

On Identity Testing and Noncommutative Rank Computation over the Free Skew Field

V. Arvind ✉

Institute of Mathematical Sciences (HBNI), Chennai, India

Abhranil Chatterjee ✉🏠

National Institute of Science Education and Research (HBNI), Bhubaneswar, India

Utsab Ghosal ✉

Chennai Mathematical Institute, India

Partha Mukhopadhyay ✉🏠

Chennai Mathematical Institute, India

C. Ramya ✉

Institute of Mathematical Sciences (HBNI), Chennai, India

Abstract

The identity testing of rational formulas (RIT) in the free skew field efficiently reduces to computing the rank of a matrix whose entries are linear polynomials in noncommuting variables [23]. This rank computation problem has deterministic polynomial-time white-box algorithms [20, 25] and a randomized polynomial-time algorithm in the black-box setting [14]. In this paper, we propose a new approach for efficient derandomization of *black-box* RIT. Additionally, we obtain results for matrix rank computation over the free skew field and construct efficient linear pencil representations for a new class of rational expressions. More precisely, we show:

- Under the hardness assumption that the ABP (algebraic branching program) complexity of every polynomial identity for the $k \times k$ matrix algebra is $2^{\Omega(k)}$ [9], we obtain a subexponential-time black-box RIT algorithm for rational formulas of inversion height *almost* logarithmic in the size of the formula. This can be seen as the first “*hardness implies derandomization*” type theorem for rational formulas.
- We show that the noncommutative rank of any matrix over the free skew field whose entries have *small linear pencil representations* can be computed in deterministic polynomial time. While an efficient rank computation was known for matrices with noncommutative *formulas* as entries [19], we obtain the first deterministic polynomial-time algorithms for rank computation of matrices whose entries are noncommutative ABPs or *rational* formulas.
- Motivated by the definition given by Bergman [7], we define a new class of rational functions where a rational function of inversion height at most h is defined as a composition of a noncommutative r -skewed circuit (equivalently an ABP) with inverses of rational functions of this class of inversion height at most $h - 1$ which are also disjoint. We obtain a polynomial-size linear pencil representation for this class which gives a white-box deterministic polynomial-time identity testing algorithm for the class.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory

Keywords and phrases Algebraic Complexity, Identity Testing, Non-commutative rank

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.6

Related Version *Full Version*: <https://arxiv.org/pdf/2209.04797.pdf>

Funding *Utsab Ghosal*: Partially supported by Infosys Foundation.

Partha Mukhopadhyay: Partially supported by Infosys Foundation.

Acknowledgements We thank the anonymous reviewers for their feedback. This work was done when the second author was a postdoctoral researcher at IIT Bombay.



© V. Arvind, Abhranil Chatterjee, Utsab Ghosal, Partha Mukhopadhyay, and C. Ramya; licensed under Creative Commons License CC-BY 4.0

14th Innovations in Theoretical Computer Science Conference (ITCS 2023).

Editor: Yael Tauman Kalai; Article No. 6; pp. 6:1–6:23



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

In algebraic circuit complexity the basic arithmetic operations are additions, multiplications, and inverses. Using these arithmetic operations algebraic circuits compute either polynomials or rational functions. A subarea of algebraic complexity is *noncommutative computation* where the multiplication of variables is not commutative and the set of monomials (over the variables) form a free monoid. Noncommutative circuits/formulas with only addition and multiplication gates compute noncommutative polynomials in the free algebra.

In the commutative case, inverses are well understood but in noncommutative computation inverses are quite subtle. To elaborate, it is known that *any* commutative rational circuit can be efficiently transformed into the form fg^{-1} where f and g are polynomials that are computed by circuits [32]. However, noncommutative rational expressions 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* then 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 τ , its *domain of definition* $\text{dom}(\tau)$ is the set of all matrix tuples (of any dimension) where τ is defined. Two rational expressions τ_1 and τ_2 are *equivalent* if they agree on $\text{dom}(\tau_1) \cap \text{dom}(\tau_2)$. This defines an equivalence relation on noncommutative rational expressions (with nonempty domains of definition). Amitsur used this in defining the *universal* free skew field (denoted by $\mathbb{F}\langle x \rangle$ when the variable set is $x = \{x_1, x_2, \dots, x_n\}$) and the equivalence classes are called the *noncommutative rational functions* [1]. This object plays an important role in the study of noncommutative algebra [1, 12], control theory [27], and algebraic automata theory [34].

Computationally, rational functions are represented by noncommutative arithmetic circuits or formulas using addition, multiplication, and inverse gates [23]. 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 that the inversion height of a rational formula of size s is bounded by $O(\log s)$ [23]. Hrubeš and Wigderson introduced the *rational identity testing* problem (RIT) of testing the equivalence of two rational formulas [23]. It is the same as testing whether a rational formula computes the zero function in the free skew field. In other words, the problem is to decide whether there is a matrix tuple (of some dimension) such that the rational formula evaluates to nonzero on it. Rational expressions exhibit peculiar properties which seem to make the RIT problem quite different from polynomial identity testing. The apparent lack of *canonical representations* such as the sum of monomials representation for polynomials and the use of nested inverses in noncommutative rational expressions complicate it. 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 [24].

A second characterization of the free skew field elements is due to Cohn [12]. A *linear pencil* L of size s over noncommuting variables $x = \{x_1, \dots, x_n\}$ is an $s \times s$ matrix whose entries are linear forms in 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} . Cohn has shown that for every τ in $\mathbb{F}\langle x \rangle$, there is a linear pencil L such that τ is an entry of the matrix inverse of L . More generally, τ has a *linear pencil representation* of size s , if for vectors $\underline{c}, \underline{b} \in \mathbb{F}^s$ and an $s \times s$ linear pencil L , $\tau = \underline{c}^t L^{-1} \underline{b}$ where \underline{c}^t is the transpose of \underline{c} . Hrubeš and Wigderson give an efficient reduction from RIT to the singularity testing problem of linear pencils [23]. In particular, if τ is a rational formula of size s , they showed that τ has a linear pencil representation L of size at most $2s$ such that τ is defined on a matrix tuple if and only if L is invertible on that tuple [23]. 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 polynomial 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 in the commutative setting [26]. In contrast, the SINGULAR problem in noncommutative setting has deterministic polynomial-time algorithms in the white-box model due to the works of Garg et al. [20] which is based on *operator scaling* and that of Ivanyos et al. [25] which 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 in this area is to design an efficient *deterministic* algorithm for noncommutative SINGULAR problem in the black-box case [20]. The algorithms by Garg et al. [20] and Ivanyos et al. [25] are inherently sequential and we believe that they are unlikely to be helpful for black-box algorithm design. It is well-known [20] that an efficient black-box algorithm (via a hitting set construction) for SINGULAR would generalize the celebrated quasi-NC algorithm for bipartite matching significantly [17]. There is a randomized polynomial-time black-box algorithm for this problem [14].

Even for the RIT problem (which could be easier than the noncommutative SINGULAR problem), there is limited progress towards designing efficient deterministic black-box algorithm. In fact, only very recently a deterministic quasipolynomial-time black-box algorithm for identity testing of rational formulas of *inversion height two* has been designed [3]. It is interesting to note that in the literature of identity testing, the noncommutative SINGULAR problem and the RIT problem stand among rare examples where deterministic polynomial-time white-box algorithms are designed but for the black-box case no deterministic *subexponential-time* algorithm is known.

► **Remark 1.** For noncommutative polynomials computed by polynomial-size arithmetic circuits, efficient randomized polynomial identity testing algorithms are known either for polynomial degree bound or for exponential sparsity bound [9, 5]. In contrast, the complexity of testing the identity of rational circuits is completely open. In fact, even in the white-box setting we do not have a randomized subexponential-time algorithm.

1.1 Derandomization of RIT from the hardness of polynomial identities

In this paper, we propose a new approach to the RIT problem in the black-box case under a suitable hardness assumption, based on the following conjecture due to Bogdanov and Wee [9, Section 6.2].

► **Conjecture 2.** *The ABP complexity (i.e. the minimum size of an algebraic branching program) of a polynomial identity for the $k \times k$ matrix algebra $\mathbb{M}_k(\mathbb{F})$ is $2^{\Omega(k)}$.*

The conjecture implies that ABPs of size s cannot evaluate to zero on all $O(\log s)$ -dimensional matrices. Bogdanov and Wee [9] also observed that if the conjecture holds then there is an $s^{O(\log^2 s)}$ -time black-box PIT for noncommutative ABPs¹ and as supportive evidence showed that the conjecture is indeed true for *normal identities* (of which the *standard identity* is a special case), and the identity of *algebraicity*.

Consider the following variant of the usual hitting set definition.

¹ Independent of the conjecture, Forbes-Shpilka [18] obtained an $s^{O(\log s)}$ -time black-box PIT for noncommutative ABPs.

► **Definition 3.** For a class of rational formulas \mathcal{R} , we say that a hitting set \mathcal{H} is strong if for any nonzero formula $\tau \in \mathcal{R}$, there exists a matrix tuple $\underline{p} \in \mathcal{H}$ such that $\tau(\underline{p})$ is invertible.

In the following theorem, we show an efficient derandomization of RIT assuming Conjecture 2. This can be seen as the first “hardness implies derandomization” type result for rational formulas.

► **Theorem 4.** If Conjecture 2 is true then we can construct a strong hitting set of size $(\text{snh}(\gamma \log s)^{2h+2})^{O(h(\gamma \log s)^{2h+2})}$ for rational formulas τ of size s over n variables and inversion height h in deterministic $(\text{snh}(\gamma \log s)^{2h+2})^{O(h(\gamma \log s)^{2h+2})}$ -time for some constant $\gamma > 1$. This result holds over infinite or sufficiently large finite fields and $h \leq \beta(\log s / \log \log s)$ for any $0 < \beta < 1$.

As a special case for $h = O(1)$, this gives a quasipolynomial-size hitting set. To get a subexponential-size bound 2^{s^δ} on the hitting set, for δ in $(0, 1)$, we can allow $h \leq c_\delta(\log s / \log \log s)$. Here $c_\delta \in (0, 1)$ is a constant that depends on δ .

As already mentioned, the inversion height of size s rational formula is bounded by $O(\log s)$ [23]. Therefore, Theorem 4 solves the RIT problem in an almost general setting.

We believe that the main interesting point about Theorem 4 is that it relates the black-box RIT derandomization that involves handling *nested inverses* with a problem purely for noncommutative polynomials: we can obtain a deterministic black-box RIT algorithm by showing an exponential size lower bound for ABPs computing any polynomial identity for matrix algebras. Over the years such hardness assumptions have proved to be useful in designing deterministic algorithms for problems related to identity testing [21, 26, 6, 15, 11, 29].

Proof Sketch

The first step in proving Theorem 4 is a variable reduction step that shows the identity testing of a rational formula τ of inversion height h can be reduced to the identity testing of another rational formula τ' over $2(h+1)$ variables in a black-box manner. Notice that for noncommutative polynomials (for which $h = 0$), such a reduction is standard and given by $x_i \rightarrow y_0 y_1^i y_0$ where y_0, y_1 are new noncommutative variables. We prove it by induction on h . In fact, we use a stronger inductive hypothesis that roughly says that for every nonzero rational formula τ of inversion height h , there also exists a $2(h+1)$ -tuple of matrices $(q_{00}, \dots, q_{h0}, q_{01}, \dots, q_{h1})$ such that $\tau(p_1, \dots, p_n)$ is invertible and for each $i \in [n]$, $p_i = \sum_{j=0}^h q_{j0} q_{j1}^i q_{j0}$. Once we assume the inductive hypothesis for inversion height $h-1$, for each rational formula τ of inversion height h , we get a matrix tuple of the form $\underline{p} = (p_1, \dots, p_n)$ where $p_i = \sum_{j=0}^{h-1} q_{j0} q_{j1}^i q_{j0}$ such that τ is defined on \underline{p} . Then, using concepts from matrix coefficient realization theory we construct the nonzero generalized series $\tau(\underline{x} + \underline{p})$ [34]. Now, we can use the standard bivariate encoding trick on $\tau(\underline{x} + \underline{p})$ to complete the variable-reduction step.

The next important step that we establish is that if Conjecture 2 is true then for any rational formula τ of size s and inversion height h , one can find a matrix tuple \underline{p} of dimension $(\gamma \log s)^{h+1}$ (for some constant γ) such that $\tau(\underline{p})$ is an invertible matrix. This is done via induction on h and a bootstrapping argument. For the base case, we take $h = 0$. In this case the rational formula is also an ABP of size s and Conjecture 2 confirms that τ is nonzero on a generic matrix tuple \underline{p} of dimension $O(\log s)$. Also $\tau(\underline{p})$ is invertible by an application of Amitsur’s theorem [1]. Inductively we assume that we can find such a matrix tuple \underline{q} of dimension $d_{h-1} \leq (\gamma \log s)^h$ for any rational formula τ of inversion height at most $h-1$

and size at most s . An easy observation shows that given a rational formula τ of inversion height h , τ is defined on such a matrix tuple q . By matrix coefficient realization theory [34] we can construct the nonzero generalized series $\tau(x+q)$ by expanding τ around the point q . Substituting the variables x_1, x_2, \dots, x_n by symbolic generic matrices over noncommuting variables $Z^{(1)}, \dots, Z^{(n)}$ of dimension d_{h-1} , we observe that each entry of the output matrix $\tau(Z+q)$ is a recognizable series computed by a small size algebraic automaton. By a standard result in algebraic automata theory, generally attributed to Schützenberger [16, Corollary 8.3, Page 14], we know that this series is nonzero if and only if it is nonzero when truncated to the degree matching the size of the algebraic automaton. By Conjecture 2, we can infer that the truncated series is nonzero on generic matrices of dimension roughly $\approx \log(sd_{h-1})$. A simple scaling trick shows that the full (infinite)-series is also nonzero on generic matrices of same dimension. This determines the dimension of the generic matrices on which the rational formula τ is nonzero. Moreover the rational formula evaluates to an invertible matrix on generic matrix substitution of that dimension. This is a consequence of Amitsur’s theorem [1].

Once we have these two steps, the rest of the proof is straightforward. Given nonzero τ over the variables x_1, \dots, x_n of height h , we apply the variable reduction step to construct nonzero τ' of height h (and roughly of same size) over $2(h+1)$ variables $\{y_{00}, y_{01}, \dots, y_{h0}, y_{h1}\}$. Now we apply the second step that says that τ' is nonzero (and hence invertible) on generic matrices over Z variables of dimension $(\gamma \log s)^{h+1}$. We also make use of the fact that $\tau'(y)$ has a small-size linear pencil. To construct the final hitting set, we just need to hit two sparse polynomials of sparsity bound roughly $(snh(\gamma \log s)^{2h+2})^{O(h(\gamma \log s)^{2h+2})}$ and this can be done by applying the standard result of sparse polynomial hitting set construction [28].

1.2 Noncommutative rank of matrices over the free skew field

For a matrix $M = (g_{i,j})_{m \times m}$ over the free skew field $\mathbb{F}\langle x \rangle$, its noncommutative rank (denoted by $\text{ncrank}(M)$) is the least positive integer $r \leq m$ such that $M = PQ$ for an $m \times r$ matrix P and an $r \times m$ matrix Q over $\mathbb{F}\langle x \rangle$. This is also called the *inner rank*. If $r = m$, then M is invertible in $\mathbb{F}\langle x \rangle$.

Indeed, a fundamental result of Cohn [13] showed that for any matrix $M = (g_{i,j})_{m \times m}$ over the noncommutative ring $\mathbb{F}\langle x \rangle$ such that $\text{ncrank}(M) = r$, there exists an $m \times r$ matrix P and an $r \times m$ matrix Q over $\mathbb{F}\langle x \rangle$.

As already mentioned, the problem of computing the noncommutative rank of a linear matrix admits deterministic polynomial-time white-box algorithms [20, 25]. If the matrix entries consist of some higher degree terms, one can use Higman’s trick [22] to reduce it to computing rank of a linear matrix. Consider the following well-known example of a 2×2 matrix [19]:

$$\begin{bmatrix} 1 & x \\ y & z + xy \end{bmatrix}.$$

Higman’s trick reduces it to another 3×3 linear matrix preserving the complement of the noncommutative rank in the following way:

$$\begin{bmatrix} 1 & x \\ y & z + xy \end{bmatrix} \mapsto \begin{bmatrix} 1 & x & 0 \\ y & z + xy & 0 \\ 0 & 0 & 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 & x & 0 \\ y & z & x \\ 0 & -y & 1 \end{bmatrix}.$$

However, it would not be efficient in general. In [19, Proposition A.2], the authors showed an effective use of Higman’s trick to efficiently reduce it to the rank computation of a linear matrix when the entries are computed by noncommutative *formulas*.

In this paper, we consider matrix rank computation over the free skew field in a more general setting. In particular, we obtain an efficient reduction to the rank computation of a linear matrix even when the entries are free skew field elements computed by small linear pencils. More precisely, we show the following.

► **Theorem 5.** *Let $M = (g_{i,j})_{m \times m}$ be a matrix such that for each $i, j \in [m]$, $g_{i,j}$ in $\mathbb{F}\langle x_1, \dots, x_n \rangle$ has a linear pencil of size at most s . Then, the noncommutative rank of M can be computed in deterministic $\text{poly}(m, n, s)$ time. Moreover, in deterministic $\text{poly}(m, n, s)$ time, we can output a matrix tuple $\underline{T} = (T_1, \dots, T_n)$ of dimension d such that the matrix the rank of matrix $M(\underline{T})$ is $d \cdot \text{ncrank}(M)$. The field \mathbb{F} could be infinite or sufficiently large finite field.*

We note an earlier result of Garg et al. [19], showing efficient matrix rank computation when the entries are noncommutative polynomials that are computed by formulas. The polynomial-time deterministic algorithm of Theorem 5 is a two-fold strengthening of this: Since noncommutative ABPs have polynomial-size linear pencils [23], the algorithm can compute the rank of matrices with entries that are polynomials computed by ABPs. Moreover, since noncommutative rational formulas also have polynomial-size linear pencils [23], the algorithm computes the rank of matrices whose entries are computed by small rational formulas.

Proof Sketch

The basic principle of our proof is to reduce the problem to the rank computation of a linear matrix. However, there is no clear notion of *degree reduction* for arbitrary elements over the free skew field. This forces us to find a new approach of constructing this linear matrix efficiently that can also handle a matrix of skew field entries as input. The main idea here is to show that the linear pencil representation enjoys the following closure property. Let A be an $m \times m$ generic matrix over m^2 indeterminates. Substituting each indeterminate A_{ij} by a free skew field element that has a linear pencil of size at most s , suppose we obtain the matrix M . We show that there is an efficiently computable linear matrix L such that $\text{ncrank}(L) = m^2s + \text{ncrank}(M)$. Somewhat surprisingly, the construction of L turns out to be relatively simple and elegant.

There are many equivalent notions of noncommutative rank for linear matrices (for example, see [25, 20]). A notion of particular interest is the blow-up definition that is crucial in the algorithm of Ivanyos et al. [25]. The blow-up notion enables to find a matrix tuple on which the maximum rank is achieved. We extend this notion and introduce a blow-up definition for noncommutative rank (denoted by ncrank^*) of matrices with free skew field entries. We show that for any matrix M of free skew field entries, $\text{ncrank}(M) = \text{ncrank}^*(M)$. Introduction of the blow-up definition allows us to find efficiently the matrix tuple \underline{T} of dimension d such that the rank of $M(\underline{T})$ is $d \cdot \text{ncrank}(M)$. One can view the blow-up definition in this case as an extension of the theory developed by Derksen and Makam [14] for the linear case. This extension could be of independent mathematical interest.

1.3 Linear pencil representations for a new class of rational functions

The study of linear pencils seem to be the key in understanding several basic questions in rational function theory [23, 19, 25, 34, 14]. Our main motivation here is to understand the relation between the linear pencil representations of rational functions and their representations using basic arithmetic operations. Let RF, LR, RC denote, respectively, the class of

polynomial-size rational formulas, the class of rational functions that have polynomial-size linear pencil representations, and the class of polynomial-size rational circuits. Hrubeš and Wigderson [23] have shown an exponential size lower bound for rational formulas computing an entry of the inverse of a symbolic matrix. Moreover, they show that each entry of the inverse of a symbolic matrix is computable by a rational circuit of polynomial size. Therefore, the current known relation is $\text{RF} \subset \text{LR} \subseteq \text{RC}$.

Following Bergman [7], a noncommutative rational function $\tau(\underline{x})$ of inversion height at most h can be inductively defined as $\tau(\underline{x}) = f(x_1, \dots, x_n, g_1^{-1}, \dots, g_m^{-1})$, where f is a noncommutative polynomial and $g_1, \dots, g_m \in \mathbb{F}\langle \underline{x} \rangle$ are rational functions of inversion height $\leq h - 1$. Based on this, we define rational r -skewed circuits.

► **Definition 6.** A rational r -skewed circuit of inversion height 0 is a noncommutative r -skewed circuit² which is also a noncommutative ABP. Inductively, we define $\tau(\underline{x}) = f(x_1, \dots, x_n, g_1^{-1}, \dots, g_m^{-1})$ as a rational r -skewed circuit of inversion height at most h if $f(\underline{x}, y_1, \dots, y_m)$ is a noncommutative r -skewed circuit ($m \geq 0$) and for each $i \in [m]$, $g_i(\underline{x})$ is a rational r -skewed circuit of inversion height $\leq h - 1$.

Let R-rSC denote the class of all rational functions computable by polynomial-size rational r -skewed circuits. Inspecting the polynomial size rational circuit for symbolic matrix inverse [23], we notice that each entry of the inverse of a polynomial-size symbolic matrix can indeed be computed by a polynomial-size rational r -skewed circuit. Hence $\text{LR} \subseteq \text{R-rSC}$. What is the exact expressive power of the class LR ? In particular, is it true that $\text{LR} = \text{R-rSC}$? It now suffices to show that $\text{R-rSC} \subseteq \text{LR}$. While we are unable to answer this completely, we show such a containment under additional structural restrictions on the circuit.

► **Definition 7.** An inversely disjoint rational r -skewed circuit of inversion height 0 is a noncommutative r -skewed circuit (which is also an ABP). Inductively, we define $\tau(\underline{x}) = f(x_1, \dots, x_n, g_1^{-1}, \dots, g_m^{-1})$ as an inversely disjoint rational r -skewed circuit of inversion height at most h if $f(\underline{x}, y_1, \dots, y_m)$ is a noncommutative r -skewed circuit ($m \geq 0$) and for each $i \in [m]$, $g_i(\underline{x})$ is a inversely disjoint rational r -skewed circuit of inversion height $\leq h - 1$ and for all $i \neq j$, the circuits of g_i and g_j are disjoint.

Let ID-R-rSC be the class of rational functions computed by polynomial-size *inversely disjoint* r -skewed circuits. This class contains rational formulas, ABPs. We are able to give polynomial-size linear pencil representations for this class.

► **Theorem 8.** Over any field, an inversely disjoint rational r -skewed circuit of size s has a linear pencil representation of size $O(s^2)$ which can be computed in deterministic polynomial time from the given circuit.

This gives the following containment:

$$\text{RF} \subseteq \text{ID-R-rSC} \subseteq \text{LR} \subseteq \text{R-rSC} \subseteq \text{RC},$$

where we know at least one of the first two containment is proper. We do not know any unconditional separation between RF and ID-R-rSC . This question is somewhat similar in spirit to the separation of noncommutative formulas and ABPs which is still open [30, 10, 33]. However, a simple inductive argument shows that functions of inversion height h

² Usually in the literature they are called right-skew circuits. In this paper, we refer to them as *right-skewed circuits* and reserve the word “skew” for *skew fields*.

in ID-R-rSC can be computed by rational formulas of size $s^{O(h \log s)}$. By standard techniques, a noncommutative r-skewed circuit of size s can be computed by a formula of size $s^{O(\log s)}$. Consider an inversely disjoint r-skewed circuit $\tau(\underline{x}, g_1^{-1}, \dots, g_m^{-1})$ where each $g_i \in \text{ID-R-rSC}$ of inversion height $\leq h - 1$ for each $1 \leq i \leq m$. Inductively, each g_i has a rational formula of size $s^{O((h-1) \log s)}$. Therefore, the size of the rational formula computing τ can be at most $s^{O(h \log s)}$. If $h = O(\log s)$, we then have a quasipolynomial-size formula simulation for this class. However, unlike rational formulas [23], it is not clear whether h can be taken as $O(\log s)$ for a general inversely disjoint r-skewed circuit of size s .

Using Theorem 8, the following corollary is obtained by the application of rank computation algorithm in [25]. For the black-box case, we can apply the algorithm in [14]. In the proof of the corollary we also mention how to apply the algorithm in [25] for the black-box case and get an efficient randomized algorithm over finite fields as well.

► **Corollary 9.** *Let \mathbb{F} be infinite or a sufficiently large field. For an inversely disjoint rational r-skewed circuit of size at most s and over n variables, we can decide whether or not it computes zero in $\mathbb{F}\langle\!\langle x \rangle\!\rangle$ in deterministic $\text{poly}(s, n)$ time in white-box, and in randomized $\text{poly}(s, n)$ time in black-box.*

Proof Sketch

The proof is based on a composition lemma that computes an efficient linear pencil for $f(\underline{x}, g_1^{-1}, \dots, g_m^{-1})$ from the linear pencils of $f(\underline{x}, \underline{y})$ and $g_1^{-1}, \dots, g_m^{-1}$. It turns out that the proof of this composition result is more subtle than the usual proofs of the linear pencil constructions for rational formulas [23, 34].

We first elaborate on the composition lemma. Let L be an $s \times s$ linear pencil over x_1, \dots, x_n and y_1, \dots, y_m . Let $f_{i,j} = (L^{-1})_{i,j}$ for $i, j \in [s]$. Let g_1, \dots, g_m be rational functions over x_1, \dots, x_n such that each g_k has a linear pencil L_k of size at most s' . Then we can construct a single linear pencil \tilde{L} of size at most $ms' + m + 2s^2 + s$ in $\text{poly}(s', s, m, n)$ -time such that

$$(\tilde{L}^{-1})_{2s^2 + \hat{s} + i, 2s^2 + \hat{s} + j} = f_{i,j}(\underline{x}, g_1^{-1}, \dots, g_m^{-1}) \quad \text{for } i, j \in [s], \text{ where } \hat{s} = ms' + m.$$

Given a rational function τ computed by an inversely disjoint rational r-skewed circuit of size at most s , we consider the rational function τ^{-1} which is still in the same class (with inversion height increased by one). Using the composition result, we construct a linear pencil of size $O(s^2)$ for τ^{-1} . Notice that $\tau(\underline{x}, \underline{y})$ is a polynomial computed by an ABP or a r-skewed circuit and it has a polynomial-size linear pencil [23]. Using a standard idea, τ^{-1} also has a small linear pencil L which we use as the input to the composition lemma along with the inductively constructed linear pencils for g_1, \dots, g_m .

The final linear pencil \tilde{L} which is the outcome of the composition lemma has the additional property that for any matrix tuple $\tau(p)$, $\tau^{-1}(p)$ is defined if and only if $\tilde{L}(p)$ is invertible. Since $\tau \neq 0$ if and only if τ^{-1} is defined [1], we can now use the algorithm for noncommutative SINGULAR problem [25] on the linear pencil \tilde{L} to check the identity of τ .

Organization

In Section 2, we mainly provide brief background on linear pencils and its connection with the rational identity testing problem, and also present some results in matrix coefficient realization theory. We prove Theorem 4 in Section 3. The proof of Theorem 5 is given in Section 4.1. We give the proof of Theorem 8 in Section 5. We state some open questions in Section 6.

2 Preliminaries

2.1 Linear pencils and rational functions

Let \mathbb{F} be a field. A linear pencil L of size s over noncommuting $\underline{x} = \{x_1, \dots, x_n\}$ variables is a $s \times s$ matrix where each entry is a linear form in \underline{x} . That is, $L = A_0 + \sum_{i=1}^n A_i x_i$ where each A_i in $\mathbb{M}_s(\mathbb{F})$. Evaluation of a linear pencil at a matrix tuple $\underline{p} = (p_1, \dots, p_n)$ in $\mathbb{M}_m^n(\mathbb{F})$ is defined using the Kronecker (tensor) product: L evaluated at \underline{p} is $A_0 \otimes I_m + \sum_{i=1}^n A_i \otimes p_i$.

Given a linear pencil L , the noncommutative SINGULAR problem is to decide whether there is a tuple \underline{p} in $\mathbb{M}_m^n(\mathbb{F})$ of $m \times m$ matrices for some m such that the output matrix L evaluated at \underline{p} is invertible.

A rational function τ in $\mathbb{F}\langle \underline{x} \rangle$ has a linear pencil representation L of size s if $\tau = \underline{c}^t L^{-1} \underline{b}$ for vectors $\underline{c}, \underline{b} \in \mathbb{F}^s$. Following is the re-statement of Proposition 7.1 proved in [23].

► **Proposition 10.** *Let τ be a rational function given by a rational formula of size s . Then τ can be represented $(L^{-1})_{i,j}$ for $i, j \in [s]$ where L is a linear pencil of size at most $2s$. Moreover, τ is nonzero if and only if L is invertible.*

Clearly in the above proposition the choice for $\underline{c}, \underline{b}$ are the indicator vectors e_i and e_j .

We also use the following classical result of Amitsur [1] in this paper.

► **Theorem 11** ([1]). *Let τ be a rational function which is nonzero on $\mathbb{M}_k(\mathbb{F})$ where \mathbb{F} is infinite or any sufficiently large field. Then $\tau(Y_1, \dots, Y_n)$ is an invertible matrix in $\mathbb{M}_k(\mathbb{F}\langle \underline{Y} \rangle)$ where Y_1, \dots, Y_n are generic indeterminate matrices of dimension k .*

► **Remark 12.** Usually Theorem 11 is stated over infinite fields. However it can be adapted over any sufficiently large finite field \mathbb{F} using the techniques in [25]. We briefly discuss it here. For details we refer the reader to ALGORITHM 1 in [25]. Define the field \mathbb{F}' by adjoining a k^{th} root ζ to \mathbb{F} i.e. $\mathbb{F}' = \mathbb{F}[\zeta]$. Let Z_1 and Z be distinct formal variables and consider the function field $\mathbb{F}'(Z_1, Z)$. Construct a $\mathbb{F}'(Z_1, Z)$ -linear basis $\Gamma = \{C_1, \dots, C_{k^2}\}$ of $\mathbb{M}_k(\mathbb{F}'(Z_1, Z))$ such that the $\mathbb{F}'(Z_1, Z^k)$ -linear span of Γ is a central division algebra over $\mathbb{F}'(Z_1, Z^k)$. It can be shown that τ is invertible on a generic linear combination of Γ . Now by a standard argument the generic variables can be fixed from \mathbb{F} (assuming that \mathbb{F} is sufficiently large) to obtain a matrix tuple \underline{T} such that $\tau(\underline{T})$ is invertible. This also implies that $\tau(\underline{Y})$ is invertible where \underline{Y} is a generic matrix tuple of dimension k .

2.2 Algebraic branching programs (ABPs)

► **Definition 13.** *An algebraic branching program (ABP) is a layered directed acyclic graph with one in-degree-0 vertex called source, and one out-degree-0 vertex called sink. Its vertex set is partitioned into layers $0, 1, \dots, d$, with directed edges only between adjacent layers (i to $i+1$). The source and the sink are in layers zero and d , respectively. Each edge is labeled by a linear form over \mathbb{F} in variables $\{x_1, \dots, x_n\}$. The polynomial computed by the ABP is the sum over all source-to-sink directed paths of the product of linear forms that label the edges of the path. The maximum number of nodes in any layer is called the width of the algebraic branching program. The size of the branching program is taken to be the total number of nodes.*

Equivalently, an ABP of width w and d many layers can be defined as an entry of a product of d many linear matrices of size at most w . Therefore, the polynomial f computed by an ABP is of form $(M_1 \cdots M_d)_{i,j}$ for some $i, j \in [w]$.

► **Proposition 14.** *An ABP of size s has a linear pencil of size at most $2s$ from the following construction:*

$$L_f = \begin{bmatrix} I_w & -M_1 & & & & \\ & I_w & -M_2 & & & \\ & & \ddots & \ddots & & \\ & & & I_w & -M_d & \\ & & & & I_w & \end{bmatrix}.$$

The ABP is computed in the upper right corner.

This construction is well-known and also used in [23].

2.3 Matrix Inverse

Let P be a 2×2 block matrix shown below.

$$P = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

where p_1 is invertible and p_2 and p_3 can be any rectangular matrices and $(p_4 - p_3p_1^{-1}p_2)$ is also invertible. Then we note that the inverse of P has the following structure [23].

$$P^{-1} = \begin{bmatrix} p_1^{-1}(I + p_2(p_4 - p_3p_1^{-1}p_2)^{-1}p_3p_1^{-1}) & -p_1^{-1}p_2(p_4 - p_3p_1^{-1}p_2)^{-1} \\ -(p_4 - p_3p_1^{-1}p_2)^{-1}p_3p_1^{-1} & (p_4 - p_3p_1^{-1}p_2)^{-1} \end{bmatrix} \quad (1)$$

If $p_3 = 0$, then P^{-1} has a simpler structure.

$$P^{-1} = \begin{bmatrix} p_1^{-1} & -p_1^{-1}p_2p_4^{-1} \\ 0 & p_4^{-1} \end{bmatrix}. \quad (2)$$

Hrubeš and Wigderson use Equation 1 to compute each entry of the matrix inverse recursively by a small rational circuit.

► **Theorem 15** ([23, Theorem 2.4]). *Each entry of the inverse of an $s \times s$ symbolic matrix is computable by a rational circuit of size $O(s^\omega)$ where ω is the exponent of matrix multiplication.*

► **Remark 16.** We observe that the same construction also yields a polynomial-size rational r-skewed circuit as defined in Definition 6 for the matrix inverse. Inspecting Equation 1, we just need to compute the entries of p_1^{-1} and $(p_4 - p_3p_1^{-1}p_2)^{-1}$ and after that the remaining computation is straightforward. Notice that, in the composition step while replacing each y_i by g_i^{-1} , Definition 6 allows any g_i to be a sub-circuit of some g_j . Therefore, we can reuse the r-skewed circuit computing each entry of p_1^{-1} and follow the same recursive construction to obtain a rational r-skewed circuit of size $O(s^\omega)$.

2.4 Recognizable series

A comprehensive treatment is in the book by Berstel and Reutenauer [8]. We will require the following concepts. Recall that $\mathbb{F}\langle\langle x \rangle\rangle$ is the formal power series ring over a field \mathbb{F} . A series S in $\mathbb{F}\langle\langle x \rangle\rangle$ is *recognizable* if it has the following linear representation: for some integer s , there exists two column vectors $\underline{c}, \underline{b} \in \mathbb{F}^s$ and an $s \times s$ matrix M whose entries are homogeneous linear forms over x_1, \dots, x_n i.e. $\sum_{i=1}^n \alpha_i x_i$ such that $S = \underline{c}^t \left(\sum_{k \geq 0} M^k \right) \underline{b}$. Equivalently, $S = \underline{c}^t (I - M)^{-1} \underline{b}$. We say, S has a representation $(\underline{c}, M, \underline{b})$ of size s^3 .

³ In the language of weighted automata, the matrix M is the transition matrix for the series S .

The following theorem is a basic result in algebraic automata theory, generally attributed to Schützenberger. It has a simple linear algebraic proof [16, Corollary 8.3, Page 145].

► **Theorem 17.** *A recognizable series with representation $(\underline{c}, M, \underline{b})$ of size s is nonzero if and only if $\underline{c}^t \left(\sum_{k \leq s-1} M^k \right) \underline{b}$ is nonzero.*

In this paper, the theorem is used to argue that the truncated series is computable by a small noncommutative ABP, thereby reducing zero-testing of recognizable series to the identity testing of noncommutative ABPs.

2.5 Matrix coefficient realization theory

We do not know any canonical form for noncommutative rational functions. However, if a noncommutative rational function is analytic (or defined) at a matrix point, then (matrix coefficient)-realization theory Volčič [34] offers a power series representation of the noncommutative rational function around that point. This is found useful in automata theory and control theory.

A *generalized word* or a *generalized monomial* in x_1, \dots, x_n over the matrix algebra $\mathbb{M}_m(\mathbb{F})$ allows matrices to interleave between variables. More formally, a generalized word over $\mathbb{M}_m(\mathbb{F})$ is of the form $a_0 x_{k_1} a_2 \cdots a_{d-1} x_{k_d} a_d$, where $a_i \in \mathbb{M}_m(\mathbb{F})$. A generalized polynomial over $\mathbb{M}_m(\mathbb{F})$ is a finite sum of generalized monomials in the ring $\mathbb{M}_m(\mathbb{F})\langle x \rangle$. Similarly, a generalized series over $\mathbb{M}_m(\mathbb{F})$ is an infinite sum of generalized monomials in the ring $\mathbb{M}_m(\mathbb{F})\langle\langle x \rangle\rangle$.

Let $E = \{e_{i,j}, 1 \leq i, j \leq m\}$ be the set of matrix units which forms a linear basis for $\mathbb{M}_m(\mathbb{F})$. A generalized monomial m of degree d over $\mathbb{M}_m(\mathbb{F})$ can be expressed as a linear combination of generalized monomials of the form $e_{i_0, j_0} x_{k_1} e_{i_1, j_1} x_{k_2} \cdots e_{i_{d-1}, j_{d-1}} x_{k_d} e_{i_d, j_d}$ by expressing each matrix a occurring in m as an \mathbb{F} -linear combination in the E -basis. Hence, we can express any generalized series S over $\mathbb{M}_m(\mathbb{F})$ as a sum of generalized monomials over only E and x which we call its canonical representation. We say the series (resp. polynomial) S is identically zero if and only if it is zero under such expansion i.e. the coefficient of each generalized monomial in the canonical representation 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 $k' \geq 1$ [34]. To match the dimension of the coefficient matrices with the matrix substitution, we use an inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{k'm}(\mathbb{F})$, for example, ι can be defined as $\iota(a) = a \otimes I_{k'}$ or $\iota(a) = I_{k'} \otimes a$. We now define the evaluation of a generalized series (resp. polynomial) over $\mathbb{M}_m(\mathbb{F})$ in the following way. Any degree- d generalized word $a_0 x_{k_1} a_1 \cdots a_{d-1} x_{k_d} a_d$ over $\mathbb{M}_m(\mathbb{F})$ on a matrix substitution $(p_1, \dots, p_n) \in \mathbb{M}_{k'm}^n(\mathbb{F})$ 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 $\iota : \mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{k'm}(\mathbb{F})$. In ring theory, all such inclusions are known to be compatible by the Skolem-Noether theorem [31, Theorem 3.1.2]. Therefore, if a series S is zero with respect to some inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{k'm}(\mathbb{F})$, then it must be zero for any such inclusions. The equivalence of the two notions of zeroness follows from the proof of [34, Proposition 3.13].

We now recall the definition of a recognizable generalized series from the same paper.

► **Definition 18.** *A generalized series S in $\mathbb{M}_m(\mathbb{F})\langle\langle x \rangle\rangle$ is recognizable if it has the following linear representation. For some integer s , there exists a row-tuple of matrices $\underline{c} \in (\mathbb{M}_m(\mathbb{F}))^{1 \times s}$, and $\underline{b} \in (\mathbb{M}_m(\mathbb{F}))^{s \times 1}$ and an $s \times s$ matrix M whose entries are homogeneous generalized linear forms over x_1, \dots, x_n i.e. $\sum_{i=1}^n p_i x_i q_i$ where each $p_i, q_i \in \mathbb{M}_m(\mathbb{F})$ such that $S = \underline{c}(I - M)^{-1} \underline{b}$. We say, S has a linear representation $(\underline{c}, M, \underline{b})$ of size s over $\mathbb{M}_m(\mathbb{F})$.*

6:12 On Identity Testing and Noncommutative Rank Computation

In [34], Volčič shows the following result.

► **Theorem 19** ([34, Corollary 5.1, Proposition 3.13]). *Given a noncommutative rational formula τ of size s over x_1, \dots, x_n and a matrix tuple $\underline{p} \in \mathbb{M}_m^n(\mathbb{F})$ in the domain of definition of τ , $\tau(\underline{x} + \underline{p})$ is a recognizable generalized series with a representation of size at most $2s$ over $\mathbb{M}_m(\mathbb{F})$. Additionally, $\tau(\underline{x})$ is zero in the free skew field if and only if $\tau(\underline{x} + \underline{p})$ is zero as a generalized series.*

Proof. For the first part, see Corollary 5.1 and Remark 5.2 of [34].

To see the second part, let $\tau(\underline{x})$ is zero in the free skew field. Then the fact that $\tau(\underline{x} + \underline{p})$ is a zero series follows from Proposition 3.13 of [34]. If $\tau(\underline{x})$ is nonzero in the free skew field, then there exists a matrix tuple $(q_1, \dots, q_n) \in \mathbb{M}_l^n(\mathbb{F})$ such that $\tau(\underline{q})$ is nonzero. W.l.o.g. we can assume $l = k'm$ for some integer k' . Fix an inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{k'm}(\mathbb{F})$. Define a matrix tuple $(q'_1, \dots, q'_n) \in \mathbb{M}_{k'm}^n(\mathbb{F})$ such that $q'_i = q_i - \iota(p_i)$. Therefore, the series $\tau(\underline{x} + \underline{p})$ on (q'_1, \dots, q'_n) evaluates to $\tau(\underline{q})$ under the inclusion map ι , hence nonzero [34, Remark 5.2]. Therefore, $\tau(\underline{x} + \underline{p})$ is also nonzero. ◀

► **Remark 20.** More explicitly we can say the following which is already outlined in [34, Section 5]. For any inclusion map $\iota : \mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{k'm}(\mathbb{F})$

$$\tau(\underline{q} + \iota(\underline{p})) = \iota(\underline{c}) \left(I_{2sk'm} - \sum_{j=1}^n \iota(A^{x_j})(\underline{q}) \right)^{-1} \iota(\underline{b}).$$

We also note down a few basic facts. The following is easy to show and also noted in [34].

► **Fact 21.** *Let $\tau(\underline{x} + \underline{p})$ be a generalized series where \underline{p} consists of matrices in $\mathbb{M}_m(\mathbb{F})$. If we replace each x_i by a generic matrix over noncommuting variables $(y_{j,k}^i)_{1 \leq j, k \leq m}$, then we get a nonzero matrix over the \underline{y} variables. More precisely, the map $\psi(x_i) = (y_{j,k}^i)_{1 \leq j, k \leq m}$ is identity preserving.*

Another easy fact is the following.

► **Fact 22.** *Let $\tau(\underline{x} + \underline{p})$ has a linear representation $\underline{c}(I - M)^{-1}\underline{b}$ of size s . Then each entry of $\tau(\psi(\underline{x}) + \underline{p})$ is a recognizable series with transition matrix $M(\psi(x_1), \dots, \psi(x_n))$ of size sm . More precisely, the $(i, j)^{\text{th}}$ entry of $\tau(\psi(\underline{x}) + \underline{p})$ has a representation $(\underline{c}_i, M(\psi(\underline{x})), \underline{b}_j)$ where \underline{c}_i and \underline{b}_j are the i^{th} row and j^{th} column of \underline{c} and \underline{b} respectively.*

3 Derandomization of RIT from the Hardness of Polynomial Identities

In this section, we present a new approach to derandomize (almost general) RIT efficiently in the black-box setting and prove Theorem 4. Given a noncommutative polynomial $P(x_1, \dots, x_n) \in \mathbb{F}\langle x_1, \dots, x_n \rangle$, there is a well-known trick to reduce the identity testing of P to the identity testing of a bivariate polynomial $P'(y_0, y_1)$ over the noncommuting variables y_0, y_1 by the substitution $x_i \leftarrow y_0 y_1^i y_0$ for $1 \leq i \leq n$.

For a rational formula $\tau(\underline{x})$, such a variable reduction step preserving identity is not immediate. Our first result in this section reduces the identity testing of an n -variate rational formula of inversion height h to the identity testing of a rational formula of inversion height h over $2(h + 1)$ variables. But before that we record a simple fact.

► **Fact 23.** *Given any rational formula τ' of inversion height at most $h - 1$ and size at most s , if we can find a matrix tuple such that τ' is invertible on that matrix tuple, then for a rational formula τ of size at most s and inversion height h , we can find a matrix tuple where τ is defined.*

Proof. Let \mathcal{F} be the collection of all those inverse gates in the formula τ such that for every $g \in \mathcal{F}$, the path from the root to g does not contain any inverse gate. For each $g_i \in \mathcal{F}$, let h_i be the sub-formula input to g_i . Consider the formula $\tau' = h_1 h_2 \cdots h_k$ (where $k = |\mathcal{F}|$) which is of size at most s since for each i and j , h_i and h_j are disjoint. Clearly, τ' is of inversion height at most $h - 1$. So if we find a point \underline{q} such that $\tau'(\underline{q})$ is invertible then τ is defined at that point \underline{q} . \blacktriangleleft

Now we state and prove the variable reduction lemma for rational formulas.

► **Lemma 24.** *Let $\tau(x_1, \dots, x_n)$ be a rational formula of inversion height h . Then, there exists a $2(h + 1)$ variate rational formula τ' of inversion height h over the variables $\{y_{j0}, y_{j1} : 0 \leq j \leq h\}$ such that τ is zero in $\mathbb{F}\langle\langle x \rangle\rangle$ if and only if τ' is zero in $\mathbb{F}\langle\langle y \rangle\rangle$. Moreover, τ' is obtained from τ by substituting x_i by $\sum_{j=0}^h y_{j0} y_{j1}^i y_{j0}$ for $1 \leq i \leq n$.*

Proof. The proof is by induction on the inversion height h . In fact we use a stronger inductive hypothesis: For every nonzero rational formula of inversion height h , there also exists a matrix tuple (p_1, \dots, p_n) and a collection of matrices $\{q_{00}, \dots, q_{h0}, q_{01}, \dots, q_{h1}\}$ such that $\tau(p_1, \dots, p_n)$ is invertible and for each $i \in [n]$, $p_i = \sum_{j=0}^h q_{j0} q_{j1}^i q_{j0}$.

It is true for noncommutative polynomials for which $h = 0$. It is already mentioned that the substitution $x_i \leftarrow y_0 y_1^i y_0$ reduces the identity testing of $P(x)$ to the identity testing of $P'(y_0, y_1)$. Moreover, by Theorem 11, we know that we can find matrices q_0, q_1 such that the bivariate polynomial $P'(q_0, q_1)$ evaluates to an invertible matrix. Since $P(q_0 q_1 q_0, q_0 q_1^2 q_0, \dots, q_0 q_1^n q_0) = P'(q_0, q_1)$, we establish the base case of the induction.

Inductively, suppose that it is true for any formula of inversion height $h - 1$. Now consider a nonzero rational formula $\tau(x_1, \dots, x_n)$ of inversion height h . From the inductive hypothesis and Fact 23, there exists a matrix tuple (p_1, \dots, p_n) and a collection of matrices $\{\tilde{q}_{00}, \dots, \tilde{q}_{(h-1)0}, \tilde{q}_{01}, \dots, \tilde{q}_{(h-1)1}\}$ such that $\tau(p_1, \dots, p_n)$ is defined and for each $i \in [n]$, $p_i = \sum_{j=0}^{h-1} \tilde{q}_{j0} \tilde{q}_{j1}^i \tilde{q}_{j0}$. Let the dimension of each p_i be m . Therefore, $\tau(x + p)$ is also a nonzero generalized series by Theorem 19. Replacing each x_i by $y_0 y_1^i y_0$, we obtain a nonzero bivariate generalized series and suppose it is nonzero for $y_0 = \widehat{q}_{h0}$ and $y_1 = \widehat{q}_{h1}$ of some dimension km for an integer k . Notice from Section 2 that a generalized series is zero if and only if the coefficient of every monomial in the canonical representation is zero. Therefore the bivariate substitution $x_i \rightarrow y_0 y_1^i y_0$ preserves the nonzeroness of a generalized series. Therefore,

$$\tau(\widehat{q}_{h0} \widehat{q}_{h1} \widehat{q}_{h0} + \iota(p_1), \dots, \widehat{q}_{h0} \widehat{q}_{h1}^n \widehat{q}_{h0} + \iota(p_n))$$

is also nonzero. Notice that, $\iota(p_i) = \sum_{j=0}^{h-1} \iota(\tilde{q}_{j0}) (\iota(\tilde{q}_{j1}))^i \iota(\tilde{q}_{j0})$ for the inclusion map ι from $\mathbb{M}_m(\mathbb{F}) \rightarrow \mathbb{M}_{km}(\mathbb{F})$. We can now define τ' substituting each x_i in τ by $\sum_{j=0}^h y_{j0} y_{j1}^i y_{j0}$. Clearly, τ' is nonzero. By Theorem 11, τ' is also invertible for some matrix tuple \underline{q} of same dimension. Hence $\tau(p_1, \dots, p_n)$ is invertible for $p_i = \sum_{j=0}^h q_{j0} q_{j1}^i q_{j0}$. \blacktriangleleft

Next we show that if Conjecture 2 is true then any rational formula of size s and inversion height $h \leq \beta(\log s / \log \log s)$ for $\beta \in (0, 1)$, is nonzero on a matrix tuple of dimension $(\gamma \log s)^{h+1}$ for some constant γ .

► **Lemma 25.** *Let $\tau(x_1, \dots, x_n)$ be a nonzero rational formula of size s and inversion height $h \leq \beta(\log s / \log \log s)$ for any constant $0 < \beta < 1$. Then, Conjecture 2 implies that there is a matrix tuple $(p_1, \dots, p_n) \in \mathbb{M}_m^n(\mathbb{F})$ such that $\tau(p_1, \dots, p_n)$ is invertible and $m = (\gamma \log s)^{h+1}$ for some constant $\gamma > 1$.*

6:14 On Identity Testing and Noncommutative Rank Computation

Proof. The proof is by induction on h . For the base case $h = 0$, Conjecture 2 implies that the noncommutative formula is nonzero on generic $c \log s$ (for some constant c) dimensional matrix tuple (Z_1, \dots, Z_n) where $Z_i = (z_{\ell,k}^{(i)})_{1 \leq \ell, k \leq c \log s}$. Also Theorem 11 says that the formula evaluates to an invertible matrix $M(Z)$ on substituting x_i by Z_i . Now using standard idea, random substitution to the variables in Z_1, \dots, Z_n yields such a matrix tuple.

Inductively assume that we have already proved the dimension bound on the witness of the invertible image for rational formulas of inversion height at most $h - 1$. Let the dimension of the matrices be d_{h-1} . Now given a rational formula \mathfrak{r} of size s and inversion height h , observe that \mathfrak{r} is defined on some $d_{h-1} \times d_{h-1}$ matrix tuple \underline{q} using Fact 23.

Then by Theorem 19, $\mathfrak{r}(\underline{x} + \underline{q})$ can be represented by a recognizable generalized series of size at most $2s$ such that $\mathfrak{r}(\underline{x})$ is nonzero if and only if $\mathfrak{r}(\underline{x} + \underline{q})$ is nonzero. Using Fact 21, apply the ψ map on the variables such that $\psi(x_i)$ substitutes the variable x_i by a matrix of fresh noncommuting variables $z_{j,k}^{(i)}$ for $1 \leq j, k \leq d_{h-1}$.

Using Fact 22, observe that we get a matrix of recognizable series and each such recognizable series can be represented by an automaton of size at most $\hat{s} \leq 2sd_{h-1}$. Since ψ preserves identity, one of such recognizable series will be nonzero. So w.l.o.g, let the series be $S_{1,1}$ computed at $(1, 1)^{th}$ entry is nonzero. Let the transition matrix for $S_{1,1}$ is $M_{1,1}$. Then using Theorem 17, the truncated finite series $\tilde{S}_{1,1} = \underline{c}^t \left(\sum_{k \leq \hat{s}-1} M_{1,1}^k \right) \underline{b}$ is nonzero, which is a noncommutative ABP.

If Conjecture 2 is true then $\tilde{S}_{1,1}$ will be nonvanishing on a matrix tuple \underline{p} of dimension $O(\log \hat{s})$. Now by the following simple scaling trick, we show that the infinite series $S_{1,1}$ is nonzero at a matrix tuple of dimension $c \log \hat{s}$.

▷ **Claim 26.** We can find a matrix tuple \underline{p}' which is a scalar multiple of \underline{p} such that $S_{1,1}(\underline{p}')$ is nonzero.

Proof. Let τ be a commutative variable and consider the matrix tuple,

$$\tau \underline{p} = (\tau p_{1,1}^{\{1\}}, \dots, \tau p_{d_{h-1}, d_{h-1}}^{\{1\}}, \dots, \tau p_{1,1}^{\{n\}}, \dots, \tau p_{d_{h-1}, d_{h-1}}^{\{n\}}).$$

Observe that $M_{1,1}(\tau \underline{p}) = \tau M_{1,1}(\underline{p})$. From the definition of the series $S_{1,1}$,

$$S_{1,1}(\underline{z}) = \tilde{S}_{1,1}(\underline{z}) + \sum_{i \geq \hat{s}} \underline{c}^t M_{1,1}^i \underline{b}.$$

Let d be the dimension of the matrices in the tuple \underline{p} . We now evaluate $S_{1,1}$ at $\tau \underline{p}$ to get the following:

$$S_{1,1}(\tau \underline{p}) = \tilde{S}_{1,1}(\tau \underline{p}) + \sum_{i \geq \hat{s}} \tau^i \cdot ((\underline{c} \otimes I_d)^t \cdot M_{1,1}^i(\underline{p}) \cdot (\underline{b} \otimes I_d)).$$

Since $\tilde{S}_{1,1}(\underline{p}) \neq 0$, we have that $S_{1,1}(\tau \underline{p})$ evaluates to a nonzero matrix whose entries are power series in the variable τ .

It is also true that $S_{1,1}(\tau \underline{p}) = (\underline{c} \otimes I_d)^t \cdot (I - M_{1,1}(\tau \underline{p}))^{-1} \cdot (\underline{b} \otimes I_d)$ which is rational expression in τ where the degrees of the numerator and denominator polynomials are bounded by $\text{poly}(\hat{s}, d)$. Hence we need to avoid only $\text{poly}(\hat{s}, d)$ values for τ such that $S_{1,1}(\tau \underline{p})$ is defined and nonzero. ◁

The above argument shows that for a specific value τ_0 for the parameter τ , the generalized series $\mathfrak{r}(\underline{x} + \underline{q})$ evaluates to nonzero on a matrix tuple $(N_1(\tau_0) + \iota(q_1), \dots, N_n(\tau_0) + \iota(q_n))$ where N_i is obtained from the matrix $(z_{j,k}^{(i)})_{1 \leq j, k \leq d_{h-1}}$ by substituting the variables $(z_{j,k}^{(i)})_{1 \leq j, k \leq d_{h-1}}$ by $\tau_0 p_{j,k}^{(i)}$. Also ι is the inclusion map $\iota : \mathbb{M}_{d_{h-1}}(\mathbb{F}) \rightarrow \mathbb{M}_{dd_{h-1}}(\mathbb{F})$ defined as $\iota(q_i) = q_i \otimes I_d$.

Hence τ is nonzero on generic matrix tuples of dimension $d_h = dd_{h-1} \leq cd_{h-1} \log(sd_{h-1})$. Inductively assume that $d_{h-1} \leq (2c \log s)^h$. Since $h \leq \beta(\log s / \log \log s)$, we can observe that $s \geq d_{h-1}$. Using this we get that $d_h \leq c(2c \log s)^h \log(s^2)$ and that yields $d_h \leq (2c \log s)^{h+1}$. We take $\gamma = 2c$.

Therefore by Theorem 11 $\tau(x)$ evaluates to an invertible matrix on substituting x_i by generic matrices of dimension $(\gamma \log s)^{h+1}$. ◀

Now we are ready to show that if Conjecture 2 is true, then we can find a subexponential-size hitting set for rational formulas of size s and inversion height up to $c'(\log s / \log \log s)$ for a suitable constant c' that depends on the exponent of the subexponential function.

Proof of Theorem 4. Let $\tau(x_1, \dots, x_n)$ be a rational formula of inversion height h and size s . Consider, $\tau'(y_{00}, y_{01}, \dots, y_{h0}, y_{h1})$ obtained from τ by substituting x_i by $\sum_{j=0}^h y_{j0} y_{j1}^i y_{j0}$ for $1 \leq i \leq n$. From Lemma 24, we know that $\tau(x)$ is nonzero if and only if $\tau'(y)$ is nonzero. Moreover, τ' has a rational formula of size at most s' which is of $O(sn)$. Therefore, τ' must be invertible on $d_h \times d_h$ generic matrix substitution where $d_h \leq (\gamma \log s')^{h+1}$ from Lemma 25. Using Proposition 10, we know that τ' has a linear pencil L' of size at most $2s'$. W.l.o.g, assume that τ' is computed at the $(1, 1)^{th}$ entry of L'^{-1} .

Hence, if we substitute the variables $y_{00}, y_{01}, \dots, y_{h0}, y_{h1}$ by $d_h \times d_h$ generic matrices $\{Z^{(i,0)}, Z^{(i,1)} : 0 \leq i \leq h\}$ (over commuting variables), the $(1, 1)^{th}$ block of $L'^{-1}(Z)$ will be of form $\frac{M'(Z)}{\det(L'(Z))}$ where $\det(L'(Z))$ is a polynomial of degree at most $2s'(\gamma \log s')^{h+1}$. Further, each entry of the matrix M' is a cofactor of $L'(Z)$ and therefore it is a polynomial over the Z variables of degree at most $2s'(\gamma \log s')^{h+1}$. This shows that $\det(M'(Z))$ is a nonzero polynomial of degree at most $2s'(\gamma \log s')^{2h+2}$. The sparsity of $\det(L'(Z))$ and $\det(M'(Z))$ are bounded by

$$\kappa = (s'(\gamma \log s)^{2h+2})O(h(\gamma \log s)^{2h+2}).$$

Now we can use standard sparse polynomial hitting set for κ -sparse polynomials to hit both the polynomials [28]. This gives us a strong hitting set \mathcal{H}' for τ' . Consequently, we get a strong hitting set of same size for τ by using the substitutions of x_i variables by the $y_{00}, y_{01}, \dots, y_{h0}, y_{h1}$ described in Lemma 24. More formally, we define

$$\mathcal{H}_{n,h,s} = \{(p_1, \dots, p_n) : \underline{q} \in \mathcal{H}'; p_i = \sum_{j=0}^h q_{j0} q_{j1}^i q_{j0}\}. \quad \blacktriangleleft$$

An immediate corollary is the following.

► **Corollary 27.** *The hitting set size and the construction time is $s^{(\log s)^{O(1)}}$ for $h = O(1)$. If we want to maintain a subexponential-size hitting set of size 2^{s^δ} for $\delta \in (0, 1)$, then h can be taken to be at most $c_\delta \left(\frac{\log s}{\log \log s}\right)$ where c_δ is a constant depending on δ .*

4 Computing the Matrix Rank over the Free Skew Field

In this section, we give an efficient algorithm to compute the rank of any matrix over the free skew field whose entries are rational functions with small linear pencils. Additionally, we output a (witness) matrix tuple on which the rank is achieved. This is done in two steps. We first introduce a blow-up definition for matrix rank over the free skew field extending the results for linear pencils and show that it is equivalent to the usual definition of noncommutative rank. Next, we show an efficient reduction from rank computation over the free skew field to the linear case in Section 4.1. A discussion on the blow-up definition of noncommutative rank can be found in the full version [2].

4.1 The Rank Computation

In this section, we prove Theorem 5. The idea is to reduce rank computation of a matrix with skew field entries from $\mathbb{F}\langle x \rangle$ to rank computation of a linear matrix over x incurring a small blow-up in the matrix size.

► **Lemma 28.** *Let $P \in \mathbb{F}\langle x \rangle^{m \times m}$ such that,*

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where $A \in \mathbb{F}\langle x \rangle^{r \times r}$ is invertible. Then,

$$\text{ncrank}(P) = r + \text{ncrank}(D - CA^{-1}B),$$

Proof. If Q is an $n \times n$ invertible matrix over $\mathbb{F}\langle x \rangle$ then

$$\text{ncrank}(QP) = \text{ncrank}(PQ) = \text{ncrank}(P).$$

For if $P = MN$ then $QP = (QM)N$ and if $QP = MN$ then $P = (Q^{-1}M)N$. Similarly for PQ .

The matrix

$$\begin{bmatrix} A^{-1} & 0 \\ 0 & I_{m-r} \end{bmatrix}$$

is full rank. Similarly, the matrix

$$\begin{bmatrix} I_r & 0 \\ -C & I_{m-r} \end{bmatrix}$$

is full rank because

$$\begin{bmatrix} I_r & 0 \\ -C & I_{m-r} \end{bmatrix} \begin{bmatrix} I_r & 0 \\ C & I_{m-r} \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & I_{m-r} \end{bmatrix}.$$

Hence, $\text{ncrank}(P)$ equals $\text{ncrank}(R)$ where

$$R = \begin{bmatrix} I_r & 0 \\ -C & I_{m-r} \end{bmatrix} \cdot \begin{bmatrix} A^{-1} & 0 \\ 0 & I_{m-r} \end{bmatrix} \cdot \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_r & A^{-1}B \\ 0 & D - CA^{-1}B \end{bmatrix}$$

Post-multiplying by the invertible matrix $\begin{bmatrix} I_r & -A^{-1}B \\ 0 & I_{m-r} \end{bmatrix}$ we obtain $\begin{bmatrix} I_r & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}$.

It is easy to see that its inner rank is $r + \text{ncrank}(D - CA^{-1}B)$. ◀

Next, we relate the noncommutative rank of a matrix with skew field entries with small linear pencils to the noncommutative rank of a linear matrix.

► **Lemma 29.** *Let $M \in \mathbb{F}\langle x \rangle^{m \times m}$ be a matrix whose $(i, j)^{th}$ entry g_{ij} is computed as the $(1, 1)^{th}$ entry of the inverse of a linear pencil L_{ij} of size at most s , for each i and j . Then, one can construct a linear pencil L of size $m^2s + m$ such that,*

$$\text{ncrank}(L) = m^2s + \text{ncrank}(M).$$

Proof. We first describe the construction of the linear pencil L and then argue the correctness. W.l.o.g. we may assume that each linear matrix L_{ij} is $s \times s$ (by padding it, if required, with an identity matrix of suitable size).

$$\text{Let } L = \left[\begin{array}{cccc|c} L_{11} & 0 & \cdots & 0 & B_{11} \\ 0 & L_{12} & \cdots & 0 & B_{12} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & L_{mm} & B_{mm} \\ \hline -C_{11} & -C_{12} & \cdots & -C_{mm} & 0 \end{array} \right], \quad (3)$$

where each C_{ij} is an $m \times s$ and B_{ij} is an $s \times m$ rectangular matrix defined below. Let e_i denote the column vector with 1 in the i^{th} entry and the remaining entries are zero. We define

$$C_{ij} = \left[\begin{array}{c|c|c|c} e_i & 0 & \cdots & 0 \end{array} \right] \quad \text{and,} \quad B_{ij} = \left[\begin{array}{c} \hline e_j^T \\ \hline 0 \\ \hline \vdots \\ \hline 0 \end{array} \right].$$

To argue the correctness of the construction, we write L as a 2×2 block matrix. As each L_{ij} is invertible (otherwise g_{ij} would not be defined), the top-left block entry is invertible. Therefore, we can find two invertible matrices U, V implementing the required row and column operations such that,

$$L = U \left[\begin{array}{cccc|c} L_{11} & 0 & \cdots & 0 & 0 \\ 0 & L_{12} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & L_{mm} & 0 \\ \hline 0 & 0 & \cdots & 0 & \tilde{D} \end{array} \right] V,$$

for some $m \times m$ matrix \tilde{D} .

▷ **Claim 30.** The matrix \tilde{D} is exactly the input matrix M .

Proof. From the 2×2 block decomposition we can write,

$$\tilde{D} = [C_{11}C_{12}\cdots C_{mm}] \left[\begin{array}{cccc} L_{11}^{-1} & 0 & \cdots & 0 \\ 0 & L_{12}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_{mm}^{-1} \end{array} \right] \left[\begin{array}{c} B_{11} \\ B_{12} \\ \vdots \\ B_{mm} \end{array} \right] = \sum_{i,j} C_{ij}L_{ij}^{-1}B_{ij}.$$

Observe that, for each i, j , $C_{ij}L_{ij}^{-1}B_{ij}$ is an $m \times m$ matrix with g_{ij} as the $(i, j)^{\text{th}}$ entry and remaining entries are 0. Hence, $\tilde{D} = M$. ◀

Notice that the top-left block of L in Equation 3 is invertible as for each $i, j \in [m]$, L_{ij} is invertible. Now the proof follows from Lemma 28. ◀

Proof of Theorem 5. For any matrix $M = (g_{i,j})_{m \times m}$ such that for each $i, j \in [m]$, g_{ij} in $\mathbb{F}\langle x_1, \dots, x_n \rangle$ has a linear pencil of size at most s , construct a linear matrix L of size $m^2s + m$ from the previous lemma. We can now compute the noncommutative rank of L

using the algorithm of [25] in deterministic $\text{poly}(s, m, n)$ -time. Let the rank be r . We now output $r - m^2s$ to be the noncommutative rank of M . The correctness of the algorithm follows from Lemma 29.

By the equivalence of the inner rank and blow-up rank, we know that $\text{ncrank}^*(M) = r - m^2s$. Now we use the algorithm in [25] to compute a matrix tuple $\underline{p} \in \mathbb{M}_d(\mathbb{F})$ such that the rank of $L(\underline{p}) = rd$ for some $d = O(m^2s)$. Clearly $\text{rank}(M(\underline{p})) = (r - m^2s)d$. Therefore, the matrix tuple \underline{p} is also a witness of the rank of M . ◀

5 Efficient Linear Pencils for Inversely Disjoint r -Skewed Circuits

We now prove that an inversely disjoint rational r -skewed circuit of size s has a linear pencil representation of size $O(s^2)$. We first prove a more general result, a composition lemma for linear pencils which implies Theorem 8.

► **Lemma 31.** *Let L be an $s \times s$ linear pencil over x_1, \dots, x_n and y_1, \dots, y_m . Let $f_{i,j} = (L^{-1})_{i,j}$ for $i, j \in [s]$. Let g_1, \dots, g_m be rational functions over x_1, \dots, x_n such that each g_k has a linear pencil L_k of size at most s_k . Then we can construct a single linear pencil \tilde{L} of size $\sum_{i=1}^m s_i + m + 2s^2 + s$ in $\text{poly}(s_1, \dots, s_m, s, m, n)$ -time such that*

$$(\tilde{L}^{-1})_{2s^2 + \hat{s} + i, 2s^2 + \hat{s} + j} = f_{i,j}(x, g_1^{-1}, \dots, g_m^{-1}) \quad \text{for } i, j \in [s], \text{ where } \hat{s} = \sum_{i=1}^m s_i + m.$$

Proof. For each $i, j \in [s]$, $f_{i,j} = (L^{-1}(x, y_1, \dots, y_m))_{(i,j)}$. We now define for each $i, j \in [s]$, $h_{i,j} = f_{i,j}(x, g_1^{-1}, \dots, g_m^{-1})$. As the variables y_1, \dots, y_m are indeterminates, we can rewrite each $h_{i,j}$ as the following:

$$h_{i,j} = (L^{-1}(x, g_1^{-1}, \dots, g_m^{-1}))_{(i,j)}.$$

We first describe the construction of the linear pencil \tilde{L} and then prove the correctness of the construction. Let \hat{L} be a linear pencil over x_1, \dots, x_n of size \hat{s} where for each $k \in [m]$, there exists $i_k, j_k \in [\hat{s}]$ such that $g_k^{-1} = (\hat{L}^{-1})_{i_k, j_k}$. The description of \hat{L} is given later.

Let us first define two $s \times s$ linear pencils L' and L'' as follows. Fix $i, j \in [s]$. Let $(L)_{i,j} = \alpha_0 + \sum_{k=1}^n \alpha_{k,i,j} x_k + \sum_{k=1}^m \beta_{k,i,j} y_k$. Write $L = L' + L''$ such that $(L')_{i,j} = \alpha_0 + \sum_{i=k}^n \alpha_{k,i,j} x_k$ and $(L'')_{i,j} = \sum_{k=1}^m \beta_{k,i,j} y_k$. We now define \tilde{L} as a 4×4 block linear matrix of size $\hat{s} + 2s^2 + s$,

$$\tilde{L} = \left[\begin{array}{c|c|c|c} I_{s^2} & A_1 & 0 & 0 \\ \hline 0 & \hat{L} & A_2 & 0 \\ \hline 0 & 0 & I_{s^2} & A_3 \\ \hline A_4 & 0 & 0 & L' \end{array} \right], \quad (4)$$

where I_{s^2} is the identity matrix of size s^2 and A_1, A_2, A_3 and A_4 are some rectangular matrices of dimension $s^2 \times \hat{s}$, $\hat{s} \times s^2$, $s^2 \times s$ and $s \times s^2$ respectively. We now define the construction A_1, A_2, A_3 and A_4 . Subsequently in this proof I is used for I_{s^2} .

$$\text{Let } \tilde{L}_1 = \left[\begin{array}{c|c} I & A_1 \\ \hline 0 & \hat{L} \end{array} \right]. \quad \text{Then } \tilde{L}_1^{-1} = \left[\begin{array}{c|c} I & -A_1 \hat{L}^{-1} \\ \hline 0 & \hat{L}^{-1} \end{array} \right].$$

We now consider the top-left 3×3 block matrix.

$$\text{Let } \tilde{L}_2 = \left[\begin{array}{c|c|c} I & A_1 & 0 \\ \hline 0 & \hat{L} & A_2 \\ \hline 0 & 0 & I \end{array} \right]. \quad \text{Then } \tilde{L}_2^{-1} = \left[\begin{array}{c|c} \tilde{L}_1^{-1} & B_1 \\ \hline 0 & I \end{array} \right],$$

$$\text{where } B_1 = - \begin{bmatrix} I & A_1 \\ 0 & \widehat{L} \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_1 \widehat{L}^{-1} A_2 \\ -\widehat{L}^{-1} A_2 \end{bmatrix}.$$

Define the $s^2 \times s^2$ matrix $A_1 \widehat{L}^{-1} A_2 = B_2$. Recall that, $(L'')_{i,j} = \sum_{k=1}^m \beta_{k,i,j} y_k$. We index the rows of A_1 and columns of A_2 as a pair (i,j) for some $i, j \in [s]$. Define for each $(i,j) \in [s] \times [s]$ and $k \in [m]$, $(A_1)_{(i,j),i_k} = \beta_{k,i,j}$, $(A_2)_{j_k,(i,j)} = 1$ and the other entries are zero. Then,

$$(B_2)_{(i,j),(i,j)} = \sum_{i_k, j_k} (A_1)_{(i,j),i_k} (\widehat{L}^{-1})_{i_k, j_k} (A_2)_{j_k,(i,j)} = \sum_{k=1}^m \beta_{k,i,j} g_k^{-1}.$$

We now define, for each $i, j \in [s]$, $(A_4)_{i,(i,j)} = -1$ and 0 otherwise and $(A_3)_{(i,j),j} = 1$ and 0 otherwise. Since

$$\widetilde{L} = \left[\begin{array}{ccc|c} I & A_1 & 0 & 0 \\ 0 & \widehat{L} & A_2 & 0 \\ 0 & 0 & I & A_3 \\ \hline A_4 & 0 & 0 & L' \end{array} \right], \quad (5)$$

$$\text{Now, } \widetilde{L}^{-1} = \left[\begin{array}{c|c} * & * \\ * & B_3 \end{array} \right] \quad \text{where, } B_3 = \left[L' - (A_4 \ 0 \ 0) \widetilde{L}_2^{-1} \begin{pmatrix} 0 \\ 0 \\ A_3 \end{pmatrix} \right]^{-1}.$$

Simplifying further,

$$B_3 = (L' - A_4 B_2 A_3)^{-1} = L^{-1}(x, g_1^{-1}, \dots, g_k^{-1}).$$

Therefore, for each $i, j \in [s]$, $(B_3)_{i,j} = (L^{-1}(x, g_1^{-1}, \dots, g_k^{-1}))_{i,j} = h_{i,j}$.

Now we construct the linear pencil \widehat{L} of size $\widehat{s} = \sum_{k=1}^m s_k + m$.

For $k \in [m]$, let there are indices $i'_k, j'_k \in [s_k]$ such that $g_k = (L_k^{-1})_{i'_k, j'_k}$. We now define for each $k \in [m]$,

$$\widetilde{L}_k := \left[\begin{array}{c|c} L_k & e_{j'_k} \\ \hline -e_{i'_k}^T & 0 \end{array} \right].$$

Here the vectors e_i are the unit vector. The construction of \widehat{L} is now as follows:

$$\widehat{L} = \begin{bmatrix} \widetilde{L}_1 & 0 & \dots & 0 \\ 0 & \widetilde{L}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widetilde{L}_m \end{bmatrix}. \quad (6)$$

Considering the \widehat{L}^{-1} as an $m \times m$ block matrix where the i^{th} block is of size $s_i + 1$, it is easy to see that for each $k \in [m]$, the bottom-right corner entry of the k^{th} block of \widehat{L}^{-1} is g_k^{-1} . To see this apply Equation 1 with $p_4 = 0$. ◀

Now the proof of Theorem 8 follows easily from Lemma 31.

Proof of Theorem 8. We show that inversely disjoint r-skewed rational functions of height h and of size s have linear pencils of size at most cs^2 for some constant c . We prove it by induction on the inversion height h . For the base case $h = 0$, the input circuit is a noncommutative ABP and the theorem holds by Proposition 14.

Let $f(x, g_1^{-1}, \dots, g_m^{-1})$ be an input inversely disjoint r -skewed rational function of height h computed by the circuit C' . Replacing g_i^{-1} by new variable y_i we get a noncommutative ABP $C'(x, y)$ of size $s' \leq s$. Again by Proposition 14, C' can be represented by a linear pencil of size at most $2s'$. Let g_1, \dots, g_m are computed by inversely disjoint r -skewed circuits of size s_1, \dots, s_m and inversion heights $\leq h - 1$. By the inductive hypothesis each g_k is computable by a linear pencil of size at most cs_k^2 .

Hence by Lemma 31, there is a linear pencil of size S representing $C'(x, g_1^{-1}, \dots, g_m^{-1})$ which satisfies the following condition.

$$S \leq c \sum_{k=1}^m s_k^2 + m + 8s'^2 + 2s'.$$

Simplifying further,

$$S \leq c \left(\sum_{k=1}^m s_k^2 + m + s'^2 \right),$$

for sufficiently large c . Since the sub-circuits for g_1, \dots, g_m are disjoint, we get that $(\sum_{k=1}^m s_k^2 + m + s'^2) \leq (\sum_{k=1}^m s_k + m + s')^2 \leq s^2$. So, $S \leq cs^2$ for some large constant c . ◀

We now prove the following property of the linear pencil constructed in Theorem 8.

► **Proposition 32.** *For any inversely disjoint rational r -skewed circuit computing $\tau \in \mathbb{F}\langle x \rangle$ and a tuple of matrix $\underline{p} \in \mathbb{M}_m^n(\mathbb{F})$ for some finite m , the following are equivalent.*

1. τ is defined at \underline{p} .
2. For every gate u which is an output gate or a child of an inverse gate, the pencil constructed in Theorem 8 corresponding to the rational expression computed at u is invertible at \underline{p} .

The proof of Proposition 32 can be found in the full version [2].

Proof of Corollary 9. Let $\tau(x, g_1^{-1}, \dots, g_m^{-1})$ be the input inversely disjoint r -skewed circuit of size s . By Theorem 8, we construct a linear pencil \tilde{L} of size $O(s^2)$ for τ^{-1} . Now by Proposition 32, τ^{-1} is defined at \underline{p} if and only if $\tilde{L}(\underline{p})$ is invertible. But τ is nonzero if and only if τ^{-1} is defined [1]. So for nonzero testing of τ , it is enough to apply the singularity testing algorithms in [25] on the linear pencil \tilde{L} in white-box case. For the black-box case one can use the algorithm in [14]. In fact the result in [25] also gives the dimension upper bound of $O(s^2)$ for the tensoring matrices on which \tilde{L} should be tested for singularity. This also leads to randomized polynomial-time black-box algorithm that simply substitutes the variables randomly from matrices of dimension $O(s^2)$ over sufficiently large fields. ◀

6 Future Directions

Our work raises the following questions for further research:

- The most important question is to obtain an *unconditional derandomization* of the black-box RIT problem. The current best known result is a quasipolynomial-time black-box RIT algorithm for rational formulas of inversion height at most two [3].
- Theorem 4 opens up a new motivation to further study the Conjecture 2. In [5], it is shown that a nonzero noncommutative polynomial of sparsity s can not be an identity for some $k = O(\log s)$ dimensional matrix algebra. This solves a special case of the conjecture and the proof uses automata theoretic ideas very crucially. Can we improve these techniques to settle the conjecture completely?

- The effective use of Higman’s trick has found new applications in randomized polynomial-time factorization algorithm for noncommutative formulas [4]. The proof of Theorem 5 does not use Higman’s trick. It would be interesting to see whether such ideas can be applied elsewhere.
- Can we exactly characterize (up to a polynomial-size equivalence) the expressive power of linear pencil representations for some sub-class of rational circuits? In this paper, we show that inversely disjoint r-skewed circuits have polynomial-size linear pencils. This gives $ID-R-rSC \subseteq LR$. It would be very interesting to prove that rational r-skewed circuits can be expressed by polynomial-size linear pencils. In other words, prove that $R-rSC = LR$.

References

- 1 S.A Amitsur. Rational identities and applications to algebra and geometry. *Journal of Algebra*, 3(3):304–359, 1966.
- 2 Vikraman Arvind, Abhranil Chatterjee, Utsab Ghosal, Partha Mukhopadhyay, and C. Ramya. On identity testing and noncommutative rank computation over the free skew field. *CoRR*, abs/2209.04797, 2022. doi:10.48550/arXiv.2209.04797.
- 3 Vikraman Arvind, Abhranil Chatterjee, and Partha Mukhopadhyay. Black-box identity testing of noncommutative rational formulas of inversion height two in deterministic quasipolynomial-time. *CoRR*, abs/2202.05693 (to appear in RANDOM 2022), 2022. URL: <https://arxiv.org/abs/2202.05693>.
- 4 Vikraman Arvind and Pushkar S. Joglekar. On efficient noncommutative polynomial factorization via higman linearization. In Shachar Lovett, editor, *37th Computational Complexity Conference, CCC 2022, July 20-23, 2022, Philadelphia, PA, USA*, volume 234 of *LIPICs*, pages 12:1–12:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.CCC.2022.12.
- 5 Vikraman Arvind, Pushkar S. Joglekar, Partha Mukhopadhyay, and S. Raja. Randomized polynomial time identity testing for noncommutative circuits. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 831–841, 2017. doi:10.1145/3055399.3055442.
- 6 Vikraman Arvind, Partha Mukhopadhyay, and Srikanth Srinivasan. New results on noncommutative and commutative polynomial identity testing. *Computational Complexity*, 19(4):521–558, 2010. doi:10.1007/s00037-010-0299-8.
- 7 George M Bergman. Rational relations and rational identities in division rings. *Journal of Algebra*, 43(1):252–266, 1976. URL: <http://www.sciencedirect.com/science/article/pii/0021869376901599>.
- 8 J. Berstel and C. Reutenauer. *Noncommutative Rational Series with Applications*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2011. URL: https://books.google.co.in/books?id=LL8Nhn72I_8C.
- 9 Andrej Bogdanov and Hoeteck Wee. More on noncommutative polynomial identity testing. In *20th Annual IEEE Conference on Computational Complexity (CCC 2005), 11-15 June 2005, San Jose, CA, USA*, pages 92–99, 2005. doi:10.1109/CCC.2005.13.
- 10 Prerona Chatterjee. Separating abps and some structured formulas in the non-commutative setting. In Valentine Kabanets, editor, *36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference)*, volume 200 of *LIPICs*, pages 7:1–7:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.
- 11 Chi-Ning Chou, Mrinal Kumar, and Noam Solomon. Hardness vs randomness for bounded depth arithmetic circuits. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*, volume 102 of *LIPICs*, pages 13:1–13:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPICs.CCC.2018.13.

- 12 P. M. Cohn. The embedding of firs in skew fields. *Proceedings of The London Mathematical Society*, pages 193–213, 1971.
- 13 P. M. Cohn. *Skew Fields: Theory of General Division Rings*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995.
- 14 Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017.
- 15 Zeev Dvir, Amir Shpilka, and Amir Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. Comput.*, 39(4):1279–1293, 2009. doi:10.1137/080735850.
- 16 Samuel Eilenberg. *Automata, Languages, and Machines (Vol A)*. Pure and Applied Mathematics. Academic Press, 1974.
- 17 Stephen A. Fenner, Rohit Gurjar, and Thomas Thierauf. Bipartite perfect matching is in quasi-nc. *SIAM J. Comput.*, 50(3), 2021. doi:10.1137/16M1097870.
- 18 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. doi:10.1109/FOCS.2013.34.
- 19 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Operator scaling: Theory and applications. *Found. Comput. Math.*, 20(2):223–290, 2020. doi:10.1007/s10208-019-09417-z.
- 20 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117, 2016.
- 21 Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In Raymond E. Miller, Seymour Ginsburg, Walter A. Burkhard, and Richard J. Lipton, editors, *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 262–272. ACM, 1980. doi:10.1145/800141.804674.
- 22 Graham Higman. *Units in group rings*. PhD Thesis. Balliol College, 1940.
- 23 Pavel Hrubeš and Avi Wigderson. Non-commutative arithmetic circuits with division. *Theory of Computing*, 11(14):357–393, 2015. URL: <http://www.theoryofcomputing.org/articles/v011a014>.
- 24 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. URL: <http://www.jstor.org/stable/88328>.
- 25 Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *computational complexity*, 27(4):561–593, December 2018.
- 26 Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complex.*, 13(1-2):1–46, 2004.
- 27 Dmitry S. Kaliuzhnyi-Verbovetskyi and Victor Vinnikov. Singularities of rational functions and minimal factorizations: The noncommutative and the commutative setting. *Linear Algebra and its Applications*, 430(4):869–889, 2009. URL: <https://www.sciencedirect.com/science/article/pii/S0024379508003893>.
- 28 Adam R. Klivans and Daniel Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing, STOC '01*, pages 216–223, New York, NY, USA, 2001. ACM.
- 29 Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 804–814. IEEE, 2021. doi:10.1109/FOCS52979.2021.00083.

- 30 Noam Nisan. Lower bounds for non-commutative computation (extended abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 410–418, 1991. doi:10.1145/103418.103462.
- 31 Louis Halle Rowen. *Polynomial identities in ring theory*. Pure and Applied Mathematics. Academic Press, 1980. URL: <http://gen.lib.rus.ec/book/index.php?md5=bde982110d09e6199643e04da0558459>.
- 32 Volker Strassen. Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202, 1973. URL: <http://eudml.org/doc/151394>.
- 33 Sébastien Tavenas, Nutan Limaye, and Srikanth Srinivasan. Set-multilinear and non-commutative formula lower bounds for iterated matrix multiplication. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 416–425. ACM, 2022. doi:10.1145/3519935.3520044.
- 34 Jurij Volčič. Matrix coefficient realization theory of noncommutative rational functions. *Journal of Algebra*, 499:397–437, April 2018.