Constructive non-commutative rank computation is in deterministic polynomial time

Gábor Ivanyos * Youming Qiao † K. V. Subrahmanyam ‡

February 6, 2018

Abstract

We extend the techniques developed in [IQS17] to obtain a deterministic polynomial-time algorithm for computing the non-commutative rank of linear spaces of matrices over any field.

The key new idea that causes a reduction in the time complexity of the algorithm in [IQS17] from exponential time to polynomial time is a reduction procedure that keeps the blow-up parameter small, and there are two methods to implement this idea: the first one is a greedy argument that removes certain rows and columns, and the second one is an efficient algorithmic version of a result of Derksen and Makam [DM17b], who were the first to observe that the blow-up parameter can be controlled. Both methods rely crucially on the regularity lemma from [IQS17]. In this note we improve that lemma by removing a coprime condition there.

1 Introduction

This paper builds on the work reported in our previous paper [IQS17]. In the interest of keeping this paper self contained we introduce the problem again, recall its connections to invariant theory and operator theory, and describe recent progress on this problem including our work, [IQS17], the work of Garg, Gurvits, Oliviera and Wigderson [GGOW16], and that of Derksen and Makam [DM17b]. As a result this introduction overlaps with the introduction in [IQS17]. Readers who are familiar with [IQS17] can skip straight to 1.2 where we describe the new results in this paper.

Let $X = \{x_1, \ldots, x_m\}$ be a set of variables. Given an $n \times n$ matrix T whose entries are homogeneous linear polynomials from $\mathbb{Z}[X]$, determining the rank of T over the rational function field $\mathbb{Q}(X)$ is a fundamental open problem. This problem, denoted $\mathrm{rk}(T)$, was introduced by J. Edmonds [Edm67]. The decision version of this problem, deciding whether T has rank n is known as the Symbolic Determinant Identity Testing problem (SDIT). It is natural to consider the problem over any field $\mathbb{F}$. If $|\mathbb{F}|$ is constant, this problem was shown to be NP-hard [BFS99]. This is not the setting we will be concerned with – we will always assume $|\mathbb{F}|$ to be at least $\Omega(n)$.

When $|\mathbb{F}| \geq 2n$, the Schwartz-Zippel lemma provides a randomized efficient algorithm for SDIT. Devising a deterministic efficient algorithm for this problem has a long history and is of fundamental

*Institute for Computer Science and Control, Hungarian Academy of Sciences , Budapest, Hungary. E-mail: Gabor.Ivanyos@sztaki.mta.hu

†Centre for Quantum Software and Information, University of Technology Sydney, Sydney, Australia. E-mail: jimmyqiao86@gmail.com

‡Chennai Mathematical Institute, Chennai, India. E-mail: kv@cmi.ac.in

importance in complexity theory. In 2003, Kabanets and Impagliazzo [KI04] showed a remarkable connection between deterministic efficient algorithms for SDIT and circuit lower bounds. This endows SDIT with fundamental importance in computational complexity, but the problem still remains hugely open. Improving on the results in [KI04], Carmosino et al [CIKK15] showed that an efficient algorithm for SDIT implies the existence of an explicit multilinear polynomial family such that its graph is computable in NE, but the polynomial family cannot be computed by polynomial-size arithmetic circuits.

It is also natural to consider this problem in the non-commutative setting. The *free skew field* is the non-commutative analogue of the rational function field. We do not define the free skew field in this paper and only point out that the free skew field was first constructed by Amitsur [Ami66], and alternative constructions were given subsequently by Bergman [Ber70], Cohn [Coh85], and Malcolmson [Mal78]. We refer the reader to [HW15] by Hrubeš and Wigderson for a nice introduction to the free skew field from the perspective of algebraic computations. Cohn's books [Coh85, Coh95] serve as a comprehensive introduction to this topic. By the *non-commutative Edmonds problem* we mean the problem of computing the non-commutative rank of T, denoted $\mathrm{ncrk}(T)$, and by the *non-commutative full rank problem* (NCFullRank) we mean the problem of deciding whether $\mathrm{ncrk}(T)$ is full or not. Cohn and Reutenauer [CR99] showed that NCFullRank is in PSPACE.

In order to talk about further progress on $\mathrm{ncrk}(T)$ and NCFullrank we need to describe the various *avatars* of the non-commutative rank. We give four equivalent formulations of the non-commutative rank. We do not give full proofs that these are equivalent formulations since the proofs were already sketched in [IQS17]. We recall some important definitions from [IQS17] needed to describe these formulations.

First some notation. Let $M(n, \mathbb{F})$ denote the linear space of $n \times n$ matrices over $\mathbb{F}$. A linear subspace of $M(n, \mathbb{F})$ is called a *matrix space*. Given T a matrix of linear forms in variables $X = \{x_1, \ldots, x_m\}$ write $T = x_1B_1 + x_2B_2 + \cdots + x_mB_m$, where $B_i \in M(n, \mathbb{F})$. Let $\mathcal{B} := \langle B_1, \ldots, B_m \rangle$, where $\langle \cdot \rangle$ denotes linear span. The rank of $\mathcal{B}$, denoted as $\mathrm{rk}(\mathcal{B})$, is defined as $\max\{\mathrm{rk}(B) \mid B \in \mathcal{B}\}$. We call $\mathcal{B}$ *singular*, if $\mathrm{rk}(\mathcal{B}) < n$. When $|\mathbb{F}| > n$, as we will assume throughout, $\mathrm{rk}(T) = \mathrm{rk}(\mathcal{B})$; this is because when the field size is large enough, the complement of the zero set of a nonzero polynomial is non-empty.

Shrunk subspaces:

Definition 1.1. Given $\mathcal{B} = \langle B_1, \ldots, B_m \rangle \leq M(n, \mathbb{F})$, a subspace $U \leq \mathbb{F}^n$ is called a *c-shrunk subspace* of $\mathcal{B}$ for $c \in \mathbb{N}$, if there exists $W \leq \mathbb{F}^n$, such that $\dim(W) \leq \dim(U) - c$ and for every $B \in \mathcal{B}$, $B(U) \leq W$. U is called a shrunk subspace of $\mathcal{B}$, if it is a c-shrunk subspace for some $c \in \mathbb{Z}^+$.

Cohn showed that the non-commutative rank is not full if and only if there is a shrunk subspace [Coh95]. This was generalized by Fortin and Reutenauer [FR04, Theorem 1], where the authors showed

$$\mathrm{ncrk}(T) = n - \max\{c \in \{0, 1, \ldots, n\} \mid \exists c\text{-shrunk subspace of } \mathcal{B}\}.$$

It follows that the non-commutative rank of the operator T is a property of the matrix space $\mathcal{B}$ and does not depend upon its presentation T. So it is natural to consider the problem of determining the maximum c such that $\mathcal{B}$ has a c-shrunk subspace.

Rank decreasing operator: When the underlying field $\mathbb{F}$ is the field of complex numbers $\mathbb{C}$, given $B_1, \ldots, B_n$, consider the following positive operator P, $P : M(n, \mathbb{C}) \to M(n, \mathbb{C})$, sending $A \to \sum_{i \in [m]} B_i A B_i^\dagger$. For $c \in \mathbb{N}$, the operator P is said to be rank c-decreasing if there exists a positive semidefinite matrix A such that $\mathrm{rk}(A) - \mathrm{rk}(P(A)) = c$. Gurvits[Gur04] considered the problem of determining the maximum c such that P is rank c-decreasing. It can be easily seen that P is rank c-decreasing iff $\mathcal{B}$ has a c-shrunk subspace - it was this formulation of the non-commutative rank which Gurvits was interested, in his attempt to generalize the alternating minimization algorithm of Linial, Samorodnistky and Wigderson [LSW00] for computing the permanent of a matrix. Gurvits proved that his algorithm runs in polynomial time when the commutative and non-commutative ranks of $\mathcal{B}$ coincide.

The null cone for the left right action: Shrunk subspaces also appear naturally in a problem of classical invariant theory. Consider the action of $\mathrm{SL}(n, \mathbb{F}) \times \mathrm{SL}(n, \mathbb{F})$ on $M(n, \mathbb{F})^{\oplus m}$ with (A, C) sending a tuple $(B_1, \ldots, B_m)$ to $(AB_1C^{\mathrm{T}}, \ldots, AB_mC^{\mathrm{T}})$.[1] Index the coordinates of the matrices by variables $(x_{i,j}^k)$, $1 \le k \le m$, $1 \le i, j \le n$. Let $R(n, m) \subseteq \mathbb{F}[x_{i,j}^{(k)}]$ be the $\mathbb{F}$-algebra of polynomials in the variables $x_{i,j}^k$, invariant with respect to this action. In the literature this ring is also called the ring of matrix semi-invariants. The *nullcone* of $R(n, m)$ is locus of m-tuples of matrices where all homogeneous positive-degree polynomials in $R(n, m)$ vanish. The null-cone is the set of points that need to be discarded when one constructs the GIT quotient of the action of $SL(n, \mathbb{C}) \times SL(n, \mathbb{C})$ on m-tuples of matrices. This motivates the question of deciding whether an m-tuple $(B_1, \ldots, B_m)$ is in the nullcone of $R(n, m)$. Burgin and Draisma[BD06] and, independently, Adsul et al [ANS07] showed that an m-tuple of matrices is in the null cone precisely when $\mathcal{B}$ has a shrunk subspace.

It is known that $R(n, m)$ is finitely generated and there is also a good description of the homogenous invariant polynomials, which follows from several independent works, including Derksen and Weyman [DW00], Schofield and Van den Bergh [SVdB01], Domkos and Zubkov [DZ01], and Adsul et al [ANS07]. Invariants exists only in degrees nd, as d runs over all positive integers. To obtain invariants of degree nd take matrices $A_1, \ldots, A_m \in M(d, \mathbb{F})$. Then $\det(A_1 \otimes X_1 + \cdots + A_m \otimes X_m)$ is a matrix semi-invariant, and every matrix semi-invariant of degree nd is a linear combination of such polynomials. Therefore $(B_1, \ldots, B_m)$ is in the nullcone if and only if, for all $d \in \mathbb{Z}^+$ and all $(A_1, \ldots, A_m) \in M(d, \mathbb{F})^{\oplus m}$, $A_1 \otimes B_1 + \cdots + A_m \otimes B_m$ is singular. This motivates the following definition and leads us to the last formulation of the non-commutative rank.

Blow-ups:

Definition 1.2. Given $\mathcal{B} = \langle B_1, \ldots, B_m \rangle \le M(n, \mathbb{F})$, the dth tensor blow-up of $\mathcal{B}$ is defined to be $\mathcal{B}^{[d]} := M(d, \mathbb{F}) \otimes \mathcal{B} \le M(dn, \mathbb{F})$, the linear span of matrices $A_1 \otimes B_1 + \cdots + A_m \otimes B_m$, with $A_i \in M(d, \mathbb{F})$.

It is clear that $\mathrm{rk}(\mathcal{B}^{[d]}) \ge d \cdot \mathrm{rk}(\mathcal{B})$. Furthermore, if $\mathcal{B}$ has no shrunk subspace, then there is some d for which $rk(\mathcal{B}^{[d]}) = nd$; this follows from the descriptions of the nullcone and the invariants of the left right action. Hence NCFullRank is equivalent to deciding whether $\mathrm{rk}(\mathcal{B}^{[d]}) = nd$ for some d. This was also shown by Hrubeš and Wigderson [HW15]. Hrubeš and Wigderson's interest in knowing whether the non-commutative rank of a matrix family is full, was motivated by their study of non-commutative arithmetic formulas *with divisions*. In [IQS17] we showed that when the field

[1] This action can also be written as: (A, C) sending $(B_1, \ldots, B_m)$ to $(AB_1C^{-1}, \ldots, AB_mC^{-1})$.

size $|\mathbb{F}|$ is large then d divides $\mathrm{rk}(\mathcal{B}^{[d]})$. We refer to this as the regularity lemma, and defer the exact statement to a later point (4.1 in 4).

So, when $|\mathbb{F}|$ is large, we can define the *non-commutative rank* of $\mathcal{B}$ to be the maximum over d of $\frac{1}{d}$ times the maximum rank of a matrix from the blow-up $\mathcal{B}^{\{d\}}$.

From the last formulation above, an important question is to determine bounds on the blow-up parameter d (as a function of n) which achieves the desired maximum. We define $\sigma(R(n,m))$ to be the smallest $d \in \mathbb{N}$, such that those non-constant homogeneous invariants of degree $\leq d$ define the nullcone of $R(n,m)$. From the work of Derksen [Der01] it follows that $\sigma(R(n,m)) \leq O(n^4 \cdot 4^{n^2})$, over algebraically closed fields of characteristic zero.[2] In [IQS17] we showed that $\sigma(R(n,m)) \leq 2^{O(n \log n)}$ over large fields of arbitrary characteristic. We also gave an algorithm to compute $\mathrm{ncrk}(T)$ and output a witnessing shrunk subspace with running time $2^{O(n \log n)}$ over large fields.

We describe this algorithm in the next section. After that we describe further progress on the non-commutative rank from the works of Garg et al [GGOW16] and Derksen and Makam [DM17b]. We then state the main theorem of the paper.

1.1 Outline of the algorithm in [IQS17]

The algorithm in [IQS17] can be viewed as an analogue of the augmenting path algorithm for the bipartite maximum matching problem. However, due to the failure of the analogue of Hall's marriage theorem in the matrix space setting, there are a couple of new and sophisticated components.

Let us briefly review some features of the augmenting path algorithm. Given a matching T for the input bipartite graph $G = (L \cup R, E)$, the algorithm tries to find an augmenting path for T. If an augmenting path is found, T is replaced by a larger matching T'. If no augmenting paths can be found, the algorithm can output a shrunk subset as the certificate of the maximality of T.

We hope to implement the above idea for the non-commutative rank problem. Given a matrix $A \in \mathcal{B} = \mathrm{span}(B_1, \ldots, B_m) \leq M(n, \mathbb{F})$, we would like to either find an "augmenting path" for it and increase its rank, or output a c-shrunk subspace where $c = \mathrm{cork}(A)$.

A linear algebraic analogue of augmenting paths was developed in [IKQS15]. Given a subspace $U \leq \mathbb{F}^n$, let $A^{-1}(U)$ be the preimage of U under A, namely the subspace $\{v \in \mathbb{F}^n : A(v) \in U\}$. We also define $\mathcal{B}(U) := \mathrm{span}(\cup_{i \in [m]} B_i(U))$. Given $A \in \mathcal{B} \leq M(n, \mathbb{F})$, we apply $\mathcal{B}$ and A^{-1} iteratively to $V_0 = \ker(A)$, to get $W_1 = \mathcal{B}(V_0)$, $V_1 = A^{-1}(W_1)$, $W_2 = \mathcal{B}(V_1)$, $\ldots$, $V_i = A^{-1}(W_i)$, $W_{i+1} = \mathcal{B}(V_i)$, $\ldots$. It can be shown that for some $\ell \in [n]$, $W_1 < W_2 < \cdots < W_\ell = W_{\ell+1} = \ldots$. This sequence of subspaces is called *the second Wong sequence* of $(A, \mathcal{B})$. [3] W_ℓ is called the *limit subspace* of this sequence. We state as a fact the following important lemma from [IKQS15].

Fact 1.3 ([IKQS15, Lemmas 9 and 10])**.** *Let $A \in \mathcal{B} \leq M(n, \mathbb{F})$, and let W^* be the limit of the second Wong sequence of $(A, \mathcal{B})$. Then there exists a $\mathrm{cork}(A)$-shrunk subspace of $\mathcal{B}$ if and only if $W^* \leq \mathrm{im}(A)$.* [4] *If this is the case then $A^{-1}(W^*)$ is a $\mathrm{cork}(A)$-shrunk subspace of $\mathcal{B}$. In the algebraic RAM model, as well as over $\mathbb{Q}$, we can detect whether $W^* \subseteq \mathrm{im}(A)$, and in that case we can compute a shrunk subspace in deterministic polynomial time.*

[2] Derksen's result applies to a wide class of invariant rings.

[3] The first Wong sequence is the dual of the second one. The sequences are named after Wong who defined them in [Won74] for the special case when $\mathcal{B}$ is of dimension 1. Over $\mathbb{Q}$ the straightforward implementation of the second Wong sequence may lead to a bit size explosion. To avoid that some tricks are needed. See [IKQS15] for more details.

[4] At the time of writing the first version of [IKQS15], the authors were unaware of [FR04] where this had already appeared.

Therefore when $W_\ell \leq \mathrm{im}(A)$, we can conclude that the non-commutative rank is $\mathrm{rk}(A)$. On the other hand, when $W_\ell \not\leq \mathrm{im}(A)$, following the bipartite maximum matching algorithm it seems natural to try to obtain $A' \in \mathcal{B}$ with $\mathrm{rk}(A') > \mathrm{rk}(A)$. However this is not always possible, as it can be the case that $\mathrm{rk}(A) = \mathrm{crk}(\mathcal{B})$ and $\mathrm{crk}(\mathcal{B}) < \mathrm{ncrk}(\mathcal{B})$. But for a matrix space $\mathcal{B}$ of dimension 2, $\mathrm{rk}(\mathcal{B}) = \mathrm{ncrk}(\mathcal{B})$ for large enough $\mathbb{F}$; this follows from the Kronecker-Weierstrass theory of matrix pencils – alternate proofs may be found in [EH88, AL81].

The key observation in [IKQS15] was that, in certain special cases, when $W_\ell \not\leq \mathrm{im}(A)$ the second Wong sequence could be used to find an "augmenting" matrix B from $\mathcal{B}$ such that $\mathrm{rk}(\mu A + \lambda B) > \mathrm{rk}(A)$ for some scalars λ and μ. This included the case of two-dimensional matrix spaces. The authors showed

Fact 1.4 ([IKQS15, Fact 11])**.** *Assume that* $|\mathbb{F}| > n$*, and let* $\mathcal{B} = \langle A, B\rangle \leq M(n, \mathbb{F})$*. Then* $\mathrm{rk}(A) = \mathrm{rk}(\mathcal{B})$ *if and only if for any* $i \in [n]$*,* $(\mathcal{B}A^{-1})^i(\mathbf{0}) \leq \mathrm{im}(A)$*.*

The key idea in [IQS17] is to reduce the general problem to the rank two situation. The idea is to find $A' \in \mathcal{B}^{\{d\}}$ of rank $\geq (r+1)d$ with some not too large d (so that the scaled-down rank $\mathrm{rk}(A')/d$ is larger than r), and iterating this procedure. We give the key steps of that algorithm.

A: Incrementing the scaled-down rank. This is achieved in two steps.

1 Incrementing rank: The first step is to obtain a matrix $\widehat{A} \in \mathcal{B}^{\{d\}}$ of rank $\geq rd + 1$ where $d = r + 1$. To see how this step works, notice first that by multiplying A and $\mathcal{B}$ with an appropriate matrix, one can arrange A to be idempotent. In that case, as long as $W_1, \ldots, W_{j-1}$ remain inside $\mathrm{im}(A)$, we have $W_j = \mathcal{B}^j \ker(A)$. Let l be the smallest index j with $W_j \not\leq \mathrm{im}(A)$. Then $l \leq r + 1$. Then there exist matrices $B_1, B_2, \ldots, B_l$ such that $B_l \cdots B_1 \ker(A) \not\leq \mathrm{im}(A)$. It would be nice if one could find a *single* matrix $B \in \mathcal{B}$ such that $B^l \ker(A) \not\leq \mathrm{im}(A)$: indeed if this happens then for some λ and μ from a subset of the base field of size at least $r + 1$ one would have for $\widehat{A} = \mu A + \lambda B$, $\mathrm{rk}(\widehat{A}) > \mathrm{rk}(A)$. This follows from Fact 1.4.

The main ingredient of the algorithm in [IKQS15] was a method to find such a $B \in \mathcal{B}$ in certain special cases. The idea in [IQS17] is that, if we relax ourselves to work with $\mathcal{B}^{\{d\}}$, then this can be achieved for every matrix space $\mathcal{B}$.

2 Rounding up the rank: For the second step, starting with $\widehat{A}$, we wish to get the desired $A' \in \mathcal{B}^{\{d\}}$ of rank $\geq (r+1)d$. This is accomplished in [IQS17] by the regularity lemma. An efficient, constructive version of this lemma is required in the algorithm. And to accomplish this we need an efficient construction of central division algebras of degree d^2 over $\mathbb{F}$ with an explicit matrix representation of such a division algebra. In [IQS17] we were able to construct explicit division algebras when the characteristic of $\mathbb{F}$ and d are coprime.

We reproduce the constructive regularity lemma from [IQS17] below.

Lemma 5.7 in [IQS17] (Regularity of blow-ups, constructive). For $\mathcal{B} \leq M(n, \mathbb{F})$ and $\mathcal{A} = \mathcal{B}^{\{d\}}$, assume that $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) \nmid d$, and $|\mathbb{F}| > (nd)^{\Omega(1)}$. Then, given a matrix $A \in \mathcal{A}$ with $\mathrm{rk}A > rd$, there exists a deterministic algorithm that returns $\widetilde{A} \in \mathcal{A}$ of rank $\geq (r+1)d$. This algorithm uses $\mathrm{poly}(nd)$ arithmetic operations and over $\mathbb{Q}$, all intermediate numbers have bit lengths polynomial in the input size.

This $A' \in \mathcal{B}^{\{d\}}$ of rank $\geq (r+1)d$ where $d = r + 1$ certifies that $\mathrm{ncrk}(\mathcal{B}) \geq r + 1$. (From the viewpoint of shrunk subspaces, it is easy to see that $\mathrm{ncrk}(\mathcal{B}) \leq r$ then $\mathrm{crk}(\mathcal{B}^{\{d\}}) \leq rd$ for any

d; see e.g. [IQS17, Proposition 5.2].) So after these two steps we obtain A' of rank $r'd$ where $r' > r$.

B: Iterating over. In the next phase, we need to use A' and $\mathcal{B}^{\{d\}}$ to restart the above procedure, hoping either to find a $\mathrm{cork}(A')$-shrunk subspace, or to obtain some A'' in $\mathcal{B}^{\{dd'\}}$ of rank $r''dd'$ where $r'' > r'$. We then apply the second Wong sequence to work with the blow-up space $\mathcal{B}^{\{d\}}$ and A'.[5] If $\mathrm{cork}(A')$-shrunk subspace U' is found for $\mathcal{B}^{\{d\}}$, then this naturally induces a $\mathrm{cork}(A')/d$-shrunk subspace U for $\mathcal{B}$ [IQS17, Proposition 5.2]. In this case we conclude that the non-commutative rank is r', and A' and U together serve as witnesses for this fact. If the limit subspace goes out of $\mathrm{im}(A')$ we need to go to an even larger blow-up space $(\mathcal{B}^{\{d\}})^{\{d'\}} \cong \mathcal{B}^{\{dd'\}}$ where $d' = r' + 1$, to find a matrix $A'' \in \mathcal{B}^{\{dd'\}}$ of rank $r''dd'$ for some $r'' > r'$.

We reproduce the following theorem from [IQS17] which summarizes the above discussion.

Theorem 5.10 in [IQS17]. Let $\mathcal{B} \leq M(n, \mathbb{F})$ and $\mathcal{A} = \mathcal{B}^{\{d\}}$. Assume that we are given a matrix $A \in \mathcal{A}$ with $\mathrm{rk}(A) = rd$. Let d' be an integer $> r$. Suppose that $|\mathbb{F}|$ is $(ndd')^{\Omega(1)}$, and if $\mathrm{char}(\mathbb{F}) = p > 0$ then assume $p \nmid dd'$. There exists a deterministic algorithm that returns either an $(n-r)d$-shrunk subspace for $\mathcal{A}$ (equivalently, an $(n-r)$-shrunk subspace for $\mathcal{B}$), or a matrix $A^* \in \mathcal{A} \otimes M(d', \mathbb{F})$ of rank at least $(r+1)dd'$. This algorithm uses $\mathrm{poly}(ndd')$ arithmetic operations and, over $\mathbb{Q}$, all intermediate numbers have bit lengths polynomial in the input size.

The main point is that to carry out the augmenting path idea for the bipartite maximum matching problem in the non-commutative rank setting, the right approach is to play with shrunk subspaces on the one hand, and matrices in the blow-up spaces on the other.

The alert reader may now notice that the above strategy leads to an exponential-time algorithm. Recall that we start with $A \in \mathcal{B}$ of rank r. If $\mathrm{ncrk}(\mathcal{B}) = n$, then we may end up finding $A^* \in \mathcal{B}^{\{d^*\}}$ of rank nd^* where d^* can be as large as $n!/r!$. This is because, increasing the scaled-down rank from r' to $r'+1$ would lead to a multiplicative factor of $r'+1$ in the size of the blow-up space. This is why the algorithm in [IQS17] runs in time $\mathrm{poly}(n!)$. We reproduce that result below.

Theorem 5.11 in [IQS17]. Suppose we are given $\mathcal{B} := \langle B_1, \ldots, B_m \rangle \leq M(n, \mathbb{F})$, and $A \in \mathcal{B}$ with $\mathrm{rk}(A) = s < n$. Let $d = (n+1)!/(s+1)!$, and assume that $|\mathbb{F}| = \Omega(nd)$. Then there exists a deterministic algorithm, that computes a matrix $B \in \mathcal{B} \otimes M(d', \mathbb{F})$ of rank rd' for some $d' \leq d$ and, if $r < n$, an $(n-r)$-shrunk subspace for $\mathcal{B}$. The algorithm uses $\mathrm{poly}(n, d)$ arithmetic operations, and when working over $\mathbb{Q}$, has bit complexity polynomial in n, d and the input size.

1.2 Progress on non-commutative rank since 2015.

Recall that an important question was to upper bound $\sigma(R(n,m))$, and exponential bounds were established in [Der01] and [IQS17]. These turned out to be sufficient for [GGOW16] to compute the non-commutative rank in deterministic polynomial time, over fields of characteristic zero, by a more refined analysis of Gurvits' algorithm in [Gur04]. After [GGOW16], the following problems were still open:

(1) a polynomial-time algorithm for the problem over finite fields, and

[5] When the second Wong sequence is applied to such blow-up spaces then it has some nice properties; cf. the proof for Theorem 5.10 in [IQS17].

(2) a search version of the problem, that is, explicitly exhibiting a matrix of rank rd in the d-th blow-up and a proof that the non-commutative rank is at most r, even over fields of characteristic 0.

Recently, Derksen and Makam[DM17b] proved that it suffices to take the maximum over d between 1 and $n-1$, for sufficiently large fields, by discovering a concavity property of blow-ups, and using the regularity lemma of blow-ups from [IQS17]. In the first version of this note, by showing that the concavity property can be made constructive, and building on the techniques from [IQS17], we obtained a deterministic polynomial-time algorithm for the non-commutative rank problem, which is constructive and works over large enough fields regardless of the characteristic. This answers the two open problems just mentioned.

After the first version of this note appeared on the arXiv, we discovered that a very simple observation already gives us the result, without having to use the results from Derksen and Makam. This argument also gives a different proof that the nullcone of the matrix semi-invariants is generated by polynomials in $R(n,m)$ of degree less than or equal to $O(n^2)$. We should point out that recently Derksen and Makam [DM17a] also gave a second proof of the regularity lemma. However their proof is not known to be constructive.

We now state our main result and the contributions of this paper.

Theorem 1.5. *Let $\mathcal{B} \leq M(n,\mathbb{F})$ be a matrix space given by a linear basis, and suppose $|\mathbb{F}| = n^{\Omega(1)}$. Suppose that $\mathcal{B}$ has (a priori unknown) non-commutative rank r. Then there is a deterministic algorithm using $n^{O(1)}$ arithmetic operations over $\mathbb{F}$ that constructs a matrix of rank rd in a blow-up $\mathcal{B}^{\{d\}}$ for some $d \leq r+1$ as well as an $(n-r)$-shrunk subspace of $\mathbb{F}^n$ for $\mathcal{B}$. When $\mathbb{F} = \mathbb{Q}$, the final data as well as all the intermediate data have size polynomial in the size of the input data and hence the algorithm runs in polynomial time.*

Compared with the algorithm in [GGOW16], our algorithm has the advantages of (1) working with arbitrary large enough fields, and (2) outputting a shrunk subspace and a matrix in a blow-up space certifying that the non-commutative rank is r. Note that the second feature is new even over $\mathbb{Q}$. We also show that the small finite fields case can be handled as well.

Remark 1.6. (a) If the constructivized version of Derksen and Makam [DM17b] is used, then in the above theorem we can improve the parameter slightly to $d \leq r-1$ instead of $d \leq r+1$.

(b) Polynomial running time of the algorithm can also be proved for a wide range "concrete" base fields $\mathbb{F}$. These include sufficiently large finite fields, and also number fields and transcendental extensions of constant degree over finite fields and over number fields.

(c) In particular, the non-commutative rank can be computed in deterministic polynomial time in positive characteristic as well, assuming that the ground field is sufficiently large.

Our result also settles a question of Gurvits [Gur04], asking if it is possible to decide efficiently, over fields of positive characteristic, whether or not there exists a non-singular matrix in a matrix space having the Edmonds-Rado property. Recall that a matrix space has the *Edmonds-Rado property* if it satisfies the promise that it either contains a non-singular matrix, or it shrinks some subspace. Since the algorithm in 1.5 efficiently tells whether the given matrix space has a shrunk subspace (e.g. the non-commutative rank is not full), it settles Gurvits' question, when the field size is as stated in the hypothesis.

Over small finite fields. From the above, we have seen a polynomial upper bound on $\sigma(R(n,m))$, and settled the non-commutative rank problem as well as SDIT for the Edmonds-Rado class, provided that the underlying field is large enough. However we can say more, even when the base field is a "too small" finite field.

Corollary 1.7. *Let $\mathbb{F}$ be a finite field of size $s < n^{O(1)}$.*

1. *Let $R(n,m)$ be the ring of matrix semi-invariants over $\mathbb{F}$. Then $\sigma(R(n,m)) \leq O((n^2 - n)\log_s n)$.*
2. *Let $\mathcal{B} \leq M(n,\mathbb{F})$ be a matrix space given by a linear basis with a priori unknown non-commutative rank r. There is a deterministic polynomial-time algorithm that constructs a matrix of rank rd in a blow-up $\mathcal{B}^{\{d\}}$ for some $d \leq O(r\log_s n)$, as well as an $(n-r)$-shrunk subspace of $\mathbb{F}^n$ for $\mathcal{B}$.*
3. *Let $\mathcal{B} \leq M(n,\mathbb{F})$ be a matrix space given by a linear basis satisfying the Edmonds-Rado property. Then there exists a deterministic polynomial-time algorithm that can decide whether $\mathcal{B}$ has a non-singular matrix, or a shrunk-subspace.*

Techniques. As described in the the iterating over step in Section1.1, the algorithm in [IQS17] takes exponential time because we increase the blow-up size in an iterative way, and in each iteration the blow-up size is increased multiplicatively by the "scaled" rank. The key new insight is that we can keep the blow-up size small: when the scaled rank is r, then the blow-up size can be brought back to $O(r)$. As mentioned, we offer two methods to realize this reduction idea: a simpler method from us, and a method based on the technique of Derksen and Makam [DM17b].

We also provide a technical improvement to the constructive regularity lemma used in the rounding up the rank step of the algorithm described in 1.1. Recall that we use it in the algorithm in the following situation: given $A \in \mathcal{B} \otimes M(d,\mathbb{F})$ of rank $(r-1)d+k$ where $1 < k < d$, we want to construct $A' \in \mathcal{B} \otimes M(d,\mathbb{F})$ of rank $\geq rd$ efficiently. This was achieved under the condition that, if $\mathrm{char}(\mathbb{F}) = p > 0$, then p and d are coprime. In this note, we remove this coprime condition.

Organization. In Section 2 we first discuss algorithmic issues that arise when working over finite extensions of fields and how they are solved. Since all this appears with detailed proofs in our previous paper we only provide pointers to these issues and refer to [IQS17] for details. In Section 3 we give an efficient construction of cyclic field extensions of arbitrary degrees. In Section 4 we use this to prove the full regularity lemma. In Section 5.1 we prove the main Theorem 1.5 using our blow-up reduction method. In Section 5.2 we give the proof for Corollaryr̃efcor:small. Finally in Section 6 we show that the Derksen–Makam technique can be constructivized to provide another blow-up reduction method.

2 Preparations on certain algorithmic issues

In this section we highlight algorithmic issues which need to be addressed to ensure that our algorithms run in polynomial time. All these issues have been addressed in our earlier paper. So we only indicate briefly where these issues arise and what needs to be done. For details and proofs the really interested reader should refer to [IQS17].

From the extension field to the original field. Assume that for some extension field $\mathbb{K}$ of $\mathbb{F}$ we are given a matrix $A' \in \mathcal{B} \otimes_{\mathbb{F}} \mathbb{K} \leq M(n, \mathbb{K})$ of rank r. Then, if $|\mathbb{F}| > r$, using the method of [dGIR96, Lemma 2.2], we can efficiently find a matrix $A \in \mathcal{B}$ of rank at least r. This procedure is also useful to keep sizes of the occurring field elements small. This is how it gets used in Lemma 4.3 and in Theorem 1.5. We give details for this procedure alone.

Let $S \subseteq \mathbb{F}$ with $|S| = r+1$ and let $B_1, \ldots, B_\ell$ be an $\mathbb{F}$-basis for $\mathcal{B}$. Then $A' = a'_1 B_1 + \ldots + a'_\ell B_\ell$, where $a'_i \in \mathbb{K}$. As A' is of rank r, there exists an $r \times r$ sub-matrix of A with nonzero determinant. Assume that $a'_1 \notin S$. Then we consider the determinant of the corresponding sub-matrix of the polynomial matrix $xB_1 + a'_2 B_2 + \ldots a'_\ell B_\ell$. This determinant is a nonzero polynomial of degree at most r in x. Therefore there exists an element $a_1 \in S$ such that $a_1 B_1 + a'_2 B_2 + \ldots a'_\ell B_\ell$ has rank at least r. Continuing with $a'_2, \ldots, a'_\ell$, we can ensure that all the a_i's are from S. Since the B_i's span $\mathcal{B}$, the resulting matrix of rank at least r is in $\mathcal{B}$. We record this as a fact.

Lemma 2.1 (Data reduction, [dGIR96, Lemma 2.2])**.** *Let $\mathcal{B} \leq M(k \times \ell, \mathbb{F})$ be given by a basis $B_1, \ldots, B_m$, and let $\mathbb{K}$ be an extension field of $\mathbb{F}$. Let S be a subset of $\mathbb{F}$ of size at least $r+1$. Suppose that we are given a matrix $A' = \sum a'_i B_i \in \mathcal{B} \otimes_{\mathbb{F}} \mathbb{K}$ of rank at least r. Then we can find $A = \sum a_i B_i \in \mathcal{B}$ of rank also at least r with $a_i \in S$. The algorithm uses* $\mathrm{poly}(k, \ell, r)$ *rank computations for matrices of the form $\sum a''_i B_i$ where $a''_i \in \{a'_1, \ldots, a'_m\} \cup S$.*

Dealing with the need for a primitive root of unity. Lemma 3.2 assumes the field $\mathbb{F}'$ contains a known primitive dth root of unity ζ. In actual applications, we start with a field $\mathbb{F}$ without a primitive dth root of unity in it, and attach one symbolically, which we still denote by ζ. However, this may cause some problem. Namely, constructing $\mathbb{F}' = \mathbb{F}[\zeta]$ would require factoring the polynomial $x^d - 1$ over $\mathbb{F}$, a task which cannot be accomplished using basic arithmetic operations. To see that this is indeed an issue notice that a black-box field may contain certain "hidden" parts of cyclotomic fields. Of course, over certain concrete fields, such as the rationals, number fields or finite fields of small characteristics, this can be done in polynomial time. However, even over finite fields of large characteristic no deterministic polynomial time solution to this task is known at present.

To get around this issue, one can perform the required computations over an appropriate factor algebra R of the algebra $C = \mathbb{F}[x]/(x^d - 1)$ in place $\mathbb{F}'$ as if R were a field. To be specific, as d is not divisible by the characteristic, we know that C is semisimple – actually it is isomorphic to a direct sum of ideals, each of which is isomorphic to the splitting field $\mathbb{F}[\sqrt[e]{1}]$ of the polynomial $x^e - 1$ for some divisor e of d, and the projection of x to such an ideal is a primitive eth root of unity. It follows that if we compute the ideal J generated by annihilators of $x^e - 1$, for all e a proper divisor of d, then $R = C/J$ is isomorphic to the direct sum of copies of the splitting field $\mathbb{F}'$ of $x^d - 1$, and the projection of x to each component is a primitive dth root of unity. And this property is inherited by any proper factor of R. A computation using R instead of $\mathbb{F}'$ may fail only at a point where we attempt to invert an non-invertible element of R. However, such an element must be a zero divisor. When this situation occurs, we replace R with the factor of R by its ideal generated by the zero divisor and restart the computation. Such a restart can clearly happen at most $d-2$ times.

Now consider the task of computing the rank of a matrix in $M(n, \mathbb{F}')$. As described above we work instead with coefficients in R. Note that we cannot talk about the "rank" of matrices in $M(n, R)$ which is not well-defined. But since R is a direct sum of $\mathbb{F}'$, the decomposition of R induces a decomposition of $M(N, R)$ into a direct sum of copies of $M(N, \mathbb{F}')$. We call the images

of the projections of a matrix $B \in M(N, R)$ to the direct summands the *components* of B. The following lemma from [IQS17] describes how to compute the maximum rank over the components.

Lemma 2.2 ([IQS17, Lemma 4.6])**.** *Let R and $\mathbb{F}'$ be as above, and suppose we are given a matrix $B \in M(N, R)$. Then there exists a deterministic polynomial-time algorithm that computes the maximum rank over the components of B.*

We remark that the issue with the need of roots of unity and working over rings instead of fields occurs only when we apply the algorithm for the constructive regularity lemma. It has no influence of the other parts of the algorithm, as after having constructed a matrix over the ring R having sufficiently large "rank", we can apply Lemma 2.1 to obtain a matrix over the base field $\mathbb{F}$ of the same or larger rank, provided that $\mathbb{F}$ is large enough. (Cyclotomic extension fields of finite fields can be constructed deterministically in time polynomial in the field size, so over small fields such issues do not occur at all.)

Computing the rank of matrices over a rational function field in few variables. In Lemma 4.3 we will need to compute the rank of matrices over a rational function field of $\mathbb{F}'$ in two variables. The following proposition from [IQS17] describes how when the field size $\mathbb{F}'$ is large we can find a matrix over the base field with the same rank as the matrix we start with.

Proposition 2.3 ([IQS17, Lemma 4.8])**.** *Let $\mathbb{F}'$ be a field and $\mathbb{K} = \mathbb{F}'(X_1, \ldots, X_k)$ be a pure transcendental extension of $\mathbb{F}'$. Let A be an $N \times N$ matrix with entries as quotients of polynomials from $\mathbb{F}'[X_1, X_2, \ldots, X_k]$, where the polynomials are explicitly given as sums of monomials. Assume that the degrees of the polynomials appearing in A are upper bounded by D. If $|\mathbb{F}'| = (ND)^{\Omega(k)}$, then we can find in time $(ND)^{O(k)}$ a matrix $B \in M(N, \mathbb{F}')$ with $\mathrm{rk}(B) = \mathrm{rk}(A)$.*

3 Efficient construction of cyclic field extensions of arbitrary degrees

A cyclic extension of a field $\mathbb{K}$ is a finite Galois extension of $\mathbb{K}$ having a cyclic Galois group. By constructing a cyclic extension $\mathbb{L}$ we mean constructing the extension as an algebra over $\mathbb{K}$, e.g., by giving an array of *structure constants* with respect to a $\mathbb{K}$-basis for $\mathbb{L}$ defining the multiplication on $\mathbb{L}$ as well as specifying a generator of the Galois group, e.g, by its matrix with respect to a $\mathbb{K}$-basis.

Lemma 3.1. *Given a prime p and an integer $s \geq 1$, one can construct in time $\mathrm{poly}(p^s)$ a cyclic extension K_s of $\mathbb{F}_p(Z)$ of degree p^s such that $\mathbb{F}_p$ is algebraically closed in K_s. The field K_s will be given in terms of structure constants with respect to a basis over $\mathbb{F}_p(Z)$, and the generator σ for the Galois group will be given by its matrix in terms of the same basis. The structure constants as well as the entries of the matrix for σ will be polynomials in $\mathbb{F}_p[Z]$ of degree $\mathrm{poly}(p^s)$.*

Proof. First we briefly recall the general construction given in Section 6.4 of [Ram54]. This, starting from a field K_0 of characteristic p, recursively builds a tower $K_0 < K_1 < \ldots < K_s$ of fields such that K_j is a cyclic extension of K_0 of degree p^j. Assume that K_s together with a K_0-automorphism σ_s of order p^s has already been constructed. (Initially let σ_0 be the identity map on K_0.) Then for any element $\beta_s \in K_s$ with $\mathrm{Tr}_{K_s:K_0}(\beta_s) = 1$ and for any $\alpha_s \in K_s$ such that $\alpha_s^{\sigma_s} - \alpha_s = \beta_s^p - \beta_s$ the polynomial $X^p - X - \alpha_s$ is irreducible in $K_s[X]$. (Existence of α_s with the required property follows from the additive Hilbert 90.) Put $K_{s+1} = K_s[X]/(X^p - X - \alpha_s)$ and let $\omega_{s+1} \in K_{s+1}$ be

the image of of X under the projection $K_s[X] \to K_{s+1}$. Then σ_s extends to a K_0-automorphism σ_{s+1} of degree p^{s+1} of K_{s+1} such that $\omega_{s+1}^{\sigma_{s+1}} = \omega_{s+1} + \beta_s$. This gives a cyclic extension of degree p^{s+1}.

Now we specify some details of a polynomial time construction for $K_0 = \mathbb{F}_p(Z)$ following the method outlined above. In the first step we take $\beta_0 = 1$, and, in order to guarantee that the only elements in K_1 which are algebraic over $\mathbb{F}_p$ is F_p (we also use the phrase F_p is algebraically closed in K_1 when this property holds), we take $\alpha_0 = Z$. Then K_1 is a pure transcendental extension of $\mathbb{F}_p$. As K_s/K_0 is a cyclic extension of oder p^s, it has a unique subfield which is an order p extension of K_0. This must be K_1. Then $\mathbb{F}_p$ has no proper finite extension in K_s as otherwise K_0 would also have another degree p extension.

We consider the following K_0-basis for K_s:

$$\Gamma_s = \left\{ \prod_{j=1}^{s} \omega_j^k, \quad (k = 0, \ldots, p-1) \right\},$$

where ω_j is a root of $X^p - X - \alpha_{j-1}$ in K_j. We claim that $\mathrm{Tr}_{K_j:K_{j-1}}(\omega_j^{p-1}) = -1$. Indeed, in the K_{j-1}-basis $\omega_j^0, \ldots, \omega_j^{p-1}$ for K_j, in the matrix of multiplication by ω_j^{p-1} the diagonal entries consist of $p-1$ ones and one zero. Therefore $\mathrm{Tr}_{K_j:K_{j-1}}(\omega_j^{p-1}\gamma) = -\gamma$ for every $\gamma \in K_{j-1}$, whence $\mathrm{Tr}_{K_j:K_0}(\omega_j^{p-1}\gamma) = -\mathrm{Tr}_{K_{j-1}:K_0}(\gamma)$. Now by induction we obtain $\mathrm{Tr}_{K_j:K_0} \prod_{i=1}^{j} \omega_i^{p-1} = (-1)^j$. Therefore in each step (when $j > 0$) we can choose $\beta_j = (-1)^j \prod_{i=1}^{j} \omega_i^{p-1}$ and α_j thereafter, following the construction in the standard proof of the additive Hilbert 90. Specifically, we set

$$\alpha_j = (-1)^{j+1} \sum_{k=1}^{p^j-1} \beta_j^{\sigma_j^k} \left(\sum_{\ell=0}^{k-1} (\beta_j^p - \beta_j)^{\sigma_j^\ell} \right). \tag{3.1}$$

Then $\alpha_j^{\sigma_j} - \alpha_j = \beta_j^p - \beta_j$. Notice that α_j is a sum of terms with each of which, up to a sign, is a product of at most $p+1$ conjugates $\beta_j^{\sigma_j^\ell}$ (with various ℓs) of β_j ($\ell \le p^j$)

Assume by induction that the structure constants of K_j with respect to the basis Γ_j are polynomials from $\mathbb{F}_p[Z]$ of degree at most Δ_j and the same holds for the entries of the matrix of σ_j^ℓ for every $1 \le \ell < p^j$ (written in the same basis). For $j = 1$ this holds with $\Delta_1 = 1$. (To see this, observe that for $0 \le k, \ell < p$, the product $\omega_1^k \omega_1^\ell$ is the basis element of $\omega_1^{k+\ell}$ if $k + \ell < p$, while otherwise it equals the sum $\omega_1^{k+\ell-p+1} + Z\omega_1^{k+\ell-p}$.) Then, if we express α_j in terms of the basis Γ_j using Eq. 3.1, we obtain that its coordinates are polynomials of degree at most $(2p+1)\Delta_j$. This is because $(-1)^j \beta_j \in \Gamma_j$, whence $\beta_j^{\sigma^\ell}$ has coordinates of polynomials of degree bounded by Δ_j. In Eq. 3.1, we have the products of at most $p+1$ such elements, so the result will have polynomial coordinates of degree at most $(2p+1)\Delta_j$.

Now consider the product of two elements $\omega_{j+1}^k \gamma_1$ and $\omega_{j+1}^\ell \gamma_2$ of Γ_{j+1}. Here $k, \ell < p$ and $\gamma_1, \gamma_2 \in \Gamma_j$. The coordinates of the product $\gamma_1\gamma_2$ with respect to Γ_j are polynomials of degree at most Δ_j. The same holds for the product $\omega_{j+1}^{k+\ell}\gamma_1\gamma_2$ if $k + \ell < p$. If $k + \ell > p$, then $\omega_{j+1}^{k+\ell} = \omega_{j+1}^p \omega_{j+1}^{k+\ell-p} = (\omega_{j+1} + \alpha_j)\omega_{j+1}^{k+\ell-p}$, whence $\omega_{j+1}^{k+\ell}\gamma_1\gamma_2$ is the sum of $\omega_{j+1}^{1+k+\ell-p}\gamma_1\gamma_2$ and $\alpha_j\gamma_1\gamma_2$. The former term has coordinates of degree at most Δ_j, the coordinates of the latter are polynomials of degree at most $(2p+1)\Delta_j + \Delta_j + \Delta_j = (2p+3)\Delta_j$.

Now consider the conjugate of $\omega_{j+1}^k\gamma$ by σ_{j+1}^ℓ, where $1 \le \ell < p^{j+1}$, $1 \le k \le p-1$ and $\gamma \in \Gamma_j$. This conjugate is $(\omega_{j+1}^{\sigma_{j+1}^\ell})^k \gamma^{\sigma_{j+1}^\ell}$. The second term equals $\gamma^{\sigma_j^\ell}$ which has coordinates of degree at most Δ_j. To investigate the first term, recall that $\omega_{j+1}^{\sigma_{j+1}} = \omega_{j+1} + \beta_j$, whence

$$\omega_{j+1}^{\sigma_{j+1}^\ell} = \omega_{j+1} + \sum_{r=0}^{\ell-1} \beta_j^{\sigma_j^r}$$

The element $\delta = \sum_{r=0}^{\ell-1} \beta_j^{\sigma_j^r}$, expressed in terms of Γ_j, has again polynomial coordinates of degree at most Δ_j. Then $(\omega_{j+1}^{\sigma_{j+1}^\ell})^k$ is the sum (with binomial coefficients) of terms of the form $\omega_{j+1}^r \delta^{k-r}$. The power δ^{k-r} has coordinates of degree at most $(k-r)\Delta_j + (k-r-1)\Delta_j \le (2p-1)\Delta_j$ in terms of Γ_j, whence we conclude that $(\omega_{j+1}^{\sigma_{j+1}^\ell})^k$ has, in terms of Γ_{j+1} polynomial coordinates of degree at most $(2p-1)\Delta_j$. It follows that the matrix of any power of σ_{j+1} has polynomial entries of degree at most $2p\Delta_j$.

We obtained that the function $(2p+3)^s = \mathrm{poly}(p^s)$ is an upper bound for both the structure constants and for the matrices of the powers of σ_s. □

Lemma 3.2. *Let $\mathbb{F}'$ be a field. Let d be any non-negative integer. If $\mathrm{char}(\mathbb{F}') = 0$ then $d_1 = d$. If $\mathrm{char}(\mathbb{F}') = p > 0$ then let d_1 be the p-free part of d, that is, $d = d_1 p^s$, where $p \nmid d_1$ and $s \in \mathbb{N}$. Assume that $\mathbb{F}'$ contains a known d_1th root of unity ζ. Then a cyclic extension $\mathbb{L}$ degree d of $\mathbb{K} := \mathbb{F}'(X)$ can be computed using $\mathrm{poly}(d)$ arithmetic operations. $\mathbb{L}$ will be given by structure constants with respect to a basis, and the matrix for a generator of the Galois group in terms of the same basis will also be given. All the output entries (the structure constants as well as the entries of the matrix representing the Galois group generator) will be polynomials of degree $\mathrm{poly}(d)$ in $\mathbb{F}'[X]$. Furthermore for $\mathbb{F}' = \mathbb{Q}[\sqrt[d_1]{1}]$, the bit complexity of the algorithm (as well as the size of the output) is $\mathrm{poly}(d)$.*

Proof. Put $\mathbb{L}_1 = \mathbb{F}'(Y)$ and $X = Y_1^{d_1}$. Then $1, Y_1, \ldots, Y_1^{d_1}$ are a $\mathbb{F}'(X)$-basis for $\mathbb{L}_1$ with $Y_1^i Y_1^j = Y_1^{i+j}$ if $i + j \le d_1$ and $XY_1^{i+j-d_1}$ otherwise. Further note that the linear extension σ_1 of the map sending Y_1^j to $\zeta^j Y_1^j$ is an automorphism of degree d_1. Then $\mathbb{L}_1$ is a cyclic extension of $\mathbb{F}'(X)$ of degree d_1. This procedure has been used in [IQS17].

We can compute whether $\mathrm{char}(\mathbb{F}')$ is a divisor of d by testing the multiples of the identity element up to d. If $\mathrm{char}(\mathbb{F}') = 0$, or if $\mathrm{char}(\mathbb{F}') = p > 0$ and $p \nmid d$, we are done. Note that in the following $p \le d$.

If $\mathrm{char}(\mathbb{F}') = p > 0$ and $p \mid d$, let d_1 be in the statement, so $d = d_1 p^s$. Let $d_2 = p^s$, and $\mathbb{F}_p$ be the prime field of $\mathbb{F}'$. Construct the cyclic extension of degree d_2 of $\mathbb{F}_p(X)$ over $\mathbb{F}_p$ by 3.1, and let the resulting field be $\mathbb{L}_2$. We also obtain the matrix a generator σ_2 of the Galois group. Then put $\mathbb{L} = \mathbb{L}_1 \otimes_{\mathbb{F}_p(X)} \mathbb{L}_2$. It contains a copy of $\mathbb{K} = \mathbb{F}'(X) \cong \mathbb{F}'(X) \otimes_{\mathbb{F}_p(X)} \mathbb{F}_p(X)$. We take the product basis for the structure constants and for matrix representation of the automorphism $\sigma_1 \otimes \sigma_2$. □

4 The complete constructive regularity lemma

We first present the formal statement of the regularity lemma in its full generality. We also add a technical notion that will be useful for the proof of Theorem 1.5. Let $n \in \mathbb{N}$, and let $\mathbf{i} = (i_1, \ldots, i_r)$, $\mathbf{j} = (j_1, \ldots, j_r)$ be two sequences of integers, where $1 \le i_1 < \cdots < i_r \le n$ and $1 \le j_1 < \cdots < j_r \le n$.

For a matrix $A \in M(n,\mathbb{F}) \otimes M(d,\mathbb{F})$, the $r \times r$ *window* indexed by $\mathbf{i}$, $\mathbf{j}$ is the sub-matrix of A consisting of the blocks indexed by (i_k, j_ℓ), $k, \ell \in [r]$.

Lemma 4.1 (Regularity of blow-ups)**.** *For $\mathcal{B} \leq M(n,\mathbb{F})$ and $\mathcal{A} = \mathcal{B}^{\{d\}}$, assume that $|\mathbb{F}| = (rd)^{\Omega(1)}$. Given a matrix $A \in \mathcal{A}$ with $\mathrm{rk}A > (r-1)d$, there exists a deterministic algorithm that returns $\widetilde{A} \in \mathcal{A}$ and an $r \times r$ window W in $\widetilde{A}$ such that W is nonsingular (of rank rd). This algorithm uses $\mathrm{poly}(nd)$ arithmetic operations and, over $\mathbb{Q}$, the algorithm runs in polynomial time. In particular, all intermediate numbers have bit lengths polynomial in the input size.*

The cases (a) $\mathrm{char}(\mathbb{F}) = 0$, (b) $\mathrm{char}(\mathbb{F})$ and d are coprime, and $|\mathbb{F}| = (rd)^{\Omega(1)}$ were settled in [IQS17, Lemma 5.7] which was reproduced in 1.1. The main issue with the case when d is not coprime to $\mathrm{char}(\mathbb{F})$ was that we did not have an efficient construction of an appropriate Artin-Schreier-Witt extension of $\mathbb{F}_p(x)$, Now we have such a construction in Lemma 3.1.

The proof makes use of the following two results from [IQS17].

Proposition 4.2 ([IQS17, Proposition 4.4])**.** *Let $\mathbb{L}$ be a cyclic extension of degree d of a field $\mathbb{K}$, and suppose that $\mathbb{L}$ is given by structure constants w.r.t. a $\mathbb{K}$-basis $A_1, \ldots, A_d$. Similarly, a generator σ for the Galois group is assumed to be given by its matrix in terms of the same basis. Let Y be a formal variable. Then one can construct a $\mathbb{K}(Y)$-basis Γ of $M(d, \mathbb{K}(Y))$ such that the $\mathbb{K}(Y^d)$-linear span of Γ is a central division algebra over $\mathbb{K}(Y^d)$ of index d, using $\mathrm{poly}(d)$ arithmetic operations in $\mathbb{K}$. Furthermore for $\mathbb{K} = \mathbb{Q}[\sqrt[d]{1}]$, the bit complexity of the algorithm (as well as the size of the output) is also $\mathrm{poly}(d)$.*

Lemma 4.3 (Conditional regularity [IQS17, Lemma 5.4])**.** *Assume that we are given a matrix $A \in \mathcal{B}^{\{d\}} \leq M(dn,\mathbb{F})$ with $\mathrm{rk}(A) = (r-1)d + k$ for some $1 < k < d$. Let X and Y be formal variables and put $\mathbb{K} = \mathbb{F}'(X)$, where $\mathbb{F}'$ is a finite extension of $\mathbb{F}$ of degree at most d. Suppose further that $|\mathbb{F}| > (nd)^{O(1)}$ and that we are also given a $\mathbb{K}(Y)$-basis Γ of $M(d, \mathbb{K}(Y))$ such that the $\mathbb{K}(Y^d)$-linear span of Γ is a central division algebra D' over $\mathbb{K}(Y^d)$. Let δ be the maximum of the degrees of the polynomials appearing as numerators or denominators of the entries of the matrices in Γ. Then, using $(nd+\delta)^{O(1)}$ arithmetic operations in $\mathbb{F}$, one can find a matrix $A'' \in \mathcal{B}^{\{d\}}$ with $\mathrm{rk}(A'') \geq rd$. Furthermore, over $\mathbb{Q}$ the bit complexity of the algorithm is polynomial in the size of the input data (that is, the total number of bits describing the entries of matrices and in the coefficients of polynomials).*

Proof of Lemma 4.1. The statement, except the window part, readily follows by plugging Lemma 3.2 of the previous section to Proposition 4.2 and the using that in Lemma 4.3. To see that such a window can be computed, we first observe that the lemma applies to d-blow-ups of rectangular matrices, by simple zero padding. Second, apply the lemma and find an $rd \times rd$ nonsingular sub-matrix of the given matrix A. If the column indices include some such that not all of its $d-1$ siblings are included, then (1) delete the corresponding column from the original matrix space; (2) let A' be the matrix obtained by deleting the corresponding d columns from A. Then $\mathrm{rk}(A') > \mathrm{rk}(A) - (d-1)$. So we apply the regularity lemma in the rectangular space with A', to round up the rank to $\mathrm{rk}(A)$ again. Do the same for row indices. Iterate until we obtain a full window. □

5 Proof of the main theorem

In Section 5.1 we prove Theorem 1.5, and in Section 5.2 we deal with the small field case. The main drawback of our earlier algorithm discussed in Section 1.1 was that the blow-up size increases

exponentially. However, a simple reduction procedure as described in Lemma 5.2 below readily implies that, once we find A' of rank $r'd$ in $\mathcal{B}^{\{d\}}$, we can efficiently reduce d to be no more than $r'+1$. This means that we can always ensure that the blow-up factor is small, which is the key to reducing the complexity of the algorithm from exponential time to polynomial time. We shall make the above idea rigorous in the next subsection.

5.1 The algorithm for the main theorem

We first recall some preparation material from [IQS17].

Finding an sd-shrunk subspace for the $\mathcal{B}^{\{d\}}$ is equivalent to finding an s-shrunk subspace for $\mathcal{B}$ because of the following simple observations ([IQS17, Proposition 5.2]). Firstly, for every s-shrunk subspace U of $\mathbb{F}^n$ the subspace $U \otimes \mathbb{F}^d$ for $\mathcal{B}$ is an sd-shrunk subspace for $\mathcal{B}^{\{d\}}$. Conversely, a s'-shrunk subspace for $\mathcal{B}^{\{d\}}$ can be embedded into a subspace of the form $U \otimes \mathbb{F}^d$ where U is an s-shrunk subspace for $\mathcal{B}$ with $sd \geq s'$.

The main technical ingredient of our algorithm is an improvement of [IQS17, Theorem 5.10], discussed in Section 1.1 . It states that either a shrunk subspace witnessing that the (scaled-down) rank of a matrix in a blow-up reaches the non-commutative rank or a matrix in a larger blow-up having larger scaled-down rank can be efficiently constructed. For completeness we give all the details and also the proof even though it is identical to that in our earlier paper excepting for the last step.

Theorem 5.1. *Let $\mathcal{B} \leq M(n,\mathbb{F})$ and let $\mathcal{A} = \mathcal{B}^{\{d\}}$. Assume that we are given a matrix $A \in \mathcal{A}$ with $\mathrm{rk}(A) = rd$, and $|\mathbb{F}|$ is $(ndd')^{\Omega(1)}$, where $d' = r+1$. There exists a deterministic algorithm that returns either an $(n-r)d$-shrunk subspace for $\mathcal{A}$ (equivalently, an $(n-r)$-shrunk subspace for $\mathcal{B}$), or a matrix $B \in \mathcal{A} \otimes M(d',\mathbb{F})$ of rank at least $(r+1)dd'$. Furthermore, in the latter case an $(r+1)\times(r+1)$ window is also found such that the corresponding $(r+1)dd' \times (r+1)dd'$ sub-matrix of B has full rank. This algorithm uses $\mathrm{poly}(ndd')$ arithmetic operations and, over $\mathbb{Q}$, all intermediate numbers have bit lengths polynomial in the input size.*

Proof. Starting with the kernel V_0 of the linear map A we compute the image W_1 of V_0 under $\mathcal{A}$. If W_1 is not in the image of A we stop and declare $W^* = W_1$. Otherwise we define V_1 to be the preimage of W_1 under A and define W_2 to be the image of V_1 under $\mathcal{A}$. We continue doing so, at each step checking if W_i is in the image of A or not. Since at each step the dimension of W_i increases by d it is clear that we halt in l steps with l at most $r+1$, obtaining the limit subspace $W^* = W_l$. If W^* is in the image of A, it follows from Fact 1.3 that the preimage of W_l under A is an $(n-r)d$-shrunk subspace. In either case in at most $r+1$ steps we find a shrunk subspace or find that W^* is not in the image of A.

When the limit subspace is not in $\mathrm{im}(A)$ we proceed as follows. Let B_l be an element of $\mathcal{A}$ and $v_l \in V_{l-1}$ such that $B_l(v_l) \notin \mathrm{im}(A)$. Then find matrices $B_{l-1} \in \mathcal{A}$ and vector $v_{l-1} \in V_{l-2}$ such that $B_{l-1}(v_{l-1}) = A(v_l)$. Walking backwards, we find matrices $B_{l-2}, \ldots, B_1$ and vectors $v_{l-3}, \ldots, v_1$, $v_i \in V_{i-1}$ such that $A(v_i) = B_{i-1}(v_{i-1})$. In particular $v_1 \in \ker(A)$.

Now let $A' = A \otimes I_{d'}$. Clearly A' is a matrix of rank rdd' in $\mathcal{A}^{d'} = \mathcal{B}^{dd'}$. Now let $E_{i,j}$ be the elementary matrix in $M(d',\mathbb{F})$ with the (i,j)th entry being 1 and others 0. Put $\widehat{B} = B_1 \otimes E_{1,2} + B_2 \otimes E_{2,3} + \ldots B_{l-1} \otimes E_{l-1,l} + B_l \otimes E_{l,1} \in \mathcal{B}^{\{dd'\}}$. If the rank of $\widehat{B}$ is more than rdd' we set A'' to be $\widehat{B}$. Otherwise consider the vectors $w_1 = v_1 \otimes u_1$, $w_2 = v_2 \otimes u_2$, ... , $w_l = v_l \otimes u_l$. It is clear that $A'(w_1) = 0$ and that $A'w_j = \widehat{B}(w_{j-1})$ for $2 \leq j \leq l$. Furthermore, $\widehat{B}(w_l) = B_l(v_l) \otimes u_{l+1}$

and this is not in $A'(\mathbb{F}^{nd} \otimes \mathbb{F}^{d'})$ since $B_l(v_l)$ is not in the image of A. So if we were to compute the second Wong sequence starting with the matrix A' in the rank two linear space of $\mathcal{B}^{dd'}$ spanned by matrices $\{A', \widehat{B}\}$, the second Wong sequence runs out of the image of A'. So by Fact 1.4 A' is not of maximal rank in the linear space spanned by $\{A', \widehat{B}\}$. So there exists $\mu \in \mathbb{F}$ such that $A' + \mu\widehat{B}$ has rank strictly bigger than rdd'. As the determinant of an $(rdd'+1) \times (rdd'+1)$ submatrix of $A' + \mu\widehat{B}$ is a polynomial of degree at most $rdd'+1$ in μ, we can find μ by running over all of elements of a subset of $\mathbb{F}$ of size $rdd'+2$ till we find one.

We then invoke Lemma 4.1 with A'' to obtain a matrix B over the base field $\mathbb{F}$ of rank $(r+1)dd'$ and the $(r+1) \times (r+1)$ window as required, completing the proof.

It is clear that the matrices $B_1, \ldots, B_l$ as well as μ can be determined in the given polynomial time. □

To obtain the algorithm for Theorem 1.5, the regularity lemma needs to be accompanied with a reduction procedure that keeps the blow-up parameter small. We mentioned in the introduction that there are two methods for this purpose, and in this section we use our method. The method based on the Derksen-Makam technique is presented in Section 6.

Lemma 5.2. *Let $\mathcal{B} \leq M(n, \mathbb{F})$, and $d > n+1$. Assume we are given a matrix $A \in \mathcal{B}^{\{d\}}$ of rank dn. Then there exists a deterministic polynomial-time procedure that constructs $A' \in \mathcal{B}^{\{d-1\}}$ of rank $(d-1)n$.*

Proof. Let A'' be an appropriate $(d-1)n \times (d-1)n$ sub-matrix of A corresponding to a matrix in $\mathcal{B}^{\{d-1\}}$. We claim A'' is of rank $> (d-1)(n-1)$. Suppose not, as A is obtained from A'' from adding n rows and then n columns, and $d > n+1$, we have $\mathrm{rk}(A) \leq \mathrm{rk}(A'') + 2n \leq dn - d - n + 1 + 2n < dn$, a contradiction. Now that $\mathrm{rk}(A'') > (d-1)(n-1)$, using Lemma 4.1, we obtain $A' \leq \mathcal{B}^{\{d-1\}}$ of rank $(d-1)n$. □

Proof of Theorem 1.5. Let $B_1, \ldots, B_m$ be the input basis for $\mathcal{B}$. The algorithm is an iteration based on Theorem 5.1. In each round we start with a matrix $A = \sum_i B_i \otimes T_i \in \mathcal{B}^{\{d\}}$ of rank rd for some integer $d \leq r+1$. In the first round, $d = 1$ and A can be taken as any matrix in $\mathcal{B}$. The procedure behind Theorem 5.1 either returns an $(n-r)$-shrunk subspace (in which case we are done), or a new matrix (denoted also by A) in a blow-up $\mathcal{B}^{\{d'\}}$ of rank $\geq (r+1)d'$ for some $d' \leq (r+1)^2$, together with a square window of size $r+1$ so that the corresponding sub-matrix of A is of rank $(r+1)d'$. If $d' > r+2$ we apply Lemma 5.2 as follows. The n in the statement of Lemma 5.2 will be $r+1$, and we use it repeatedly to get a matrix in the $(r+2)$-blow-up, the main content of which consists of $(r+2) \times (r+2)$ matrices $T'_1, \ldots, T'_m$ such that the corresponding $(r+1)(r+2) \times (r+1)(r+2)$ sub-matrix of $A' = \sum_i B_i \otimes T'_i$ has full rank. Then we replace A with A' and apply the size reduction procedure in Lemma 2.1 to arrange that the entries of T_i fall into the prescribed subset of $\mathbb{F}$, and continue the iteration with this new matrix A. □

5.2 Proof of Corollary 1.7: the case of small finite fields

We only need to prove Corollary 1.7 (2), from which (1) and (3) are immediate.

Given a matrix space $\mathcal{B} \leq M(n, \mathbb{F})$ and a field extension $\mathbb{K}/\mathbb{F}$, $\mathcal{B}$ can be viewed naturally as a matrix space in $M(n, \mathbb{K})$. For convenience we use $\mathrm{ncrk}_{\mathbb{F}}(\mathcal{B})$ to signal that we consider the non-commutative rank of $\mathcal{B}$ over $\mathbb{F}$. We first observe that the non-commutative rank does not change under field extensions. This is classical, and can be seen from the perspective of the second Wong

sequences (see e.g. [IKQS15, Section 2]). Note that the commutative rank may get larger if we go to an extension field from a too-small field.

Lemma 5.3. *Given $\mathcal{B} \leq M(n,\mathbb{F})$ and a field extension $\mathbb{K}/\mathbb{F}$, we have* $\mathrm{ncrk}_{\mathbb{F}}(\mathcal{B}) = \mathrm{ncrk}_{\mathbb{K}}(\mathcal{B})$.

Suppose $\mathcal{B} \leq M(n,\mathbb{F})$ is given by a linear basis $\{B_1,\ldots,B_m\}$. Let $\mathbb{K}/\mathbb{F}$ be a field extension of degree g so that $|\mathbb{K}| = n^{\Omega(1)}$ satisfies the field size condition of Theorem 1.5. Note that $g \leq O(\log_{|\mathbb{F}|} n)$. Viewing $\mathcal{B}$ as a matrix space over $\mathbb{K}$, we apply Theorem 1.5 to compute $\mathrm{ncrk}_{\mathbb{K}}(\mathcal{B})$, which is equal to $r = \mathrm{ncrk}_{\mathbb{F}}(\mathcal{B})$ by Lemma 5.3. We also obtain the following: (1) $A_1,\ldots,A_m \in M(d,\mathbb{K})$ such that $A = \sum_{i\in[m]} A_i \otimes B_i$ is of rank rd, and (2) $U \leq \mathbb{K}^n$ such that U is a shrunk subspace of $\mathcal{B}$ a matrix space in $M(n,\mathbb{K})$. We fix an embedding ϕ of $\mathbb{K}$ into $M(g,\mathbb{F})$ using the regular representation. For $i \in [m]$, construct $\widetilde{A_i} \in M(gd,\mathbb{F})$ by replacing each entry α of A_i with $\phi(\alpha)$, and form $\widetilde{A} = \sum_{i\in[m]} \widetilde{A_i} \otimes B_i$. Note that $\widetilde{A}$ is in $M(gd,\mathbb{F}) \otimes \mathcal{B}$, and it can be seen easily that $\mathrm{rk}(\widetilde{A}) = g \cdot \mathrm{rk}(A)$. Since $\mathrm{rk}(\widetilde{A})/gd = r = \mathrm{ncrk}_{\mathbb{F}}(\mathcal{B})$, we have $\mathrm{crk}_{\mathbb{F}}(M(gd,\mathbb{F}) \otimes \mathcal{B}) = \mathrm{ncrk}_{\mathbb{F}}(M(gd,\mathbb{F}) \otimes \mathcal{B})$. This implies that we can apply the second Wong sequence to $(\widetilde{A}, M(gd,\mathbb{F}) \otimes \mathcal{B})$ to obtain an $(n-r)gd$-shrunk subspace of $M(gd,\mathbb{F}) \otimes \mathcal{B}$ which then induces an $(n-r)$-shrunk subspace of $\mathcal{B}$.

6 Constructivizing the result of Derksen and Makam

Here is an algorithmic version of Lemma 2.7 of [DM17b]. Although the most relevant blow-ups in the context of the non-commutative rank problem are square (e.g, of the form $\mathcal{B}^{\{k\}}$, described earlier), non-square blow-ups turned out to be crucial in the reduction techniques in [DM17b]. So we use a different notation for blow-ups from what was used so far. Given a matrix space $\mathcal{B} \leq M(n,\mathbb{F})$, its (k,ℓ)-*blow-up* $\mathcal{B}^{\{k,\ell\}}$ is defined as the matrix space $\mathcal{B} \otimes M(k \times \ell, \mathbb{F})$ in $M(nk \times n\ell, \mathbb{F})$.

Lemma 6.1. *Let $\mathcal{B} \leq M(n,\mathbb{F})$. Assume that for $k,\ell = 1,\ldots,N$ we are given matrices $M_0(k,\ell) \in \mathcal{B}^{\{k,\ell\}}$ of rank $r_0(k,\ell)$, and suppose that $|\mathbb{F}| \geq 2nN+1$. Then for every $k,\ell = 0,\ldots,N$ we can efficiently (that is, by an algorithm that uses* $\mathrm{poly}(Nn)$ *arithmetic operations and, over e.g. $\mathbb{Q}$, produces intermediate and final data of size polynomial in the input size) construct matrices $M(k,\ell) \in \mathcal{B}^{\{k,\ell\}}$ of rank $r(k,\ell) \geq r_0(k,\ell)$ such that*

(1) $r(k,\ell+1) \geq r(k,\ell)$ *$(0 \leq \ell < N)$;*

(2) $r(k+1,\ell) \geq r(k,\ell)$ *$(0 \leq k < N)$;*

(3) $r(k,\ell+1) \geq \frac{1}{2}(r(k,\ell) + r(k,\ell+2))$ *$(0 \leq \ell < N-1)$;*

(4) $r(k+1,\ell) \geq \frac{1}{2}(r(k,\ell) + r(k+2,\ell))$ *$(0 \leq k < N-1)$;*

(5) $r(k,k)$ *is divisible by k.*

For $k = 0$ (resp. $\ell = 0$) we assume that $M_0(k,\ell)$ is the empty matrix having ℓ columns (resp. k rows), and $r(k,\ell) = 0$.

Proof. Initially put $M(k,\ell) = M_0(k,\ell)$ for every pair (k,ℓ). For a $k \times \ell$ matrix T let T^+ denote the $(k+1)\times\ell$ matrix obtained form T by appending a zero ($(k+1)$st) row, T^{++} is obtained by appending two zero rows. For $M = \sum_{i=1}^m B_i \otimes T_i$ we use M^+ for $\sum_{i=1}^m B_i \otimes T_i^+$, while $M^{++} = \sum_{i=1}^m B_i \otimes T_i^{++}$.

Let (k,ℓ) be a pair such that any of (1)–(5) is violated. Then we will replace some of the matrices $M(k',\ell')$ with matrices having larger rank. Over an infinite base field like $\mathbb{Q}$, each such

replacement step (or each small group consisting of a few them) can be followed by an application of the data reduction procedure in Lemma 2.1 to keep intermediate (as well as the final) data small.

If (1) is violated then, like in [DM17b], replace $M(k+1,\ell)$ with $M(k,\ell)^+$. We can treat a violation of (2) symmetrically.

When (3) is violated we consider the matrix $A = A(t) = M(k+2,\ell)+tM(k,\ell)^{++}$ as a $(k+2)\times\ell$ block matrix consisting of square blocks of size n from $\mathcal{B}$. We can choose t from any subset S of size $2nN+1$ of the base field so that A has rank at least $r(k+2,\ell)$, while the first kn rows form a matrix of rank at least $r(k,\ell)$. This is because a necessary condition for violating either of these two conditions is that the determinant of an appropriate (but unknown) sub-matrix vanishes which determinant is, as a polynomial of degree at most nN in t is not identically zero. The product of these polynomials has degree at most $2nN$ therefore it cannot have more that $2nN$ zeros.

If A has rank larger than $r(k+2,\ell)$ then we replace $M(k+2,\ell)$ with A. Otherwise, like in [DM17b], let U be the span of the first kn rows of A, V be the span of the first $(k+1)n$ rows and W be the span of the first kn rows and the last n rows. Note that these collections rows correspond to matrices of the form $A_0 = \sum B_i\otimes T_i$, $A_1 = \sum B_i\otimes T_i'$ and $A_2 = \sum B_i\otimes T_i''$ where T_i are $k\times\ell$ matrices, while T_i' and T_i'' have $(k+1)$ rows and ℓ columns. As $U \le V\cap W$ and the row space of A is $V+W$, we have $r(k,\ell) \le \dim U \le \dim(V\cap W) = \dim V+\dim W-\dim V+W = \dim V+\dim W-r(k+2,\ell)$. It follows that $\dim V + \dim W \ge r(k,\ell) + r(k+2,\ell)$, whence violation of (3) is only possible if either $\dim V$ or $\dim W$ is strictly larger than $\frac{1}{2}(r(k,\ell)+r(k+2,\ell))$. Then we replace $M(k+1,\ell)$ with A_1 or A_2, according to which one has larger rank. A violation of (4) is treated symmetrically.

When (5) is violated then we can apply 4.1.

As in each round when violation of (1),...,(4) or (5) occurs the rank of at least one of the matrices $M(k,\ell)$ is incremented, the total number of rounds for achieving (1)–(5) is at most N^3n. □

And here is essentially Proposition 2.10 of [DM17b]. We include a proof (which is almost literally the same as the proof in [DM17b]) here for completeness. We note that this lemma deals only with the property of certain families of functions, without referring to matrices.

Lemma 6.2 ([DM17b, Proposition 2.10]). *Assume that $N > n > 0$, $r : \{0,1,\dots,N\}^2 \to \mathbb{Z}$ is a function with $0 \le r(k,\ell) \le \min(k,\ell)n$ for $k,\ell \in \{0,1,\dots,N\}$ also satisfying (1)–(5) of 6.1. Suppose further that $r(1,1) > 1$, and there exists d such that $n \le d+1 \le N$ and $r(d+1,d+1) = n(d+1)$. Then, $r(d,d) = nd$ as well.*

Proof. By $r(d+1,d+1) = n(d+1)$, for $1 \le a < d+1$,

$$r(d+1,a) \ge \frac{(d+1-1)\cdot r(d+1,0) + a\cdot r(d+1,d+1)}{d+1} = an.$$

As by assumption $r(d+1,a) \le an$, we have $r(d+1,a) = an$. Similarly $r(a,d+1) = an$ for $1 \le a < d+1$.

Then we bound $r(1,d)$ as follows:

$$\begin{aligned} r(1,d) &\ge \frac{(d-1)\cdot r(1,d+1) + 1\cdot r(1,1)}{d} \\ &\ge \frac{(d-1)n+2}{d} = n - \frac{n-2}{d} > n-1. \end{aligned}$$

Note that we use $r(1,1) > 1$ and $d \ge n-1$. Since $r(1,d) \in \mathbb{Z}$, $r(1,d) = n$.

We are ready to bound $r(d,d)$ then.

$$\begin{aligned} r(d,d) &\geq \frac{(d-1)\cdot r(d+1,d)+1\cdot r(1,d)}{d} \\ &= \frac{(d-1)dn+n}{d} = nd-n+\frac{n}{d}. \end{aligned}$$

From $d \geq n-1$ it is inferred easily that $-n+\frac{n}{d} > -d$. Therefore $nd-n+\frac{n}{d} > (n-1)d$. By (5) we conclude that $r(d,d) = nd$. □

We finally remark that, if we use Lemma 6.1 in the proof of Theorem 1.5, then n in the statement of the lemma will be $r+1$, N will be d', $M_0(d',d')$ is the nonsingular $(r+1)d' \times (r+1)d'$ block of A and $M_0(p,q)$ can be actually even the zero matrix for $(p,q) \neq (d',d')$. It will prepare matrices in several not necessarily square blow-ups, among others, most importantly, one in an (r,r)-blow-up with a similar content as described in the proof of Theorem 1.5.

Acknowledgements. We would like to thank the authors of [GGOW16] and of [DM17b] for sharing their ideas with us and making us possible to read early versions of their manuscripts. Part of the work was done when Gábor and Youming were visiting the Centre for Quantum Technologies at the National University of Singapore. Research of the first author was also supported in part by the Hungarian National Research, Development and Innovation Office NKFIH Grant 115288. Youming's research was supported by the Australian Research Council DECRA DE150100720. KV's research was supported by a grant from the Infosys foundation.

References

[AL81] MD Atkinson and S Lloyd. Primitive spaces of matrices of bounded rank. *Journal of the Australian Mathematical Society (Series A)*, 30(04):473–482, 1981.

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

[ANS07] B. Adsul, S. Nayak, and K. V. Subrahmanyam. A geometric approach to the Kronecker problem II: rectangular shapes, invariants of matrices and the Artin–Procesi theorem. preprint, 2007.

[BD06] M. Bürgin and J. Draisma. The Hilbert null-cone on tuples of matrices and bilinear forms. *Mathematische Zeitschrift*, 254(4):785–809, 2006.

[Ber70] George W. Bergman. Skew fields of noncommutative rational functions (preliminary version). *Séminaire Schützenberger*, 1:1–18, 1969-1970.

[BFS99] Jonathan F. Buss, Gudmund S. Frandsen, and Jeffrey O. Shallit. The computational complexity of some problems of linear algebra. *J. Comput. Syst. Sci.*, 58(3):572–596, 1999.

[CIKK15] Marco Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Tighter connections between derandomization and circuit lower bounds. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques,*

APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA, pages 645–658, 2015.

[Coh85] P. M. Cohn. *Free Rings and Their Relations*. L.M.S. Monographs. Acad. Press, 1985. First edition 1971.

[Coh95] P. M. Cohn. *Skew Fields: Theory of General Division Rings*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995.

[CR99] P. M. Cohn and C. Reutenauer. On the construction of the free field. *International Journal of Algebra and Computation*, 9(3-4):307–323, 1999.

[Der01] Harm Derksen. Polynomial bounds for rings of invariants. *Proceedings of the American Mathematical Society*, 129(4):955–964, 2001.

[dGIR96] Willem A. de Graaf, Gábor Ivanyos, and Lajos Rónyai. Computing Cartan subalgebras of Lie algebras. *Applicable Algebra in Engineering, Communication and Computing*, 7(5):339–349, 1996.

[DM17a] H. Derksen and V. Makam. On non-commutative rank and tensor rank. *Linear and Multilinear Algebra*, pages 1–16, 2017. Article in Press.

[DM17b] H. Derksen and V. Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017.

[DW00] Harm Derksen and Jerzy Weyman. Semi-invariants of quivers and saturation for littlewood-richardson coefficients. *Journal of the American Mathematical Society*, 13(3):467–479, 2000.

[DZ01] M. Domokos and A. N. Zubkov. Semi-invariants of quivers as determinants. *Transformation groups*, 6(1):9–24, 2001.

[Edm67] Jack Edmonds. Systems of distinct representatives and linear algebra. *J. Res. Nat. Bur. Standards Sect. B*, 71:241–245, 1967.

[EH88] David Eisenbud and Joe Harris. Vector spaces of matrices of low rank. *Advances in Mathematics*, 70(2):135 – 155, 1988.

[FR04] M. Fortin and C. Reutenauer. Commutative/noncommutative rank of linear matrices and subspaces of matrices of low rank. *Séminaire Lotharingien de Combinatoire*, 52:B52f, 2004.

[GGOW16] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 109–117, 2016.

[Gur04] Leonid Gurvits. Classical complexity and quantum entanglement. *J. Comput. Syst. Sci.*, 69(3):448–484, 2004.

[HW15] Pavel Hrubeš and Avi Wigderson. Non-commutative arithmetic circuits with division. *Theory of Computing*, 11:357–393, 2015.

[IKQS15] Gábor Ivanyos, Marek Karpinski, Youming Qiao, and Miklos Santha. Generalized wong sequences and their applications to edmonds' problems. *J. Comput. Syst. Sci.*, 81(7):1373–1386, 2015.

[IQS17] G. Ivanyos, Y. Qiao, and K.V. Subrahmanyam. Non-commutative edmonds problem and matrix semi-invariants. *Computational Complexity*, 26(3):717–763, 2017.

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

[LSW00] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4):545–568, 2000.

[Mal78] Peter Malcolmson. A prime matrix ideal yields a skew field. *Journal of the London Mathematical Society*, s2-18(2):221–233, 1978.

[Ram54] K.G. Ramanathan. *Lectures on the Algebraic Theory of Fields*. Tata Institute of Fundamental Research, Bombay, 1954.

[SVdB01] Aidan Schofield and Michel Van den Bergh. Semi-invariants of quivers for arbitrary dimension vectors. *Indagationes Mathematicae*, 12(1):125–138, 2001.

[Won74] Kai-Tak Wong. The eigenvalue problem $\lambda Tx + Sx$. *Journal of Differential Equations*, 16(2):270 – 280, 1974.