 6 Proof of the Main Result

In this section, we prove Theorem 2. For ease of exposition and clarity, we divide the proof into two subsections. In the first subsection, we prove a weaker statement that yields a quasipolynomial-size hitting set for rational formulas of constant inversion height. Building on this, in the next subsection, we explain the steps to strengthen the result and obtain a quasipolynomial-size hitting set for the general case i.e. for all rational formulas of polynomial size.

### 6.1 Hitting set for rational formulas of constant inversion height

**Theorem 28** (Black-box RIT for constant inversion height). *For the class of  $n$ -variate noncommutative rational formulas of size  $s$  and inversion height  $\theta$ , we can construct a hitting set  $\mathcal{H}'_{n,s,\theta} \subseteq \text{Mat}_{\ell_\theta}(\mathbb{Q})^n$  of size  $(ns)^{2^{O(\theta^2)} \log ns}$  in deterministic time  $(ns)^{2^{O(\theta^2)} \log ns}$ , where  $\ell_\theta \leq (ns)^{2^{O(\theta^2)}}$ .*

*Proof.* The proof is by induction on the inversion height of a rational formula. We will show that for every inversion height  $\theta$  we can construct a hitting set  $\mathcal{H}_{n,s,\theta} \subseteq D_\theta^n$  as claimed, where  $D_\theta$  is a cyclic division algebra. The base case  $\theta = 0$  is for noncommutative formulas (which have inversion height 0). Such a construction is given for noncommutative formulas without inversions, in fact even for noncommutative ABPs [ACM22].

Inductively assume that we have such a construction for rational formulas of size  $s$  and inversion height  $\theta - 1$ . Let  $r(\underline{x})$  be any rational formula of inversion height  $\theta$  in  $\mathbb{Q}\langle\langle \underline{x} \rangle\rangle$  of size  $s$ . We first show the following.

**Claim 29.** *For every rational formula  $r$  of inversion height  $\theta$  in  $\mathbb{Q}\langle\langle \underline{x} \rangle\rangle$  of size  $s$ , there exists a  $\underline{p} \in \mathcal{H}_{n,s,\theta-1}$  such that  $r(\underline{p})$  is defined.*

*Proof.* Let  $\mathcal{F}$  be the collection of all those inverse gates in the formula such that for every  $g \in \mathcal{F}$ , the path from the root to  $g$  does not contain any inverse gate. For each  $g_i \in \mathcal{F}$ , let  $h_i$  be the subformula input to  $g_i$ . Consider the formula  $h = h_1 h_2 \cdots h_k$  (where  $k = |\mathcal{F}|$ ) which is of size at most  $s$  since for each  $i, j$ ,  $h_i$  and  $h_j$  are disjoint. Note that  $h$  is of inversion height  $\theta - 1$ . Therefore, for some  $\underline{p} \in \mathcal{H}_{n,s,\theta-1}$ ,  $h(\underline{p})$  is nonzero and hence invertible as it is a division algebra hitting set. Therefore, each  $h_i$  is also invertible at  $\underline{p}$ . By definition, the path from the root to each  $g$  does not contain any inverse gate. Hence,  $r(\underline{x})$  is defined at  $\underline{p}$ . ■

If the rational formula  $r$  has size  $s$ , it is shown in [HW15, Theorem 2.6] that  $r$  can be represented as the top right corner of the inverse of a linear matrix of size at most  $2s$ . More precisely,  $r(\underline{x}) = uL^{-1}v^t$  where  $L$  is a linear matrix of size at most  $2s$  and  $u, v \in \mathbb{Q}^{2s}$  are  $2s$ -dimensional vectors whose first (resp. last) entry is 1 and others are zero. Therefore,  $r^{-1}$  can be written as the following [HW15, Equation 6.3]:

$$r^{-1}(\underline{x}) = [1 \ 0 \ \dots \ 0] \cdot \widehat{L}^{-1} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad \text{where} \quad \widehat{L} = \left[ \begin{array}{c|c} v^t & L \\ \hline 0 & -u \end{array} \right].$$

By Claim 29 the formula  $r$  is defined for some  $\underline{p} \in \mathcal{H}_{n,s,\theta-1}$ . Therefore,  $L$  is invertible at  $\underline{p}$  (see [HW15, Proposition 7.1]).

Our goal is to find a division algebra tuple such that  $r$  is nonzero and hence invertible. Equivalently, the goal is to find a division algebra tuple such that  $r^{-1}$  is defined, and therefore  $\widehat{L}$  is invertible [HW15, Proposition 7.1].

Notice that  $\widehat{L}$  is of size at most  $2s + 1$ . Moreover, we know a tuple  $\underline{p} \in \mathcal{H}_{n,s,\theta-1}$  such that a submatrix  $L$  of  $\widehat{L}$  of size  $2s$  is invertible. We can now use the construction of  $\widetilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^{\underline{p}}$  (where  $\ell_{\theta-1}$  is the index of the cyclic division algebra  $D_{\theta-1}$ ), as described in Theorem 25, to find a tuple  $\underline{q}$  inside a division algebra of dimension  $\ell_\theta$  such that  $\widehat{L}(\underline{q})$  is invertible, therefore  $r(\underline{q})$  is nonzero.

We now obtain the following hitting set:

$$\mathcal{H}_{n,s,\theta} = \left\{ \widetilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^p : \underline{p} \in \mathcal{H}_{n,s,\theta-1} \subseteq D_{\theta-1}^n \right\},$$

where  $\widetilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^p = \left\{ (tq_1 + p_1 \otimes I_{\ell_2}, \dots, tq_n + p_n \otimes I_{\ell_2}) : \underline{q} \in \widehat{\mathcal{H}}_{n,2s,\ell_{\theta-1}(2s+1)}^{D_1} \text{ and } t \in \Gamma \right\}$

and  $\widehat{\mathcal{H}}_{n,2s,\ell_{\theta-1}(2s+1)}^{D_1} = \left\{ (q_1, \dots, q_n) : q_i = \sum_{j,k} C_{jk} \otimes q_{ijk} : (q_{111}, \dots, q_{n\ell_{\theta-1}\ell_{\theta-1}}) \in \widehat{\mathcal{H}}_{\ell_{\theta-1}n,2\ell_{\theta-1}s,\ell_{\theta-1}(2s+1)} \right\}.$

Using our construction, we get that  $\ell_\theta = \ell_{\theta-1} \rho_\theta^{O(\rho_\theta \log(\ell_{\theta-1} sn))} = \ell_{\theta-1} (\ell_{\theta-1} sn)^{O(\rho_\theta \log \rho_\theta)}$ . We choose  $\rho_\theta$  to be the  $(\theta + 1)^{th}$  prime selected at the  $\theta^{th}$  stage. By prime number theorem  $\rho_\theta \leq \theta \log \theta$ . Now we want to argue that  $\ell_\theta \leq (ns)^{c\theta^2}$  for sufficiently large constant  $c$ .

To see that, note that  $\ell_\theta \leq (\ell_{\theta-1})^{1+O(\theta^2)} (ns)^{O(\theta^2)}$ . Inductively,  $\ell_{\theta-1} \leq (ns)^{c(\theta-1)^2}$ . Therefore,

$$\ell_\theta \leq (ns)^{c(\theta-1)^2(1+O(\theta^2))} \cdot (ns)^{O(\theta^2)} \leq (ns)^{c\theta^2},$$

for sufficiently large constant  $c$ . Note that at the base case,  $\ell_0 = (ns)^{O(1)}$  [ACM22]. Similarly, by unfolding the recursion, we get

$$|\mathcal{H}_{n,s,\theta}| = |\mathcal{H}_{n,s,\theta-1}| \cdot |\widetilde{\mathcal{H}}_{\ell_{\theta-1}n,2\ell_{\theta-1}s,2s+1}^2|.$$

Solving it, we get that  $|\mathcal{H}_{n,s,\theta}| \leq (ns)^{2^{O(\theta^2)} \log ns}$ .

Note that  $\mathcal{H}_{n,s,\theta} \subseteq D_\theta^n$ . That is, the entry of each matrix in the hitting set is in  $\mathbb{Q}(z, \omega)$  where  $\omega$  is a complex  $\rho_\theta^{\ell_\theta}$  root of unity. We now discuss how to obtain a hitting set over  $\mathbb{Q}$  itself. In the hitting set points suppose we replace  $\omega$  and  $z$  by commuting indeterminates  $t_1, t_2$  of degree bounded by  $\ell_\theta$ . Then, for any nonzero rational formula  $r$  of size  $s$  there is a matrix tuple in the hitting set on which  $r$  evaluates to a nonzero matrix  $M(t_1, t_2)$  of dimension  $\text{poly}(n, s, 2^{\theta^2})$  over the commutative function field  $\mathbb{Q}(t_1, t_2)$ . It is easy to show that each entry of  $M(t_1, t_2)$  is a commutative rational function of the form  $a/b$ , where  $a$  and  $b$  are polynomials in  $t_1$  and  $t_2$  and the degrees of both  $a$  and  $b$  are bounded by  $\text{poly}(n, s, 2^{\theta^2})$ . We can now vary the parameters  $t_1, t_2$  over a sufficiently large set  $\widetilde{T} \subseteq \mathbb{Q}$  of size  $\text{poly}(n, s, 2^{\theta^2})$  such that we avoid the roots of the numerator and denominator polynomials involved in the computation. This gives our final hitting set  $\mathcal{H}'_{n,s,\theta} \subseteq \text{Mat}_{\ell_\theta}^n(\mathbb{Q})$  defined as:

$$\left\{ \underline{q}'(\alpha_1, \alpha_2) : \underline{q}'(\omega, z) \in \mathcal{H}_{n,s,\theta} \subseteq D_\theta^n, (\alpha_1, \alpha_2) \in \widetilde{T} \times \widetilde{T} \right\}. \quad \blacksquare$$

## 6.2 Hitting set construction for all rational formulas

In this section, our goal is to improve the upper bound of Theorem 28 and obtain a quasipolynomial-size hitting set for the general case. We first analyze the source of blow-up (incurred by the inversion height  $\theta$ ) and then figure out the means to control it.

Recall from the last theorem that,  $\mathcal{H}_{n,s,\theta} \subseteq D_\theta^n$  is the hitting set for  $n$ -variate size- $s$  rational formulas of inversion height  $\theta$  where  $D_\theta$  is a cyclic division algebra of index  $\ell_\theta$ . From the

hitting set construction of [Theorem 28](#),  $\ell_\theta = \ell_{\theta-1} \cdot \ell$  where  $\ell$  is the dimension of the hitting set  $\widehat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1} s, \ell_{\theta-1}(2s+1)}^4$ .

What is the value of  $\ell$ ? Recall from the proof of [Theorem 28](#) that  $\ell = \rho_\theta^{O(\rho_\theta \log(\ell_{\theta-1} s n))}$ . The source of blow-up is the presence of  $\ell_{\theta-1}$  term in  $\log(\ell_{\theta-1} s n)$  on the exponent of  $\rho_\theta$ . Since, by unfolding the recursion, we get that  $\ell_\theta$  has  $\ell_0^{2^{O(\theta^2)}}$  dependency. Thus our (re-defined) goal is to construct a hitting set in which the dependence of  $\ell_{\theta-1}$  in  $\ell$  is only logarithmic. Now look at the parameters in  $\widehat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1} s, \ell_{\theta-1}(2s+1)}$ . All the parameters (i.e. the number of variables, width, degree) contain  $\ell_{\theta-1}$ . Thus it is important to control their dependence in  $\ell$ . More precisely, these three parameters enter from the following sources:

1. The degree of the truncated generalized ABP in [Claim 26](#).
2. Dependency on width in the hitting set dimension in [Theorem 17](#).
3. Dependency on the number of variables in the hitting set dimension in [Theorem 17](#).

We now explain how to modify the hitting set construction to deal with each of these.

### 6.2.1 Degree Improvement

First we analyze the degree bound of the truncated generalized ABP as obtained in [Claim 26](#) and show how it can be improved.

**Claim 30.** Consider the generalized  $D_1$ -series  $P_{ij}$  as defined in [Equation \(7\)](#) where  $D_1 = (K_1/F, \sigma_1, z)$ .

$$P_{ij}(\underline{x}) = C_{ij} - B_i \left( \sum_{k \geq 0} L^k \right) A_j,$$

$$\text{Define its truncation: } \widetilde{P}_{ij}(\underline{x}) = C_{ij} - B_i \left( \sum_{0 \leq k \leq s-1} L^k \right) A_j.$$

$$\text{Then } P_{ij}(\underline{x}) = 0 \iff \widetilde{P}_{ij}(\underline{x}) = 0.$$

*Proof.* Suppose  $P_{ij}$  is nonzero. Substitute each  $\{x_i : 1 \leq i \leq n\}$  by the following map used in the proof of [Theorem 22](#):

$$x_i \mapsto \sum_{j,k} C_{jk} \otimes y_{ijk}.$$

Let  $L = \sum_i L_i x_i$ . Since  $C_{jk}, 1 \leq j, k \leq \ell_1$  is a basis for the division algebra  $D_1$ , we can write each entry of  $L_i$  as  $\sum_{j,k} \beta_{ijk} C_{jk}$  for some  $\beta_{ijk} \in F$ . Substituting for each  $x_i$  as above, it follows that each entry of  $L$  can be expressed as  $\sum_{j,k} (C_{jk} \otimes \sum_i \alpha_{ijk} y_{ijk})$ , where each  $\alpha_{ijk} \in F$ . Therefore, it now computes a series  $\sum C_{jk} \otimes f_{jk} \in D_1 \otimes_F F\langle\langle y \rangle\rangle$ . We first observe the following claim. Its proof is omitted as it is a straightforward generalization of the proof of [Claim 23](#).

**Claim 31.**  $P_{ij}(\underline{x}) = 0 \iff \sum C_{jk} \otimes f_{jk} = 0$ .

---

<sup>4</sup>By dimension of the hitting set we mean here the dimension of the matrices in the matrix representation of the division algebra elements.



Recall that,  $D_1\langle\langle y \rangle\rangle$  denotes the formal power series in noncommuting  $y$  variables where the coefficients are in  $D_1$  and  $y$  variables commute with the elements in  $D_1$ . We now define the following map:

$$\begin{aligned}\psi : D_1 \otimes_F F\langle\langle y \rangle\rangle &\rightarrow D_1\langle\langle y \rangle\rangle, \\ C_{jk} \otimes y_{ijk} &\mapsto C_{jk}y_{ijk}.\end{aligned}$$

Note that,  $\psi$  is an isomorphism. Each entry of the matrix  $L$  is now of form  $\sum_{i,j,k} \beta_{ijk}y_{ijk}$  and computes a series in  $D_1\langle\langle y \rangle\rangle$ . Therefore, we can apply [Fact 13](#) and truncate it to degree  $s - 1$  preserving the nonzeroness.

Clearly, applying the substitution  $x_i \mapsto \sum_{j,k} C_{jk} \otimes y_{ijk}$  and then the  $\psi$ -map on  $\tilde{P}_{ij}$  will have the same effect. Therefore,  $\tilde{P}_{ij}$  is also nonzero.  $\blacksquare$

## 6.2.2 Improving the dependency of dimension on the width

In this section, we modify the hitting set construction of [Theorem 17](#) and make the dimension of the hitting set *independent* of the ABP width. More precisely, we show the following.

**Theorem 32.** *Let  $p$  be any prime number. For the class of  $n$ -variate degree  $\tilde{d}$  noncommutative polynomials computed by homogeneous ABPs of width  $r$ , we can construct a hitting set  $\widehat{\mathcal{H}}_{n,r,\tilde{d}} \subseteq D_2^n$  of size  $(nr\tilde{d})^{O(p \log p \log \tilde{d})}$  in  $(nr\tilde{d})^{O(p \log p \log \tilde{d})}$  time. Here  $D_2$  is a cyclic division algebra of index  $\ell_2 = p^L$  where  $L \leq O(p \log(n\tilde{d}))$ .*

*Proof.* The proof is along the same lines as the proof of [Theorem 17](#) with a few crucial modifications. Let  $\Lambda = 2^\tau$ , the order of the root of unity  $\omega_0$ , be sufficiently large (indeed, it suffices to choose  $\tau$  such that  $\Lambda$  is larger than all the values of  $r^{3p}$  that will arise in the recursive hitting set construction. The actual value of  $\Lambda$ , that will turn out to be quasipolynomially bounded, we shall fix later in the analysis). Define  $\omega_0$  as the primitive  $\Lambda^{\text{th}}$  root of unity. We set the base field for the cyclic division algebra constructions as  $F = \mathbb{Q}(z, \omega_0)$ .

The following lemma is, *mutatis mutandis*, the same as [Lemma 16](#) except the value of  $\mu$  we set.

**Lemma 33.** *Consider  $p$  many families of  $r \times r$  matrices  $\mathcal{M}_1 = \{M_{1,1}, M_{1,2}, \dots, M_{1,p^{d-1}}\}, \dots, \mathcal{M}_p = \{M_{p,1}, M_{p,2}, \dots, M_{p,p^{d-1}}\}$  where for the  $j^{\text{th}}$  family the entries are univariate polynomials over  $F[u_j]$  of degree less than  $n$ . Let  $(f_1(u), f_2(u), \dots, f_{p^{d-1}}(u)) \in F[u]$  be polynomials of degree at most  $m$ . Let  $\omega \in \overline{F}$  be a root of unity of order more than  $(p^{d-1}nm)^p$ , and let  $K = F(\omega)$ . Define polynomials in indeterminate  $v$ :*

$$\begin{aligned}f'_i(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_1}) q_\ell(v), \quad 1 \leq i \leq p^{d-1} \\ f'_{i+p^{d-1}}(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_2}) q_\ell(v), \quad 1 \leq i \leq p^{d-1} \\ &\vdots \\ f'_{i+(p-1)p^{d-1}}(v) &= \sum_{\ell=1}^{r^2} f_i((\omega^\ell \alpha_d)^{\mu_p}) q_\ell(v), \quad 1 \leq i \leq p^{d-1}\end{aligned}$$

where  $\mu_j = \mu^{j-1}$ ,  $\mu = 1 + \Lambda p^{d-1}nm$ , and  $q_t(v)$  is the corresponding Lagrange interpolation polynomial.

Then, for all but  $(p^{d-1}nmr)^p$  many values of  $\alpha_d$ , the  $K$ -linear span of the matrix coefficients of the matrix product  $\prod_{j=1}^p \prod_{i=1}^{p^{d-1}} M_{j,i}(f_i(u_j))$  is contained in the  $K$ -linear span of the matrix coefficients of the product  $\prod_{j=1}^p \prod_{i=1}^{p^{d-1}} M_{j,i}(f'_i(v))$ .

Now we are ready to prove [Theorem 32](#). We will set  $\ell_2 = p^L$  as the index of the division algebra  $D_2$ , where  $p \neq 2$  is the given prime and  $L$  will be determined in the analysis below. One of the necessary conditions is that  $p^L > \tilde{d}$ .

As mentioned before, an important step in [\[FS13\]](#) is to convert the given ABP into a set-multilinear form and eventually a read-once form. More specifically, they replace the noncommutative variable  $x_i$  by the matrix  $M(x_i)$ :

$$M(x_i) = \begin{bmatrix} 0 & x_{i1} & 0 & \cdots & 0 \\ 0 & 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & x_{i\tilde{d}} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

and the variables  $x_{i1}, x_{i2}, \dots, x_{i\tilde{d}}$  will be replaced by  $u_1^i, u_2^i, \dots, u_{\tilde{d}}^i$ . Obviously, these matrices are nilpotent matrices and they are not elements of any division algebra. These variables will be finally substituted by the output of a generator  $\mathcal{G}_{\log \tilde{d}}$  that stretches a seed  $(\alpha_1, \alpha_2, \dots, \alpha_{\log \tilde{d}+1})$  to  $(f_1(\alpha), f_2(\alpha), \dots, f_{\tilde{d}}(\alpha))$ .

Here, our plan will be to replace  $x_i$  by the following matrix  $M(x_i)$ :

$$M(x_i) = \left[ \begin{array}{ccccc|ccc} 0 & f_1^i(\alpha) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & f_2^i(\alpha) & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f_{\tilde{d}}^i(\alpha) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & f_{\tilde{d}+1}^i(\alpha) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & f_{\ell_2-1}^i(\alpha) \\ z f_{\ell_2}^i(\alpha) & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array} \right],$$

where the tuple  $(f_1(\alpha), \dots, f_{\ell_2}(\alpha))$  will be the output of a generator of seed length  $O(\log \ell_2)$ . Additionally, if we can maintain the property that each such matrix is a circulant matrix that represent a cyclic division algebra element of the form shown in [Proposition 9](#), we will be in good shape. Now we discuss the implementation of these ideas.

Choose  $\omega = e^{\frac{2\pi i}{p^L}}$ , a primitive root of unity of order  $p^L$ . Let  $F = \mathbb{Q}(z, \omega_0)$  and  $K = F(\omega)$  be its (finite) extension by  $\omega$ . Using the construction described in [Section 2.3](#), we consider the cyclic division algebra  $D_2 = (K/F, \sigma, z)$ . We fix the  $K$ -automorphism  $\sigma$  as

$$\sigma(\omega) = \omega^{\Lambda p^k + 1},$$

where the positive integer  $\kappa$  will be suitably chosen in the following analysis, fulfilling the constraints of [Lemma 33](#) and some additional requirements. Note that, as  $\sigma$  fixes  $F$  pointwise,  $\sigma(\omega_0) = \omega_0$ .

Let  $d = \log_p \tilde{d}$ , where we assume (without loss of generality) that  $\tilde{d}$  is a power of  $p$ . Let  $\omega_i = \omega^{p^{a_i}}$  for  $a_1 > a_2 > \dots > a_d > a_{d+1} > 0$ , where  $a_i$  are positive integers to be chosen. We denote by  $K_i$  the cyclic Galois extension  $K_i = F(\omega_i)$  of  $F$  by  $\omega_i$ , for  $1 \leq i \leq d+1$ . This gives a tower of field extensions

$$F \subset F(\omega_1) \subset F(\omega_2) \subset \dots \subset F(\omega_d) \subset F(\omega_{d+1}) \subset F(\omega) = K.$$

We require two properties of  $\omega_i, 1 \leq i \leq d+1$ .

1. For the hitting set generator  $\mathcal{G}_i$  we will choose the root of unity as  $\omega_i$  and the variable  $\alpha_i$  will take values only in the set

$$W_i = \{\omega_0^{\hat{j}} \omega_i^j \mid 1 \leq \hat{j} \leq \Lambda, 1 \leq j \leq p^{L-a_i}\}.$$

2. We require that the  $K$ -automorphism  $\sigma$  has the property that for all  $1 \leq i \leq d+1$  the map  $\sigma^{p^i}$  fixes  $\omega_i$ . It is enough to ensure that  $\sigma^{p^i}$  has  $F(\omega_i)$  as its fixed field.

We take up the second property. As  $\sigma(\omega) = \omega^{\Lambda p^\kappa + 1}$ , we have  $\sigma(\omega_i) = \omega^{p^{a_i}(\Lambda p^\kappa + 1)}$ . Therefore,

$$\sigma^{p^i}(\omega_i) = \omega^{p^{a_i}(\Lambda p^\kappa + 1)^{p^i}}.$$

Now,  $(\Lambda p^\kappa + 1)^{p^i} = \sum_{j=0}^{p^i} \binom{p^i}{j} \Lambda^j p^{\kappa j}$ . Choosing  $\kappa = L/2$ , we have  $\omega^{p^{\kappa j}} = 1$  for  $j \geq 2$ . Therefore,

$$\sigma^{p^i}(\omega_i) = \omega^{p^{a_i}(\Lambda p^{i+\kappa} + 1)} = \omega_i \cdot \omega^{\Lambda p^{a_i+i+\kappa}}.$$

We can set  $a_i + i + \kappa = L$  for  $1 \leq i \leq d+1$  to ensure that  $\sigma^{p^i}$  fixes  $\omega_i$ . Putting  $L = 2\kappa$ , we obtain

$$a_i = \kappa - i \quad \text{for } 1 \leq i \leq d+1. \tag{9}$$

It remains to choose  $\kappa$ . In the construction of our hitting set generator  $\mathcal{G}_i$ , the parameter  $\alpha_i$  will take values only in  $W_i$  defined above. We note that  $|W_i| = \Lambda p^{L-a_i} = \Lambda p^{\kappa+i}$  (because for two different pairs  $(j_1, j_2)$  and  $(j'_1, j'_2)$ ,  $\omega_0^{j_1} \omega_i^{j_2} \neq \omega_0^{j'_1} \omega_i^{j'_2}$  since the orders of  $\omega_i$  and  $\omega_0$  are relatively prime). By [Lemma 33](#) there are at most  $(p^d n m r)^p$  many bad values of  $\alpha_i$  for any  $i$ . Thus, it suffices to choose  $\kappa$  such that  $\Lambda p^\kappa > (p^d n m r)^p$ . As  $m \leq r^2$  and  $\Lambda > r^{3p}$ , it suffices to set

$$\kappa = \lceil p d \rceil + \lceil p \log_2 n \rceil + 1.$$

The choice of  $\kappa$  determines the value of parameter  $\mu$  in [Lemma 33](#). Since  $L = 2\kappa$ , notice that  $p^L > \tilde{d}$  is satisfied.

Coming back to the modified construction of  $\mathcal{G}_d$ , inductively, we can assume that we have already constructed hitting set generators for each *window* of length  $p^{d-1}$ . More precisely, let  $\mathcal{G}_{d-1} : (\alpha_1, \dots, \alpha_{d-1}, u) \mapsto (f_1(u), f_2(u), \dots, f_{p^{d-1}}(u))$  (where the polynomial  $f_i(u) \in K_{d-1}[u]$ , for  $1 \leq i \leq p^{d-1}$ ) with the above two properties has already been constructed. Namely, for each

window suppose  $f_{i+1}(u) = \sigma(f_i(u))$  holds for all  $i \leq \rho^{d-1} - 1$ . Now define  $\mathcal{G}_d : (\alpha_1, \dots, \alpha_d, v) \mapsto (f'_1(v), f'_2(v), \dots, f'_{\rho^d}(v))$  using [Lemma 33](#).

Since the Lagrange interpolation polynomial  $q_\ell(v)$  has only integer coefficients,  $\sigma(q_\ell(v)) = q_\ell(v)$ . Therefore, for every  $j^{\text{th}}$  window (where  $j \in \{1, 2, \dots, \rho\}$ ) we have that  $1 + (j-1)\rho^{d-1} \leq i \leq j\rho^{d-1} - 1$ , we have  $f'_{i+1}(v) = \sigma(f'_i(v))$ .

Now, consider each boundary condition, i.e.  $i = j\rho^{d-1}$ . We need to ensure that  $\sigma(f'_{j\rho^{d-1}}(v)) = f'_{1+j\rho^{d-1}}(v)$ . Equivalently, we need to ensure that

$$\sigma \left( \sum_{\ell=1}^{r^2} f_{\rho^{d-1}}((\omega_d^\ell \alpha_d)^{\mu_{j-1}}) q_\ell(v) \right) = \sum_{\ell=1}^{r^2} f_1((\omega_d^\ell \alpha_d)^{\mu_j}) q_\ell(v).$$

We prove it by induction on  $j$ . Inductively, we can enforce it by requiring that

$$\sigma^{(j-1)\rho^{d-1}} \left( \sum_{\ell=1}^{r^2} f_1(\omega_d^\ell \alpha_d) q_\ell(v) \right) = \sum_{\ell=1}^{r^2} f_1((\omega_d^\ell \alpha_d)^{\mu_j}) q_\ell(v).$$

Since  $\alpha_d$  will be chosen from  $W_d$ , we can write  $\omega_d^\ell \alpha_d = \omega_0^{j_1} \omega_d^{j_2}$  for some  $j_1, j_2$ . Now,  $\sigma^{(j-1)\rho^{d-1}}(f_1(\omega_0^{j_1} \omega_d^{j_2})) = f_1(\sigma^{(j-1)\rho^{d-1}}(\omega_0^{j_1} \omega_d^{j_2}))$  as  $\sigma^{\rho^{d-1}}$  fixes all coefficients of  $f_1$  (because  $f_1(u) \in K_{d-1}[u]$ ). Now,

$$\sigma^{(j-1)\rho^{d-1}}(\omega_0^{j_1} \omega_d^{j_2}) = \omega_0^{j_1} \cdot \omega_d^{j_2 \cdot (\Lambda \rho^\kappa + 1)^{(j-1)\rho^{d-1}}} = \omega_0^{j_1} \omega_d^{j_2(1 + \Lambda \rho^{d-1+\kappa})^{j-1}} = (\omega_0^{j_1} \omega_d^{j_2})^{\mu^{j-1}},$$

since  $\omega_0^{\mu^{j-1}} = \omega_0$ . It verifies that the choice of  $\mu$  in [Lemma 33](#) is  $1 + \Lambda \rho^{d-1+\kappa}$ .

As already discussed, the parameter  $v$  (whose place holder is  $\alpha_{d+1}$  in the description of  $\mathcal{G}_d$ ) should vary over a set of size  $O((\rho^d n m r)^\rho)$ . This way we ensure that  $f_{i+1} = \sigma(f_i)$  for  $1 \leq i \leq \rho^d - 1$ . Now define  $f_{\rho^d+j} = \sigma(f_{\rho^d+j-1})$  for  $1 \leq j \leq \ell_2 - \rho^d$ . The fact that  $\mathcal{G}_d$  is indeed a generator follows from the span preserving property and the proof is identical to the proof given in of [\[FS13\]](#). For our case it uses [Lemma 33](#). To see the final hitting set size, we note that the seed  $(\alpha_1, \dots, \alpha_d, \alpha_{d+1}) \in S_1 \times S_2 \times \dots \times S_{d+1}$ , where  $S_i \subseteq W_i$  and  $|S_i| = \rho^\kappa$ . Each seed  $(\alpha_1, \dots, \alpha_{d+1})$  defines a  $n$ -tuple over  $D_2^n$  in the hitting set. So the size of the hitting set is  $(\rho^d n m r)^{O(d\rho)}$ . After simplification,  $|\widehat{H}_{n,r,\tilde{d}}| \leq (nr\tilde{d})^{O(\rho \log \rho \log \tilde{d})}$ .  $\blacksquare$

### 6.2.3 Improving the dependency on the number of variables

In the hitting set construction of [Theorem 32](#), we ensure that the dimension of the hitting set is independent of the width of the input ABP. Recall that, the number of variables is now the only source of dependency of  $\ell_\theta$  on  $\ell$ . In this subsection, we modify the hitting set construction further that improves the dimension of the hitting set sacrificing in the hitting set size.

**Theorem 34.** *Let  $\rho$  be any prime number. For the class of  $n$ -variate degree  $\tilde{d}$  noncommutative polynomials computed by homogeneous ABPs of width  $r$ , we can construct a hitting set  $\widehat{\mathcal{H}}_{n,r,\tilde{d}} \subseteq D_2^n$  of size  $(nr\tilde{d})^{O(\rho \log \rho \log(\tilde{d} \log n))}$  in  $(nr\tilde{d})^{O(\rho \log \rho \log(\tilde{d} \log n))}$  time. Here  $D_2$  is a cyclic division algebra of index  $\ell_2 = \rho^L$  where  $L \leq O(\rho \log \tilde{d} + \rho \log \log n)$ .*

*Proof.* The proof is exactly same as the proof of [Theorem 32](#) with an additional trick to reduce the number of variables in the ABP. Introduce two new noncommuting variables  $y_0$  and  $y_1$ . Now use the following mapping:

$$\text{For each } 1 \leq i \leq n : \quad x_i \mapsto \prod_{j=1}^{\log n} y_{b_j},$$

where  $b_{\log n} \cdots b_2 b_1$  is the binary representation of  $i$ . This modification will increase the degree (and hence the ABP depth) to  $\tilde{d} \log n$ . The width of the resulting ABP increases to  $nr^2$ . We now apply [Theorem 32](#) on this bivariate degree  $\tilde{d} \log n$  ABP of width  $nr^2$  to obtain the desired bounded on the dimension  $\ell_2$  of the division algebra.  $\blacksquare$

### 6.3 Final hitting set

We now explicitly define the final hitting set where the base field  $F = \mathbb{Q}(\omega_0, z)$ . As before, we can express it as:

$$\begin{aligned} \mathcal{H}_{n,s,\theta} &= \left\{ \tilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^p : \underline{p} \in \mathcal{H}_{n,s,\theta-1} \subseteq D_{\theta-1}^n \right\}, \\ \text{and } \tilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^p &= \left\{ (tq_1 + p_1 \otimes I_{\ell_2}, \dots, tq_n + p_n \otimes I_{\ell_2}) : \underline{q} \in \hat{\mathcal{H}}_{n,2s,2s+1}^{D_1} \text{ and } t \in \Gamma \right\}, \\ \text{and } \hat{\mathcal{H}}_{n,2s,2s+1}^{D_1} &= \left\{ (q_1, \dots, q_n) : q_i = \sum_{j,k} C_{jk} \otimes q_{ijk} : (q_{111}, \dots, q_{n\ell_{\theta-1}\ell_{\theta-1}}) \in \hat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1}s, 2s+1} \right\}, \end{aligned}$$

where we recall that  $\mathcal{H}_{n,s,\theta}$  is the hitting set for  $n$ -variate rational formulas of size  $s$ ,  $\tilde{\mathcal{H}}_{n,2s+1,\ell_{\theta-1}}^p$  is the hitting set, as defined in [Equation \(8\)](#), for  $n$ -variate linear matrices of dimension  $2s+1$  with witness tuple  $\underline{p}$  from an  $\ell_{\theta-1}$  dimensional cyclic division algebra, and  $\hat{\mathcal{H}}_{n,2s,2s+1}^{D_1}$ , as defined in [Equation \(5\)](#), is the hitting set for  $n$ -variate  $D_1$ -ABP of width  $2s$  and degree  $2s+1$ .

Let  $\ell$  be the dimension of  $\hat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1}s, 2s+1}$ . By [Theorem 34](#) we obtain  $\ell = (s \log(n\ell_{\theta-1}))^{O(\rho \log \rho)}$  where  $\rho$  is the prime number used for the construction for the inversion height  $\theta$ . As we can choose  $\rho$  to be the  $(\theta+2)^{\text{th}}$  prime for this stage which is bounded by  $(\theta+2) \log(\theta+2)$ , noting that  $\ell_{\theta} = \ell \cdot \ell_{\theta-1}$ , we have the bound

$$\ell_{\theta} \leq \ell_{\theta-1} (s \log n + s \log \ell_{\theta-1})^{O(\theta \log(\theta))}.$$

We claim  $\ell_{\theta} = (ns)^{O(\theta^3)}$ . The base case holds as  $\ell_0 = (ns)^{O(1)}$ . Now  $\ell_{\theta-1} = (ns)^{c(\theta-1)^3}$  for some sufficiently large constant  $c$ , from the inductive hypothesis. Therefore,

$$\begin{aligned} \ell_{\theta} &\leq \ell_{\theta-1} (s \log n + (\theta-1)^3 s \log(ns))^{O(\theta \log(\theta))} \\ &\leq \ell_{\theta-1} (ns)^{O(\theta^2)} \leq (ns)^{c\theta^3}. \end{aligned}$$

We also have,  $|\mathcal{H}_{n,s,\theta}| = |\mathcal{H}_{n,s,\theta-1}| \cdot |\hat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1}s, 2s+1}| \cdot |\Gamma|$ . Recall that in [Theorem 25](#) we bounded  $\Gamma$  by  $(ns\ell_{\theta-1})^{\rho}$ . Combined with [Theorem 34](#), we obtain

$$\begin{aligned} |\hat{\mathcal{H}}_{\ell_{\theta-1}^2 n, 2\ell_{\theta-1}s, 2s+1}| \cdot |\Gamma| &\leq (ns\ell_{\theta-1})^{O(\theta^2 \log s \log \log(\ell_{\theta-1}n))} \\ &\leq (ns)^{O(\theta^5 \log s \log(c\theta^3 + \log(ns)))}. \end{aligned}$$

Unfolding the recursion, we now obtain,

$$\mathcal{H}_{n,s,\theta} = (ns)^{O(\theta^6 \log s \log(c\theta^3 + \log(ns)))} \leq (ns)^{O(\theta^6 \log^2(ns))}.$$

**Final steps** To complete the description we need to bound the parameter  $\Lambda = 2^\tau$  which is the order of the root of unity  $\omega_0$  in the base field  $F$  (as described in the construction of [Theorem 32](#)). As observed there, it suffices to choose  $\Lambda \geq r^{3p}$  for all the ABP widths  $r$  and primes  $p$  that arise in the recursive construction. For rational formulas of inversion height  $\theta$  we have  $r \leq \ell_\theta \leq (ns)^{O(\theta^3)}$ . As  $p \leq \theta^2$ , it suffices to choose  $\Lambda \geq (ns)^{O(\theta^5)}$ .

Finally, as done in the proof of [Theorem 28](#), we can use the same trick to obtain a hitting set over  $\mathbb{Q}$  itself. In the hitting set points we replace  $\omega_0, \omega$  and  $z$  by commuting indeterminates  $t_1, t_2, t_3$ . Notice that all three are of degree  $\leq (ns)^{O(\theta^5)}$ . Then, for any nonzero rational formula  $r$  of size  $s$  there is a matrix tuple in the hitting set on which  $r$  evaluates to a nonzero matrix  $M(t_1, t_2, t_3)$  of dimension quasipolynomial over the commutative function field  $\mathbb{Q}(t_1, t_2, t_3)$ . Each entry of  $M(t_1, t_2, t_3)$  is a commutative rational expression of the form  $a/b$ , where  $a$  and  $b$  are polynomials in  $t_1, t_2$  and  $t_3$  and the degrees of both  $a$  and  $b$  are quasipolynomial. We can now vary the parameters  $t_1, t_2, t_3$  over a  $(ns)^{\theta^{O(1)}}$  large set  $\tilde{T} \subseteq \mathbb{Q}$  such that we avoid the roots of the numerator and denominator polynomials involved in the computation. Therefore, finally we obtain the hitting set  $\mathcal{H}_{n,s,\theta} \subseteq \text{Mat}_{\ell_\theta}(\mathbb{Q})$  where

$$\ell_\theta \leq (ns)^{\theta^{O(1)}} \quad \text{and,} \quad |\mathcal{H}_{n,s,\theta}| \leq (ns)^{\theta^{O(1)} \log^2(ns)}.$$

It completes the proof of [Theorem 2](#).

## 7 Conclusion

In this paper, we settle the black-box complexity of the RIT problem. However, designing a black-box algorithm for the NSINGULAR problem remains wide open. The connection of this problem to the parallel algorithm for bipartite matching [[FGT21](#)] is already discussed in [Section 1](#).

We believe that the techniques introduced in this paper might be useful in designing efficient hitting sets for the NSINGULAR problem. The construction of hitting sets inside cyclic division algebras is also reasonable to hope for, because by the result of Derksen and Makam [[DM17](#)] it suffices to evaluate linear matrices of size  $s$  on random  $s \times s$  matrix to test whether or not it is invertible over the free skew field. [Proposition 11](#) ensures that random elements from a cyclic division algebra of the same dimension suffice for the black-box case. Therefore, the existence of even a polynomial-size hitting set inside a cyclic division algebra of polynomial dimension is guaranteed by a standard counting argument.

Note that, the result of Derksen and Makam [[DM17](#)] also implies that for every nonzero rational formula of size  $s$ , there is a  $2s \times 2s$  matrix tuple such that the evaluation is nonzero. However, the dimension of the hitting set point obtained in [Theorem 2](#) is quasipolynomial. An interesting open problem is to construct a hitting set where the dimension is polynomial in the size of the formula.

## References

- [ACG<sup>+</sup>22] Vikraman Arvind, Abhranil Chatterjee, Utsab Ghosal, Partha Mukhopadhyay, and C. Ramya. On identity testing and noncommutative rank computation over the free skew field. *CoRR*, abs/2209.04797, 2022.
- [ACG<sup>+</sup>23] Vikraman Arvind, Abhranil Chatterjee, Utsab Ghosal, Partha Mukhopadhyay, and C. Ramya. On identity testing and noncommutative rank computation over the free skew field. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA*, volume 251 of *LIPICs*, pages 6:1–6:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.
- [ACM22] Vikraman Arvind, Abhranil Chatterjee, and Partha Mukhopadhyay. Black-box identity testing of noncommutative rational formulas of inversion height two in deterministic quasipolynomial time. In Amit Chakrabarti and Chaitanya Swamy, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference)*, volume 245 of *LIPICs*, pages 23:1–23:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
- [AL50] A. S. Amitsur and J. Levitzki. Minimal identities for algebras. *Proceedings of the American Mathematical Society*, 1(4):449–463, 1950.
- [Ami66] S.A Amitsur. Rational identities and applications to algebra and geometry. *Journal of Algebra*, 3(3):304 – 359, 1966.
- [BBJP19] Vishwas Bhargav, Markus Bläser, Gorav Jindal, and Anurag Pandey. *A Deterministic PTAS for the Algebraic Rank of Bounded Degree Polynomials*, pages 647–661. 01 2019.
- [Ber76] George M Bergman. Rational relations and rational identities in division rings. *Journal of Algebra*, 43(1):252–266, 1976.
- [BFG<sup>+</sup>19] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 845–861. IEEE Computer Society, 2019.
- [Coh71] P. M. Cohn. The Embedding of Firs in Skew Fields. *Proceedings of the London Mathematical Society*, s3-23(2):193–213, 10 1971.
- [Coh95] P. M. Cohn. Skew fields: Theory of general division rings. 1995.
- [DK21] Manfred Droste and Dietrich Kuske. Weighted automata. In Jean-Éric Pin, editor, *Handbook of Automata Theory*, pages 113–150. European Mathematical Society Publishing House, Zürich, Switzerland, 2021.

- [DL78] Richard A. Demillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978.
- [DM17] Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017.
- [DM20] Harm Derksen and Visu Makam. Algorithms for orbit closure separation for invariants and semi-invariants of matrices. *Algebra & Number Theory*, 14(10):2791–2813, 2020.
- [Eil74] Samuel Eilenberg. *Automata, Languages, and Machines (Vol A)*. Pure and Applied Mathematics. Academic Press, 1974.
- [FGT21] Stephen A. Fenner, Rohit Gurjar, and Thomas Thierauf. Bipartite perfect matching is in quasi-NC. *SIAM J. Comput.*, 50(3), 2021.
- [FR04] MARC FORTIN and Christophe Reutenauer. Commutative/noncommutative rank of linear matrices and subspaces of matrices of low rank. *Séminaire Lotharingien de Combinatoire [electronic only]*, 52, 01 2004.
- [FS13] Michael A. Forbes and Amir Shpilka. Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 243–252, 2013.
- [GGdOW16] Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 109–117. IEEE Computer Society, 2016.
- [GGdOW20] Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Operator scaling: Theory and applications. *Found. Comput. Math.*, 20(2):223–290, 2020.
- [GR08] Ariel Gabizon and Ran Raz. Deterministic extractors for affine sources over large fields. *Comb.*, 28(4):415–440, 2008.
- [HH21] Masaki Hamada and Hiroshi Hirai. Computing the nc-rank via discrete convex optimization on CAT(0) spaces. *SIAM J. Appl. Algebra Geom.*, 5(3):455–478, 2021.
- [Hua49] Loo-Keng Hua. Some properties of a sfield. *Proceedings of the National Academy of Sciences of the United States of America*, 35(9):533–537, 1949.
- [HW15] Pavel Hrubeš and Avi Wigderson. Non-commutative arithmetic circuits with division. *Theory of Computing*, 11(14):357–393, 2015.
- [IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *Computational Complexity*, 27(4):561–593, Dec 2018.



- [KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complex.*, 13(1-2):1–46, 2004.
- [Lam01] T.Y. Lam. *A First Course in Noncommutative Rings (Second Edition)*. Graduate Texts in Mathematics. Springer, 2001.
- [MW19] Visu Makam and Avi Wigderson. Singular tuples of matrices is not a null cone (and, the symmetries of algebraic varieties). *Electron. Colloquium Comput. Complex.*, TR19-114, 2019.
- [Pie82] Richard S. Pierce. *Associative Algebras*. Springer-Verlag, 1982.
- [Row80] Louis Halle Rowen. *Polynomial identities in ring theory*. Pure and Applied Mathematics. Academic Press, 1980.
- [Sch80] Jacob T. Schwartz. Fast probabilistic algorithm for verification of polynomial identities. *J. ACM.*, 27(4):701–717, 1980.
- [Str73] Volker Strassen. Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202, 1973.
- [Vol18] Jurij Volčič. Matrix coefficient realization theory of noncommutative rational functions. *Journal of Algebra*, 499:397–437, 04 2018.
- [Zip79] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proc. of the Int. Sym. on Symbolic and Algebraic Computation*, pages 216–226, 1979.