Electronic Colloquium on Computational Complexity, Report No. 67 (2022)

Black-box Identity Testing of Noncommutative Rational Formulas of Inversion Height Two in Deterministic Quasipolynomial-time

V. Arvind* Abhranil Chatterjee† Partha Mukhopadhyay‡

Abstract

Hrubeš and Wigderson [HW15] initiated the complexity-theoretic study of noncommutative formulas with inverse gates. They introduced the Rational Identity Testing (RIT) problem which is to decide whether a noncommutative rational formula computes zero in the free skew field. In the white-box setting, there are deterministic polynomial-time algorithms due to Garg, Gurvits, Oliveira, and Wigderson [GGdOW16] and Ivanyos, Qiao, and Subrahmanyam [IQS18].

A central open problem in this area is to design an efficient deterministic *black-box* identity testing algorithm for rational formulas. In this paper, we solve this for the first nested inverse case. More precisely, we obtain a deterministic quasipolynomial-time black-box RIT algorithm for noncommutative rational formulas of inversion height two via a hitting set construction. Several new technical ideas are involved in the hitting set construction, including concepts from matrix coefficient realization theory [Vol18] and properties of cyclic division algebras [Lam01]. En route to the proof, an important step is to embed the hitting set of Forbes and Shpilka for noncommutative formulas [FS13] inside a cyclic division algebra of small index.

*Institute of Mathematical Sciences (HBNI), Chennai, India, email: arvind@imsc.res.in
†Indian Institute of Technology Bombay, India, email: abhneil@gmail.com
‡Chennai Mathematical Institute, Chennai, India,email: partham@cmi.ac.in

ISSN 1433-8092

1 Introduction

The broad goal of *algebraic complexity* is to study the complexity of computing polynomials and rational functions using basic arithmetic operations: additions, multiplications, and inverses. *Arithmetic circuits* and *arithmetic formulas* are two extensively studied models of computation. An important sub-area of algebraic complexity is *noncommutative computation*: the set of monomials over variables X is the free monoid X^* of all words. In particular, variables in X do not commute (i.e. $xy \neq yx$). If we allow only the addition and multiplication gates in the noncommutative formulas/circuits, they compute noncommutative polynomials (similar to the commutative case) in the free algebra.

In the commutative case, the role of inverses is well understood, but in the noncommutative world 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 expressions (formulas with inverses) such as $x^{-1} + y^{-1}$ or $xy^{-1}x$ cannot be represented as fg^{-1} or $f^{-1}g$. If we have *nested inverses*, it makes the rational expression more complicated, for example $(z + xy^{-1}x)^{-1} - z^{-1}$. Moreover, a noncommutative rational expression is not always defined on a matrix substitution. For a noncommutative rational expression $\mathfrak{r}$, its *domain of definition* is the set of matrix tuples (of any dimension) where $\mathfrak{r}$ is defined. We denote it by $\mathrm{dom}(\mathfrak{r})$. Two rational expressions $\mathfrak{r}_1$ and $\mathfrak{r}_2$ are *equivalent* if they agree on $\mathrm{dom}(\mathfrak{r}_1) \cap \mathrm{dom}(\mathfrak{r}_2)$. This induces an equivalence relation on the set of all noncommutative rational expressions (with nonempty domain of definition). It was used by Amitsur in his characterization of the *universal* free skew field [Ami66] and the equivalence classes are called *noncommutative rational functions*.

The *inversion height* of a rational formula is the maximum number of inverse gates in a path from an input gate to the output gate. It is known [HW15] that the inversion height of a rational formula of size s is bounded by $O(\log s)$. Hrubeš and Wigderson [HW15] consider the *rational identity testing* problem (RIT) of testing the equivalence of two rational formulas. It is the same as testing whether a rational formula computes the zero function in the free skew field. In other words, decide whether there exists a matrix tuple (of any dimension) such that the rational formula evaluates to nonzero on that substitution. Rational expressions exhibit peculiar properties which seem to make the RIT problem quite different from polynomial identity testing. For example, Bergman has constructed an explicit rational expression, of inversion height two, which is an identity for 3×3 matrices but not an identity for 2×2 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. For example, the rational expression $(x + xy^{-1}x)^{-1} + (x + y)^{-1} - x^{-1}$ of inversion height two is a rational identity, known as Hua's identity [Hua49].

However, Hrubeš and Wigderson give an efficient reduction from the RIT problem to the singularity testing problem of linear pencils. A *linear pencil* L of size s over noncommuting variables $\underline{x} = \{x_1, \ldots, x_n\}$ is a $s \times s$ matrix whose entries are linear forms in $\underline{x}$ variables, i.e. $L = A_0 + \sum_{i=1}^{n} A_i x_i$, where each A_i is an $s \times s$ matrix over the field $\mathbb{F}$. A rational function $\mathfrak{r}$ in $\mathbb{F}\langle\langle \underline{x} \rangle\rangle$ has a *linear pencil representation* L of size s, if for some $i, j \in [s]$, $\mathfrak{r} = (L^{-1})_{i,j}$. In particular, if $\mathfrak{r}$ is a rational formula of size s, Hrubeš and Wigderson have shown that $\mathfrak{r}$ has a linear pencil representation L of size at most $2s$ such that r is defined on a matrix tuple if and only if L is invertible on that tuple [HW15]. Using this connection, they reduce the RIT problem to the problem of testing whether a given linear pencil is invertible over the free skew field in deterministic poly-

nomial time. The latter is the noncommutative SINGULAR problem, whose commutative analog is the symbolic determinant identity testing problem. The deterministic complexity of symbolic determinant identity testing is completely open [KI04] in the commutative setting. In contrast, the SINGULAR problem in noncommutative setting has deterministic polynomial-time algorithms in the white-box model due to [GGdOW16, IQS18]. The algorithm in [GGdOW16] is based on operator scaling and the algorithm in [IQS18] is based on the second Wong sequence and a constructive version of *regularity lemma*. As a consequence, a deterministic polynomial-time white-box RIT algorithm follows.

A central open problem is to design an efficient *deterministic* RIT algorithm in the black-box case [GGdOW16]. There is a randomized polynomial-time black-box algorithm for the problem [DM17]. Can we derandomize this result even in some restricted settings, for example when the inversion height of the input rational formula is small? Notice that inversion height zero rational formulas are just noncommutative formulas, and a result of Forbes and Shpilka have shown a deterministic quasipolynomial-time identity testing for those (more generally, for noncommutative ABPs) via a hitting set construction [FS13]. Whether their approach can be extended to the RIT problem for rational formulas is a natural direction and we prove the following theorem which is our main result.

Theorem 1. *For the class of rational formulas in $\mathbb{Q}\langle\langle x_1, \ldots, x_n\rangle\rangle$ of inversion height two and size at most s, we can construct a hitting set $\mathcal{H} \subseteq \mathbb{M}_d^n(\mathbb{Q})$ of size $(ns)^{O(\log ns)}$ in deterministic $(ns)^{O(\log ns)}$-time. The parameter d is* $\mathrm{poly}(s, n)$ *bounded.*

Before this work, no such hitting set construction was known that could handle nested inverses. As we discuss in the next section, even to derandomize RIT for the special case of inversion height two, we need to accumulate several ideas involving cyclic division algebras [Lam01] and matrix coefficient realization theory [Vol18] combined with the hitting set construction in [FS13].

Proof Idea

Consider the following noncommutative rational formula, $\mathfrak{r} = [x, y]^{-1} = (xy - yx)^{-1}$. Clearly there is no point in $\mathrm{dom}(\mathfrak{r})$ from the ground field, and the natural idea is to expand the series around a matrix point. Let (p_1, p_2) be a matrix pair such that $[p_1, p_2]$ is invertible and let $\mathfrak{r}(p_1, p_2) = [p_1, p_2]^{-1} = q$. Then,

$$\mathfrak{r}(x + p_1, y + p_2) = ([p_1, p_2] - [p_2, x] - [y, p_1] - [y, x])^{-1}.$$

Simplifying this we can write $\mathfrak{r}(x+p_1, y+p_2) = (I - g(x, y))^{-1}q$ where $g(x, y) = q([p_2, x] + [y, p_1] + [y, x])$. Now expanding this using $(I - g(x, y))^{-1} = \sum_{i \geqslant 0}(g(x, y))^i$, we can see that every term in the expansion looks like $a_0 z_1 a_1 z_2 \ldots a_{d-1} z_d a_d$ where each a_j is a matrix and $z_j \in \{x, y\}$. In the language of matrix coefficient realization theory [Vol18], such terms (resp. series) are called generalized words or monomials (resp. generalized series). In fact if a rational formula $\mathfrak{r}$ of size s has a defined point $\underline{u}$ in some dimension l (in other words $\underline{u} \in \mathrm{dom}(\mathfrak{r})$, and we use it interchangeably), Volčič shows that one can associate a special class of generalized series, a recognizable generalized series to the shifted rational formula [Vol18]:

$$\mathfrak{r}(\underline{x} + \underline{u}) = \boldsymbol{c}\left(I_{2ls} - \sum_{j=1}^{n} A^{x_j}\right)^{-1}\boldsymbol{b}.$$

Here $\boldsymbol{c} \in (\mathbb{M}_l(\mathbb{F}))^{1\times 2s}$ and $\boldsymbol{b} \in (\mathbb{M}_l(\mathbb{F}))^{2s\times 1}$. The matrices $A^{x_j}, 1 \leqslant j \leqslant n$ are of dimension $2s \times 2s$ as a block matrix and $(k_1, k_2)^{th}$ entry of A^{x_j} is given by a generalized linear form $C_{k_1,k_2,j}x_jC'_{k_1,k_2,j}$ where $C_{k_1,k_2,j}, C'_{k_1,k_2,j} \in \mathbb{M}_l(\mathbb{F})$.

Focusing on our problem for rational formulas of inversion height two, the first step is to construct a quasipolynomial-size set $\mathcal{H}_1$ of matrix tuples of small dimension such that for every nonzero rational formula $\mathfrak{r}$ of inversion height two, there exists a point $\underline{u} \in \mathcal{H}_1$ on which $\mathfrak{r}$ is defined. Given such a point, testing whether $\mathfrak{r}$ is zero or not reduces to testing whether the generalized series $\mathfrak{r}(\underline{x} + \underline{u})$ is zero or not. This is formally stated in Theorem 10. For a recognizable series in algebraic automata theory, a standard result by Schützenberger shows that the identity testing of such infinite series is equivalent to the identity testing of polynomial obtained by truncation of the series up to a small degree [Eil74, Corollary 8.3]. We can adapt this result in the case of generalized series too and observe that the truncated generalized polynomial (of small degree d) can be represented by an algebraic branching program with edge labels are linear forms over matrices. Such ABPs can be identity tested efficiently using an adaptation of the hitting set construction shown by Forbes-Shpilka [FS13].

Although it is not clear how to carry out the truncation in the black-box setting, we can show that a suitable scaling of the hitting set for such generalized ABPs is good enough to hit the generalized series too. To fit the dimension correctly, throughout the computation the coefficient matrices should be embedded in the matrix algebra of dimension dl using the inclusion map $\iota : a \to a \otimes I_d$. This is shown in Proposition 26.

Clearly, $\mathfrak{r}$ is defined at a point $\underline{u}$ if and only if all the maximal sub-formulas of inversion height one in $\mathfrak{r}$ evaluate to invertible matrices on $\underline{u}$. One can consider the product of all such maximal formulas and thus our goal is now re-defined: construct $\mathcal{H}_1$ such that for every size-s rational formula $\mathfrak{r}$ of inversion height one, there is a point $\underline{u} \in \mathcal{H}_1$ at which $\mathfrak{r}(\underline{u})$ is *invertible*. We call such a hitting set a *strong hitting set*. We give the formal definition.

Definition 2 (Strong hitting set)**.** For a class of rational functions (resp. polynomials) a hitting set $\mathcal{H}$ is strong if any nonzero rational function (resp. polynomial) in that class evaluates to an invertible matrix at some point in $\mathcal{H}$.

A rational formula $\mathfrak{r}$ of inversion height one is defined at a point $\underline{v}$ if and only if all sub-formulas which are input to inverse gates evaluate to invertible matrices on $\underline{v}$. These sub-formulas are just noncommutative formulas. Since the Forbes-Shpilka hitting set [FS13] for noncommutative formulas consists of tuples of nilpotent matrices, it is not directly applicable to our problem.

However, it is possible to adapt their construction and get a strong hitting set, also of quasipolynomial size, such that every size-s nonzero noncommutative formula evaluates to an invertible matrix on some matrix tuple in the strong hitting set [1]. In particular, all matrices in the hitting set construction will be invertible and have the following shape:

$$\begin{bmatrix} 0 & * & 0 & \cdots & 0 \\ 0 & 0 & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & * \\ * & 0 & \cdots & 0 & 0 \end{bmatrix},$$

[1] This was first explicitly constructed in [ACDM20].

where the dimension is determined by the depth of the noncommutative formulas. Expanding $\mathfrak{r}$ around such a point would again lead to a generalized series, and (a somewhat more involved) truncation and scaling argument show that we can get a strong hitting set for $\mathfrak{r}$ by constructing a strong hitting set for *generalized* ABPs whose edges are labeled by linear forms over matrices. This is the essence of the second part of Proposition 26.

At this point, we face a serious obstacle. How do we find invertible matrices in the image of the generalized ABPs? In other words, how to construct a strong hitting set for generalized ABPs? The main insight is that, if the matrices present in the linear forms of the generalized ABPs are from a division algebra, then one can construct a strong hitting set from a hitting set. To implement this, we construct the hitting set for noncommutative formulas (which are of inversion height zero) over a division algebra of small index and expand the rational formula with respect to the points in that hitting set. Why does it work? Roughly speaking, as already mentioned it is easier to find a nonzero in the image of generalized ABPs and if the computation occurs in a division algebra then a computed nonzero element is also invertible.

Section 4 elaborates on this idea. In particular, Lemma 20 provides an existential argument showing that if the linear forms of the generalized ABP are defined over a division algebra of dimension ℓ, then there *exists* a substitution to the variables from D such that the generalized ABP evaluates to an invertible matrix. The proof uses two ideas. Firstly, we show that such a point exists in the full matrix algebra of dimension ℓ. Then we use Proposition 15 to find such a certificate in D. Once we establish the existential argument, we can use a reduction to the hitting set construction of ROABPs (in unknown order) [AGKS15] to construct the hitting set in quasi-polynomial time. To work out the technical details we need to employ the inclusion map $\iota' : a \to I_d \otimes a$ for the coefficients which are now elements of division algebra. In ring theory the maps ι and ι' are compatible: by the Skolem-Noether theorem [Row80, Theorem 3.1.2] there is an invertible matrix q_0 such that $q_0(I_d \otimes a)q_0^{-1} = a \otimes I_d$ for all a. However, in our case, we give a simple explicit construction of a permutation matrix q_0.

In the remaining part of the proof sketch, we informally describe how to find a hitting set for noncommutative formulas (more generally for noncommutative ABPs) in a division algebra of a small index. For simplicity, suppose the ABP degree is 2^d. The Forbes-Shpilka hitting set [FS13] has a recursive construction and it is by a reduction to the hitting set construction for ROABPs (read-once algebraic branching programs) over the commutative variables $u_1, u_2, \ldots, u_{2^d}$. The recursive step in the 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^{th} step. More precisely, $\mathcal{G}_d$ is a map from $\mathbb{F}^{d+1} \to \mathbb{F}^{2^d}$ that stretches the seed $(\alpha_1, \ldots, \alpha_{d+1})$ to a 2^d tuple for the read-once variables.

For our purpose, we take a classical construction of cyclic division algebras [Lam01, Chapter 5]. The division algebra $D = (K/F, \sigma, z)$ is defined using a indeterminate x as the ℓ-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 $b \in K$. Here $\sigma : K \to K$ is an automorphism of the Galois group $\mathrm{Gal}(K/F)$. The field $F = \mathbb{Q}(z)$ and $K = F(\omega)$, where z is an indeterminate and ω is an ℓ^{th} primitive root of unity. The matrix

representation of a general element in D is of the following form:

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

To embed the hitting set of [FS13], we need to choose $\ell = 2^L$ appropriately larger than 2^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 $\omega_i = \omega^{2^{a_i}}$ for $a_1 > a_2 > \cdots > a_d > 0$, where a_i are positive integers suitably chosen. Let $K_i = F(\omega_i)$ be the cyclic Galois extension for $1 \leqslant i \leqslant d$ giving a tower of extension fields

$$F \subset F(\omega_1) \subset F(\omega_2) \subset \cdots \subset F(\omega_d) \subset F(\omega).$$

As we show in Section 3 that we require two properties of $\omega_i, 1 \leqslant i \leqslant d$. Firstly, for the hitting set generator $\mathcal{G}_i$ we will choose the root of unity as ω_i and the variable α_i will take values only in the set $W_i = \{\omega_i^j \mid 1 \leqslant j \leqslant 2^{L-a_i}\}$. We also require that the K-automorphism σ has the property that for all $1 \leqslant i \leqslant d$ the map σ^{2^i} fixes ω_i. In fact we will ensure that σ^{2^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.

Implementing all these steps we get a quasipolynomial-size hitting set over $\mathbb{Q}(\omega, z)$. Then we show how to transfer the hitting set over $\mathbb{Q}$ itself by a relatively standard idea that treats the parameters ω and z as *fresh indeterminates* t_1, t_2 and vary them over a suitably chosen polynomial-size set. This is sketched in Section 5.

We include a brief discussion in Section 6 about possibly extending our approach to any constant inversion height formula.

Organization

In Section 2, we collect some background results from algebraic complexity theory, matrix coefficient realization theory, and cyclic division algebra. Section 3 contains the proof that the Forbes-Shpilka hitting set can be embedded in a cyclic division algebra of small index. In Section 4, we construct a quasipolynomial-size strong hitting set for generalized ABPs over division algebra. Finally, in Section 5 we combine the results developed in Section 3 and Section 4 to obtain our main result which gives a quasipolynomial-size hitting set for rational formulas of inversion height two. In Section 6, we mainly discuss the possibility of extending our method to higher inversion heights.

2 Background and Notation

Throughout the paper, we use $\mathbb{F}, F, K$ for fields. The notation $\mathbb{M}_m(\mathbb{F})$ (respectively, $\mathbb{M}_m(F)$, $\mathbb{M}_m(K)$) are used for m dimensional matrix algebra over $\mathbb{F}$ (respectively over F, K) where m is clear from the context. D is used to denote cyclic division algebras. Let $\underline{x}$ be the set of variables $\{x_1, \ldots, x_n\}$. Sometime we use notation like $\underline{u}, \underline{v}, \underline{p}, \underline{q}$ to denote the matrix tuples in suitable matrix algebras. The free noncommutative ring of polynomials over a field $\mathbb{F}$ is denoted by $\mathbb{F}\langle \underline{x} \rangle$. The

ring of *formal power series* is denoted by $\mathbb{F}\langle\langle \underline{x} \rangle\rangle$. For a series (or polynomial) S, the coefficient of a monomial (word) in S is denoted by $[m]S$.

2.1 Algebraic Complexity

Definition 3 (Algebraic Branching Program)**.** An *algebraic branching program* (ABP) is a layered directed acyclic graph. The vertex set is partitioned into layers $0, 1, \ldots, d$, 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. 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. The ABP model is defined for computing commutative or noncommutative polynomials. ABPs of width r can also be seen as iterated matrix multiplication $\boldsymbol{c} \cdot M_1 M_2 \cdots M_\ell \cdot \boldsymbol{b}$, where $\boldsymbol{c}, \boldsymbol{b}$ are $1 \times r$ and $r \times 1$ vectors respectively and each M_i is a $r \times r$ matrix, whose entries are affine linear forms over $\underline{x}$.

We also consider commutative set-multilinear ABPs and read-once oblivious ABPs (ROABPs). For the set-multilinear case, the (commutative) variable set is partitioned as $Y = Y_1 \sqcup Y_2 \sqcup \cdots \sqcup Y_d$ where for each $j \in [d]$, $Y_j = \{y_{ij}\}_{i=1}^n$. An ABP B is homogeneous set-multilinear if each edge in the j^{th} layer of the ABP is labelled by linear forms over Y_j. For ROABP, a different variable is used for each layer, and the edge labels are univariate polynomials. Therefore, an ROABP of d layers can be represented as $\boldsymbol{c} \cdot M_1(v_1) M_2(v_2) \cdots M_{v_d}(d) \cdot \boldsymbol{b}$. We say that the ROABP respects the variable order $v_1 < v_2 < \cdots < v_d$.

Identity testing results

We say a set $\mathcal{H} \subseteq \mathbb{F}^n$ is a hitting set for a circuit class $\mathcal{C}$ if for every n-variate polynomial f in $\mathcal{C}$, $f \not\equiv 0$ if and only if $f(\underline{a}) \neq 0$ for some $\underline{a} \in \mathcal{H}$. For the class of ROABPs, Forbes and Shpilka [FS13] obtained the first quasipolynomial-time black-box algorithm by constructing a hitting set of the same size.

Theorem 4. *For the class of polynomials computable by a width r, depth D, individual degree $< n$ ROABPs of knwon order, if $|\mathbb{F}| \geqslant (2Dnr^3)^2$, there is a poly$(D, n, r)$-explicit hitting set of size at most $(2Dn^2r^4)^{\lceil \log D+1 \rceil}$.*

Indeed, they proved something more general.

Definition 5 (Hitting Set Generator)**.** A polynomial map $\mathcal{G} : \mathbb{F}^t \to \mathbb{F}^n$ 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 6. [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 $\mathcal{G} : \mathbb{F}^{\lceil \log D+1 \rceil} \to \mathbb{F}^D$ of degree Dnr^4 efficiently.*

The hitting set is defined as $\mathcal{H} \subseteq \mathbb{M}_d^n(\mathbb{F})$ for any class of noncommutative polynomials. For the black-box case, Forbes and Shpilka [FS13], have shown an efficient construction of quasipolynomial-size hitting set for noncommutative ABPs. 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 $\mathcal{H}_{w,d,n}$ contained in $\mathbb{M}_{d+1}(\mathbb{F})$, such that for any ABP (with parameters w and d) computing a nonzero polynomial f, there always exists $(p_1,\ldots,p_n) \in \mathcal{H}_{w,d,n}$ such that $f(\underline{p}) \neq 0$.

Theorem 7 (Forbes and Shpilka [FS13])**.** *For all $w,d,n \in \mathbb{N}$, if $|\mathbb{F}| \geqslant \mathrm{poly}(d,n,w)$, then there is a hitting set $\mathcal{H}_{w,d,n} \subset \mathbb{M}_{d+1}(\mathbb{F})$ for noncommutative ABPs of parameters w,d,n such that $|\mathcal{H}_{w,d,n}| \leqslant (wdn)^{O(\log d)}$ and there is a deterministic algorithm to output the set $\mathcal{H}_{w,d,n}$ in time $(wdn)^{O(\log d)}$.*

Recognizable series

A comprehensive treatment is in the book by Berstel and Reutenauer [BR11]. We will require the following concepts. Recall that $\mathbb{F}\langle\langle \underline{x}\rangle\rangle$ is the formal power series ring over a field $\mathbb{F}$. A series S in $\mathbb{F}\langle\langle \underline{x}\rangle\rangle$ is *recognizable* if it has the following linear representation: for some integer s, there exists a row vector $\underline{c} \in \mathbb{F}^{1\times s}$, a column vector $\underline{b} \in \mathbb{F}^{s\times 1}$ and an $s \times s$ matrix M whose entries are homogeneous linear forms over $x_1,\ldots,x_n$ i.e. $\sum_{i=1}^{n} \alpha_i x_i$ such that $S = \underline{c}\left(\sum_{k\geqslant 0} M^k\right)\underline{b}$. Equivalently, $S = \underline{c}(I-M)^{-1}\underline{b}$. We say, S has a representation $(\underline{c}, M, \underline{b})$ of size s.

The following theorem is a basic result in algebraic automata theory.

Theorem 8. *A recognizable series with representation $(\underline{c}, M, \underline{b})$ of size s is nonzero if and only if $\underline{c}\left(\sum_{k\leqslant s-1} M^k\right)\underline{b}$ is nonzero.*

It has a simple linear algebraic proof [Eil74, Corollary 8.3, Page 145]. This result is generally attributed to Schützenberger. For this paper, the theorem is used to apply that the truncated series is computable by a small noncommutative ABP, therefore, reducing zero-testing of recognizable series to the identity testing of noncommutative ABPs.

2.2 Matrix Coefficient Realization Theory

We require some basic notions and results about generalized automata and generalized recognizable series from Volčič's article [Vol18]. A detailed exposition is given in it [Vol18].

A *generalized word* or a *generalized monomial* in $x_1,\ldots,x_n$ over the matrix algebra $\mathbb{M}_m(\mathbb{F})$ 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 $a_i \in \mathbb{M}_m(\mathbb{F})$, 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 $\mathbb{M}_m(\mathbb{F})\langle\underline{x}\rangle$. A *generalized formal power series* over $\mathbb{M}_m(\mathbb{F})$ 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 $\mathbb{M}_m(\mathbb{F})$ is denoted $\mathbb{M}_m(\mathbb{F})\langle\langle\underline{x}\rangle\rangle$.

A generalized series (resp. polynomial) S over $\mathbb{M}_m(\mathbb{F})$ admits the following canonical description. Let $E = \{e_{i,j}, 1 \leqslant i,j \leqslant m\}$ 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 $\mathbb{M}_m(\mathbb{F})$ is defined on any $k'm \times k'm$ matrix algebra for some integer $k' \geqslant 1$ [Vol18]. To match the dimension of the coefficient matrices with the matrix

substitution, we use an inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{k'm}(\mathbb{F})$, for example, ι 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 $\mathbb{M}_m(\mathbb{F})$ on matrix substitution $(p_1, \ldots, p_n) \in \mathbb{M}^n_{k'm}(\mathbb{F})$ 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 $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{k'm}(\mathbb{F})$. 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 $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{k'm}(\mathbb{F})$, then it is zero w.r.t. any such inclusion map.

The two notions of zeroness are equivalent [Vol18, Proposition 3.13].

Definition 9. [Vol18] A generalized series S in $\mathbb{M}_m(\mathbb{F})\langle\langle \underline{x} \rangle\rangle$ is said to be *recognizable* if it has the following linear representation. For some integer s, there exists a row-tuple of matrices $\boldsymbol{c} \in (\mathbb{M}_m(\mathbb{F}))^{1\times s}$, and $\boldsymbol{b} \in (\mathbb{M}_m(\mathbb{F}))^{s\times 1}$ and an $s \times s$ matrix M whose entries are homogeneous generalized linear forms over $x_1, \ldots, x_n$ i.e. $\sum_{i=1}^{n} \widetilde{p_i} x_i \widehat{p_i}$ where each $\widetilde{p_i}, \widehat{p_i} \in \mathbb{M}_m(\mathbb{F})$ such that $S = \boldsymbol{c}(I - M)^{-1}\boldsymbol{b}$. We say, S has a linear representation $(\boldsymbol{c}, M, \boldsymbol{b})$ of size s over $\mathbb{M}_m(\mathbb{F})$.

The linear representation is said to be over a subalgebra $\mathfrak{A} \subseteq \mathbb{M}_m(\mathbb{F})$ if $\boldsymbol{c} \in \mathfrak{A}^{1\times s}$, and $\boldsymbol{b} \in \mathfrak{A}^{s\times 1}$ and each $\widetilde{p_i}, \widehat{p_i} \in \mathfrak{A}$.

Theorem 10. [Vol18, Corollary 5.1, Proposition 3.13]

1. *Given a noncommutative rational formula $\mathfrak{r}$ of size s over $x_1, \ldots, x_n$ and a matrix tuple $\underline{p} \in \mathbb{M}^n_m(\mathbb{F})$ in the domain of definition of $\mathfrak{r}$, $\mathfrak{r}(\underline{x} + \underline{p})$ is a recognizable generalized series with a representation of size at most $2s$ over $\mathbb{M}_m(\mathbb{F})$. Moreover, if $\mathfrak{A} \subseteq \mathbb{M}_m(\mathbb{F})$ is the subalgebra generated by the matrices $p_1, \ldots, p_n$ then $\mathfrak{r}(\underline{x} + \underline{p})$ has, in fact, a linear representation over the subalgebra $\mathfrak{A}$.*

2. *Additionally, $\mathfrak{r}(\underline{x})$ is zero in the free skew field if and only if $\mathfrak{r}(\underline{x} + \underline{p})$ is zero as a generalized series.*

Proof. For the first part, see Corollary 5.1 and Remark 5.2 of [Vol18].

To see the second part, suppose $\mathfrak{r}(\underline{x})$ is zero in the free skew field. Then the fact that $\mathfrak{r}(\underline{x} + \underline{p})$ is a zero series follows from [Vol18, Proposition 3.13]. If $\mathfrak{r}(\underline{x})$ is nonzero in the free skew field, then there exists a matrix tuple $(q_1, \ldots, q_n) \in \mathbb{M}^n_l(\mathbb{F})$ such that $\mathfrak{r}(\underline{q})$ is nonzero. W.l.o.g. we can assume $l = k'm$ for some integer k'. Fix an inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{k'm}(\mathbb{F})$. Define a matrix tuple $(q'_1, \ldots, q'_n) \in \mathbb{M}^n_{k'm}(\mathbb{F})$ such that $q'_i = q_i - \iota(p_i)$. Therefore, the series $\mathfrak{r}(\underline{x} + \underline{p})$ on $(q'_1, \ldots, q'_n)$ evaluates to $\mathfrak{r}(\underline{q})$, under the inclusion map ι, which is nonzero [Vol18, Remark 5.2]. Therefore, $\mathfrak{r}(\underline{x} + \underline{p})$ is nonzero. □

Remark 11. Moreover we have the following [Vol18, Section 5]. Let $\mathfrak{r}(\underline{x})$ be a rational formula of size s and $\underline{p} \in \mathbb{M}^n_m(\mathbb{F})$ be in the domain of definition of $\mathfrak{r}$. Then $\mathfrak{r}(\underline{x} + \underline{p})$ has a linear representation $(\boldsymbol{c}, M, \boldsymbol{b})$ of size $2s$ over $\mathbb{M}_m(\mathbb{F})$. Then M is a $2s \times 2s$ matrix with entries of the form $\sum_{i=1}^{n} \widetilde{p_i} x_i \widehat{p_i}$, $\widetilde{p_i}, \widehat{p_i} \in \mathbb{M}_m(\mathbb{F})$. For an inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{k'm}(\mathbb{F})$ and a matrix tuple $\underline{q} \in \mathbb{M}^n_{k'm}(\mathbb{F})$, replacing each $\sum_{i=1}^{n} \widetilde{p_i} x_i \widehat{p_i}$ by $\sum_{i=1}^{n} \iota(\widetilde{p_i}) q_i \iota(\widehat{p_i})$, we obtain a $2sk'm \times 2sk'm$ matrix $\iota(M)(\underline{q})$. Then,

$$\mathfrak{r}(\underline{q} + \iota(\underline{p})) = \iota(\boldsymbol{c}) \left(I_{2sk'm} - \iota(M)(\underline{q})\right)^{-1} \iota(\boldsymbol{b}),$$

where $\iota(\boldsymbol{c})$ and $\iota(\boldsymbol{b})$ are an $k'm \times k'ms$ and an $k'ms \times k'm$ matrix respectively obtained by applying ι on every $m \times m$ blocks of $\boldsymbol{c}$ and $\boldsymbol{b}$.

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 the context of 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 ω be an ℓ^{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 ω. The elements of K are polynomials in ω (of degree at most $\ell - 1$) with coefficients from F.

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

The division algebra $D = (K/F, \sigma, z)$ is defined using a new indeterminate x as the ℓ-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 $b \in K$. Then D is a division algebra of dimension ℓ^2 over F [Lam01, Theorem 14.9]. The *index* of D is defined to be the square root of the dimension of D over F. In our example, D is of index ℓ. Its elements have matrix representations in $K^{\ell\times\ell}$ (the regular matrix representation defined by multiplication from the left) given below:

The matrix representation $M(x)$ of x is:

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

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

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

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

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

Remark 12. We note that $M(x)$ has a "circulant" matrix structure and $M(b)$ is a diagonal matrix. For a vector $v \in K^\ell$, it is convenient to write $\mathrm{circ}(v_1, v_2, \ldots, v_\ell)$ for the $\ell \times \ell$ matrix with $(i, i+1)^{th}$ entry v_i for $i \leqslant \ell - 1$, $(\ell, 1)^{th}$ entry as v_ℓ and remaining entries zero. Thus, we have $M(x) = \mathrm{circ}(1, 1, \ldots, 1, z)$. Similarly, we write $\mathrm{diag}(v_1, v_2, \ldots, v_\ell)$ for the diagonal matrix with entries v_i.

Fact 13. *The F-algebra generated by $M(x)$ and $M(b), b \in K$ is an isomorphic copy of the cyclic division algebra in the matrix algebra $\mathbb{M}_\ell(K)$.*

Proposition 14. *For all $b \in K$, $\mathrm{circ}(b, \sigma(b), \ldots, z\sigma^{\ell-1}(b)) = M(b) \cdot M(x)$.*

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

A standard fact is the following.

Proposition 15. *[Lam01, Section 14(14.13)] Then K linear span of $\mathfrak{B}$ is the entire matrix algebra $\mathbb{M}_\ell(K)$.*

3 Embedding Forbes-Shpilka Hitting Set in a Division Algebra

Given any noncommutative algebraic branching program of size s computing a polynomial $h \in \mathbb{F}\langle x_1, \ldots, x_n\rangle$ of degree $\tilde{d}$, the hitting set $\mathcal{H}$ contains a matrix tuple $(p_1, \ldots, p_n)$ such that $h(p_1, \ldots, p_n)$ is nonzero. Forbes and Shpilka [FS13] have shown a quasipolynomial-size hitting set construction contained in $\mathbb{M}^n_{\tilde{d}+1}(\mathbb{F})$. For ABPs over $\mathbb{Q}$, we will show the construction of a hitting set $\mathcal{H}$ which is contained in D^n such that D is a cyclic division algebra of index ℓ where ℓ is suitably chosen depending on $n, \tilde{d}$ and s.

Before we present our construction, we recall the matrix substitutions from the Forbes-Shpilka hitting set construction. Their idea is to reduce PIT for noncommutative ABPs to PIT for commutative read-once oblivious ABP (ROABP) and to design a hitting set generator for the latter. Without loss of generality, we can assume that the given ABP is a $\tilde{d}$-product of $r \times r$ matrices $M = A_1 \cdot A_2 \cdots A_{\tilde{d}}$, where the entries of each matrix A_i are homogeneous linear forms in $x_1, x_2, \ldots, x_n$. The matrix A_i corresponds to the i^{th} of the ABP. The polynomial f in $\mathbb{F}\langle x_1, \ldots, x_n\rangle$ that the ABP computes is of degree $\tilde{d} = 2^d$, and f is an entry of this matrix product M.

We can write $A_j = \sum_{i=1}^n A_{ij} x_i$, $1 \leqslant j \leqslant \tilde{d}$, where $A_{ij} \in \mathbb{F}^{r \times r}$. The entries M_{ij} of the matrix M are homogeneous polynomials in $\mathbb{F}\langle \underline{x}\rangle$. The polynomial f is computed at some entry of M as the output polynomial. Let $\{u_1, \ldots, u_{\tilde{d}}\}$ be distinct commuting indeterminates. In [FS13], the authors make the following $(\tilde{d}+1) \times (\tilde{d}+1)$ matrix substitution for each x_i, where the variable index i is encoded as the exponent of the commuting variables:

$$M(x_i) = \begin{bmatrix} 0 & u_1^i & 0 & \cdots & 0 \\ 0 & 0 & u_2^i & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & u_{\tilde{d}}^i \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Evaluating the ABP for f on this matrix substitution $x_i \leftarrow M(x_i)$ produces a $(\tilde{d}+1) \times (\tilde{d}+1)$

matrix whose $(1, \tilde{d}+1)^{th}$ entry is an ROABP as it effectively replaces each x_i variable at layer j by u_j^i. Therefore, the index j of u_j encodes the layer of the noncommutative ABP.

The black-box PIT algorithm then follows from the construction of a hitting set generator for commutative ROABPs:

$$\mathcal{G}_d : (\alpha_1, \alpha_2, \ldots, \alpha_d, \alpha_{d+1}) \mapsto (f_0(\alpha_1, \ldots, \alpha_d, \alpha_{d+1}), f_1(\alpha_1, \ldots, \alpha_d, \alpha_{d+1}), \ldots, f_{2^d-1}(\alpha_1, \ldots, \alpha_d, \alpha_{d+1})),$$

where each f_i is a polynomial of degree $\mathrm{poly}(2^d, r, n)$. The actual points of the hitting set are obtained by choosing values for each variable α_i from a subset of scalars $U \subseteq \mathbb{F}$ of $\mathrm{poly}(2^d, r, n)$ size. This makes the size of the hitting set quasipolynomial. The final substitution for each x_i variable in the noncommutative ABP is the following:

$$M(x_i) = \begin{bmatrix} 0 & f_0^i(\alpha_1, \ldots, \alpha_{d+1}) & 0 & \cdots & 0 \\ 0 & 0 & f_1^i(\alpha_1, \ldots, \alpha_{d+1}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & f_{2^d-1}^i(\alpha_1, \ldots, \alpha_{d+1}) \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{1}$$

Therefore, one approach to embedding the matrix substitutions in a cyclic division algebra $D = (K/F, \sigma, z)$ (where $F = \mathbb{Q}(z)$) of index ℓ (where ℓ is the index of D which is larger than 2^d that we fix later) would be to find a hitting set generator

$$\mathcal{G}_d : (\alpha_1, \alpha_2, \ldots, \alpha_d, \alpha_{d+1}) \mapsto (f_0(\alpha_1, \ldots, \alpha_d, \alpha_{d+1}), f_2(\alpha_1, \ldots, \alpha_d, \alpha_{d+1}), \ldots, f_{2^d-1}(\alpha_1, \ldots, \alpha_d, \alpha_{d+1})),$$

with the following additional property: $f_{i+1}(\alpha_1, \ldots, \alpha_{d+1}) = \sigma(f_i(\alpha_1, \ldots, \alpha_{d+1}))$ for each $0 \leqslant i \leqslant \ell - 2$. In that case, consider the following $\ell \times \ell$ matrix substitutions:

$$M(x_i) = \left[\begin{array}{ccccc|ccc} 0 & f_0^i(\underline{\alpha}) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & f_1^i(\underline{\alpha}) & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f_{\tilde{d}-1}^i(\underline{\alpha}) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & f_{\tilde{d}}^i(\underline{\alpha}) & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & f_{\ell-2}^i(\underline{\alpha}) \\ zf_{\ell-1}^i(\underline{\alpha}) & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right].$$

Notice that the top-left $(\tilde{d}+1) \times (\tilde{d}+1)$ submatrix of this substitution is exactly the substitution described in Equation 1. Therefore, evaluating a degree-$\tilde{d}$ noncommutative ABP B over $\{x_1, \ldots, x_n\}$ on these matrices will output the evaluation of corresponding ROABP in the $(1, \tilde{d}+1)^{th}$ entry as in [FS13]. Moreover, by Proposition 14, we can ensure that each $M(x_i)$ is in the cyclic division algebra D assuming that each $f_i(\underline{\alpha}) \in K$. Therefore, the output will also be in the division algebra D only. To conclude, for a nonzero noncommutative ABP, the image will be nonzero and in a division algebra, hence invertible.

Our goal is now to find a cyclic division algebra $D = (K/F, \sigma, z)$ (where $F = \mathbb{Q}(z)$) of index ℓ (more than $\tilde{d}$) and to construct a hitting set generator $\mathcal{G}_d : \underline{\alpha} \mapsto (f_0(\underline{\alpha}), \ldots, f_{2^d-1}(\underline{\alpha}))$ for commutative ROABPs with the additional property that $f_{i+1}(\alpha_1, \ldots, \alpha_{d+1}) = \sigma(f_i(\alpha_1, \ldots, \alpha_{d+1}))$ for each $0 \leqslant i \leqslant \ell - 2$.

We now examine the Forbes-Shpilka construction to incorporate these aspects. The construction is recursive. Suppose that we have the construction for degree 2^{d-1}.

The hitting set for degree 2^d is obtained in [FS13] by combining two copies of the hitting set for degree 2^{d-1} using the following key technical lemma, [FS13, Lemma 3.7], rephrased below in somewhat different notation.

Let $p_{\ell'}(v), 1 \leqslant \ell' \leqslant r^2$ denote the Lagrange interpolation polynomials, defining a basis for univariate polynomials interpolating values from $[r^2]$. Given $\beta_1, \ldots, \beta_{r^2} \in \mathbb{F}$, the Lagrange interpolation polynomials with respect to r^2 and the β_i's are the unique polynomials $p_{\ell'}(v)$ of degree less than r^2 such that

$$p_{\ell'}(\beta_i) = \begin{cases} 1 & \text{if } \ell' = i \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 16. [FS13, Lemma 3.7] *Let M_i and $N_i, 0 \leqslant i \leqslant 2^{d-1}-1$, be $r \times r$ matrices with entries from $\mathbb{F}[x]$ of degree less than n. Let $(f_0(u), f_1(u), \ldots, f_{2^{d-1}-1}(u)) \in \mathbb{F}[u]$ be polynomials of degree at most m. Let $\omega \in \mathbb{F}$ (or in an extension field) be an element of order at least $(2^d nm)^2$. Define polynomials in one indeterminate v:*

$$\begin{aligned} f'_i &= \sum_{\ell'=1}^{r^2} f_i(\omega^{\ell'}\alpha_d) p_{\ell'}(v),\ 0 \leqslant i \leqslant 2^{d-1}-1 \\ f'_{i+2^{d-1}} &= \sum_{\ell'=1}^{r^2} f_i((\omega^{\ell'}\alpha_d)^\mu) p_{\ell'}(v),\ 0 \leqslant i \leqslant 2^{d-1}-1, \end{aligned}$$

where $\mu = 2^{\kappa+d-1}+1$ and κ is chosen such that $2^\kappa \geqslant 2^d nm$.

Then, for all but at most $(2^d nmr)^2$ many values of α_d, the $\mathbb{F}$-linear span of the matrix coefficients of the matrix product $\prod_{i=0}^{2^{d-1}-1} M_i(f_i(x)) \prod_{i=0}^{2^{d-1}-1} N_i(f_i(y))$ is contained in the $\mathbb{F}$-linear span of the matrix coefficients of the product $\prod_{i=0}^{2^{d-1}-1} M_i(f'_i(v)) \prod_{i=2^{d-1}}^{2^d-1} N_i(f'_i(v))$.

Lemma 16 essentially gives the construction for going from the degree 2^{d-1} hitting set generator to the degree 2^d hitting set generator as proved in [FS13].

Remark 17. In our modified construction we will use different roots of unity (for the element ω) for different stages of the recursive construction. In particular, roots of unity ω_i, $i < d$, used in stages $i < d$ will be of lower order. We explain below in detail, the choice of the parameters: ℓ, κ, ω_i, and α_i for the modified construction.

We now adapt Lemma 16 to ensure the additional properties that will guarantee that the points of the hitting set are from D^n, for a suitably large cyclic division algebra D.

Theorem 18. *In deterministic quasipolynomial-time, we can construct a hitting set $\mathcal{H}$ of size $(nr\tilde{d})^{O(\log \tilde{d})}$ in D^n for the class of noncommutative polynomials in $\mathbb{Q}\langle x_1, \ldots, x_n\rangle$ computed by ABPs of width at most r with $\tilde{d}$ many layers where the index of the cyclic division algebra D, the parameter $\ell (> \tilde{d})$ is bounded by* $\mathrm{poly}(r, n, \tilde{d})$.

Proof. Let ℓ be the index of D. We set $\ell = 2^L$, where L is to be determined below. Thus, $\omega = e^{\frac{2\pi}{2^L}}$ is a 2^L-th primitive root of unity. Let $F = \mathbb{Q}(z)$ and $K = F(\omega, z)$ which gives the cyclic division algebra $D = (K/F, \sigma, z)$ where we fix the K-automorphism σ as

$$\sigma(\omega) = \omega^{2^\kappa+1},$$

and κ will be suitably chosen in the following analysis, fulfilling the constraints of Lemma 16 and some additional requirements.

Let $\omega_i = \omega^{2^{a_i}}$ for $a_1 > a_2 > \cdots > a_d > 0$, where a_i are positive integers to be chosen. Let $K_i = F(\omega_i)$ be the cyclic Galois extension for $1 \leqslant i \leqslant d$. This gives a tower of extension fields

$$F \subset F(\omega_1) \subset F(\omega_2) \subset \cdots \subset F(\omega_d) \subset F(\omega).$$

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

1. For the hitting set generator $\mathcal{G}_i$ we will choose the root of unity as ω_i and the variable α_i will take values only in the set $W_i = \{\omega_i^j \mid 1 \leqslant j \leqslant 2^{L-a_i}\}$.
2. We require that the K-automorphism σ has the property that for all $1 \leqslant i \leqslant d$ the map σ^{2^i} fixes ω_i. In fact we will ensure that σ^{2^i} has $F(\omega_i)$ as its fixed field.

We take up the second property. As $\sigma(\omega) = \omega^{2^\kappa+1}$, we have $\sigma(\omega_i) = \omega^{2^{a_i}(2^\kappa+1)}$. Therefore

$$\sigma^{2^i}(\omega_i) = \omega^{2^{a_i}(2^\kappa+1)^{2^i}}.$$

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

$$\sigma^{2^i}(\omega_i) = \omega^{2^{a_i}(2^{i+\kappa}+1)} = \omega_i \cdot \omega^{2^{a_i+i+\kappa}}.$$

We can set $a_i + i + \kappa = L$ for $1 \leqslant i \leqslant d$ to ensure that σ^{2^i} fixes ω_i. Putting $L = 2\kappa$, we obtain

$$a_i = \kappa - i \text{ for } 1 \leqslant i \leqslant d. \tag{2}$$

It remains to choose κ. In the construction of our hitting set generator $\mathcal{G}_i$, the parameter α_i will take values only in W_i defined above. We note that $|W_i| = 2^{L-a_i} = 2^{\kappa+i}$. By Lemma 16 there are at most $(2^d nmr)^2$ many bad values of α_i for any i. Thus, it suffices to choose κ such that $2^\kappa > (2^d nmr)^2$. It suffices to set

$$\kappa = 2d + \lceil 2\log_2(nmr) \rceil + 1.$$

The choice of κ determines the value of parameter μ in Lemma 16.

Coming back to the modified construction of $\mathcal{G}_d$, inductively, we can assume that the hitting set generator $\mathcal{G}_{d-1} : (\alpha_1, \ldots, \alpha_{d-1}, u) \mapsto (f_0(u), f_1(u), \ldots, f_{2^{d-1}-1}(u))$ (where for $0 \leqslant i \leqslant 2^{d-1}-1$, the polynomial $f_i(u) \in K_{d-1}[u]$) has that property. Namely, suppose $f_{i+1}(u) = \sigma(f_i(u))$ holds for all $i \leqslant 2^{d-1}-2$. Now define $\mathcal{G}_d$ using Lemma 16. Since $p_{\ell'}(v)$ has only integer coefficients, $\sigma(p_{\ell'}(v)) = p_{\ell'}(v)$. Therefore, for $0 \leqslant i \leqslant 2^{d-1}-2$ and for $2^{d-1} \leqslant i \leqslant 2^d - 2$ we have $f'_{i+1}(v) = \sigma(f'_i(v))$.

Now, consider $i = 2^{d-1}-1$. We need to ensure that $\sigma(f'_{2^{d-1}-1}(v)) = f'_{2^{d-1}}(v)$. Equivalently, we need to ensure that

$$\sigma\left(\sum_{\ell'=1}^{r^2} f_{2^{d-1}-1}(\omega_d^{\ell'}\alpha_d)p_{\ell'}(v)\right) = \sum_{\ell'=1}^{r^2} f_1((\omega_d^{\ell'}\alpha_d)^\mu)p_{\ell'}(v).$$

This is enforced by requiring that

$$\sigma^{2^{d-1}}\left(\sum_{\ell'=1}^{r^2} f_1(\omega_d^{\ell'}\alpha_d)p_{\ell'}(v)\right) = \sum_{\ell'=1}^{r^2} f_1((\omega_d^{\ell'}\alpha_d)^{\mu})p_{\ell'}(v).$$

Since α_d will be chosen from W_d (all powers of ω_d), we can write $\omega_d^{\ell'}\alpha_d = \omega_d^j$ for some j. Now, $\sigma^{2^{d-1}} f_1(\omega_d^j) = f_1(\sigma^{2^{d-1}}(\omega_d^j))$ as $\sigma^{2^{d-1}}$ fixes all coefficients of f_1 (because $f_1(u) \in K_{d-1}[u]$). Now,

$$\sigma^{2^{d-1}}(\omega_d^j)) = \omega_d^{j\cdot(2^\kappa+1)^{2^{d-1}}} = \omega_d^{j(1+2^{d-1+\kappa})} = (\omega_d^\ell\alpha_d)^\mu,$$

which verifies the choice of μ in Lemma 16 is $1 + 2^{d-1+\kappa}$.

As shown in [FS13], the parameter v (whose place holder is α_{d+1} in the description of $\mathcal{G}_d$) should vary over a set of size $\text{poly}(2^d, n, m, r)$. This way we ensure that $f_{i+1} = \sigma(f_i)$ for $0 \leqslant i \leqslant 2^d - 2$. Now define $f_{2^d+j} = \sigma(f_{2^d+j-1})$ for $0 \leqslant j \leqslant \ell - 2^d - 1$. The fact that $\mathcal{G}_d$ is indeed a generator follows from the span preserving property and the proof is identical to the proof of [FS13, Lemma 3.19]. □

Note that $\mathcal{H}$ is a strong hitting set for any such noncommutative ABP.

4 Strong Hitting Set for Generalized ABPs over Division Algebra

In this section, we first define the notion of generalized ABPs and ABPs over a division algebra. Then we show the construction of a quasipolynomial-size strong hitting set for generalized ABPs over a division algebra such that any nonzero generalized ABP will evaluate to an invertible matrix at some point in the hitting set.

Definition 19. A *generalized ABP* over the matrix algebra $\mathbb{M}_m(\mathbb{F})$ is defined in the same way as a noncommutative ABP, except for the fact that the linear forms labeling the edges are of the form $\sum_{i=1}^n a_i x_i b_i$, where $a_i, b_i \in \mathbb{M}_m(\mathbb{F})$. Such an ABP computes a generalized polynomial in the generalized polynomial ring $\mathbb{M}_m(\mathbb{F})\langle X\rangle$, where the polynomial is defined as the sum of products of the linear forms along all s-to-t paths of the ABP, where s is the source node and t is the sink node of the directed acyclic graph underlying the ABP.

For a division algebra D, if the linear forms labeling the edges of the ABP are of the form $\sum_{i=1}^n a_i x_i b_i, a_i, b_i \in D$ then it is a generalized ABP over the division algebra D.

Let $D = (K/F, \sigma, z)$ (here $F = \mathbb{Q}(z)$) be a cyclic division algebra of index ℓ as defined in Section 2.3. Let $\mathfrak{B} = \{C_{ij}\}_{i,j\in[\ell]}$ be the F-basis of D for $i, j \in [\ell]$ as described in Section 2. Informally, our idea is to reduce the problem of finding strong hitting set for generalized ABPs over division algebra to the hitting set construction of a product of commutative ROABPs.

Lemma 20. *For any nonzero generalized ABP B of degree d over $D\langle \underline{x}\rangle$, there exists a substitution for each x_k of the following form:*

$$M(x_k) = \begin{bmatrix} 0 & p_{k1} & 0 & \cdots & 0 \\ 0 & 0 & p_{k2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & p_{k(d-1)} \\ p_{kd} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

such that for each $l \in [d]$, p_{kl} is in D and image of B is invertible on that substitution under the inclusion map $a \mapsto I_d \otimes a$ where $a \in D$.

Proof. Let ℓ be the index of the division algebra D. We first prove that for any nonzero generalized ABP B of degree d over $D\langle \underline{x} \rangle$, there exists a substitution for each x_k of the following form:

$$M(x_k) = \begin{bmatrix} 0 & q_{k1} & 0 & \cdots & 0 \\ 0 & 0 & q_{k2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & q_{k(d-1)} \\ q_{kd} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

such that for each $l \in [d]$, q_{kl} is in $\mathbb{M}_\ell(K)$ and the image of B is nonzero on that substitution with a block-diagonal structure. To evaluate B on such matrix substitution the coefficients $a \in D$ (which have matrix representations in $\mathbb{M}_\ell(K)$) are fit to the correct dimension using the inclusion map $\iota' : \mathbb{M}_\ell(K) \to \mathbb{M}_{d\ell}(K)$ where $\iota'(a) = I_d \otimes a$.

Let ψ be the substitution map that replaces each variables $\{x_k\}_{k\in[n]}$ by an $\ell \times \ell$ matrix of noncommuting variables $\{z_{ijk}\}_{i,j\in[\ell],k\in[n]}$. One can naturally extend the definition of $\psi : \mathbb{M}_\ell(K)\langle \underline{x} \rangle \to \mathbb{M}_\ell(K\langle \underline{z} \rangle)$ i.e. ψ maps a generalized polynomial over matrix algebra $\mathbb{M}_\ell(K)$ to an $\ell \times \ell$ matrix of noncommutative polynomials in $K\langle \underline{z} \rangle$. Indeed, the map ψ is identity preserving (see [Vol18, Equation 3.10] for example).

Introduce a new set of commuting variables $\widetilde{Z} = \{\tilde{z}_{ijkl}\}$ where $i, j \in [\ell]$, $k \in [n]$ and $l \in [d]$ and consider the following substitution for each x_k:

$$\widetilde{Z}_k = \begin{bmatrix} 0 & \widetilde{Z}_{k1} & 0 & \cdots & 0 \\ 0 & 0 & \widetilde{Z}_{k2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \widetilde{Z}_{k(d-1)} \\ \widetilde{Z}_{kd} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

where $\widetilde{Z}_{kl} = (\tilde{z}_{ijkl})_{1\leqslant i,j\leqslant \ell}$. In effect, the substitution of the x_k variables by the matrices $\widetilde{Z}_k$ is just set-multilinearization of $\psi(B)$ position-wise and hence identity preserving.

What is the effect of this substitution on a degree-d generalized word? To understand that consider a generalized word $w = a_0 x_{k_1} a_1 x_{k_2} \cdots a_{d-1} x_{k_d} a_d$ where each $a_i \in \mathbb{M}_\ell(K)$. Observe that $w(\widetilde{Z_1}, \ldots, \widetilde{Z_n})$ is a block-diagonal matrix (using the inclusion map ι') with $(i,i)^{th}$ block entry $a_0 \widetilde{Z}_{k_1\pi_i(1)} a_1 \cdots a_{d-1} \widetilde{Z}_{k_d\pi_i(d)} a_d$ where $\pi_i, 1 \leqslant i \leqslant d$ is the cyclic permutations on $[d]$ such that $\pi_i(1) = i$, $\pi_i(2) = i+1$ and so on. For example, consider the case $d = 3$. For a generalized word $a_0 x_1 a_1 x_2 a_2 x_3 a_3$, the image will be the following product:

$$\begin{bmatrix} a_0 & & \\ & a_0 & \\ & & a_0 \end{bmatrix} \begin{bmatrix} & \widetilde{Z}_{11} & \\ & & \widetilde{Z}_{12} \\ \widetilde{Z}_{13} & & \end{bmatrix} \begin{bmatrix} a_1 & & \\ & a_1 & \\ & & a_1 \end{bmatrix} \begin{bmatrix} & \widetilde{Z}_{21} & \\ & & \widetilde{Z}_{22} \\ \widetilde{Z}_{23} & & \end{bmatrix} \begin{bmatrix} a_2 & & \\ & a_2 & \\ & & a_2 \end{bmatrix} \begin{bmatrix} & \widetilde{Z}_{31} & \\ & & \widetilde{Z}_{32} \\ \widetilde{Z}_{33} & & \end{bmatrix} \begin{bmatrix} a_3 & & \\ & a_3 & \\ & & a_3 \end{bmatrix},$$

which outputs the following diagonal matrix:

$$\begin{bmatrix} a_0\widetilde{Z}_{11}a_1\widetilde{Z}_{22}a_2\widetilde{Z}_{33}a_3 & & \\ & a_0\widetilde{Z}_{12}a_1\widetilde{Z}_{23}a_2\widetilde{Z}_{31}a_3 & \\ & & a_0\widetilde{Z}_{13}a_1\widetilde{Z}_{21}a_2\widetilde{Z}_{32}a_3 \end{bmatrix}.$$

Let $B = \sum a_0 x_{k_1} a_1 x_{k_2} a_2 \dots a_{d-1} x_{k_d} a_{i_d}$. So the $(i,i)^{th}$ entry of $B(\widetilde{Z_1}, \dots, \widetilde{Z_n})$ is

$$B^{\pi_i} = \sum a_0 \widetilde{Z}_{k_1 \pi_i(1)} a_1 \widetilde{Z}_{k_2 \pi_i(2)} \dots a_{d-1} \widetilde{Z}_{k_d \pi_i(d)} a_{i_d}.$$

Hence the final output matrix will be the following:

$$B(\widetilde{Z}) = \begin{bmatrix} B^{\pi_1} & & & \\ & B^{\pi_2} & & \\ & & \ddots & \\ & & & B^{\pi_d} \end{bmatrix}.$$

We now claim the following.

Claim 21. *For each $i \in [d]$, B^{π_i} is nonzero.*

Proof. As B in $D\langle \underline{x} \rangle$ is nonzero and ψ is an identity preserving substitution, $\psi(B) \in \mathbb{M}_\ell(K\langle \underline{z} \rangle)$ is also nonzero. We now consider the entry-wise set-multilinearization of $\psi(B)$ with respect to the cyclic permutation π_i i.e. encoding any word using $\pi_i(j)$ as the position index for the j^{th} position for each entry of $\psi(B)$. Notice that, it outputs the matrix B^{π_i}. Moreover, as $\psi(B)$ is nonzero, B^{π_i} must be nonzero as set-multilinearization preserves identity. □

Hence, there exist substitutions q_{kl} from $\mathbb{M}_\ell(K)$ for the $\widetilde{Z}$ variables such that B is nonzero.

Now we use Proposition 15 which says that K-linear span of $\mathfrak{B}$ is the entire matrix algebra $\mathbb{M}_\ell(K)$. The above argument shows that if we replace each q_{kl} in $M(x_k)$ by a linear combination

$$\sum_{i,j} y_{ijkl} C_{ij},$$

each diagonal block matrix of the output matrix obtained from the image of B on this evaluation is still nonzero over the $\{y_{ijkl}\}$ variables. We now find substitutions for the Y variables from the ground field F to make each diagonal block matrix nonzero. As any F-linear combination of C_{ij} is in the division algebra, each such linear combinations is in D. So, define $p_{kl} = \sum_{i,j} \beta_{ijkl} C_{ij} \in D$ where β_{ijkl} are the substitutions for y_{ijkl} variables from F. In fact the values for the variables β_{ijkl} can be found from $\mathbb{Q}$ itself by a standard use of Polynomial Identity Lemma [DL78, Zip79, Sch80]. Notice that, each diagonal block will also be in D. Since each diagonal block matrix is nonzero and in D it is invertible. Therefore, the image of B is also invertible on the chosen matrix tuple. □

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

Theorem 22. *Given the parameters n, ℓ, r, d, in deterministic quasipolynomial-time we can construct strong hitting set $\mathcal{H}'$ of size $(nrd\ell)^{O(\log nd\ell)}$ for any nonzero generalized ABP B of degree d and width r over $D\langle \underline{x} \rangle$ where ℓ is the index of D.*

Proof. By Lemma 20, we know that there exists matrix tuple $(p_1, \dots, p_n)$ in $\mathbb{M}^n_{d\ell}(K)$ of the following form

$$p_k = M(x_k) = \begin{bmatrix} 0 & p_{k1} & 0 & \cdots & 0 \\ 0 & 0 & p_{k2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & p_{k(d-1)} \\ p_{kd} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

where each $p_{kl} \in D : 1 \leqslant k \leqslant n, 1 \leqslant l \leqslant d$ such that $B(p_1, p_2, \ldots, p_n)$ is an invertible matrix.

Write each p_{kl} as $p_{kl} = \sum_{i,j \in [\ell]} y_{ijkl} C_{ij}$ for some commuting indeterminates $Y = \{y_{ijkl}\}$ whose values we need to determine. On such a substitution, B evaluates to the following matrix:

$$\begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_d \end{bmatrix}.$$

where each $B_l, 1 \leqslant l \leqslant d$ is nonzero by Lemma 20 (using the inclusion map ι'). We now observe the following.

Claim 23. *For each $l \in [d]$, B_l is a matrix of commutative set-multilinear ABPs each of width $r\ell$.*

Proof. To see this, consider the matrix B_1. We can think of B_1 as the matrix obtained by substituting p_{kl} for x_k in layer l of the input generalized ABP B over D of index ℓ. This computation can also be thought of by making ℓ-many copies of each node in B.

More precisely, each coefficient $a \in D$ in B has a $\ell \times \ell$ matrix representation over K. Now consider each edge $\sum_{k=1}^{n} a_k x_k b_k$ between the layer l and $l+1$. Since x_k is replaced by p_{kl} and $a_k, b_k \in D$, this edge can be replaced by an $\ell \times \ell$ bipartite graph such that for each $i, j \in [\ell]$, the edge connecting the i^{th} node (from left) to the j^{th} node (to right) is labeled by the $(i,j)^{th}$ entry of the product of $a_k p_{kl} b_k$, a linear form over $K[Y]$. Clearly, it produces an ℓ-input ℓ-output setmultilinear ABP of width $r\ell$. Therefore, each entry in B_1 is computed by a set-multilinear ABP of width $r\ell$ and degree d. The situation for other $B_l : 2 \leqslant l \leqslant d$ are similar. □

Therefore we can use a hitting set generator for commutative set-multilinear ABPs of width $r\ell$ and degree d to obtain a point such that the image for each B_l is nonzero on that evaluation.

However, our goal is to obtain an invertible image for the image of B. In other words, we want a substitution of Y variables for which each B_l would be invertible. Notice that if for substitution of Y variables from F at least one entry of B_l is nonzero, then the matrix B_l is also invertible as the image of B_l is a nonzero element in D. Hence, to obtain a strong hitting set for the input generalized ABP over D (equivalently, to obtain a substitution on which the product of the matrices $B_l, 1 \leqslant l \leqslant d$ is invertible), it suffices to obtain a hitting set for the product of set-multilinear ABPs (product of one of the nonzero entries of each B_l).

We do this by first converting each set-multilinear ABP to an ROABP encoding each y_{ijkl} to $v_l^{(\ell+1)^2 i + (\ell+1) j + k}$[2]. By construction each encoded B_l yields a known variable partition for the corresponding ROABP. More precisely, for each l the ROABP computed in the $(l,l)^{th}$ diagonal block follows the variable partition:

$$v_l < v_{l+1} < \ldots < v_d < v_1 < \ldots < v_{l-1}.$$

Therefore, for each $nd\ell^2$-variate ROABP of degree d and of width ℓr computed in each diagonal block, we can use the hitting set generator of Theorem 6. Now, for a d-fold product of such ROABPs of different but known orders, the same hitting set generator will also work. This is because we can ensure that the hitting set generator of Theorem 6 has the property that more than $1 - 1/d$

[2]Note that by the choice, ℓ is larger than n and d.

fraction of seeds for the generator works for a given ROABP with the above parameters. Then, by a standard union bound argument it follows that there is a choice of seed for the generator that will hit the product of these d many ROABPs. Notice that the ROABPs are all $nd\ell^2$-variate, d-degree, and of width ℓr. Thus by Theorem 6, the size of the hitting set for them is $(nd\ell r)^{O(\log nd\ell)}$. Hence, we can now find a substitution for the v_l variables such that each B_l is invertible, hence B is also invertible.

This gives us a hitting set $\mathcal{H}$ under the inclusion map $\iota' : \mathbb{M}_\ell(K) \to \mathbb{M}_{d\ell}(K)$ where $\iota'(a) = I_d \otimes a$. However for the purpose of Section 5, we find a hitting set $\mathcal{H}'$ under the inclusion map $\iota : \mathbb{M}_\ell(K) \to \mathbb{M}_{d\ell}(K)$ where $\iota(a) = a \otimes I_d$. Although it is technically possible to work with two inclusion maps thanks to Remark 11, we find it mathematically nicer to work with a single inclusion map. For this we explicitly find a permutation matrix q_0 of dimension $d\ell$ such that $q_0(I_d \otimes a)q_0^{-1} = a \otimes I_d$ for all $a \in \mathbb{M}_\ell(K)$. Once we find q_0, the final hitting set can be defined as $\mathcal{H}' = \{(q_0 p_1 q_0^{-1}, \ldots, q_0 p_n q_0^{-1}) \mid \underline{p} \in \mathcal{H}\}$. To see this, let

$$B = \sum a_0 x_{k_1} a_1 \cdots a_{d-1} x_{k_d} a_d.$$

Let $M = B(q_1, \ldots, q_n)$ is an invertible matrix for $\underline{q} \in \mathcal{H}$. We know that,

$$\sum (I_d \otimes a_0) q_{k_1} (I_d \otimes a_1) \cdots (I_d \otimes a_{d-1}) q_{k_d} (I_d \otimes a_d) = M. \tag{3}$$

By conjugating M with q_0, obtain the following:

$$\sum q_0(I_d \otimes a_0)q_0^{-1} q'_{k_1} q_0(I_d \otimes a_1)q_0^{-1} \cdots q_0(I_d \otimes a_{d-1})q_0^{-1} q'_{k_d} q_0(I_d \otimes a_d)q_0^{-1} = q_0 M q_0^{-1}, \tag{4}$$

where $q'_{k_j} = q_0 q_{k_j} q_0^{-1}$. In other words $B(q'_1, \ldots, q'_n)$ is the invertible matrix $M' = q_0 M q_0^{-1}$ under the inclusion map ι. In the following, we show that the permutation matrix q_0 can be constructed explicitly.

Proposition 24. *For any $a \in \mathbb{M}_\ell(\mathbb{F})$, one can construct a $d\ell \times d\ell$ permutation matrix q_0 such that $q_0(I_d \otimes a)q_0^{-1} = a \otimes I_d$.*

Proof. We can obtain $a \otimes I_d$ from $I_d \otimes a$ just by rearranging the rows and columns properly. Let us divide the $d\ell$ rows in group of d rows as $1, \ldots, d, d+1, \ldots, 2d, \ldots, (\ell-1)d, \ldots, d\ell$. We define q_0 such that for the group of rows $id+1$ to $id+d$ (for $0 \leqslant i \leqslant (\ell-1)$), the $(id+j, (j-1)\ell + (i+1))^{th}$ entry for $1 \leqslant j \leqslant d$ is 1 and remaining entries to be zero. Notice that, q_0^{-1} is the transpose of q_0 as it is a permutation matrix. Now the proof follows from the definition of $I_d \otimes a$ and $a \otimes I_d$.

To elaborate, we consider the case where $\ell = 2$ and $d = 3$ and give an illustrative example. In this case, q_0 and q_0^{-1} are the following matrices.

$$q_0 = \begin{bmatrix} 1 & & & & & \\ & & 1 & & & \\ & & & & 1 & \\ & 1 & & & & \\ & & & 1 & & \\ & & & & & 1 \end{bmatrix}, \qquad q_0^{-1} = \begin{bmatrix} 1 & & & & & \\ & & & 1 & & \\ & 1 & & & & \\ & & & & 1 & \\ & & 1 & & & \\ & & & & & 1 \end{bmatrix}.$$

$$\text{Let, } a = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad \text{then, } I_3 \otimes a = \left[\begin{array}{cc|cc|cc} 1 & 2 & & & & \\ 3 & 4 & & & & \\ \hline & & 1 & 2 & & \\ & & 3 & 4 & & \\ \hline & & & & 1 & 2 \\ & & & & 3 & 4 \end{array}\right].$$

Consider the effect of q_0.

$$q_0(I_3 \otimes a)q_0^{-1} = \left[\begin{array}{ccc|ccc} 1 & & & 2 & & \\ & 1 & & & 2 & \\ & & 1 & & & 2 \\ \hline 3 & & & 4 & & \\ & 3 & & & 4 & \\ & & 3 & & & 4 \end{array}\right] = a \otimes I_3.$$

□

□

5 Putting all together

In this section we prove our main result, the construction of a hitting set for noncommutative rational formulas of inversion height two. An intermediate step is to construct a strong hitting set for rational formulas of inversion height one. En route to our proof, we crucially use the connection of rational identity testing with the identity testing of generalized ABPs. We make it explicit in Proposition 26. But before this, we note a basic result that we use throughout the section.

Lemma 25. *Let $\mathfrak{r} \in \mathbb{F}\langle\langle \underline{x} \rangle\rangle$ be a rational formula of size s. Let $\underline{p} = (p_1, \ldots, p_n) \in \mathbb{M}_m^n(\mathbb{F}(t_1, t_2))$ be an n-tuple of matrix of bivariate rational functions where the degrees of the numerator and denominator polynomials over t_1, t_2 at each entry are at most d' and $\mathfrak{r}$ is defined at $\underline{p}$. Then, evaluating $\mathfrak{r}$ on $\underline{p}$ outputs $\mathfrak{r}(\underline{p}) \in \mathbb{M}_m(\mathbb{F}(t_1, t_2))$ such that each entry of the output matrix is of form $\frac{P(t_1,t_2)}{Q(t_1,t_2)}$ where P and Q are bivariate polynomials of degree at most $O(smd')$.*

Proof. As already stated in Section 1 that $\mathfrak{r}$ has a linear pencil L of size (at most) $2s$ such that for any tuple $\underline{p}$, $\mathfrak{r}(\underline{p})$ is defined if and only if $L(\underline{p})$ is invertible [HW15, Proposition 7.1]. Moreover, $\mathfrak{r}(\underline{p}) = L_{i,j}^{-1}(\underline{p})$ for some $(i,j)^{th}$ entry of L i.e. $\mathfrak{r}(\underline{p})$ is the $(i,j)^{th}$ block of $L^{-1}(\underline{p})$ thinking of it as a $2s \times 2s$ block matrix where each block is of size m. Notice that, if $L = \sum_{i=1}^n A_i x_i$, then $L(\underline{p}) = \sum_{i=1}^n A_i \otimes p_i$. Therefore, $L(\underline{p})$ is a $2sm \times 2sm$ matrix such that each entry is a polynomial over t_1, t_2 of degree at most d'. From the standard computation of matrix inverse, it is immediate that each entry of $L^{-1}(\underline{p})$ (therefore, each entry of $\mathfrak{r}(\underline{p})$) is a commutative rational function such that the numerator and the denominator are bivariate polynomials over t_1, t_2 with degree bound $O(smd')$. □

Now we are ready to prove the main proposition.

Proposition 26. *Let $\mathfrak{r}$ be a noncommutative rational formula over $x_1, \ldots, x_n$ of size s and $(q_1, \ldots, q_n) \in \mathbb{M}_m^n(\mathbb{F})$ be a matrix tuple such that $\mathfrak{r}$ is defined on $\underline{q}$. Suppose, $\mathfrak{r}(\underline{x} + \underline{q})$ is a recognizable generalized series over $\mathbb{M}_m(\mathbb{F})\langle\langle \underline{x} \rangle\rangle$ with a linear representation $(\boldsymbol{c}, M, \boldsymbol{b})$ of size at most $2s$ over $\mathbb{M}_m(\mathbb{F})$. Define $S^{\{d\}} = \boldsymbol{c} \cdot M^d \cdot \boldsymbol{b}$ computing a generalized polynomial in $\mathbb{M}_m(\mathbb{F})\langle \underline{x} \rangle$. Then $\mathfrak{r}$ is nonzero in $\mathbb{F}\langle\langle \underline{x} \rangle\rangle$ if and only if $S^{\{d\}}$ is nonzero for some $d \leqslant 2sm - 1$. Additionally, there exists some $N \leqslant \text{poly}(ksm)$ such that if $|\mathbb{F}| > N$, then for any $T \subseteq \mathbb{F}$, $|T| = N$ and for some matrix tuple $(p_1, \ldots, p_n) \in \mathbb{M}_{km}^n(\mathbb{F})$, evaluating $S^{\{d\}}$ at $\underline{p}$ under the inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{km}(\mathbb{F})$ (where $\iota(a) = a \otimes I_k$) for each $d \leqslant 2sm - 1$,*

1. *If $S^{\{d\}}$ evaluates nonzero for some d , then there exists an $\alpha \in T$ such that $\mathfrak{r}$ is nonzero at the following matrix tuple:*

$$(\alpha p_1 + q_1 \otimes I_k, \ldots, \alpha p_n + q_n \otimes I_k).$$

2. *If $S^{\{d\}}(\underline{p})$ is invertible for some d, there exists an $\alpha \in T$ such that $\mathfrak{r}$ is invertible at the following matrix tuple:*

$$(\alpha p_1 + q_1 \otimes I_k, \ldots, \alpha p_n + q_n \otimes I_k).$$

Proof. By Theorem 10, we know that $\mathfrak{r}(\underline{x})$ is zero if and only if $\mathfrak{r}(\underline{x} + \underline{q})$ is zero. Let $Z = \{z_{i,j,k'}\}_{1 \leqslant i,j \leqslant m, 1 \leqslant k' \leqslant n}$ be a set of noncommuting variables. Consider a substitution map ψ that substitutes each variable $x_{k'}, 1 \leqslant k' \leqslant n$ of $\mathfrak{r}(\underline{x} + \underline{q})$ by an $m \times m$ matrix $Z_{k'}$ consisting of fresh noncommutative variables $\{z_{i,j,k'}\}_{1 \leqslant i,j \leqslant m}$. Consider $\mathfrak{r}(\psi(\underline{x}) + \underline{q})$ and observe that, ψ is an identity preserving and degree preserving substitution.

From the definition, $\mathfrak{r}(\underline{x} + \underline{q}) = \boldsymbol{c}(I - M)^{-1}\boldsymbol{b}$ where M is of size at most $2s$ by Theorem 10. Therefore, $\mathfrak{r}(\psi(\underline{x}) + \underline{q}) = C(I - \psi(M))^{-1}B$, where it is convenient to think of $\boldsymbol{c}$ (respectively $\boldsymbol{b}$) as an $m \times 2ms$ (resp. $2ms \times m$) rectangular matrix C (resp. B), and $\psi(M)$ as $2ms \times 2ms$ matrix.

Observe that, for the matrix $\mathfrak{r}(\psi(\underline{x}) + \underline{q})$, the $(i,j)^{th}$ entry is the following recognizable series which has linear representation of size at most $2sm$:

$$\boldsymbol{C_i}(I - \psi(M))^{-1}\boldsymbol{B_j}$$

where $\boldsymbol{C_i}$ is the i^{th} row of C and $\boldsymbol{B_j}$ is the j^{th} column of B. If $\mathfrak{r}(\underline{x} + \underline{q})$ is nonzero, then some $(i,j)^{th}$ entry of $\mathfrak{r}(\psi(\underline{x}) + \underline{q})$ is also nonzero. Clearly, the degree-d truncated part of the matrix $\mathfrak{r}(\psi(\underline{x}) + \underline{q})$ is $\psi(S^{\{d\}})$. Moreover, for the matrix $\psi(S^{\{d\}})$, each entry is computed by a noncommutative ABP of width $2sm$ and depth d over Z variables. By Theorem 8, there exists a minimum $d \leqslant 2sm - 1$ such that $\psi(S^{\{d\}})$ and thus $S^{\{d\}}$ is nonzero. Clearly $S^{\{d\}}$ is computable by a generalized ABP.

Proof of part(1): Now, for some matrix tuple $(p_1, \ldots, p_n) \in \mathbb{M}_{km}(\mathbb{F})$, let $d \leqslant 2sm - 1$ such that $S^{\{d\}}$ is nonzero at $\underline{p}$ under the inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{km}(\mathbb{F})$ given by $\iota : a \to a \otimes I_k$. Consider the evaluation of $\mathfrak{r}$ at $(tp_1 + q_1 \otimes I_k, \ldots, tp_n + q_n \otimes I_k)$ where t is some commuting indeterminate. Let $M(t) = \mathfrak{r}(tp_1 + q_1 \otimes I_k, \ldots, tp_n + q_n \otimes I_k)$. We now interpret $M(t)$ in two ways. First, think of $M(t)$ as the evaluation of the generalized series $\mathfrak{r}(\underline{x} + \underline{q})$ at $(tp_1, \ldots, tp_n)$ under the inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{km}(\mathbb{F})$ given by $\iota : a \to a \otimes I_k$. We can write $M(t) = t^d S^{\{d\}}(\underline{p}) + M'(t)$ where t-degree of each term of the matrix $M'(t)$ is strictly more than d. Therefore, $M(t)$ is nonzero.

Another way to interpret $M(t)$ is to evaluate the rational formula $\mathfrak{r}$ on $(tp_1 + q_1 \otimes I_k, \ldots, tp_n + q_n \otimes I_k)$. Since $\mathfrak{r}$ is a rational formula of size s, each entry of the matrix $M(t)$ is an element of

the function field $\mathbb{F}(t)$. Moreover by Lemma 25, the t-degrees of the numerator and denominator polynomials of each such commutative rational expression computed at all the nodes, are bounded by $\hat{d} = \text{poly}(ksm)$. Therefore, the final choice of the parameter t should be such that it avoids the zeros of the numerator and denominator polynomials involved in the computation of $M(t)$. This is clearly possible by varying t over a $\text{poly}(ksm)$ size set $T \subseteq \mathbb{F}$.

Proof of part(2): The proof of the second part is similar. For some matrix tuple $(p_1, \ldots, p_n) \in \mathbb{M}_{km}(\mathbb{F})$, let $d \leqslant 2sm - 1$ such that $S^{\{d\}}$ is invertible at $\underline{p}$ under the inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{km}(\mathbb{F})$ given by $\iota : a \to a \otimes I_k$. Let $M(t) = \mathfrak{r}(tp_1 + q_1 \otimes I_k, \ldots, tp_n + q_n \otimes I_k)$. As before, consider two interpretations of $M(t)$. Think of $M(t)$ as the evaluation of the generalized series $\mathfrak{r}(\underline{x} + \underline{q})$ at $(tp_1, \ldots, tp_n)$ again under the inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \to \mathbb{M}_{km}(\mathbb{F})$ given by $\iota : a \to a \otimes I_k$. We write $\det M(t) = t^{mkd} \det S^{\{d\}}(\underline{p}) + M''(t)$ where t-degree of each term of the matrix $M''(t)$ is strictly more than mkd. Therefore, $\det M(t)$ is nonzero.

Interpret $M(t)$ as the evaluation of the rational formula $\mathfrak{r}$ on $(tp_1 + q_1 \otimes I_k, \ldots, tp_n + q_n \otimes I_k)$. Since $\mathfrak{r}$ is a rational formula of size s, each entry of the matrix $M(t)$ is an element of the function field $\mathbb{F}(t)$. Again by Lemma 25, the t-degrees of each numerator and denominator polynomial involved in the computation of $M(t)$ and $\det M(t)$ is also bounded by $\text{poly}(ksm)$. Therefore, the final choice of the parameter t should be such that it avoids the zeros of all such the numerator and denominator polynomials involved in the computation of $M(t)$ and $\det(M(t))$. This is clearly possible by varying t over any $\text{poly}(ksm)$ size set $T \subseteq \mathbb{F}$.

Final substitution is of the following form in both the cases:

$$\{(\alpha p_1 + q_1 \otimes I_k, \ldots, \alpha p_n + q_n \otimes I_k)\}, \tag{5}$$

for some suitably chosen $\alpha \in T \subseteq \mathbb{F}$. □

Strong hitting set for rational formulas of inversion height one

We now show the construction of a strong hitting set for noncommutative rational formulas of inversion height one.

Theorem 27. *Given n, s, we can construct a strong hitting set $\widetilde{\mathcal{H}}_1$ of size $(ns)^{O(\log ns)}$ over $\mathbb{M}^n_{d'}(K)$ for the class of noncommutative rational formulas $\mathfrak{r} \in \mathbb{Q}\langle\!\langle x_1, \ldots, x_n \rangle\!\rangle$ of size s and of inversion height one. The parameter d' is $\text{poly}(n, s)$ and $K = \mathbb{Q}(\omega, z)$ is the extension field by adjoining a primitive root of unity ω of order ℓ where $\ell = \text{poly}(n, s)$.*

Proof. Let $\mathfrak{r}(\underline{x})$ be a rational formula of inversion height one in $\mathbb{Q}\langle\!\langle \underline{x} \rangle\!\rangle$ of size s. Let $h_1, \ldots, h_k$ be all the sub-formulas input to the inverse gates in the rational formula for $\mathfrak{r}$. Consider the noncommutative formula $h = h_1 h_2 \cdots h_k$ in $\mathbb{Q}\langle \underline{x} \rangle$ which is of size at most s and degree is also bounded by s.

By Theorem 18, we construct a hitting set $\mathcal{H}_0$ in D^n where $D = (K/F, \sigma, z)$ is a cyclic division algebra of index $\ell = \text{poly}(n, s)$ for noncommutative ABPs in $\mathbb{Q}\langle \underline{x} \rangle$ of width and layers at most s. Then there is a point $\underline{q} \in \mathcal{H}_0$ such that $h(\underline{q})$ is invertible and hence $\mathfrak{r}(\underline{q})$ is defined.

Following Theorem 10, if $\mathfrak{r}(\underline{x})$ is nonzero then $\mathfrak{r}(\underline{x} + \underline{q})$ can be represented as a nonzero recognizable generalized series. Indeed, it is a recognizable generalized series over D following Theorem 10. Moreover, using the second part of Proposition 26, to obtain a strong hitting set for $\mathfrak{r}(\underline{x})$, it suffices

to find a strong hitting set of a generalized ABP over D of width $r \leqslant 2s$ and degree $d \leqslant 2s\ell - 1$. We now use the strong hitting set $\mathcal{H}_1$ in $\mathbb{M}^n_{d\ell}(K)$ (recall that $K = \mathbb{Q}(z, \omega)$ where ω is the primitive root of unity of order ℓ) for generalized ABPs of degree d over D (here ℓ is the index of D) obtained in Theorem 22. Inspecting the proof of Proposition 26, we can now find a subset $T \subseteq \mathbb{Q}$ of size $\text{poly}(n, s)$ and the final quasipolynomial-size hitting set is the following:

$$\widehat{\mathcal{H}_1} = \{\alpha \underline{p} + \underline{q} \otimes I_d : \underline{p} \in \mathcal{H}_1, \underline{q} \in \mathcal{H}_0, \alpha \in T\} \subseteq \mathbb{M}^n_{d\ell}(K).$$

□

Hitting set for rational formulas of inversion height two

We are now ready to prove our main theorem.

Proof of Theorem 1. Let $\mathfrak{r}(\underline{x})$ be a rational formula of inversion height two in $\mathbb{Q}\langle\langle \underline{x} \rangle\rangle$ of size s. Let $\mathcal{F}$ be the collection of all those inverse gates in the formula such that for every $\mathfrak{g} \in \mathcal{F}$, the path from the root to $\mathfrak{g}$ does not contain any inverse gate. For each $\mathfrak{g}_i \in \mathcal{F}$, let h_i be the sub-formula input to $\mathfrak{g}_i$. 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. h is of inversion height one. By Theorem 27, we construct a strong hitting set $\widehat{\mathcal{H}_1}$ in $\mathbb{M}_d(K)$ where $d = \text{poly}(n, s)$. Then there is a point $\underline{q} \in \widehat{\mathcal{H}_1}$ such that $h(\underline{q})$ is invertible and hence $\mathfrak{r}(\underline{q})$ is defined.

Following Theorem 10, if $\mathfrak{r}(\underline{x})$ is nonzero then $\mathfrak{r}(\underline{x} + \underline{q})$ can be represented as a nonzero recognizable generalized series over $\mathbb{M}_d(K)$. Moreover, using the first part of the proof of Proposition 26, to obtain a hitting set for $\mathfrak{r}(\underline{x})$, it suffices to find a hitting set for generalized ABP B over $\mathbb{M}_d(K)$ of width $r \leqslant 2s$ and degree $\hat{d} \leqslant 2sd - 1$, the degree-$\hat{d}$ truncated part of the generalized series $\mathfrak{r}(\underline{x} + \underline{q})$. We recall the substitution map ψ from Proposition 26 and consider $\psi(B)$. Each entry of $\psi(B)$ is computable by a noncommutative ABP of width $2sd$ and degree $\hat{d}$ over $Z = \{z_{i,j,k'}\}$ variables. Let $\mathcal{H}_{FS} \subseteq \mathbb{M}^{nd^2}_{\hat{d}+1}(K)$ be the hitting set for ABPs of width $2sd$ and of degree $\hat{d}$ over nd^2 many variables obtained from Theorem 7. We now define $\widetilde{\mathcal{H}}_{FS} \in \mathbb{M}^n_{d(\hat{d}+1)}(K)$ in the following way. For every matrix substitution in $\mathcal{H}_{FS}$, define a matrix substitution for each $x_{k'}$ as a $d(\hat{d}+1) \times d(\hat{d}+1)$ matrix which can be thought of as a $d \times d$ block matrix whose $(i, j)^{th}$ block is the matrix substituted for $z_{i,j,k'}$ variable from $\mathcal{H}_{FS}$. It follows that $\mathcal{H}_{FS}$ is a hitting set of B under the inclusion map $a \mapsto a \otimes I_{\hat{d}+1}$.

Remark 28. To explain the purpose of the inclusion map $a \mapsto a \otimes I_{\hat{d}+1}$, we illustrate with a small example. Consider a generalized monomial $a_1 x_1 a_2 x_2 a_3$ where a_1, a_2, a_3 are 2×2 matrices. Now the substitution map ψ replaces the variables x_1, x_2 by 2×2 symbolic matrices over noncommutative Z variables. So the entries of the output 2×2 matrix are noncommutative polynomials over Z variables. Now substituting the Z variables by 3×3 matrices is equivalent to substituting x_1, x_2 by 6×6 matrices putting the 3×3 matrices in the corresponding blocks and evaluating it under the inclusion map that blows up the 2×2 matrices $a_i : 1 \leqslant i \leqslant 3$ to $a_i \otimes I_3$.

Inspecting the proof of Proposition 26, we can now find a subset $T \subseteq \mathbb{Q}$ of size $\text{poly}(n, s)$ and the final quasipolynomial-size hitting set is the following:

$$\mathcal{H}_2 = \{\alpha \underline{p} + \underline{q} \otimes I_{\hat{d}+1} : \underline{p} \in \widetilde{\mathcal{H}}_{FS}, \underline{q} \in \widehat{\mathcal{H}}_1, \alpha \in T\}. \tag{6}$$

Now we discuss how to obtain a hitting set over $\mathbb{Q}$ itself. In the hitting set points suppose we replace ω and z by commuting indeterminates t_1, t_2 of degree bounded by ℓ. Then, for any nonzero rational formula $\mathfrak{r}$ of size s there is a matrix tuple in the hitting set on which $\mathfrak{r}$ evaluates to a nonzero matrix $M(t_1, t_2)$ of dimension $\mathrm{poly}(n, s)$ over the commutative function field $\mathbb{Q}(t_1, t_2)$. By Lemma 25, each entry of $M(t_1, t_2)$ is a rational expression 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 $\mathrm{poly}(n, s)$. Hence by the argument sketched in Proposition 26, we can vary the parameters t_1, t_2 over a sufficiently large set $\widetilde{T} \subseteq \mathbb{Q}$ of size $\mathrm{poly}(n, s)$ such that we avoid the roots of the numerator and denominator polynomials involved in the computation. This gives our final hitting set $\widetilde{\mathcal{H}}_2 = \{\underline{q}'(\alpha_1, \alpha_2) : \underline{q}'(\omega, z) \in \mathcal{H}_2, (\alpha_1, \alpha_2) \in \widetilde{T} \times \widetilde{T}\}$. □

6 Discussion

We have presented a deterministic quasipolynomial-time RIT algorithm for rational formulas of inversion height two in the black-box model by a quasipolynomial size hitting set construction. We briefly discuss a possible approach to obtain a quasipolynomial-size hitting set for constant inversion height formulas. An inspection of the proof of Theorem 1 shows the following.

Let $\widehat{\mathcal{H}_h}$ be a strong hitting set of rational formulas of height h and $\widetilde{H}_{FS}$ be the set of matrix tuples as defined in that proof. Then, for a small set T we can define,

$$\mathcal{H}_{h+1} = \{\alpha \underline{p} + \underline{q} \otimes I_{\hat{d}+1} : \underline{p} \in \widetilde{\mathcal{H}}_{FS}, \underline{q} \in \widehat{\mathcal{H}_h}, \alpha \in T\}.$$

Following the same proof, one can show $\mathcal{H}_{h+1}$ is a hitting set for rational formulas of height $h+1$. Therefore, we can conclude the following.

Lemma 29. *If we have a quasipolynomial-size* strong *hitting set for rational formulas of inversion height h, then we can efficiently construct a hitting set of quasipolynomial-size for rational formulas of inversion height $h+1$ with a small blow-up in the dimension of matrices in the hitting set.*

One way to construct a strong hitting set (say, for rational formulas of inversion height h) is to construct a hitting set $\mathcal{H}$ such that the matrices occurring in each $\underline{p} \in \mathcal{H}$ come from some finite-dimensional division algebra D. We will refer to such hitting sets as division algebra hitting sets. This is the approach we have taken. We can construct such a hitting set for rational formulas of inversion height 0 (essentially for ABPs computing polynomials) which yields a strong hitting set for inversion height 1 rational formulas which, in turn, by Lemma 29 yields a hitting set for inversion height 2 rational formulas.

Indeed, in general, given a quasipolynomial-size construction of a division algebra hitting set for inversion height h formulas, we can construct a strong hitting set for inversion height $h+1$ formulas with a small increase in dimension of the matrices (the proof is along similar lines as the construction in Section 4). If we could ensure that this construction yields not just a strong hitting set but a division algebra hitting set then we will obtain a quasipolynomial-size hitting set construction for all constant inversion height formulas.

We conjecture that it is possible to construct quasipolynomial-size hitting sets for rational formulas of any constant inversion height in a division algebra of polynomially bounded index, and we believe that generalized cyclic division algebras [Jac96] could be useful for the construction. This is a reasonable conjecture because Derksen and Makam's randomized RIT algorithm [DM17]

shows that it suffices to evaluate rational formulas of size s on random $2s \times 2s$ matrices. By Proposition 15, we can ensure that random elements from a cyclic division algebra of small index also suffices for the black-box RIT. Therefore, by a standard counting argument, the existence of even a polynomial-size hitting set inside a cyclic division algebra of polynomial dimension is guaranteed.

Acknowledgement

We thank the anonymous reviewers of an earlier version for valuable comments that have helped us to improve the presentation.

References

[ACDM20] V. Arvind, Abhranil Chatterjee, Rajit Datta, and Partha Mukhopadhyay. A Special Case of Rational Identity Testing and the Brešar-Klep Theorem. In Javier Esparza and Daniel Kráľ, editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[AGKS15] Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Hitting-sets for ROABP and sum of set-multilinear circuits. *SIAM J. Comput.*, 44(3):669–697, 2015.

[Ami66] S.A Amitsur. Rational identities and applications to algebra and geometry. *Journal of Algebra*, 3(3):304 – 359, 1966.

[Ber76] George M Bergman. Rational relations and rational identities in division rings. *Journal of Algebra*, 43(1):252–266, 1976.

[BR11] J. Berstel and C. Reutenauer. *Noncommutative Rational Series with Applications*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2011.

[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.

[Eil74] Samuel Eilenberg. *Automata, Languages, and Machines (Vol A)*. Pure and Applied Mathematics. Academic Press, 1974.

[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.

[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.

[Jac96] Nathan Jacobson. *Finite-Dimensional Division Algebras Over Fields*. Volume 233 of Grundlehren der Mathematischen Wissenschaften Series. Springer, 1996.

[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.

[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.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il