Constructive Non-Commutative Rank Computation Is in Deterministic Polynomial Time*

Gábor Ivanyos[1], Youming Qiao[2], and K Venkata Subrahmanyam[3]

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

2 **Centre for Quantum Software and Information,University of Technology Sydney, Australia**
Youming.Qiao@uts.edu.au

3 **Chennai Mathematical Institute, Chennai, India**
kv@cmi.ac.in

Abstract

Let $\mathcal{B}$ be a linear space of matrices over a field $\mathbb{F}$ spanned by $n \times n$ matrices $B_1, \ldots, B_m$. The non-commutative rank of $\mathcal{B}$ is the minimum $r \in \mathbb{N}$ such that there exists $U \leq \mathbb{F}^n$ satisfying $\dim(U) - \dim(\mathcal{B}(U)) \geq n - r$, where $\mathcal{B}(U) := span(\cup_{i\in[m]} B_i(U))$.

Computing the non-commutative rank generalizes some well-known problems including the bipartite graph maximum matching problem and the linear matroid intersection problem.

In this paper we give a deterministic polynomial-time algorithm to compute the non-commutative rank over any field $\mathbb{F}$. Prior to our work, such an algorithm was only known over the rational number field $\mathbb{Q}$, a result due to Garg et al, [20]. Our algorithm is constructive and produces a witness certifying the non-commutative rank, a feature that is missing in the algorithm from [20].

Our result is built on techniques which we developed in a previous paper [24], with a new reduction procedure that helps to keep the blow-up parameter small. There are two ways to realize this reduction. The first involves constructivizing a key result of Derksen and Makam [12] which they developed in order to prove that the null cone of matrix semi-invariants is cut out by generators whose degree is polynomial in the size of the matrices involved. We also give a second, simpler method to achieve this. This gives another proof of the polynomial upper bound on the degree of the generators cutting out the null cone of matrix semi-invariants.

Both the invariant-theoretic result and the algorithmic result rely crucially on the regularity lemma proved in [24]. In this paper we improve on the constructive version of the regularity lemma from [24] by removing a technical coprime condition that was assumed there.

1998 ACM Subject Classification F.2.0 General, F.2.1 Numerical Algorithms and Problems, I.1.2 Algorithms , F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases invariant theory, non-commutative rank, null cone, symbolic determinant identity testing, semi-invariants of quivers

Digital Object Identifier 10.4230/LIPIcs.ITCS.2017.55

* Research of the first author was also supported in part by the Hungarian National Research, Development and Innovation Office – NKFIH Grant K115288. Youming's research was supported by the Australian Research Council DECRA DE150100720. KV's research was supported by a grant from the Infosys foundation.

8th Innovations in Theoretical Computer Science Conference (ITCS 2017).
Editor: Christos H. Papadimitrou; Article No. 55; pp. 55:1–55:19

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

1.1 From the bipartite perfect matching problem to the commutative and non-commutative rank problems

Given a bipartite graph $G = (L \cup R, E)$ where $|L| = |R|$, the celebrated Hall's marriage theorem states that G has a perfect matching if and only if G has no *shrunk subsets*: $S \subseteq L$ is called a shrunk subset, if $|N(S)| < |S|$ where $N(S)$ denotes the set of neighbours of S.

Consider the following linear algebraic analogue of the bipartite perfect matching problem. Let $\mathbb{F}$ be a field, and let $U \cong V \cong \mathbb{F}^n$ be two vector spaces. We assume $\mathbb{F}$ is large (larger than a fixed polynomial in n is enough). U and V may be thought of as corresponding to sets of vertices. In a bipartite graph, each edge may be viewed as a "partial function" from the left vertex set to the right vertex set. Carrying out this analogue, in the linear algebraic setting we shall think of one linear map from U to V as one edge. After fixing bases of U and V this linear map is represented as a matrix. Let $M(n, \mathbb{F})$ denote the linear space of $n \times n$ matrices over $\mathbb{F}$. Then the edge set in the linear algebraic setting is just a set of matrices, and suppose we have m matrices $\{B_1, \ldots, B_m\} \subseteq M(n, \mathbb{F})$.

In a bipartite graph, a perfect matching is a subset of edges that form a bijective function from the left vertex set to the right vertex set. Choosing a subset of edges can be viewed as assigning 0 and 1 to the edges and then selecting those edges with 1's. This leads us to consider linear combinations of the matrices in the linear algebraic setting. That is, we are interested in the linear span of the given matrices $B_1, \ldots, B_m \in M(n, \mathbb{F})$, denoted as $\mathcal{B} = \text{span}(B_1, \ldots, B_m) \leq M(n, \mathbb{F})$. We call a linear subspace of $M(n, \mathbb{F})$ a *matrix space*. A "perfect matching" in the linear algebraic setting is then naturally a full-rank (non-singular) matrix as it is a bijective linear map between the left and the right vector spaces. A matrix space is *non-singular* if it contains a non-singular matrix, and *singular* otherwise.

It is also easy to obtain an analogue of shrunk subsets as in the Hall's marriage theorem. For $c \in \mathbb{N}$, we call $W \leq U$ a *c-shrunk subspace* of $\mathcal{B} = \text{span}(B_1, \ldots, B_m) \leq M(n, \mathbb{F})$, if $\dim(W) - \dim(\mathcal{B}(W)) \geq c$ where $\mathcal{B}(W) := \text{span}(\cup_{i \in [m]} B_i(W))$. We also call $W \leq U$ a shrunk subspace if it is a c-shrunk subspace for some $c \geq 1$. Clearly, if $\mathcal{B}$ possesses a shrunk subspace then it is singular.

A natural question is then whether the analogue of Hall's theorem holds, that is, whether every singular matrix space has a shrunk subspace. A counter-example is not hard to find, as evidenced by the linear space of 3×3 skew-symmetric matrices. That is, while perfect matchings and shrunk subsets are two sides of the same coin for bipartite graphs, non-singular matrices and shrunk subspaces are not in the linear algebraic setting. This gives rise to two natural algorithmic problems.

Input Given $B_1, \ldots, B_m \in M(n, \mathbb{F})$, let $\mathcal{B}$ be the matrix space spanned by these matrices.

The commutative rank problem Call the maximum rank over matrices in $\mathcal{B}$ the *commutative rank* of $\mathcal{B}$, denoted as $\text{crk}(\mathcal{B})$. The commutative rank problem asks to compute $\text{crk}(\mathcal{B})$.

The non-commutative rank problem Define the non-commutative corank to be the maximum $c \in \mathbb{N}$ such that there exists a c-shrunk subspace of $\mathcal{B}$, and the *non-commutative rank* to be $n - c$, denoted as $\text{ncrk}(\mathcal{B})$. The non-commutative rank problem asks to compute $\text{ncrk}(\mathcal{B})$.

The names of these problems were coined in Fortin and Reutenauer [19],[1] though both problems have been around since 1970's (see Section 1.2).

While these two problems are different in general, as suggested by the 3×3 skew-symmetric matrix space, they do coincide in some special cases. For this let us recall how the bipartite perfect matching problem and the linear matroid intersection problem can be cast as special instances of both problems.

Given a bipartite graph $G = (L \cup R, E)$ where $L = R = \{1, 2, \ldots, n\}$, for each edge $e = (i, j) \in E$ construct the elementary matrix $E_{i,j}$, which span a matrix space $\mathcal{B}$.

It is easy to see that the size of a maximum matching equals the maximum rank over matrices in $\mathcal{B}$ (over a large enough field). Hall's marriage theorem then implies that if the maximum rank is $n - c$ then there exists a c-shrunk subspace, and it is clear that if there exists a c-shrunk subspace then the maximum rank can be no larger than $n - c$. This shows that the commutative rank and the non-commutative rank coincide in this special case.

Lovász [27] observed that the linear matroid intersection problem (LMIP) can be cast as an instance of both the commutative and the non-commutative rank problems as follows. The input to LMIP is two tuples of vectors $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_m)$ where $a_i, b_j \in \mathbb{F}^n$. It asks to compute the maximum $s \in \mathbb{N}$ such that there exist $1 \le i_1 < \cdots < i_s \le m$ with both $\{a_{i_1}, \ldots, a_{i_s}\}$ and $\{b_{i_1}, \ldots, b_{i_s}\}$ being linearly independent. Construct m rank-1 matrices $a_i b_i^t$ spanning a matrix space $\mathcal{B}$. Using Edmonds' matroid intersection theorem, Lovász showed that in this case the commutative rank and the non-commutative rank of $\mathcal{B}$ coincide, and both are equal to the matroid intersection number.

1.2 Backgrounds to the commutative and non-commutative rank problems

As far as we are aware the commutative rank problem was first proposed by Edmonds [17] in 1967. It was then realized that it admits an efficient randomized algorithm via the Schwartz-Zippel lemma. Lovász [27] cast several problems from matroid theory, including the linear matroid intersection problem and the linear matroid parity problem, as special instances of the commutative rank problem. To decide whether the commutative rank is full is now better known as the *symbolic determinant identity testing* (SDIT) problem.

At present, SDIT stands at the frontier of proving arithmetic circuit lower bounds due to the celebrated work of Kabanets and Impagliazzo [25], which was recently improved by Carmosino et al [5]. They show that a deterministic efficient algorithm for SDIT implies the existence of a polynomial family such that its graph is in NE, but it cannot be computed by polynomial-size arithmetic circuits. Also, since determinants with affine entries are equivalent[33] to weakly-skew arithmetic circuits,[2], up to a polynomial overhead, SDIT is just another way of formulating the more well-known *polynomial identity testing* (PIT) problem for weakly-skew arithmetic circuits. The research into PIT has received a lot of attention since early 2000 (see the surveys [32, 30]).

[1] We explain the name choices here: from $B_1, \ldots, B_m \in M(n, \mathbb{F})$ one can construct a symbolic matrix $T = x_1B_1 + \cdots + x_mB_m$ where each entry is a linear form. When the variables x_i's commute, the rank of T over the rational function field is equal to the commutative rank (as we assumed $\mathbb{F}$ is large enough). If x_i's are non-commutative, there exists a non-commutative analogue of the rational function field, called the free skew field. Fortin and Reutenauer [19] (building on [8]) proved that the rank of T over the free skew field is equal to the non-commutative rank defined above.

[2] An arithmetic circuit is weakly skew if each product gate is of fan-in 2 and has at least one child labelled by a variable or a field element. The computation power of weakly skew circuit is between arithmetic formulas and arithmetic circuits.

The non-commutative rank problem can be traced back to 1970's, when Cohn[6, 7] raised this problem in his research on the free skew field.[3] There Cohn showed that this problem is decidable. This is already non-trivial, partly because his starting point was the free skew field, which itself is a complex object. This was improved to PSPACE by Cohn and Reutenauer [9] by reducing to testing the solvability of a system of multivariate polynomial equations. Fortin and Reutenauer [19] realized the connection between the non-commutative rank problem and the structure of matrix spaces,[4] and the definition of the non-commutative rank in terms of shrunk subspaces we have given above is due to them. Around the same time, Gurvits [21] came to the non-commutative rank problem in his study of the commutative rank problem, and endowed both the commutative and non-commutative rank problems with quantum information theoretic interpretations. In 2014, Hrubeš and Wigderson [22] studied non-commutative arithmetic formulas with divisions, and reduced the identity testing problem in this model to the non-commutative rank problem. In particular, they discovered that a good upper bound of a quantity in invariant theory (which will be explained below) implies the efficient elimination of divisions in arithmetic formulas with divisions, as well as an efficient randomized algorithm for the non-commutative rank problem.

We refer readers to [24, 20] for an introduction to the various avatars of the non-commutative rank problem.

Our main motivation to examine the non-commutative rank problem comes from its relation with the commutative one. For the purpose of getting arithmetic circuit lower bounds following [5], it is actually enough to put SDIT in NP, that is, to find a short witness which testifies the singularity of singular matrix spaces. Shrunk subspaces just form one natural class of singularity witnesses, as evidenced in its analogue with the shrunk subsets in Hall's marriage theorem and its crucial role in the linear matroid intersection problem. If we can efficiently distinguish matrix spaces with shrunk subspaces from those without then, for SDIT it would be enough to focus on singular matrix spaces without shrunk subspaces. While such spaces are known to have a rich structure, see [2, 18, 16], to further understand them and see how they can be efficiently recognized, is a natural approach to study SDIT.

1.3 Certifying matrix spaces without shrunk subspaces

Since 2015 there has been impressive progress on the non-commutative rank problem and the relevant invariant theoretic problem. To introduce this progress, let us review the link between the non-commutative rank problem and invariant theory.

Recall that the non-commutative rank problem asks to compute, given a matrix space $\mathcal{B} \leq M(n, \mathbb{F})$, the maximum c such that $\mathcal{B}$ has a c-shrunk subspace. For convenience, let us consider a decision version of this problem, that is to decide whether $\mathcal{B}$ has a shrunk subspace. One conceptual difficulty in tackling this problem is to find a short witness certifying that a matrix space $\mathcal{B}$ has no shrunk subspaces. It is clear that, if $\mathcal{B}$ has a non-singular matrix, then this non-singular matrix can be used to certify that $\mathcal{B}$ has no shrunk subspaces. But we've seen that, the space of 3×3 skew-symmetric matrices sk_3, has no non-singular matrices, nor shrunk subspaces. Our approach to resolving this difficulty is to introduce the following blow-up operation for matrix spaces. Given a matrix space $\mathcal{B} \leq M(n, \mathbb{F})$, the dth *blow-up* of

[3] The free skew field may be thought of as the non-commutative analogue of the rational function field.

[4] Fortin and Reutenauer were inspired by a paper by Eisenbud and Harris [18] in algebraic geometry. In fact matrix spaces with commutative rank bounded from above roughly correspond to certain torsion-free sheaves on projective spaces.

$\mathcal{B}$, denoted as $\mathcal{B}^{[d]}$, is defined to be the matrix space $M(d,\mathbb{F})\otimes\mathcal{B}\leq M(nd,\mathbb{F})$. It is easy to show the following.

▶ **Proposition 1.** *If $\mathcal{B}\leq M(n,\mathbb{F})$ has a c-shrunk subspace, then $\mathcal{B}^{[d]}$ has a cd-shrunk subspace.*

That is, the blow-up operation preserves shrunk subspaces. On the other hand, it is not immediately clear what the blow-up operation's effect is on matrix spaces without shrunk subspaces, though it can be shown that for sk_k, the space of skew-symmetric matrices of an odd size k, $\mathrm{sk}_k{}^{[2]}$ contains a non-singular matrix. Therefore the first question is whether for any matrix space $\mathcal{B}\leq M(n,\mathbb{F})$ without shrunk subspaces, $\mathcal{B}^{[d]}$ contains a non-singular matrix for some finite d. This turns out to be true, but to see it we need to go through several results in invariant theory and algebraic geometry.

Consider the group action of $(A,C)\in \mathrm{SL}(n,\mathbb{F})\times\mathrm{SL}(n,\mathbb{F})$ on $M(n,\mathbb{F})^{\oplus m}$ sending $(B_1,\ldots,B_m)$ to $(AB_1C^t,\ldots,AB_mC^t)$. This induces an action on the ring of polynomial functions on $M(n,\mathbb{F})^{\oplus m}$. Let $R(n,m)$ be the ring of those polynomials invariant under this action. $R(n,m)$ is called *the ring of matrix semi-invariants* (for m matrices of size $n\times n$) [24, 12]. The common zeros of all polynomials in $R(n,m)$, denoted as $N(R(n,m))$, is called the *nullcone* of this invariant ring in the invariant theory literature. The first link is the following result from invariant theory, proved using the celebrated Hilbert-Mumford criterion.

▶ **Theorem 2** ([4, 1]). *$(B_1,\ldots,B_m)$ is in $N(R(n,m))$ if and only if* $\mathrm{span}(B_1,\ldots,B_m)$ *has a shrunk subspace.*

Theorem 2 shows that matrix spaces with shrunk subspaces are characterized by those polynomials in $R(n,m)$. It is then desirable to know what polynomials in $R(n,m)$ look like. This task is usually resolved in the so-called first fundamental theorem for $R(n,m)$. In fact, it is this theorem that leads to the use of the blow-up operation.

▶ **Theorem 3** ([14, 31, 15, 1]). *Every homogeneous polynomial in $R(n,m)$ is of degree dn for some $d\in\mathbb{N}$, and is a linear combination of polynomials of the form* $\det(A_1\otimes X_1+\cdots+A_m\otimes X_m)$ *where X_i's are $n\times n$ variable matrices, and A_i's are $d\times d$ matrices over $\mathbb{F}$.*

Therefore, if $\mathrm{span}(B_1,\ldots,B_m)$ does not possess a shrunk subspace, then $(B_1,\ldots,B_m)$ is not in the null cone of $R(n,m)$ (Theorem 2). This implies that there exists some $(A_1,\ldots,A_m)\in M(d,\mathbb{F})^{\oplus m}$ such that $\det(A_1\otimes B_1+\cdots+A_m\otimes B_m)\neq 0$ (Theorem 3), which just says that $\mathcal{B}^{[d]}$ contains a non-singular matrix. This almost suggests that the blow-up operation would resolve the difficulty of certifying matrix spaces without shrunk subspaces, except that it is not immediately clear how large d needs to be for $\mathcal{B}^{[d]}$ to contain a non-singular matrix. To see that d is finite is classical: by Hilbert's basis theorem, $N(R(n,m))$ can be defined by finitely many polynomials, so d is also finite.

But if our hope is that the blow-up operation would *efficiently* certify matrix spaces without shrunk subspaces, just knowing d to be finite is not very useful – in fact, for this purpose we would require d to be upper bounded by a polynomial in n. Let $\sigma(R(n,m))$ be the smallest d such that $N(R(n,m))$ is defined by polynomials in $R(n,m)$ of degree no more than d. (This definition is valid for any invariant ring S, and to get an explicit upper bound on $\sigma(S)$ for invariant rings satisfying certain general conditions is an important research topic in invariant theory [28, 11].) Digging into the literature, from Derksen's work [11] it follows that $\sigma(R(n,m))\leq 1/4\cdot n^2\cdot 4^{n^2}$ for algebraically closed fields of characteristic 0.

Since 2015 there has been considerable progress towards a better upper bound on $\sigma(R(n,m))$. In [24] we show that $\sigma(R(n,m))\leq n!$ for all large enough fields. The key is the following so-called regularity lemma.

▶ **Lemma 4** (Regularity lemma for blow-ups, [24, Lemma 5.6]). *For $\mathcal{B} \leq M(n, \mathbb{F})$, assume that $|\mathbb{F}| > (nd)^{\Omega(1)}$. Then $\mathrm{crk}(\mathcal{B}^{[d]})$ is divisible by d.*

After [24] appeared, Derksen and Makam [12] discovered a concavity property of blow-ups, and showed that combining the regularity lemma with this property gives the following surprising result.

▶ **Theorem 5** ([12]). *Suppose $|\mathbb{F}| > n^{\Omega(1)}$. Then $\sigma(R(n,m)) \leq n^2 - n$.*

Theorem 5 immediately suggests an efficient approach to certify matrix spaces without shrunk subspaces: if $\mathcal{B}$ has no shrunk subspace, to see it we only need to exhibit a non-singular matrix in $\mathcal{B}^{[d]}$ for some $d \leq n-1$. A slightly more careful analysis suggests that if $\mathrm{ncrk}(\mathcal{B}) \geq r$, then to certify this we can exhibit a matrix in $\mathcal{B}^{[d]}$ of rank dr for some $d \leq r-1$.

By [11], this implies that $R(n,m)$ can be generated as a ring by those polynomials of degree $\leq O(n^8)$. Further consequences include improved degree bounds for generating semi-invariants of quivers (see [12]), since semi-invariants of quivers can be described as determinants of block matrices in general [15].

Soon after Derksen and Makam[12] announced their result, we discovered another argument based on the regularity lemma which gives $\sigma(R(n,m)) \leq n^2 + n$. While slightly worse in parameters, our argument is simpler than the one using the concavity property discovered by Derksen and Makam. We shall present this in Section 2.

1.4 Deterministic efficient algorithms for the non-commutative rank problem

As mentioned in Section 1.2, before 2015 it was only known that the non-commutative rank problem is in PSPACE. We gave a $\mathrm{poly}((n+1)!)$-time algorithm for this problem in 2015 [24]. Our algorithm can be viewed as a linear algebraic analogue of the augmenting path algorithm for the bipartite maximum matching problem, and relies heavily on a constructive proof of the regularity lemma.

Later that year Garg et al [20] showed that a careful analysis of an algorithm of Gurvits from 2003, [21], puts the non-commutative rank problem in P over $\mathbb{Q}$.

▶ **Theorem 6** ([20]). *Over $\mathbb{Q}$, the non-commutative rank problem is in* P.

Interestingly, Gurvits' algorithm is a quantum generalization of an algorithm for the bipartite maximum matching problem proposed by Linial et al,[26]: given the bipartite adjacency matrix of a bipartite graph, perform the row averaging and the column averaging operations alternatively. In [26] the authors proved that after a polynomially number of rounds, the resulting matrix is close to the identity matrix if and only if the bipartite graph has a perfect matching. Gurvits' algorithm is a beautiful and far-reaching generalization of the above idea to the setting of matrix spaces. It was originally used to tackle the commutative rank problem when the matrix space satisfies what Gurvits called the *Edmonds-Rado property*, namely those matrix spaces that are either non-singular, or have a shrunk subspace. The main innovation of Garg et al. in [20] is to realize that, by combining the exponential bounds on $\sigma(R(n,m))$ (either from [11] or [24]) with the capacity of operators introduced by Gurvits [21], Gurvits' algorithm actually solves the non-commutative rank problem over $\mathbb{Q}$. However, their algorithm fails to output a c-shrunk subspace and a matrix of rank $(n-c)d$ in $\mathcal{B}^{[d]}$ which together certify that the non-commutative rank is $n-c$.

We observed that Derksen and Makam's [12] concavity property of blow-ups can be constructivized, and this boosted our previous $\mathrm{poly}((n+1)!)$-time algorithm to polynomial

time. As mentioned later we discovered another, simpler, constructive argument, which also achieves the same bound. This leads to our main result.

▶ **Theorem 7** (Main theorem). *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 [20], our algorithm has the advantages of working with arbitrary large enough fields, and 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 $\mathbb{Q}$. In section 1.5 we show that the small finite fields case can also be handled.

▶ Remark.

(a) If the constructivized version of Derksen and Makam [12] is used (see Appendix A), then d in the above theorem can be improved 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 [21], asking if it is possible to decide efficiently, over fields of positive characteristics, whether or not there exists a non-singular matrix in a matrix space having the Edmonds-Rado property. Since the algorithm in Theorem 7 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.

1.5 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 8.** *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 decide whether $\mathcal{B}$ has a non-singular matrix, or a shrunk-subspace.*

1.6 A technical improvement of the regularity lemma

As mentioned in Section 1.3, the regularity lemma from [24] (Lemma 4) is the key to the polynomial bound on $\sigma(R(n, m))$. Furthermore, a constructive version of it from [24] is crucial

for our algorithm for the non-commutative rank problem. Specifically, we use the regularity lemma in the algorithm in the following situation: given $A \in \mathcal{B}^{[d]}$ of rank $(r-1)d+k$ where $1 < k < d$, we wish to construct $A' \in \mathcal{B}^{[d]}$ of rank $\geq rd$ efficiently. In [24, Lemma 5.7], this was achieved under the condition that, if $\mathrm{char}(\mathbb{F}) = p > 0$, then $p \nmid d$. In this paper, we remove this coprime condition. Roughly speaking, the proof of the regularity lemma in [24] made crucial use of central division algebras, and a classical construction of such algebras is based on cyclic field extensions. Therefore an efficient construction of cyclic field extensions is the basis of the constructive regularity lemma. In [24] an efficient construction of cyclic field extensions is presented assuming that the extension degree and the field characteristic are coprime. We remove this coprime condition to get a complete constructive proof of the regularity lemma. The technical details of this construction are given in Appendix B.

We note that recently Derksen and Makam obtained another proof of the regularity lemma in [13]. However they remark [13] that their proof is "less constructive".

Organization of the article

In Section 2 we present our proof of the n^2+n upper bound on $\sigma(R(n,m))$. In Section 3 we outline the algorithm in [24] and explain how the idea in Section 2 can be used to bring down its time complexity from exponential to polynomial, both intuitively (Section 3.1) and rigorously (Section 3.2). Section 4 contains the proof of Corollary 8. A proof of the full constructive regularity lemma is given in section 5. In Appendix A we show how the concavity property of Derksen and Makam can be constructivized. In Appendix B, we give an efficient construction of cyclic field extensions.

2 Another proof of a polynomial upper bound on defining the nullcone of matrix semi-invariants

In this short section we show how a simple argument based on the regularity lemma (Lemma 4) proves that $\sigma(R(n,m)) \leq n^2+n$. Let us remark again that Derksen and Makam [12] were the first to prove that $\sigma(R(n,m)) \leq n^2-n$. They too relied crucially on the regularity lemma from [24]. Here we present a simpler proof, albeit with a slightly worse parameter.

▶ **Theorem 9.** *Suppose* $|\mathbb{F}| > n^{\Omega(1)}$. *Then* $\sigma(R(n,m)) \leq n^2+n$.

Proof. Suppose $\mathcal{B} \leq M(n,\mathbb{F})$ is of dimension m. To prove the statement, we need to show that for any such $\mathcal{B}$, $\mathcal{B}^{[d]}$ contains a non-singular matrix for some $d \leq n+1$. As explained in Section 2, by Theorem 2 and 3, as well as Hilbert's basis theorem, $\mathcal{B}^{[d]}$ contains a non-singular matrix A for some $d \in \mathbb{N}$. If $d > n+1$, then note that $A \in \mathcal{B}^{[d]} \leq M(d,\mathbb{F}) \otimes M(n,\mathbb{F})$ is a $d \times d$ block matrix where each block is of size $n \times n$. Let A' be the right lower $(d-1) \times (d-1)$ block matrix of A; note that $A' \in \mathcal{B}^{[d-1]}$. Since A' is obtained from A by removing n rows and n columns, we have $\mathrm{rk}(A') \geq dn-2n = dn-n+1-(n+1) > dn-n+1-d = (d-1)(n-1)$, which gives $\mathrm{crk}(\mathcal{B}^{[d-1]}) > (d-1)(n-1)$. By Lemma 4, $\mathrm{crk}(\mathcal{B}^{[d-1]})$ has to be divisible by $(d-1)$, so $\mathrm{crk}(\mathcal{B}^{[d-1]}) = (d-1)n$ is of full rank. Continuing this way we can reduce d to be no more than $n+1$. ◀

3 Proof of the main theorem

As mentioned in Section 1.4, the algorithm for Theorem 7 is obtained by combining the algorithm in [24] with the idea in the proof of Theorem 9. For the readers convenience we give an intuitive explanation of the algorithm from [24] in Section 3.1; the reader is referred

to [24] for a rigorous treatment. We also explain why it runs in exponential time and how the idea in the proof of Theorem 9 can reduce its complexity to polynomial. Section 3.2 gives a rigorous treatment of the algorithm that proves Theorem 9.

3.1 Outline of the algorithm in [24]

The algorithm in [24] can be viewed as an analogue of the augmenting path algorithm for the bipartite 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.

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 has been developed in [23]. 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 $\ell \in [n]$, $W_1 < W_2 < \cdots < W_\ell = W_{\ell+1} = \ldots$, and this is called *the second Wong sequence* of $(A, \mathcal{B})$.[5][6] W_ℓ is called the *limit subspace* of this sequence.

In [23], it was proved that $W_\ell \leq \mathrm{im}(A)$ if and only if $\mathcal{B}$ has a $\mathrm{cork}(A)$-shrunk subspace[7]. 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 $\mathrm{crk}(\mathcal{B}) < \mathrm{ncrk}(\mathcal{B})$. Thanks to Theorem 9, the key to resolve this difficulty is to find $A' \in \mathcal{B}^{[d]}$ of rank $\geq (r+1)d$ for some not too large d. (So that the scaled-down rank $\mathrm{rk}(A')/d$ is larger than r.) This is accomplished in two steps.

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 k be the smallest index j with $W_j \not\leq \mathrm{im}(A)$. Obviously $k \leq r + 1$. Then there exist indices $1 \leq i_1, \ldots, i_k \leq m$ such that $B_1 \cdots B_k \ker(A) \not\leq \mathrm{im}(A)$. It would be nice if one could find a *single* matrix $B \in \mathcal{B}$ such that $B^k \ker(A) \not\leq \mathrm{im}(A)$: if this indeed happens, then for some λ and μ from a subset of the base field of size at least $r + 1$ one would have $\mathrm{rk}(\mu A + \lambda B) > \mathrm{rk}(A)$. This is because of the result from [3], where it is proved that for two-dimensional matrix spaces the commutative and the non-commutative ranks coincide.

---

[5] The first Wong sequence is the dual of the second one. This naming convention is due to Wong [34] who defined the two sequences for the special case when $\mathcal{B}$ is of dimension 1. See [23] for more details.

[6] 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 [23].

[7] At the time of writing the first version of [23], the authors were unaware of [19] where actually this statement has already appeared. The real value added in [23] 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.

The main ingredient of the algorithm in [23] was a method to find such a $B \in \mathcal{B}$ in certain special cases. The idea in [24] is that if we relax our requirement and instead work in $\mathcal{B}^{[d]}$, then this can be achieved in a simple way, for every matrix space $\mathcal{B}$. Observe that $A \otimes I_d$ is a matrix from $\mathcal{B}^{[d]}$ of rank rd. Let $E_{i,j}$ be the elementary matrix with the (i,j)th entry being 1 and others 0. Put $\widehat{B} = B_1 \otimes E_{1,2} + B_2 \otimes E_{2,3} + \ldots B_{k-1} \otimes E_{k-1,k} + B_k \otimes E_{k,1} \in \mathcal{B}^{[d]}$. Then $\widehat{B}^k = (B_1 \cdots B_k) \otimes E_{1,1} + (B_2 \cdots B_k B_1) \otimes E_{2,2} + \ldots + (B_k B_1 \cdots B_{k-1}) \otimes E_{k,k}$, whence $\widehat{B}^k \ker(A \otimes I_d) \not\leq \mathrm{im}(A \otimes I_d)$. Therefore $\widehat{A} = \mu(A \otimes I_d) + \lambda \widehat{B}$ will have rank larger than rd for some λ and μ from a subset of the base field of size $rd + 1$.

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 by the constructive regularity lemma ([24, Lemma 5.7]; see Lemma 11 in Section 3.2).

This $A' \in \mathcal{B}^{[d]}$ of rank $\geq (r+1)d$ where $d = r + 1$ certifies that $\mathrm{ncrk}(\mathcal{B}) \geq r + 1$. (If $\mathrm{ncrk}(\mathcal{B}) \leq r$ then $\mathrm{crk}(\mathcal{B}^{[d]}) \leq rd$ for any d by Proposition 1.) So after these two steps we obtain A' of rank $r'd$ where $r' > r$.

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'.[8] 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 $\mathcal{B}$ [24, 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 $r'' > r'$.

So the right approach to carrying out the augmenting path idea in this setting is to play with shrunk subspaces on 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 [24] runs in time $\mathrm{poly}(n!)$.

However, the idea in the proof of Theorem 9 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 from exponential time to polynomial time. We make the above idea rigorous.

3.2 The algorithm for the main theorem

In this subsection we prove Theorem 7. Here it is easier to work with $\mathcal{B}^{\{d\}} := \mathcal{B} \otimes M(d, \mathbb{F})$ instead of $\mathcal{B}^{[d]} = M(d, \mathbb{F}) \otimes \mathcal{B}$. This does not change anything, as $\mathcal{B}^{[d]}$ is isomorphic to $\mathcal{B}^{\{d\}}$. The point is that we now think of matrices in the blow-up space as $n \times n$ block matrices with blocks of size $d \times d$. We first recall some notions from [24].

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 ([24, 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 $sd \geq s'$.

[8] 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 [24].

We shall also need the following useful lemma.

▶ **Lemma 10** (Data reduction, [10] and [24, Lemma 5.3]). *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 $A' = \sum_i a'_i B_i \in \mathcal{B} \otimes_{\mathbb{F}} \mathbb{K}$ of rank at least r. Then we can find $A = \sum_i a_i B_i$ 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 $\sum \sigma_i A_i$ where $a''_i \in \{a'_1, \ldots, a'_m\} \cup S$.*

We now present the formal statement of the constructive regularity lemma in its full generality, with an addition of a technical notion that will be useful for the proof of Theorem 7. 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 $1 \leq i_1 < \cdots < i_r \leq n$ and $1 \leq j_1 < \cdots < j_r \leq n$. 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 11** (The complete constructive regularity lemma). *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 non-singular (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 [24, Lemma 5.7]. We will remove this coprime condition in .

The main technical ingredient of our algorithm will be the following result,[24, Theorem 5.10] based on Lemma 11, and the second Wong sequences introduced in [19] and [23]. 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.

▶ **Theorem 12** ([24, Theorem 5.10]). *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 $d' > r$ is any positive integer.*

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.*

The sentence on the $(r+1) \times (r+1)$ window was not explicitly stated in [24]. However, the algorithm in its proof contains, as a last step, a call to the method behind Lemma 11. Also, the theorem was stated only under the assumption that d was not divisible by $\mathrm{char}(\mathbb{F})$ because of this last call. As the algorithm up to this step constructs a matrix of rank greater than rdd', the complete constructive regularity lemma, as stated in Lemma 11, makes it possible to dispense with that assumption.

To complete the proof Theorem 7, the regularity lemma needs to be accompanied with a reduction procedure that keeps the blow-up parameter small. In this section we use our method, which follows immediately from the proof of Theorem 9. The method based on the Derksen-Makam idea is presented in Appendix A.

▶ **Lemma 13.** *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 $(d-1)n \times (d-1)n$ submatrix 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 11, we obtain $A' \leq \mathcal{B}^{\{d-1\}}$ of rank $(d-1)n$. ◀

Proof of Theorem 7. Let $B_1, \ldots, B_m$ be the input basis for $\mathcal{B}$. The algorithm is an iteration based on Theorem 12. 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 12 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 $d' > r+2$ then we apply Lemma 13 as follows. The n in the statement of Lemma 13 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 10 to ensure that the entries of T_i are from the prescribed subset of $\mathbb{F}$, and continue the iteration with this new matrix A. ◀

4 Proof of Corollary 8: the case of small finite fields

We only need to prove Corollary 8 (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 $\mathbb{F}$. Observe first 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. [23, Section 2]). Note that unlike the non-commutative rank the commutative rank may get larger if we go to an extension field of a too-small field.

▶ **Lemma 14.** *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 7. Note that $g \leq O(\log_{|\mathbb{F}|} n)$. Viewing $\mathcal{B}$ as a matrix space over $\mathbb{K}$, we apply Theorem 7 to compute $\mathrm{ncrk}_{\mathbb{K}}(\mathcal{B})$, which is equal to $r = \mathrm{ncrk}_{\mathbb{F}}(\mathcal{B})$ by Lemma 14. 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 $\mathcal{B}$.

5 The full constructive regularity lemma

In this section we prove Lemma 11. The proof makes use of the following two results from [24], which we reproduce here as Proposition 15 and Lemma 16. Roughly speaking, Proposition 15 states that from cyclic field extensions one can construct central division algebras. Lemma 16 is the conditional constructive regularity lemma, where the condition required there is an efficient construction of central division algebras.

▶ **Proposition 15** ([24, 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 σ of $M(d, \mathbb{K}(Y))$ such that the $\mathbb{K}(Y^d)$-linear span of σ 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, are also* $\mathrm{poly}(d)$.

▶ **Lemma 16** (Conditional regularity [24, 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 σ of $M(d, \mathbb{K}(Y))$ such that the $\mathbb{K}(Y^d)$-linear span of σ 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 σ. 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).*

From the above two results, the missing piece is just an efficient construction of cyclic field extensions. In [24] such a construction based on Kummer extensions was presented assuming $\mathrm{char}(\mathbb{F})$ and d are coprime. The main issue with the case when d is divisible by $\mathrm{char}(\mathbb{F})$ is that the proof requires an efficient construction of an appropriate Artin-Schreier-Witt extension of $\mathbb{F}_p(x)$, not known to us when writing [24]. Now we have such a construction leading to the following lemma whose proof we give in appendix B.

▶ **Lemma 17.** *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 of Lemma 11. The statement, except the window part, readily follows by plugging Lemma 17 and Proposition 15 to Lemma 16.

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. ◀

Acknowledgements. We would like to thank the authors of [20] and of [12] for sharing their ideas with us and making it possible for us 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.

References

1 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.

2 M. D. Atkinson. Primitive spaces of matrices of bounded rank. II. *Journal of the Australian Mathematical Society (Series A)*, 34(03):306–315, 1983.

3 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.

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

5 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. doi:10.4230/LIPIcs.APPROX-RANDOM.2015.645.

6 P. M. Cohn. The word problem for free fields. *J. Symbolic Logic*, 38(2):309–314, 06 1973. URL: http://projecteuclid.org/euclid.jsl/1183738636.

7 P. M. Cohn. The word problem for free fields: A correction and an addendum. *J. Symbolic Logic*, 40(1):69–74, 03 1975. URL: http://projecteuclid.org/euclid.jsl/1183739310.

8 P. M. Cohn. *Skew Fields: Theory of General Division Rings*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995. URL: http://books.google.com.au/books?id=u-4ADgUgpSMC.

9 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.

10 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.

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

12 Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. preprint ArXiv:1512.03393, 2015.

13 Harm Derksen and Visu Makam. On non-commutative rank and tensor rank. Preprint ArXiv:1606.06701, 2016.

14 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.

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

16 Jan Draisma. Small maximal spaces of non-invertible matrices. *Bulletin of the London Mathematical Society*, 38:764–776, 10 2006. doi:10.1112/S0024609306018741.

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

18 David Eisenbud and Joe Harris. Vector spaces of matrices of low rank. *Advances in Mathematics*, 70(2):135 – 155, 1988. doi:10.1016/0001-8708(88)90054-0.

19 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.

20 Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. Preprint ArXiv:1511.03730. To appear in FOCS 2016, 2015.

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

22 Pavel Hrubeš and Avi Wigderson. Non-commutative arithmetic circuits with division. *Theory of Computing*, 11:357–393, 2015. doi:10.4086/toc.2015.v011a014.

23 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. doi:10.1016/j.jcss.2015.04.006.

24 Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Non-commutative Edmonds' problem and matrix semi-invariants. Preprint arXiv:1508.00690, to appear in Computational Complexity, doi:10.1007/s00037-016-0143-x, 2015. doi:10.1007/s00037-016-0143-x.

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

26 Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4):545–568, 2000. doi:10.1007/s004930070007.

27 László Lovász. Singular spaces of matrices and their application in combinatorics. *Boletim da Sociedade Brasileira de Matemática-Bulletin/Brazilian Mathematical Society*, 20(1):87–99, 1989.

28 Vladimir L Popov. The constructive theory of invariants. *Izvestiya: Mathematics*, 19(2):359–376, 1982.

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

30 Nitin Saxena. Progress on polynomial identity testing - II. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:186, 2013. URL: http://eccc.hpi-web.de/report/2013/186.

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

32 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010. doi:http://dx.doi.org/10.1561/0400000039.

33 Seinosuke Toda. Classes of arithmetic circuits capturing the complexity of computing the determinant. *IEICE Transactions on Information and Systems*, 75(1):116–124, 1992.

34 Kai-Tak Wong. The eigenvalue problem $\lambda Tx + Sx$. *Journal of Differential Equations*, 16(2):270 – 280, 1974. doi:10.1016/0022-0396(74)90014-X.

A Constructivizing the result of Derksen and Makam

Here is an algorithmic version of Lemma 2.7 of [12]. Let $M(k \times \ell, \mathbb{F})$ be the space of $k \times \ell$ matrices over $\mathbb{F}$, and define $\mathcal{B}^{\{k,\ell\}} := \mathcal{B} \otimes M(k \times \ell, \mathbb{F})$.

▶ **Lemma 18.** *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 from [10] to keep intermediate (as well as the final) data small.

If (1) is violated then, like in [12], 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 [12], 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 $U \leq V \cap W$ and the row space of A is $V + W$, we have $r(k, \ell) \leq \dim U \leq \dim(V \cap W) = \dim V + \dim W - \dim V + W = \dim V + \dim W - r(k + 2, \ell)$. It follows that $\dim V + \dim W \geq r(k, \ell) + r(k + 2, \ell)$, 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 Lemma 11.

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 $N^3 n$. ◀

And here is essentially Proposition 2.10 of [12]. We include a proof (which is almost literally the same as the proof in [12]) here for completeness. We note that this lemma deals only with the property of certain families of functions, without referring to matrices.

▶ **Lemma 19** ([12, Proposition 2.10]). *Assume that $N > n > 0$, $r : \{0,1,\ldots,N\}^2 \to \mathbb{Z}$ is a function with $0 \le r(k,\ell) \le \min(k,\ell)n$ for $k,\ell \in \{0,1,\ldots,N\}$ also satisfying (1)–(5) of Lemma 18. 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.

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

From $d \ge n-1$ 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 18 in the proof of Theorem 7, 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 $(p,q) \ne (d',d')$. 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 7.

B 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 20.** *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 [29]. 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 $K_0 = \mathbb{F}_p(Z)$. 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:

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

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 $\mathrm{Tr}_{K_j:K_{j-1}}(\omega_j^{p-1}\sigma) = -\sigma$ for every $\sigma \in K_{j-1}$, whence $\mathrm{Tr}_{K_j:K_0}(\omega_j^{p-1}\sigma) = -\mathrm{Tr}_{K_{j-1}:K_0}(\sigma)$. 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

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

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 ($\ell \leq p^j$)

Assume by induction that the structure constants of K_j with respect to the basis σ_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 $1 \leq \ell < p^j$ (written in the same basis). For $j = 1$ this holds with $\Delta_1 = 1$. (To see this, observe that for $0 \leq k, \ell < p$, 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 σ_j using Eq. B.1, we obtain that its coordinates are polynomials of

degree at most $(2p+1)\Delta_j$. This is because $(-1)^j\beta_j \in \sigma_j$, whence $\beta_j^{\sigma^\ell}$ has coordinates of polynomials of degree bounded by Δ_j. In Eq. B.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 $\omega_{j+1}^k\sigma_1$ and $\omega_{j+1}^\ell\sigma_2$ of σ_{j+1}. Here $k,\ell < p$ and $\sigma_1,\sigma_2 \in \sigma_j$. The coordinates of the product $\sigma_1\sigma_2$ with respect to σ_j are polynomials of degree at most Δ_j. The same holds for the product $\omega_{j+1}^{k+\ell}\sigma_1\sigma_2$ 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 $\omega_{j+1}^{k+\ell}\sigma_1\sigma_2$ is the sum of $\omega_{j+1}^{1+k+\ell-p}\sigma_1\sigma_2$ and $\alpha_j\sigma_1\sigma_2$. 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 $\omega_{j+1}^k\sigma$ by σ_{j+1}^ℓ, where $1 \le \ell < p^{j+1}$, $1 \le k \le p-1$ and $\sigma \in \sigma_j$. This conjugate is $(\omega_{j+1}^{\sigma_{j+1}^\ell})^k\sigma^{\sigma_{j+1}^\ell}$. The second term equals $\sigma^{\sigma_j^\ell}$ 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 σ_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 σ_j, whence we conclude that $(\omega_{j+1}^{\sigma_{j+1}^\ell})^k$ has, in terms of σ_{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 17** (restated). *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, $d = d_1p^s$, 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 $Y_1^iY_1^j = Y_1^{i+j}$ 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 $\zeta^jY_1^j$ 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 [24].

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 $d = d_1p^s$. 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 Lemma 20, 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 $\sigma_1 \otimes \sigma_2$. ◀