

MATRIX EVALUATIONS OF NONCOMMUTATIVE RATIONAL FUNCTIONS AND WARING PROBLEMS

MATEJ BREŠAR AND JURIJ VOLČIČ

ABSTRACT. Let r be a nonconstant noncommutative rational function in m variables over an algebraically closed field \mathbb{K} of characteristic 0. We show that for n large enough, there exists an $X \in M_n(\mathbb{K})^m$ such that $r(X)$ has n distinct and nonzero eigenvalues. This result is used to study the linear and multiplicative Waring problems for matrix algebras. Concerning the linear problem, we show that for n large enough, every matrix in $\mathfrak{sl}_n(\mathbb{K})$ can be written as $r(Y) - r(Z)$ for some $Y, Z \in M_n(\mathbb{K})^m$. We also discuss variations of this result for the case where r is a noncommutative polynomial. Concerning the multiplicative problem, we show, among other results, that if f and g are nonconstant polynomials, then, for n large enough, every nonscalar matrix in $\mathrm{GL}_n(\mathbb{K})$ can be written as $f(Y) \cdot g(Z)$ for some $Y, Z \in M_n(\mathbb{K})^m$.

1. INTRODUCTION

Throughout, \mathbb{K} stands for an algebraically closed field of characteristic 0. Let $\underline{x} = (x_1, \dots, x_m)$ be a tuple of m freely noncommuting variables, let $\mathbb{K}\langle\underline{x}\rangle$ be the free \mathbb{K} -algebra of noncommutative polynomials in \underline{x} , and let $\mathbb{K}\langle\langle\underline{x}\rangle\rangle$ be its universal skew field of fractions [Coh06], the free field of noncommutative rational functions in \underline{x} . For example, $x_1^2 + x_1x_2 - x_2x_1 - x_1 + 2$ is a noncommutative polynomial, and $(x_1x_2^{-1}x_1 - x_2)^{-1} - x_2^{-1}$ is a noncommutative rational function. Our most general results concern elements r from $\mathbb{K}\langle\langle\underline{x}\rangle\rangle$, but are also new, and of particular interest, in the special case where $r \in \mathbb{K}\langle\underline{x}\rangle$.

Noncommutative rational functions originated in noncommutative algebra [Ami66, Coh06, GGRW05] and automata theory [BR11]. Nowadays, they prominently appear in free analysis [K-VV12, K-VV14], free probability [HMS18, CMMPY22] and free real algebraic geometry [HMV06, AHKM18, Vol21], which study analytic, geometric and asymptotic properties of their matrix evaluations. More precisely, noncommutative rational functions are (after noncommutative polynomials) the most tractable class of so-called *free functions*, which are maps on tuples of matrices of all sizes that preserve direct sums and similarities. When dealing with $r \in \mathbb{K}\langle\langle\underline{x}\rangle\rangle$, we will be interested in the matrix evaluations $r(X)$ with $X \in M_n(\mathbb{K})^m$ (more precisely, X must belong to the domain of r) for arbitrary n . Here, $M_n(\mathbb{K})$ stands for the algebra of $n \times n$ matrices over \mathbb{K} .

Our fundamental result, from which all else follows, states that given a nonconstant $r \in \mathbb{K}\langle\langle\underline{x}\rangle\rangle$, for every large enough n there exists an X in $M_n(\mathbb{K})^m$ such that $r(X)$ has

2020 *Mathematics Subject Classification.* 16R10, 16S10, 16S50, 16S85, 47A56.

Key words and phrases. Noncommutative polynomial, noncommutative rational function, Waring problem, matrix algebra.

M. Brešar was supported by Grant P1-0288, ARIS (Slovenian Research and Innovation Agency). J. Volčič was supported by the NSF grant DMS-1954709.

n distinct eigenvalues (Theorem 2.7). Moreover, these eigenvalues can be chosen to be nonzero (Corollary 2.9). The proof of this result consists of two natural parts. Firstly, one investigates the image of a noncommutative rational function on matrices of a fixed prime size using tools from central simple algebras and algebraic geometry. Then one uses ampliation technique inherent in free analysis along with the classical number theory to glue together the preceding size-dependent conclusions.

Our result on eventually attaining images with pairwise distinct eigenvalues conforms organically with the perspective of free analysis and free probability, which investigate those features of matrix evaluations of noncommutative rational functions that persist for matrices of large sizes. On the other hand, our main motivation behind this result is its applicability to Waring problems for matrix algebras. We explain this in the next paragraphs.

The classical Waring's problem, proposed by Waring in the 18th century and solved affirmatively by Hilbert in 1909, asks whether for every natural number k there exists a natural number $g(k)$ such that every natural number is a sum of $g(k)$ k th powers. Various variations of this problem occur in different areas of mathematics, including group theory. In [LST11], Larsen, Shalev, and Tiep proved that for any word $w \neq 1$, $w(G)^2 = G$ for every finite non-abelian simple group G of sufficiently high order (here, $w(G)$ denotes the image of the word map induced by w). More generally, if w, w' are two nontrivial words, then $w(G)w'(G) = G$.

Seeking for analogs of this result in noncommutative algebra, it is natural to replace the role of words by noncommutative polynomials, and the role of finite simple groups by matrix algebras. This type of Waring problem was initiated in [Bre20] and discussed further in [BŠ23a, BŠ23b, Che23, PP23]. More specifically, the following question, raised as Question 4.8 in [BŠ23a] and studied (but not solved) in [BŠ23b], emerged as central: Given a nonconstant polynomial $f \in \mathbb{K}\langle x \rangle$, does $f(M_n(\mathbb{K})) - f(M_n(\mathbb{K}))$ contain the Lie algebra of trace zero matrices $\mathfrak{sl}_n(\mathbb{K})$ for all sufficiently large n ? Here, $f(M_n(\mathbb{K}))$ stands for the image of f in $M_n(\mathbb{K})$ (i.e., $f(M_n(\mathbb{K}))$ is the set of all $f(A_1, \dots, A_m)$ with $A_i \in M_n(\mathbb{K})$), and $f(M_n(\mathbb{K})) - f(M_n(\mathbb{K}))$ for $\{B - B' : B, B' \in f(M_n(\mathbb{K}))\}$. As a byproduct of the result on distinct eigenvalues discussed above, we will solve this problem, not only for noncommutative polynomials but for noncommutative rational functions; see Theorem 3.2. In later parts of Section 3, we discuss extensions and variations of this result. In particular, we consider the question of when matrices with nonzero trace are linear combinations of two matrices from $f(M_n(\mathbb{K}))$ (see Theorem 3.5). Also, we obtain (partial) extensions of our results from matrices over \mathbb{K} to algebras \mathcal{A} satisfying $A \cong M_2(\mathcal{A})$ (see Theorem 3.10).

The problem of presenting matrices as linear combinations of matrices from the image of a rational function or polynomial may be called the *linear Waring problem*. The fundamental theorem on eigenvalues also enables us to consider a different, *multiplicative Waring problem* for polynomials and rational functions, which is even closer to the group-theoretic one and has, to the best of our knowledge, not yet been studied in the literature. This problem is the topic of Section 4. One of our results, Theorem 4.6, states that if

$f, g \in \mathbb{K}\langle x \rangle$ are nonconstant polynomials, then, for n large enough, every nonscalar matrix in $\mathrm{GL}_n(\mathbb{K})$ is contained in $f(M_n(\mathbb{K})) \cdot g(M_n(\mathbb{K}))$. We also mention Theorem 4.9 which in particular shows that if $f \in \mathbb{K}\langle x \rangle$ is a nonzero polynomial with zero constant term, then for n large enough, every matrix in $M_n(\mathbb{K})$ is a product of twelve matrices from $f(M_n(\mathbb{K}))$.

We do not know at present whether scalar matrices should really be excluded in Theorem 4.6. This is one of the problems that are left open. It is interesting that a genuinely different problem that arises in Section 3 also involves scalar matrices (see Question 3.7). We also do not know whether Theorem 4.9 holds for nonconstant polynomials with nonzero constant term (and, on the other hand, we believe that the number twelve can be lowered).

Acknowledgments. The authors thank Igor Klep and Peter Šemrl for insightful discussions.

2. EIGENVALUES OF MATRIX EVALUATIONS OF NONCOMMUTATIVE RATIONAL FUNCTIONS

For a comprehensive study of the skew field of noncommutative rational functions, also called the free skew field, we refer to [Coh06]. Analogous to how a (usual) rational function can be given as a fraction of different pairs of polynomials, a noncommutative rational function can be given by different formal rational expressions. For example, expressions $x_2(1 - x_1x_2)^{-1}x_2$ and $(x_1^{-1}x_2^{-1} - 1)^{-1}$ represent the same noncommutative rational function, in the sense that their evaluations agree on all matrix tuples where both are well-defined. Given $r \in \mathbb{K}\langle x \rangle$, let $\mathrm{dom}_n r$ be the domain of r in $M_n(\mathbb{K})^m$, defined as the union of the domains of all representatives of r (formal rational expressions) in $M_n(\mathbb{K})^m$ [K-VV12, Vol17].

Let $\Xi(n, \xi)$ be an m -tuple of $n \times n$ generic matrices, where $\Xi(n, \xi)_k = (\xi_{kij})_{i,j=1}^n \in M_n(\mathbb{K}[\xi])$ and

$$\underline{\xi} = (\xi_{kij} : 1 \leq i, j \leq n, 1 \leq k \leq m)$$

is a tuple of independent commuting indeterminates, viewed as the coordinates of the affine space $M_n(\mathbb{K})^m$. Let UD_n be the universal division \mathbb{K} -algebra of degree n generated by $\Xi(n, \xi)_1, \dots, \Xi(n, \xi)_m$ [Row80, Section 3.2].

Lemma 2.1. *For every $r \in \mathbb{K}\langle x \rangle$ there exist $\delta \in \mathbb{N}$, an affine matrix pencil $L = A_0 + A_1x_1 + \dots + A_mx_m$ with $A_j \in M_\delta(\mathbb{K})$ and $b, c \in \mathbb{K}^\delta$ such that $r = c^t L^{-1} b$ holds in $\mathbb{K}\langle x \rangle$, and*

$$\mathrm{dom}_n r \supseteq \{X \in M_n(\mathbb{K})^m : \det L(X) \neq 0\} \neq \emptyset$$

for all $n \geq \delta - 1$.

Proof. The statement holds by [CR99, Corollary 1.3] and [DM17, Propositions 1.12 and 2.10]. □

Given $r \in \mathbb{K}\langle x \rangle$, its representation as $r = c^t L^{-1} b$ as in Lemma 2.1 is called a *linear representation* or *realization* of r [CR99] of size δ . Moreover, the linear representation is *minimal* if δ is minimal; by [CR99, Corollary 1.6], a minimal linear representation of r is unique up to a left-right basis change.

Remark 2.2. Note that for $f \in \mathbb{K}\langle x \rangle$, we have that f vanishes (or is singular) on $M_n(\mathbb{K})^m$ if f vanishes (or is singular) on $M_{n+1}(\mathbb{K})^m$. On the other hand, for $r \in \mathbb{K}\langle x \rangle$ it can happen that r vanishes on $\text{dom}_{n+1} r \neq \emptyset$ and r is non-singular on $\text{dom}_n r \neq \emptyset$ [Ber76]. Nevertheless, Lemma 2.1 guarantees that this can only happen for small n (relative to the minimal size of a linear representation of r).

Lemma 2.3. *Let $r \in \mathbb{K}\langle x \rangle \setminus \mathbb{K}$ have a linear representation of size δ . Then for all $n \geq 2\delta$,*

$$(1) \quad \{X \in \text{dom } r : r(X) \notin \mathbb{K} \cdot I\}$$

is a nonempty Zariski open subset of $M_n(\mathbb{K})^m$.

If $r \in \mathbb{K}\langle x \rangle \setminus \mathbb{K}$ is of degree d , then (1) is nonempty for all $n \geq \lceil \frac{d}{2} \rceil + 1$.

Proof. Let x_0 be an auxiliary noncommuting variable. By assumption, $(x_0 r - r x_0)^{-1}$ is a well-defined noncommutative rational function in $m + 1$ variables. If $r = c^t L^{-1} b$ is a linear representation of size δ , then

$$\begin{aligned} (x_0 r - r x_0)^{-1} &= (c^t L^{-1} b x_0 - c^t x_0 L^{-1} b)^{-1} \\ &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} L & 0 & b x_0 \\ 0 & L & b \\ -c^t & c^t x_0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =: \tilde{c}^t \tilde{L}^{-1} \tilde{b} \end{aligned}$$

is a linear representation of size $\tilde{\delta} = 2\delta + 1$, and (1) contains the projection of $\text{dom}_n(x_0 r - r x_0)^{-1} \subset M_n(\mathbb{K})^{m+1}$ onto its last m components in $M_n(\mathbb{K})^m$, and

$$\text{dom}_n(x_0 r - r x_0)^{-1} \supseteq \{X \in M_n(\mathbb{K})^{m+1} : \det \tilde{L}(X) \neq 0\} \neq \emptyset$$

for $n \geq \tilde{\delta} - 1$ by [DM17, Propositions 1.12 and 2.10]. Therefore (1) is nonempty for $n \geq \tilde{\delta} - 1$; it is also evidently Zariski open.

Lastly, if r is a nonconstant noncommutative polynomial of degree d , then $x_0 r - r x_0$ is a nonzero polynomial on $M_n(\mathbb{K})^m$ for $d + 1 \leq 2n - 1$ by [Row80, Lemma 1.4.3]. In particular, (1) is nonempty (and clearly Zariski open) for $n \geq \lceil \frac{d}{2} \rceil + 1$. \square

Lemma 2.4. *Let $r \in \mathbb{K}\langle x \rangle$ with $\text{dom}_n r \neq \emptyset$ and $\lambda \in \mathbb{K}$. If λ is an eigenvalue of $r(X)$ for all X in a Zariski dense subset of $\text{dom}_n r$, then $r(\Xi(n, \xi)) = \lambda I$.*

Proof. By the assumption we have $\det(r(\Xi(n, \xi)) - \lambda I) = 0$, so $r(\Xi(n, \xi)) - \lambda I$ is a zero divisor in UD_n . Since the latter is a division algebra, it follows $r(\Xi(n, \xi)) = \lambda I$. \square

Let R be a commutative domain, and $f \in R[t]$. The *discriminant* of f [GKZ94, Section 12.1.B], $\text{disc}(f) \in R$, is a polynomial in the coefficients of f , with the property that $\text{disc}(f) = 0$ if and only if f has a repeated root in the algebraic closure of the

quotient field of R . This notion is used in the proof of the following is a rational version of [K-BRZ23, Proposition 1.7] (see also the proof of [BŠ23a, Theorem 4.1]).

Lemma 2.5. *Let $p \in \mathbb{N}$ be a prime, and let $r \in \mathbb{K}\langle x \rangle$ with $\text{dom}_p r \neq \emptyset$ be such that $r(\Xi(p, \xi))$ is not central in UD_p . Then*

$$(2) \quad \{X \in \text{dom}_p r : r(X) \text{ has } p \text{ distinct eigenvalues}\}$$

is a nonempty Zariski open subset of $M_p(\mathbb{K})^m$.

Proof. The set (2) is clearly Zariski open; one only needs to show that (2) is nonempty. Let Z_p denote the center of UD_p , and let F be the Z_p -subfield of UD_p generated by $r(\Xi(p, \xi))$. Then the inclusions $Z_p \subset F \subset \text{UD}_p$ are strict; the first one because $r(\Xi(p, \xi))$ is not central, and the second one because F is commutative and UD_p is not. Since p is prime and UD_p has dimension p^2 over Z_p , it follows that F has dimension p over Z_p . Let $\chi \in Z_p[t]$ be the characteristic polynomial of $r(\Xi(p, \xi))$. Then $\deg \chi = p$ and χ is also the minimal polynomial of $r(\Xi(p, \xi))$. Furthermore, χ is irreducible over Z_p since $F \cong Z_p[t]/(\chi)$ is a field. Since \mathbb{K} has characteristic 0, Z_p is a perfect field, so χ is separable and has only simple roots in its splitting field. Therefore $\text{disc}(\chi)$ is a nonzero element of Z_p . Since \mathbb{K} is infinite, there exists $X \in \text{dom}_p r$ such that $\text{disc}(\chi)|_{\Xi(p, \xi)=X} \neq 0$. Hence $\chi|_{\Xi(p, \xi)=X}$ has p distinct roots, so $r(X)$ has p distinct eigenvalues. \square

Lemma 2.6. *Let R, S be domains containing \mathbb{K} , and $f \in R[t], g \in S[t]$ without roots in \mathbb{K} . For $fg = f \cdot g \in (R \otimes_{\mathbb{K}} S)[t]$ we have $\text{disc}(fg) \neq 0$ if and only if $\text{disc}(f) \text{disc}(g) \neq 0$.*

Proof. Throughout the proof we identify $f = f \otimes 1, g = 1 \otimes g \in (R \otimes_{\mathbb{K}} S)[t]$. By [GKZ94, Section 12.1.B] we have

$$\text{disc}(fg) = (-1)^{\deg f \deg g} \text{disc}(f) \text{disc}(g) \text{res}(f, g)^2,$$

where $\text{res}(f, g)$ is the resultant of f and g [GKZ94, Section 12.1.A]. Since \mathbb{K} is algebraically closed, $R \otimes_{\mathbb{K}} S$ is a domain (see e.g. [Sha13, Theorem 1.6 and Example 1.1.10]). Let F_1, F_2, F be algebraic closures of the quotient fields of $R, S, R \otimes_{\mathbb{K}} S$, respectively; note that $F_1 \otimes_{\mathbb{K}} F_2$ embeds into F . The roots of polynomials f and g lie in $(F_1 \otimes 1) \setminus \mathbb{K}$ and $(1 \otimes F_2) \setminus \mathbb{K}$, respectively. In particular, f and g have no common roots, so $\text{res}(f, g) \neq 0$ by [GKZ94, Section 12.1.A]. \square

The next result shows that nonconstant noncommutative polynomials and rational functions eventually attain values with pairwise distinct eigenvalues (for a stronger conjecture, see [K-BRZ23, Problem 1.4]).

Theorem 2.7. *Let $r \in \mathbb{K}\langle x \rangle \setminus \mathbb{K}$ admit a linear representation of size δ , and let $p < q$ be primes larger than 2δ . Then for every $n \geq (p-1)(q-1)$,*

$$(3) \quad \{X \in \text{dom}_n r : r(X) \text{ has } n \text{ distinct eigenvalues}\}$$

is a nonempty Zariski open subset of $M_n(\mathbb{K})^m$.

Proof. By Lemma 2.3, $r(\Xi(n, \xi))$ is a well-defined non-central element of UD_n for every $n \geq 2\delta$. Since Zariski openness is evident, it suffices to show that (3) is nonempty. The discriminants of characteristic polynomials of $r(\Xi(p, \xi))$ and $r(\Xi(q, \xi))$ are nonzero by Lemma 2.5. If $n \geq (p-1)(q-1)$, then n can be expressed as $ap + bq$ for some $a, b \in \mathbb{N}$ by Sylvester's theorem [Alf05, Theorem 2.1.1]. Let $a, b \in \mathbb{N}$ be arbitrary, and consider the tuple of block diagonal matrices

$$\Xi := \Xi(p, \xi^{(1)}) \oplus \cdots \oplus \Xi(p, \xi^{(a)}) \oplus \Xi(q, \xi^{(a+1)}) \oplus \cdots \oplus \Xi(q, \xi^{(a+b)})$$

over $\mathbb{K}[\underline{\xi}^{(1)}] \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} \mathbb{K}[\underline{\xi}^{(a+b)}]$. Then

$$r(\Xi) = r(\Xi(p, \xi^{(1)})) \oplus \cdots \oplus r(\Xi(p, \xi^{(a)})) \oplus r(\Xi(q, \xi^{(a+1)})) \oplus \cdots \oplus r(\Xi(q, \xi^{(a+b)})),$$

and the characteristic polynomial of $r(\Xi)$ is a product of characteristic polynomials of $r(\Xi(p, \xi^{(i)}))$ for $i = 1, \dots, a$ and $r(\Xi(q, \xi^{(j)}))$ for $j = 1, \dots, b$. None of these have a root in \mathbb{K} by Lemma 2.4. Therefore the discriminant of the characteristic polynomial χ of $r(\Xi)$ is nonzero by Lemma 2.6. Since \mathbb{K} is infinite, there exists a tuple of block-diagonal matrices $X \in \text{dom}_{ap+bq} r$ such that $\text{disc}(\chi)|_{\Xi=X} \neq 0$. Hence $r(X)$ has $ap + bq$ distinct eigenvalues, so (3) is nonempty for $n = ap + bq$. \square

Remark 2.8. Using Bertrand's postulate [NZM91, Theorem 8.7], one concludes that (3) is nonempty for all $n \geq (2\delta - 3 - 1)(2(2\delta - 3) - 3 - 1) = 4(\delta - 2)(2\delta - 5)$.

Corollary 2.9. *Let r be a nonconstant noncommutative rational function. For all n large enough, the set $\mathcal{O}_r \subseteq M_n(\mathbb{K})^m$ where r attains values with distinct and nonzero eigenvalues is Zariski open and nonempty.*

Proof. Let $n \in \mathbb{N}$ be large enough so that the conclusion of Theorem 2.7 holds. Then the set $\mathcal{O}_1 \subseteq M_n(\mathbb{K})^m$ where r attains values with distinct eigenvalues is Zariski open and nonempty. In particular, r is not constantly zero on $M_n(\mathbb{K})^m$; since UD_n is a division ring, the set $\mathcal{O}_2 \subseteq M_n(\mathbb{K})^m$ where r attains invertible values is Zariski open and nonempty. Therefore $\mathcal{O}_r = \mathcal{O}_1 \cap \mathcal{O}_2$ is also Zariski open and nonempty. \square

Corollary 2.10. *Let f be a nonconstant noncommutative polynomial. For all n large enough, the set $\mathcal{O}_f \subseteq M_n(\mathbb{K})^m$ where f attains values with distinct and nonzero eigenvalues is Zariski open and nonempty.*

More precisely, the statement is valid for $n \geq (p-1)(q-1)$ where $p < q$ are primes larger than $\lceil \frac{\deg f}{2} \rceil$.

Proof. The improved lower bound on n follows from using the part of Lemma 2.3 referring to noncommutative polynomials at the beginning of the proof of Theorem 2.7. \square

Remark 2.11. Recall that a polynomial $f \in \mathbb{K}\langle x \rangle$ is said to be a polynomial identity of $M_n(\mathbb{K})$ if $f(X) = 0$ for every $X \in M_n(\mathbb{K})^m$. If $f(X)$ is always a scalar matrix, but f is not a polynomial identity, then f is said to be central for $M_n(\mathbb{K})$. It is obvious that the conclusion of Corollary 2.10 does not hold for polynomial identities and central polynomials. One may wonder whether these are the only polynomials that have to be excluded, and, accordingly, whether the assumption that n must be large enough is too

restrictive. This is not the case. Recall further that $g \in \mathbb{K}\langle x \rangle$ is 2-central for $M_n(\mathbb{K})$ if g^2 is central for $M_n(\mathbb{K})$, but g is not. For example, $[x_1, x_2]$ is 2-central for $M_2(\mathbb{K})$. It is known that 2-central polynomials for $M_n(\mathbb{K})$ exist for various large (even) n [K-BMR16, Proposition 1]. Observe that the evaluations of a 2-central polynomial have at most two distinct eigenvalues. We may also consider more general k -central polynomials, as well sums of these and central polynomials. This justifies the assumption in Corollary 2.10 (and hence also in Theorem 2.7) that n must be large enough. On the other hand, this assumption is superfluous for multilinear polynomials [K-BMR16, Theorem 3].

3. THE LINEAR WARING PROBLEM

We first record an elementary lemma, which was implicitly already noticed in [BŠ23a].

Lemma 3.1. *Let $D \in M_n(\mathbb{K})$ be a matrix with n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Then every matrix in $\mathfrak{sl}_n(\mathbb{K})$ is a difference of two matrices similar to D .*

Proof. By [HJ85, Section 2.2, Problem 3], every matrix in $M \in \mathfrak{sl}_n(\mathbb{K})$ is similar to a matrix whose all diagonal entries are zero. Therefore $M = Q(N_1 - N_2)Q^{-1}$ for some $Q \in GL_n(\mathbb{K})$, an upper-triangular matrix N_1 with diagonal entries $\lambda_1, \dots, \lambda_n$, and a lower-triangular matrix N_2 with diagonal entries $\lambda_1, \dots, \lambda_n$. Observe that QN_1Q^{-1} and QN_2Q^{-1} are similar to D . \square

Our first application of Theorem 2.7 is the following theorem.

Theorem 3.2. *For every nonconstant $r \in \mathbb{K}\langle x \rangle$ and large enough $n \in \mathbb{N}$,*

$$\mathfrak{sl}_n(\mathbb{K}) \subseteq r(\text{dom}_n r) - r(\text{dom}_n r).$$

Proof. Given r , let $n \in \mathbb{N}$ be large enough so that the conclusion of Theorem 2.7 holds; that is, there exists $X \in \text{dom}_n r$ such that $r(X)$ has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Take $M \in \mathfrak{sl}_n(\mathbb{K})$. By Lemma 3.1, there exist $P_1, P_2 \in GL_n(\mathbb{K})$ such that $M = P_1 r(X) P_1^{-1} - P_2 r(X) P_2^{-1}$, and hence $M = r(P_1 X P_1^{-1}) - r(P_2 X P_2^{-1})$. \square

As mentioned in the introduction, the special case of Theorem 3.2 where $r = f \in \mathbb{K}\langle x \rangle$ gives a positive answer to [BŠ23a, Question 4.8]. Observe that if f is a sum of commutators in $\mathbb{K}\langle x \rangle$, then all matrix evaluations of f consist of trace zero matrices. This justifies the restriction to $\mathfrak{sl}_n(\mathbb{K})$ in the conclusion of Theorem 3.2. It is natural to ask what can be said if f is not a sum of commutators. To give a (partial) answer, we first record the following lemma.

Lemma 3.3. *Let $\lambda_1, \dots, \lambda_n$ be distinct elements in \mathbb{K} having nonzero sum. Then every nonscalar matrix $A \in M_n(\mathbb{K})$ with nonzero trace can be written as $A = \alpha(D_1 + D_2)$, where $\alpha \in \mathbb{K}$ and D_1, D_2 are matrices similar to $\text{diag}(\lambda_1, \dots, \lambda_n)$.*

Proof. Let $\alpha \in \mathbb{K}$ be such that $2\alpha(\lambda_1 + \dots + \lambda_n)$ is equal to the trace of A . By [BŠ23b, Lemma 2.8], there exists a matrix which is similar to A and whose diagonal entries are

$2\alpha\lambda_1, \dots, 2\alpha\lambda_n$. Therefore, there is no loss of generality in assuming that $2\alpha\lambda_1, \dots, 2\alpha\lambda_n$ are the diagonal entries of A . Write

$$A = \text{diag}(2\alpha\lambda_1, \dots, 2\alpha\lambda_n) + U + L,$$

where U is a strictly upper triangular matrix and L is a strictly lower diagonal matrix. We remark that $\alpha \neq 0$ since A has nonzero trace. Note that the matrices

$$D_1 = \text{diag}(\lambda_1, \dots, \lambda_n) + \alpha^{-1}U, \quad D_2 = \text{diag}(\lambda_1, \dots, \lambda_n) + \alpha^{-1}L$$

are similar to $\text{diag}(\lambda_1, \dots, \lambda_n)$ and $A = \alpha(D_1 + D_2)$. \square

Remark 3.4. The assumption that A is nonscalar is necessary. In fact, the identity I is not always a linear combination of two matrices similar to $\text{diag}(\lambda_1, \dots, \lambda_n)$. Indeed, $I = \beta D_1 + \gamma D_2$ implies that the spectrum of $I - \beta D_1$ is equal to the spectrum of γD_2 , that is, $\{1 - \beta\lambda_1, \dots, 1 - \beta\lambda_n\} = \{\gamma\lambda_1, \dots, \gamma\lambda_n\}$. It is easy to find λ_i 's for which there are no such β, γ .

Let $f \in \mathbb{K}\langle x \rangle$. We simplify the notation and write $f(M_n(\mathbb{K}))$ for $f(M_n(\mathbb{K})^m)$. More generally, for any algebra \mathcal{A} , we write $f(\mathcal{A})$ for $f(\mathcal{A}^m)$. We call $f(\mathcal{A})$ the image of f in \mathcal{A} . Note that $f(\mathcal{A})$ is closed under conjugation by invertible elements in \mathcal{A} .

Theorem 3.5. *If a polynomial f is not a sum of commutators, then, for n large enough, every nonscalar matrix in $M_n(\mathbb{K})$ is a linear combination of two matrices from $f(M_n(\mathbb{K}))$.*

Proof. We claim that for n large enough, $f(M_n(\mathbb{K}))$ contains a matrix with nonzero trace. Since f is not a sum of commutators, [BK09, Corollary 4.7] tells us that there exists a natural number n_0 such that $f(M_{n_0}(\mathbb{K}))$ contains a matrix with nonzero trace. If f has constant term zero, this readily implies that the same is true for $f(M_n(\mathbb{K}))$ for every $n \geq n_0$. If, however, f has a nonzero constant term, then obviously $f(0, \dots, 0)$ has nonzero trace for each n .

Thus, for n large enough, the tuples in $M_n(\mathbb{K})^m$ in which f attains a matrix with nonzero trace form a nonempty Zariski open subset of $M_n(\mathbb{K})^m$. Corollary 2.10 shows that for n large enough, the tuples in $M_n(\mathbb{K})^m$ in which f attains a matrix with n distinct eigenvalues also form a nonempty Zariski open subset of $M_n(\mathbb{K})^m$. Their intersection is therefore nonempty. Thus, for n large enough, $f(M_n(\mathbb{K}))$ contains a matrix D with nonzero trace and n distinct eigenvalues. Since $f(M_n(\mathbb{K}))$ is closed under conjugation with invertible matrices, we may assume that D is a diagonal matrix. Lemma 3.3 therefore shows that every nonscalar matrix in $M_n(\mathbb{K})$ having nonzero trace is a scalar multiple of the sum of two matrices similar to D . By Lemma 3.1, every trace zero matrix in $M_n(\mathbb{K})$ is a difference of two matrices similar to D . Since matrices similar to D belong to $f(M_n(\mathbb{K}))$, this proves the corollary. \square

Remark 3.6. Taking a polynomial like $f = 1 + [x, y]$ we see that one really has to involve linear combinations, i.e., sums and differences are insufficient.

If f attains a nonzero value in scalars (i.e., $f(\alpha_1, \dots, \alpha_m) \neq 0$ for some $\alpha_j \in \mathbb{K}$), in particular if f has nonzero constant term, then we can omit ‘‘nonscalar’’ in the statement of Theorem 3.5. We leave the general case as an open problem.

Question 3.7. Does the conclusion of Theorem 3.5 also hold for scalar matrices?

What can be said is the following.

Corollary 3.8. *If a polynomial f is not a sum of commutators, then, for n large enough, every matrix in $M_n(\mathbb{K})$ is a linear combination of three matrices from $f(M_n(\mathbb{K}))$.*

Proof. As noticed at the beginning of the proof of Theorem 3.5, $f(M_n(\mathbb{K}))$ contains a matrix A with nonzero trace. Since every matrix in $M_n(\mathbb{K})$ is a linear combination of A and a traceless matrix, the desired conclusion follows from Theorem 3.2. \square

In view of Theorem 3.2, Question 3.7 can be equivalently asked as follows: Given any nonconstant polynomial f , is the set of linear combinations of two matrices from $f(M_n(\mathbb{K}))$ a vector space for all large enough n ? We remark that [BŠ23a, Theorem 1.2] shows that the restriction to large enough n is necessary here.

In our last theorem of this section we will consider algebras more general than $M_n(\mathbb{K})$. Here, by an algebra we mean a unital, associative and not necessarily commutative algebra over \mathbb{K} . We first record an elementary lemma.

Lemma 3.9. *Let \mathcal{A} be an algebra, let $\lambda_1, \dots, \lambda_m$ be distinct elements in \mathbb{K} , and let $d = \text{diag}(\lambda_1 1, \dots, \lambda_m 1) \in M_m(\mathcal{A})$, where 1 stands for the unity of \mathcal{A} . If b is an upper triangular (resp. lower triangular) matrix in $M_m(\mathcal{A})$ with the same diagonal as d , then there exists an upper triangular (resp. lower triangular) matrix $t \in M_m(\mathcal{A})$ with 1 's on the diagonal such that $b = tdt^{-1}$.*

The lemma can be easily proved by induction on m or directly. We leave the details to the reader.

Recall that a matrix over a field has trace zero if and only if it is a commutator of two matrices [AM57]. When considering matrices over algebras that are not fields, the role of trace zero matrices are therefore played by commutators.

Theorem 3.10. *Let \mathcal{A} be an algebra over \mathbb{K} such that $\mathcal{A} \cong M_2(\mathcal{A})$. If f is a nonconstant polynomial, then every commutator in \mathcal{A} is a sum of seven elements from $f(\mathcal{A}) - f(\mathcal{A})$.*

Proof. Observe that $\mathcal{A} \cong M_2(\mathcal{A})$ implies that $\mathcal{A} \cong M_{2^n}(\mathcal{A})$ for every positive integer n . By Corollary 2.10, there exists a positive integer n such that $f(M_{2^n}(\mathbb{K}))$ contains a diagonal matrix $\text{diag}(\lambda_1, \dots, \lambda_{2^n})$ with distinct λ_i . In light of the natural embedding of $M_{2^n}(\mathbb{K})$ into $M_{2^n}(\mathcal{A})$, this implies that $f(M_{2^n}(\mathcal{A}))$ contains the diagonal matrix

$$d = \text{diag}(\lambda_1 1, \dots, \lambda_{2^n} 1) \in M_{2^n}(\mathcal{A}).$$

Since $f(M_{2^n}(\mathcal{A}))$ is closed under conjugation with invertible elements, Lemma 3.9 shows that $f(M_{2^n}(\mathcal{A}))$ contains all matrices of the form $d + u$ where u is a strictly upper triangular matrix in $M_{2^n}(\mathcal{A})$, as well as all matrices of the form $d + \ell$ where ℓ is a strictly lower triangular matrix in $M_{2^n}(\mathcal{A})$. Therefore,

$$u - \ell \in f(M_{2^n}(\mathcal{A})) - f(M_{2^n}(\mathcal{A})),$$

meaning that $f(M_{2^n}(\mathcal{A})) - f(M_{2^n}(\mathcal{A}))$ contains all matrices in $M_{2^n}(\mathcal{A})$ that have zeros on the diagonal. In particular, all matrices $\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ with $a, b \in M_{2^{n-1}}(\mathcal{A})$ are contained in $f(M_{2^n}(\mathcal{A})) - f(M_{2^n}(\mathcal{A}))$.

In view of our assumption on \mathcal{A} , we may replace $M_{2^{n-1}}(\mathcal{A})$ by \mathcal{A} in the last statement. Thus, $f(M_2(\mathcal{A})) - f(M_2(\mathcal{A}))$ contains the subspace \mathcal{S} of $M_2(\mathcal{A})$ consisting of matrices $\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ with $a, b \in \mathcal{A}$. Take a commutator $k \in M_2(\mathcal{A})$. We have

$$\begin{aligned} k &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} - \begin{bmatrix} e & f \\ g & h \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} ae - ea + bg - fc & 0 \\ 0 & dh - hd - gb + cf \end{bmatrix} + s \end{aligned}$$

where $s \in \mathcal{S}$. Using that $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix}$ for every $x \in \mathcal{A}$, it follows that

$$\begin{aligned} k &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ ae & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} + \begin{bmatrix} 1 & -e \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ a & 0 \end{bmatrix} \begin{bmatrix} 1 & -e \\ 0 & 1 \end{bmatrix}^{-1} \\ &+ \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ g & 0 \end{bmatrix} \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}^{-1} + \begin{bmatrix} 1 & -f \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix} \begin{bmatrix} 1 & -f \\ 0 & 1 \end{bmatrix}^{-1} \\ &+ \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ dh & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} + \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ h & 0 \end{bmatrix} \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}^{-1} + s' \end{aligned}$$

where $s' \in \mathcal{S}$. Each of the first six terms is a conjugate of a matrix from \mathcal{S} . As $\mathcal{S} \subseteq f(M_2(\mathcal{A})) - f(M_2(\mathcal{A}))$ and $f(M_2(\mathcal{A})) - f(M_2(\mathcal{A}))$ is closed under conjugation, it follows that k is a sum of seven elements from $f(M_2(\mathcal{A})) - f(M_2(\mathcal{A}))$. Since $\mathcal{A} \cong M_2(\mathcal{A})$, this completes the proof. \square

We point out two cases to which Theorem 3.10 applies. The first corollary is a version of [Bre20, Corollary 3.15] which imposes milder conditions on \mathbb{K} , but in which the role of seven is played by a certain four-digit number.

Corollary 3.11. *Let V be an infinite-dimensional vector space over \mathbb{K} . If f is a non-constant polynomial, then every element in $\mathcal{A} = \text{End}_F(V)$ is a sum of seven elements from $f(\mathcal{A}) - f(\mathcal{A})$.*

Proof. From $V \cong V \oplus V$ we obtain $\mathcal{A} \cong M_2(\mathcal{A})$. Since every element in \mathcal{A} is a commutator [Mes, Proposition 12], the desired conclusion follows from Theorem 3.10. \square

By $B(X)$ we denote the algebra of all continuous linear operators on a Banach space X . If $X \cong X \oplus X$, then $\mathcal{A} = B(X)$ satisfies $\mathcal{A} \cong M_2(\mathcal{A})$. Thus the following corollary holds.

Corollary 3.12. *Let X be a Banach space such that $X \cong X \oplus X$. If f is a nonconstant polynomial, then every commutator in $\mathcal{A} = B(X)$ is a sum of seven elements from $f(\mathcal{A}) - f(\mathcal{A})$.*

Infinite-dimensional Hilbert spaces H certainly satisfy $H \cong H \oplus H$, but for them Corollary 3.12 is not entirely new; see [Bre20, Corollary 3.17] and [BŠ23a, Theorem 4.2] (since every operator in $B(H)$ is a sum of two commutators [Hal54], the conclusions of these two results are not restricted to commutators). Corollary 3.12 may thus be viewed as a generalization from Hilbert spaces to Banach spaces X isomorphic to their square $X \oplus X$. This actually covers most Banach spaces. In fact, the first infinite-dimensional Banach space not isomorphic to its square was discovered only in 1960 [BP60], thereby answering a problem posed by Banach in 1932.

It is unlikely that the number seven from the last three results is the least possible.

Question 3.13. What is the smallest number that can replace the number seven in Theorem 3.10 and Corollaries 3.11 and 3.12?

4. THE MULTIPLICATIVE WARING PROBLEM

We write \mathbb{K}^\times for the group of all nonzero scalar matrices. Our first result in this section is analogous to Theorem 3.2.

Theorem 4.1. *For every nonconstant $r \in \mathbb{K}\langle \underline{x} \rangle$ and large enough $n \in \mathbb{N}$,*

$$\mathrm{SL}_n(\mathbb{K}) \setminus \mathbb{K}^\times \subseteq r(\mathrm{dom}_n r) \cdot r(\mathrm{dom}_n r)^{-1}.$$

Proof. Let $n \in \mathbb{N}$ be large enough so that the conclusion of Corollary 2.9 holds. Then there exists $X \in \mathrm{dom}_n r$ such that $r(X)$ has n distinct and nonzero eigenvalues $\lambda_1, \dots, \lambda_n$.

Take $M \in \mathrm{SL}_n(\mathbb{K}) \setminus \mathbb{K}^\times$. By [Sou86, Theorem 1] there exists matrices $N_1, N_2 \in \mathrm{GL}_n(\mathbb{K})$ with eigenvalues $\lambda_1, \dots, \lambda_n$ such that $M = N_1 N_2^{-1}$. For $j = 1, 2$ there exists $P_j \in \mathrm{GL}_n(\mathbb{K})$ such that $N_j = P_j r(X) P_j^{-1}$. Then $M = r(P_1 X P_1^{-1}) r(P_2 X P_2^{-1})^{-1}$. \square

We do not know whether the exclusion of \mathbb{K}^\times is necessary.

Question 4.2. Is (in Theorem 4.1) $\mathrm{SL}_n(\mathbb{K}) \cap \mathbb{K}^\times$ also contained in $r(\mathrm{dom}_n r) \cdot r(\mathrm{dom}_n r)^{-1}$?

Since the identity matrix is trivially contained in $r(\mathrm{dom}_n r) \cdot r(\mathrm{dom}_n r)^{-1}$, the following corollary to Theorem 4.1 holds.

Corollary 4.3. *For every nonconstant $r \in \mathbb{K}\langle \underline{x} \rangle$ and large enough $n \in \mathbb{N}$,*

$$\mathrm{GL}_n(\mathbb{K}) \subseteq \mathbb{K}^\times \cdot r(\mathrm{dom}_n r) \cdot r(\mathrm{dom}_n r)^{-1}.$$

In the rest of the section we consider the multiplicative Waring problem for noncommutative polynomials. We start with an auxiliary result.

Proposition 4.4. *Let $f \in \mathbb{K}\langle \underline{x} \rangle$ and $n \in \mathbb{N}$ be such that f is not constant on $M_n(\mathbb{K})^m$. Then $\det f(M_n(\mathbb{K})) = \mathbb{K}$.*

Proof. It suffices to see that $\det f(\Xi(n, \xi)) \in \mathbb{K}[\xi]$ is a nonconstant polynomial. Suppose $\det f(\Xi(n, \xi)) = \alpha \in \mathbb{K}$. By the Cayley-Hamilton theorem, $f(\Xi(n, \xi)) \cdot p = \alpha$ for a trace polynomial p . The algebra of trace polynomials [Pro76, Section 2] is graded by degree, and has no zero divisors. Let \tilde{f} and \tilde{p} be the homogeneous components of $f(\Xi(n, \xi))$ and

p , respectively, of highest degree. Since f is not constant on $M_n(\mathbb{K})^m$, we have $\deg \tilde{f} > 0$. Then $f(\Xi(n, \xi)) \cdot p = \alpha$ implies $\tilde{f} \cdot \tilde{p} = 0$, and so $\tilde{p} = 0$. Hence $p = 0$, and therefore $\alpha = 0$. Thus $f(\Xi(n, \xi))$ is not invertible in UD_n , which is a division ring. Hence $f(\Xi(n, \xi)) = 0$, contradicting the assumption that f is not constant. \square

Example 4.5. Let $r = x_1 x_2 x_1^{-1} x_2^{-1} \in \mathbb{K}\langle x \rangle$. Then r is nonconstant (and in particular attains values with pairwise distinct eigenvalues by Theorem 2.7), but $\det r(X) = 1$ for every $X \in \text{dom}_n r$ and $n \in \mathbb{N}$.

Theorem 4.6. For all nonconstant $f, g \in \mathbb{K}\langle x \rangle$ and large enough $n \in \mathbb{N}$,

$$\text{GL}_n(\mathbb{K}) \setminus \mathbb{K}^\times \subseteq f(M_n(\mathbb{K})) \cdot g(M_n(\mathbb{K})).$$

Proof. Let $n \in \mathbb{N}$ be large enough so that the conclusion of Corollary 2.10 holds for both f and g . By Corollary 2.9, the set $\mathcal{O}_f \subseteq M_n(\mathbb{K})^m$ where f attains values with distinct and nonzero eigenvalues is Zariski open and nonempty. The set $\det f(\mathcal{O}_f)$ is constructible in \mathbb{K} by Chevalley's theorem [Har92, Theorem 3.16]. In particular, $\det f(\mathcal{O}_f)$ is either finite or cofinite in \mathbb{K} . The first case would contradict $\det f(M_n(\mathbb{K})) = \mathbb{K}$ which holds by Proposition 4.4, so $\det f(\mathcal{O}_f)$ is cofinite in \mathbb{K} . Of course, for an analogously defined set \mathcal{O}_g , $\det g(\mathcal{O}_g)$ is also cofinite in \mathbb{K} .

Take $M \in \text{GL}_n(\mathbb{K}) \setminus \mathbb{K}^\times$. Since $\mathbb{K} \setminus \det f(\mathcal{O}_f)$ and $\mathbb{K} \setminus \det g(\mathcal{O}_g)$ are finite, there exist $X_1, X_2 \in \mathcal{O}$ such that $\det M = \det f(X_1) \det g(X_2)$. By [Sou86, Theorem 1], there are matrices $N_1, N_2 \in \text{GL}_n(\mathbb{K})$ such that $M = N_1 N_2$, eigenvalues of N_1 coincide with those of $f(X_1)$, and eigenvalues of N_2 coincide with those of $g(X_2)$. Therefore $M = f(P_1 X P_1^{-1}) g(P_2 X P_2^{-1})$ for suitable $P_1, P_2 \in \text{GL}_n(\mathbb{K})$. \square

Question 4.7. Is (in Theorem 4.6) \mathbb{K}^\times also contained in $f(M_n(\mathbb{K})) \cdot g(M_n(\mathbb{K}))$?

At least we know that three factors are sufficient for covering the whole $\text{GL}_n(\mathbb{K})$.

Corollary 4.8. For all nonconstant $f, g, h \in \mathbb{K}\langle x \rangle$ and large enough $n \in \mathbb{N}$,

$$\text{GL}_n(\mathbb{K}) \subseteq f(M_n(\mathbb{K})) \cdot g(M_n(\mathbb{K})) \cdot h(M_n(\mathbb{K})).$$

Proof. Let n be large enough so that the conclusion of Theorem 4.6 holds for f and g as well as the conclusion of Corollary 2.9 holds for h . Let $A_1 \in h(M_n(\mathbb{K}))$ be a matrix that has n distinct and nonzero eigenvalues, and let A_2 be a matrix that is similar to but linearly independent with A_1 . Note that A_2 also lies in $h(M_n(\mathbb{K}))$. By Theorem 4.6, $f(M_n(\mathbb{K})) \cdot g(M_n(\mathbb{K})) \cdot A_i$ contains all matrices from $\text{GL}_n(\mathbb{K})$ besides possibly scalar multiples of A_i , $i = 1, 2$. This readily implies the desired conclusion. \square

It would be desirable to also cover singular matrices by products of the images of polynomials. In our final result we will obtain a result of such kind, however, by restricting to polynomials that have a root in scalars. By this we mean that there exist $\alpha_j \in \mathbb{K}$ such that $f(\alpha_1, \dots, \alpha_m) = 0$. This obviously holds if f has zero constant term.

Theorem 4.9. If a nonzero polynomial f has a root in scalars, then for n large enough, every matrix in $M_n(\mathbb{K})$ is a product of twelve matrices from $f(M_n(\mathbb{K}))$.

Proof. By Corollary 4.8, there exists an $N_0 \geq 1$ such that for every $n \geq N_0$, every invertible matrix in $M_n(\mathbb{K})$ is a product of three matrices from $f(M_n(\mathbb{K}))$. Set $N = 2N_0$ and take $n \geq N$. Since every matrix in $M_n(\mathbb{K})$ is a product of two diagonalizable matrices [Bot98], it is enough to show that every diagonalizable matrix $D \in M_n(\mathbb{K})$ is a product of six matrices from $f(M_n(\mathbb{K}))$.

Write k for the rank of D . Since $f(M_n(\mathbb{K}))$ is closed under conjugation by invertible matrices, we may assume that

$$D = \text{diag}(\lambda_1, \dots, \lambda_k, 0, \dots, 0)$$

where λ_i are nonzero elements in \mathbb{K} (so there are $n - k$ zeros).

Assume first that $k \geq N_0$. We claim that in this case D is actually a product of three matrices from $f(M_n(\mathbb{K}))$ (since this in particular implies that the identity matrix I is a product of three matrices from $f(M_n(\mathbb{K}))$, $D = DI$ is then a product of six matrices from $f(M_n(\mathbb{K}))$, as desired). As we know, our assumption on N_0 implies that $\text{diag}(\lambda_1, \dots, \lambda_k)$ is a product of three matrices from $f(M_k(\mathbb{K}))$:

$$\text{diag}(\lambda_1, \dots, \lambda_k) = f(A_{11}, \dots, A_{1m}) \cdot f(A_{21}, \dots, A_{2m}) \cdot f(A_{31}, \dots, A_{3m})$$

with $A_{ij} \in M_k(\mathbb{K})$. Let $\alpha_i \in \mathbb{K}$ be such that $f(\alpha_1, \dots, \alpha_m) = 0$. Setting

$$A'_{ij} = \begin{bmatrix} A_{ij} & 0 \\ 0 & \alpha_j I_{n-k} \end{bmatrix} \in M_n(\mathbb{K}),$$

where I_{n-k} is the $(n - k) \times (n - k)$ identity matrix and the zeros are blocks of the appropriate size, we have

$$D = f(A'_{11}, \dots, A'_{1m}) \cdot f(A'_{21}, \dots, A'_{2m}) \cdot f(A'_{31}, \dots, A'_{3m}),$$

proving our claim.

Now consider the case where $k < N_0$. Set

$$D_1 = \text{diag}(\lambda_1, \dots, \lambda_k, 1, \dots, 1, 0, \dots, 0)$$

where 1 occurs $N_0 - k$ times and 0 occurs $n - N_0$ times, and

$$D_2 = \text{diag}(1, \dots, 1, 0, \dots, 0, 1, \dots, 1)$$

where 1 occurs k times in the first sequence, 0 occurs $N_0 - k$ times, and 1 occurs $n - N_0$ times in the last sequence. Since D_1 and D_2 are diagonal matrices of rank N_0 and $n - N_0 + k \geq N_0$, respectively, each of them is a product of three matrices from $f(M_n(\mathbb{K}))$. Hence, $D = D_1 D_2$ is a product of six matrices from $f(M_n(\mathbb{K}))$. \square

Question 4.10. What is the smallest number that can replace the number twelve in Theorem 4.9?

Remark 4.11. Observe that if g is 2-central for $M_n(\mathbb{K})$, then $f = g^3$ satisfies $f(M_n(\mathbb{K})) \subseteq \text{GL}_n(\mathbb{K}) \cup \{0\}$. This may be viewed as a justification for the assumption in Theorem 4.9 that n must be large enough (compare Remark 2.11).

Question 4.12. Does Theorem 4.9 hold for all nonconstant polynomials f ?

Remark 4.13. Question 4.12 is related to the following one by Leonid Makar-Limanov: *Given a nonconstant $f \in \mathbb{K}\langle x \rangle$ let $\rho_n = \min \text{rk } f(M_n(\mathbb{K}))$ for $n \in \mathbb{N}$; is it true that $\liminf_n \frac{\rho_n}{n} = 0$?* As observed by Makar-Limanov, an affirmative answer to this question implies that the principal ideal in $\mathbb{K}\langle x \rangle$ generated by f is proper. The latter seemingly innocuous statement turns to be always true if \mathbb{K} has characteristic 0 (by the existence of an algebraically closed skew field constructed by Makar-Limanov [M-L]), but remains open for \mathbb{K} with positive characteristic. On the other hand, even in characteristic 0, Makar-Limanov’s question is compelling to the free analysis community, as its affirmative answer would mitigate the fact that not every nonconstant noncommutative polynomial has a matrix root.

As the proof of Theorem 4.9 indicates, a crucial obstruction to answering Question 4.12 is a lack of grasp on low-rank values of noncommutative polynomials. To inspire further work in this direction, we pose the following refinement of Makar-Limanov’s question: *If $f \in \mathbb{K}\langle x \rangle$ is not constant on $n \times n$ matrices, does there exist $X \in M_n(\mathbb{K})^m$ such that $\text{rk } f(X) \leq 1$?*

REFERENCES

- [AM57] A. Albert, B. Muckenhoupt, *On matrices of trace 0*, Michigan Math. J. 1 (1957) 1–3. [9](#)
- [Alf05] J. L. R. Alfonsín: *The diophantine Frobenius problem*, Oxford University Press, Oxford, 2005. [6](#)
- [Ami66] S. A. Amitsur: *Rational identities and applications to algebra and geometry*, J. Algebra 3 (1966) 304–359. [1](#)
- [AHKM18] M. Augat, J.W. Helton, I. Klep, S. McCullough: *Bianalytic Maps Between Free Spectrahedra*, Math. Ann. 371 (2018) 883–959. [1](#)
- [Ber76] G. M. Bergman: *Rational relations and rational identities in division rings. I*, J. Algebra 43 (1976) 252–266. [4](#)
- [BR11] J. Berstel, C. Reutenauer: *Noncommutative rational series with applications*, Encyclopedia of Mathematics and its Applications 137, Cambridge University Press, Cambridge, 2011. [1](#)
- [BP60] C. Bessaga, A. Pełczyński: *Banach spaces non-isomorphic to their Cartesian squares*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astr. Phys. 8 (1960) 77–80. [11](#)
- [Bot98] J. D. Botha: *Products of diagonalizable matrices*, Linear Algebra Appl. 273 (1998) 65–82. [13](#)
- [Bre20] M. Brešar: *Commutators and images of noncommutative polynomials*, Adv. Math. 374 (2020) 107346, 21 pp. [2](#), [10](#), [11](#)
- [BK09] M. Brešar, I. Klep: *Values of noncommutative polynomials, Lie skew-ideals and tracial Nullstellensätze*, Math. Res. Lett. 16 (2009) 605–626. [8](#)
- [BŠ23a] M. Brešar, P. Šemrl: *The Waring problem for matrix algebras*, Israel J. Math. 253 (2023) 381–405. [2](#), [5](#), [7](#), [9](#), [11](#)
- [BŠ23b] M. Brešar, P. Šemrl: *The Waring problem for matrix algebras, II*, Bull. London Math. Soc. 55 (2023) 1880–1889. [2](#), [7](#)
- [Che23] Q. Chen, *On Panja-Prasad conjecture*, arXiv: 2306.15118. [2](#)
- [Coh06] P. M. Cohn: *Free Ideal Rings and Localization in General Rings*, New Mathematical Monographs 3, Cambridge University Press, Cambridge, 2006. [1](#), [3](#)
- [CR99] P. M. Cohn, C. Reutenauer: *On the construction of the free field*, Internat. J. Algebra Comput. 9 (1999) 307–323. [3](#), [4](#)
- [CMMPY22] B. Collins, B., T. Mai, A. Miyagawa, F. Parraud, S. Yin: *Convergence for noncommutative rational functions evaluated in random matrices*, Math. Ann. (2022) DOI:10.1007/s00208-022-02530-5. [1](#)

- [DM17] H. Derksen, V. Makam: *Polynomial degree bounds for matrix semi-invariants*, Adv. Math. 310 (2017) 44–63. [3](#), [4](#)
- [GGRW05] I. M. Gelfand, S. Gelfand, V. Retakh, R. L. Wilson: *Quasideterminants*, Adv. Math. 193 (2005) 56–141. [1](#)
- [GKZ94] I. M. Gelfand, M. M. Kapranov, A. V. Zelevinsky: *Discriminants, resultants, and multidimensional determinants*, Birkhäuser Boston, 1994. [4](#), [5](#)
- [Hal54] P. R. Halmos: *Commutators of operators II*, Amer. J. Math. 76 (1954) 191–198. [11](#)
- [Har92] J. Harris: *Algebraic geometry: a first course*, Graduate Texts in Mathematics 133, Springer-Verlag, New York, 1992. [12](#)
- [HMS18] J. W. Helton, T. Mai, R. Speicher: *Applications of realizations (aka linearizations) to free probability*, J. Funct. Anal. 274 (2018) 1–79. [1](#)
- [HMV06] J. W. Helton, S. McCullough, V. Vinnikov: *Noncommutative convexity arises from linear matrix inequalities*, J. Funct. Anal. 240 (2006) 105–191. [1](#)
- [HJ85] R. A. Horn, C. R. Johnson: *Matrix analysis*, Cambridge University Press, London, 1985. [7](#)
- [K-VV12] D. S. Kaliuzhnyi-Verbovetskyi, V. Vinnikov: *Noncommutative rational functions, their difference-differential calculus and realizations*, Multidimens. Syst. Signal Process. 23 (2012) 49–77. [1](#), [3](#)
- [K-VV14] D. S. Kalyuzhnyi-Verbovetskyi, V. Vinnikov: *Foundations of free noncommutative function theory*, Mathematical Surveys and Monographs 199, American Mathematical Society, Providence RI, 2014. [1](#)
- [K-BMR16] A.Kanel-Belov, S.Malev, L.Rowen, *Power-central polynomials on matrices*, J. Pure Appl. Algebra 220 (2016) 2164–2176. [7](#)
- [K-BRZ23] A. Kanel-Belov, F. Razavinia, W. Zhang: *Centralizers in free associative algebras and generic matrices*, Mediterr. J. Math. 20 (2023) no 85. [5](#)
- [LST11] M. Larsen, A. Shalev, P.H. Tiep: *The Waring problem for finite simple groups*, Ann. Math. 174 (2011) 1885–1950. [2](#)
- [M-L] L. Makar-Limanov: *Algebraically closed skew fields*, J. Algebra 93 (1985) 117–135. [14](#)
- [Mes] Z. Mesyan: *Commutator rings*, Bull. Austral. Math. Soc. 74 (2006) 279–288. [10](#)
- [NZM91] I. Niven, H. S.Zuckerman, H. L. Montgomery: *An introduction to the theory of numbers*, John Wiley & Sons, New York, 1991. [6](#)
- [PP23] S. Panja, S. Prasad, *The image of polynomials and Waring type problems on upper triangular matrix algebras*, J. Algebra 631 (2023) 148–193. [2](#)
- [Pro76] C. Procesi: *The invariant theory of $n \times n$ matrices*, Adv. Math. 19 (1976) 306–381. [11](#)
- [Row80] L. H. Rowen: *Polynomial identities in ring theory*, Pure and Applied Mathematics 84, Academic Press, Inc., New York-London, 1980. [3](#), [4](#)
- [Sha13] I. R. Shafarevich: *Basic algebraic geometry. 1*, 3rd edition, Springer, Heidelberg, 2013. [5](#)
- [Sou86] A.R. Sourour: *A factorization theorem for matrices*, Linear Multilinear Algebra 19 (1986) 141–147. [11](#), [12](#)
- [Vol17] J. Volčič: *On domains of noncommutative rational functions*, Linear Algebra Appl. 516 (2017) 69–81. [3](#)
- [Vol21] J. Volčič: *Hilbert’s 17th problem in free skew fields*, Forum Math. Sigma 9 (2021) E61. [1](#)

FACULTY OF MATHEMATICS AND PHYSICS, UNIVERSITY OF LJUBLJANA & FACULTY OF NATURAL SCIENCES AND MATHEMATICS, UNIVERSITY OF MARIBOR & IMFM, LJUBLJANA, SLOVENIA

Email address: `matej.bresar@fmf.uni-lj.si`

DEPARTMENT OF MATHEMATICS, DREXEL UNIVERSITY, PHILADELPHIA, PENNSYLVANIA

Email address: `jurij.volcic@drexel.edu`