Linear Algebra and its Applications 431 (2009) 1652–1674

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Simultaneous similarity and triangularization of sets of 2 by 2 matrices

Carlos A.A. Florentino

Instituto Superior Técnico, Univ. Técnica Lisboa, Departamento de Matemática, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

A R T I C L E   I N F O

Article history:
Received 15 July 2008
Accepted 31 May 2009
Available online 4 July 2009

Submitted by R. Guralnick

AMS classification:
15A21
15A72

Keywords:
Simultaneous similarity
Simultaneous triangularization
Invariants of 2 by 2 matrices

A B S T R A C T

Let $\mathcal{A} = (A_1, \ldots, A_n, \ldots)$ be a finite or infinite sequence of 2×2 matrices with entries in an integral domain. We show that, except in a very special case, $\mathcal{A}$ is (simultaneously) triangularizable if and only if all pairs (A_j, A_k) are triangularizable, for $1 \leqslant j, k \leqslant \infty$. We also provide a simple numerical criterion for triangularization.
Using constructive methods in invariant theory, we define a map (with the minimal number of invariants) that distinguishes simultaneous similarity classes for non-commutative sequences over a field of characteristic $\neq 2$. We also describe canonical forms for sequences of 2×2 matrices over algebraically closed fields, and give a method for finding sequences with a given set of invariants.

1. Introduction

1.1. Motivation

The properties of a set of square matrices which are invariant under simultaneous conjugation have been the subject of many investigations. In the case of a pair of matrices many problems have been solved including finding criteria for simultaneous similarity, simultaneous triangularization, existence of common eigenvectors, etc. Analogous problems have been solved for subalgebras, subgroups or sub-semigroups of square matrices (see, for example, [6,9,17] and references therein).

Here, we mainly concentrate on the problems of simultaneous similarity, simultaneous triangularization and canonical forms of countable sets of 2×2 matrices with entries in an arbitrary field,

E-mail address: cfloren@math.ist.utl.pt

0024-3795/$ - see front matter
doi:10.1016/j.laa.2009.05.037

with the focus on effectively computable solutions (the triangularization results will hold over integral domains).

Even though we are concerned with 2×2 matrices, it is convenient to describe the setup more generally as follows. Fix integers $m \geqslant 1$, $n \geqslant 1$ and let $V_{m,n}$ be the vector space of n-tuples of $m \times m$ matrices with entries in a field $\mathbb{F}$. The group of invertible matrices $G := GL_m(\mathbb{F})$ acts on $V_{m,n}$ by simultaneous conjugation on every component:

$$g \cdot \mathcal{A} := (gA_1g^{-1}, \ldots, gA_ng^{-1}), \tag{1.1}$$

where $g \in G$ and $\mathcal{A} = (A_1, \ldots, A_n) \in V_{m,n}$. By analogy with the case $n = 1$, the orbit $G \cdot \mathcal{A} := \{g \cdot \mathcal{A} : g \in G\} \subset V_{m,n}$ will be called the *conjugacy class* of $\mathcal{A}$, and can be viewed as an element $[\mathcal{A}] \in V_{m,n}/G$ of the quotient space. One can consider the following problems.

(i) Identify one element for each conjugacy class in a natural way, i.e, list all possible 'canonical forms';
(ii) Construct invariants that distinguish all conjugacy classes.

Naturally, these problems are related. In general, a solution to (i) will lead to a solution of (ii); however, the answer obtained in this way might be unnatural, and the description in terms of invariants is often useful.

These are difficult questions in general. Here, we will be concerned with the description and classification of conjugacy classes in the case $m = 2$. As we will see, in this case the questions above admit simple and complete answers, given by explicitly computable numerical criteria. In order to state our results in concise terms (see Definition 1.4) we will adopt the following terminology for (ordered) sets of matrices.

Definition 1.1. An element $\mathcal{A} = (A_1, \ldots, A_n) \in V_{m,n}$ will be called a *matrix sequence of size* $m \times m$ *and length* n, or just a *matrix sequence* when the integers m and n are assumed, and the matrices $A_1, \ldots, A_n$ will be called the *components* or *terms* of $\mathcal{A}$. We say that two matrix sequences $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ in $V_{m,n}$ are *similar* and write $\mathcal{A} \sim \mathcal{B}$ if they lie in the same conjugacy class (or G orbit) i.e, if there is an element $g \in G$ such that $\mathcal{B} = g \cdot \mathcal{A}$ (equivalently, $B_j = gA_jg^{-1}$ for all $j = 1, \ldots, n$).

We will also allow sequences $\mathcal{A} = (A_1, \ldots, A_n, \ldots) \in V_{m,\infty}$ with a countable infinite number of terms.

It turns out that a concrete answer to the above problems requires the separation of all matrix sequences (of fixed size and length) into G-invariant subclasses with distinct algebraic and/or geometric properties. For instance, for the class of 2×2 matrices that pairwise commute some of these questions are easier. This case can be reduced to the case of a single matrix ($n = 1$) which is solved with the Jordan decomposition theorem (at least when $\mathbb{F}$ is algebraically closed).

One property that is often relevant is that of simultaneous triangularization (or block triangularization); it generalizes commutativity and has a natural geometric interpretation. To be precise, let us consider the following standard notions. Note that a sequence $\mathcal{A} \in V_{m,n}$ can be viewed either as an ordered set of $m \times m$ matrices, as in the definition above, or alternatively, as a *single* $m \times m$ matrix whose entries are n-dimensional vectors.

$$\mathcal{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}, \quad a_{jk} \in \mathbb{F}^n. \tag{1.2}$$

Definition 1.2. Let $\mathcal{A}$ be a matrix sequence. $\mathcal{A}$ will be called *commutative* if all its terms pairwise commute (and non-commutative otherwise). We will say that $\mathcal{A}$ is an *upper triangular matrix sequence* if all the vectors below the main diagonal are zero ($a_{jk} = 0$ for all $j > k$ in Eq. (1.2)), and that $\mathcal{A}$ is *triangularizable* if it is similar to an upper triangular matrix sequence.

In geometric terms, $\mathcal{A}$ is triangularizable if and only if there is a full flag of vector subspaces of $\mathbb{F}^m$ which is invariant under every term A_j of $\mathcal{A}$ (for the standard action of $G = GL_m(\mathbb{F})$ on $\mathbb{F}^m$). Observe also that upper and lower triangularization are equivalent (over any ring). For the reasons mentioned, we will add the following *triangularization problem* to the previous list.

(iii) Give an effective numerical criterion for a matrix sequence to be triangularizable.

Problem (iii) was solved effectively by the following refinement of McCoy's Theorem [10]. Let $[A, B] = AB - BA$ denote the commutator of two matrices.

Theorem 1.3 (McCoy, see [10,9]). *Let $\mathbb{F}$ be algebraically closed. A $m \times m$ matrix sequence $\mathcal{A}$ of length n is triangularizable if and only if*

$$p(A_1, \ldots, A_n)[A, A']$$

is nilpotent for all monomials $p(x)$ (in non-commuting indeterminants $x_1, \ldots, x_n$) of total degree not greater than m^2 and all terms A, A' of $\mathcal{A}$.

Using results of Paz [13] and Pappacena [14], the bound on the degree d of the monomials $p(x)$ in Theorem 1.3 can be improved to $d \leqslant \frac{m^2}{3} + 1$ and $d \leqslant m\sqrt{\frac{2m^2}{m-1} + \frac{1}{4}} + \frac{m}{2} - 1$, respectively, each one being more efficient for smaller (resp. larger) values of m.

To discuss simultaneous similarity, consider the notion of subsequence.

Definition 1.4. A *subsequence* of a matrix sequence $\mathcal{A} = (A_1, \ldots, A_n) \in V_{m,n}$ is a matrix sequence of the form

$$\mathcal{A}_J = (A_{j_1}, \ldots, A_{j_l}) \in V_{m,l}$$

obtained from $\mathcal{A}$ by deleting some of its terms (here $J = (j_1, \ldots, j_l) \in \{1, \ldots, n\}^l$ for some natural number $l \leqslant n$, and the indices are strictly increasing: $1 \leqslant j_1 < \cdots < j_l \leqslant n$).

Let $\mathbb{A} = \mathbb{F}[A_1, \ldots, A_n]$ be the algebra generated over $\mathbb{F}$ by the terms of $\mathcal{A}$ (of dimension $\leqslant m^2$). Suppose we rearrange the terms of $\mathcal{A}$ such that $\mathbb{A}$ is in fact generated by just the first $k \leqslant m^2$ elements. Then, for all $j > k$, $A_j = p_j(A_1, \ldots, A_k)$ for some polynomial p_j with coefficients in $\mathbb{F}$. From this one can easily obtain a similarity test.

Fact 1.5. Let $\mathcal{A}, \mathcal{B}$ be two $m \times m$ matrix sequences of the same finite or infinite length n. Then $\mathcal{A}$ and $\mathcal{B}$ are similar if and only if all corresponding subsequences of length $\leqslant m^2$ are similar (that is, $\mathcal{A}_J \sim \mathcal{B}_J$ for all $J = (j_1, \ldots, j_l)$ with $1 \leqslant j_1 < \cdots < j_l \leqslant n$ and $l \leqslant m^2$).

One can improve this statement if we restrict to a big class of matrix sequences. The Dubnov–Ivanov–Nagata–Higman [4] theorem states that, given $m \in \mathbb{N}$, there is a natural number $N(m)$ with the following property: every associative algebra over $\mathbb{F}$ satisfying the identity $x^m = 0$ is nilpotent of degree $N(m)$ (i.e, any product $x_1 \cdots x_k$ of k elements $x_1, \ldots, x_k$ of the algebra is zero, for $k \geqslant N(m)$). It is known that $\frac{m(m+1)}{2} \leqslant N(m) \leqslant \min\{2^m - 1, m^2\}$ [16], and it was conjectured that $N(m) = \frac{m(m+1)}{2}$. This was verified to be true for $m = 2, 3, 4$.

The following remarkable result of Procesi [15] relates invariants of matrices and nilpotency degrees of associative nil-algebras. It provides explicit generators for the algebra of G-invariant regular (i.e, polynomial) functions on $V_{m,n}$. For a multiindex $J = (j_1, \ldots, j_k) \in \{1, \ldots, n\}^k$ of *length* $|J| = k$, define the G-invariant regular function $t_J : V_n(\mathbb{F}) \to \mathbb{F}$ as the trace of the word in the terms of $\mathcal{A} \in V_n(\mathbb{F})$ dictated by J, that is

$$t_J(\mathcal{A}) := \operatorname{tr}(A_{j_1} \cdots A_{j_k}). \tag{1.3}$$

Theorem 1.6 [15]. *Let $\mathbb{F}$ have characteristic 0, and let $f : V_{m,n} \to \mathbb{F}$ be a G-invariant polynomial function. Then f is a polynomial in the set of functions*

$$\{t_J : |J| \leqslant N(m)\}$$

where $|J|$ denotes the length of the multiindex J.

Let us say that a matrix sequence $\mathcal{A}$ is *semisimple* if its G orbit is Zariski closed in $V_{m,n}$. Semisimple sequences form a dense subset of $V_{m,n}$. Noting that $N(2) = 3$, Procesi's theorem has the following corollary (compare with Fact 1.5).

Corollary 1.7. *Let $\mathcal{A}, \mathcal{B}$ be two 2×2 semisimple matrix sequences of the same finite or infinite length. Then $\mathcal{A} \sim \mathcal{B}$ if and only if all corresponding subsequences of length $\leqslant 3$ are similar (that is, $\mathcal{A}_J \sim \mathcal{B}_J$ for all J with length $\leqslant 3$).*

This corollary gives a numerical criterion for simultaneous similarity of semisimple sequences. In the case $m = 2$, the number of generators can be further reduced, and moreover, Drensky described all relations in terms of a generating set [3]. He states his main theorem for traceless matrices, but an easy modification yields:

Theorem 1.8 [3]. *The generators of the algebra (over $\mathbb{F}$ of characteristic 0) of G-invariant polynomial functions on $V_{2,n}$ are given by:*

$$t_j,\ t_{jj},\ u_{jk},\ s_{jkl}, \qquad 1 \leqslant j < k < l \leqslant n,$$

where $u_{jk} := 2t_{jk} - t_j t_k$, $s_{jkl} := t_{jkl} - t_{lkj}$ and t_j, t_{jk}, t_{jkl} are as in (1.3). A full set of relations is

$$s_{abc}s_{def} + \frac{1}{4}\begin{vmatrix} u_{ad} & u_{ae} & u_{af} \\ u_{bd} & u_{be} & u_{bf} \\ u_{cd} & u_{ce} & u_{cf} \end{vmatrix} = 0, \quad u_{ea}s_{bcd} - u_{eb}s_{acd} + u_{ec}s_{abd} - u_{ed}s_{abc} = 0 \tag{1.4}$$

for all appropriate indices.

In this article, we show that these same generators can be used to get even more explicit solutions to problems (i)–(iii) in the case $m = 2$. Our methods are mostly elementary and their generalization to higher m seems possible, although computationally very involved.

1.2. Statement of the main results

From now on, we restrict to the space $V_n = V_{2,n}$ of sequences of 2×2 matrices of length $n \in \mathbb{N} \cup \{\infty\}$, i.e, the case $m = 2$, except where explicitly stated.

The article can be roughly divided into two parts. In the first part, Section 2, we work with the space $V_n(R)$ of 2×2 matrix sequences with coefficients in an integral domain R, on which the group $G = GL_2(R)$ of invertible matrices over R acts by conjugation. Here, we define and study reduced sequences, consider the triangularization problem, and prove Theorems 1.11–1.13, stated below. They provide an efficient numerical criterion for triangularization of matrix sequences in $V_n(R)$.

The case $m = 2$ is simple enough that some easy arguments already improve some of the statements above, even in the more general case of 2×2 matrices over integral domains.

Proposition 1.9. *$\mathcal{A} \in V_n(R)$ is triangularizable if and only if every subsequence of length $\leqslant 3$ is triangularizable.*

For simultaneous similarity, a simple argument generalizes and improves Corollary 1.7, in the case $m = 2$, to account for all matrix sequences (not necessarily semisimple).

Proposition 1.10. *Over an integral domain $\mathcal{A} \sim \mathcal{B}$ if and only if $\mathcal{A}_J \sim \mathcal{B}_J$ for all J of length $\leqslant 3$. Moreover, under the generic condition $[A_1, A_2] \neq 0$, $\mathcal{A}$ is similar to $\mathcal{B}$ if and only if $(A_1, A_2, A_j) \sim (B_1, B_2, B_j)$, for all $j = 3, \ldots, n$.*

In other words, the conjugation action of $GL_2(R)$ on 2×2 matrix sequences of any length is completely determined by the same action on triples of 2×2 matrices. These two propositions seem to be standard, and we thank Guralnick for furnishing simple arguments leading to their proofs, included below for convenience and completeness.

The statement of Proposition 1.9 is the best possible in this generality, as there are 2×2 matrix sequences of length 3 (therefore, *a fortiori* for every size $m \times m$, $m \geqslant 2$ and every length $n \geqslant 3$) that are not triangularizable but are pairwise triangularizable (see Example 2.11). However, these sequences are very special (see Theorem 1.12 below).

Let us now consider the problem of finding a numerical criterion for triangularization of general 2×2 sequences. By Proposition 1.9, we just need to consider *triples* of 2×2 matrices. Using only the G-invariant functions given by the trace and the determinant, the following is a simple numerical criterion.

Theorem 1.11. *A 2×2 matrix sequence $\mathcal{A}$ (over R) of length n is triangularizable if and only if all its terms are triangularizable and*

$$\det (AB - BA) = \operatorname{tr}(ABC - CBA) = 0 \tag{1.5}$$

for all terms A, B, C of $\mathcal{A}$. In particular, a pair $(A, B) \in V_2(R)$ is triangularizable if and only if both A and B are triangularizable and $\det (AB - BA) = 0$.

Remark. Over an algebraically closed field $\overline{\mathbb{F}}$, every single matrix is triangularizable. Suppose $\mathcal{A}$ is a 2×2 matrix sequence over $\overline{\mathbb{F}}$ with $[A, B]$ and $C[A, B]$ nilpotent for all terms A, B, C of $\mathcal{A}$. Then, Eq. (1.5) holds and Theorem 1.11 shows that all bounds mentioned after Theorem 1.3 can be improved for $m = 2$, as the condition that p is a monomial of degree $\leqslant 1$ is already sufficient (and necessary) for triangularization.

Theorem 1.11 generalizes a result proved in [5] for algebraically closed fields. Observe that the case $n = 2$ of this theorem is a direct generalization to integral domains of a well-known criteria, obtained in [6], for a pair of 2×2 matrices over $\overline{\mathbb{F}}$ to be triangularizable.

As a consequence of our study of invariants, we can improve Proposition 1.9 under a simple non-degeneracy condition on $\mathcal{A}$. Let us define a *reduced* sequence to be one with no commuting pairs among its terms.

Theorem 1.12. *A 2×2 reduced matrix sequence $\mathcal{A}$ of length $\geqslant 4$ is triangularizable (over R) if and only if all subsequences of $\mathcal{A}$ of length $\leqslant 2$ are triangularizable (over R).*

With a little more care we get a test, computationally much more efficient, whose complexity grows *only linearly* with the number n of matrices in $\mathcal{A}$. Let us define the reduced length of a sequence $\mathcal{A}$ to be the biggest length of a reduced subsequence $\mathcal{B} \subset \mathcal{A}$.

Theorem 1.13. *Let $\mathcal{A} = (A_1, \ldots, A_n) \in V_n$ have reduced length $l \leqslant n$ and rearrange its terms so that $\mathcal{A}' = (A_1, \ldots, A_l)$ is reduced. Then*

(i) *In the case $l \leqslant 3$, $\mathcal{A}$ is triangularizable if and only if $\mathcal{A}'$ is triangularizable.*
(ii) *In the case $l \geqslant 4$, $\mathcal{A}$ is triangularizable if and only if $A_1, \ldots, A_l$ are triangularizable and* $\det ([A_j, A_k]) = 0$ *for all $j = 1, 2, 3$ and all $j < k \leqslant l$.*

In the second part of the article, we work mainly over a field $\mathbb{F}$.

Section 3 deals with simultaneous similarity for 2×2 matrix sequences. Using standard invariant theory, one sees that the values of the Drensky generators are enough to distinguish semisimple conjugacy classes. But these classes should depend on $4n - 3$ parameters only, the "dimension" of the quotient space V_n/G (V_n has dimension $4n$, and G acts as $SL_2(\mathbb{F})$, a three dimensional group) which is much less than the number, $2n + \binom{n}{2} + \binom{n}{3}$, of Drensky generators. After describing rational invariants that distinguish general triangularizable sequences, we obtain a solution, with the minimal number of invariants, to problem (ii) for non-commutative sequences as follows.

Let $\mathcal{S}'$ (resp. $\mathcal{U}'$) be the dense subsets of $V_n = V_n(\mathbb{F})$ of semisimple (resp. triangularizable) sequences such that A_1 is diagonalizable and $[A_1, A_2] \neq 0$. Using the G-invariant functions $t_J : V_n \to \mathbb{F}$ in (1.3) define the maps $\Phi' : \mathcal{S}'/G \to \mathbb{F}^{4n-3}$ and $\Psi' : \mathcal{U}'/G \to \mathbb{F}^{2n} \times \mathbb{P}^{n-2}$ ($\mathbb{P}^k$ denotes the projective space over $\mathbb{F}$ of dimension k), by

$$\Phi'([\mathcal{A}]) = (t_1, t_{11}, t_2, t_{22}, t_{12}, \ldots, t_k, t_{1k}, t_{2k}, s_{12k}, \ldots, t_n, t_{1n}, t_{2n}, s_{12n}), \tag{1.6}$$

$$\Psi'([\mathcal{A}]) = \left(t_1, t_{11}, t_2, t_{12}, \ldots, t_k, t_{1k}, \ldots, t_n, t_{1n};\ \psi'\right), \tag{1.7}$$

where ψ' is defined later in Section 4, and $[\mathcal{A}]$ denotes the conjugacy class of $\mathcal{A}$.

Theorem 1.14. *Let $\mathbb{F}$ be a field of characteristic $\neq 2$. The map $\Phi' : \mathcal{S}'/G \to \mathbb{F}^{4n-3}$ is injective and the map $\Psi' : \mathcal{U}'/G \to \mathbb{F}^{2n} \times \mathbb{P}^{n-2}$ is two-to-one.*

The last section (Section 4) concerns the classification of canonical forms for sets of 2×2 matrices over $\overline{\mathbb{F}}$ and proposes a solution for problem (i). The main result is Theorem 4.3, where five types of canonical forms are obtained for sequences with at least one non-commuting pair. It shows that, for non-commutative sequences, the restriction to $\mathcal{S}'$ and $\mathcal{U}'$ in Theorem 1.14 is only apparent (see also Remark 4.6). We also describe a simple method for finding a sequence in canonical form with a given value of Φ' or Ψ'.

Appendix A contains results on the triangularization a single 2×2 matrix over R which are crucial for Theorem 1.11, and Appendix B, for completeness, describes the well-known canonical forms of commuting matrices over $\overline{\mathbb{F}}$.

2. Simultaneous triangularization

Throughout the article, R will stand for an integral domain, $\mathbb{F}$ for a field and $\overline{\mathbb{F}}$ for an algebraically closed field. V_n (resp. G) will denote the space of matrix sequences of length $n \in \mathbb{N} \cup \{\infty\}$ (resp. the group of invertible 2×2 matrices) over the appropriate ring or field. When the coefficients need to be explicitly mentioned, we will use the notations $V_n(R)$, $G(R)$, etc. Sequences of length 1, 2, 3 and 4 will be called *singlets*, *pairs*, *triples* and *quadruples*, respectively.

2.1. Simultaneous triangularization and subtriples; reduced sequences

We start by fixing notation and recalling some well known facts about matrices over $\mathbb{F}$ and R. After this, we define *reduced sequences*, a notion which will be fundamental in the sequel.

For a given $\mathcal{A} \in V_n(R)$, a matrix sequence of length $n \in \mathbb{N}$, we will use the notation

$$\mathcal{A} = (A_1, \ldots, A_n) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in R^n,$$

$$A_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}, \quad j = 1, \ldots, n$$

and we let $e = (e_1, \ldots, e_n)$ denote the n-tuple $a - d \in R^n$. To avoid the most trivial case, we consider only matrix sequences with at least one non-scalar term.

The commutator of 2 matrices A_1 and A_2 is given by

$$[A_1, A_2] = \begin{pmatrix} b_1c_2 - c_1b_2 & e_1b_2 - b_1e_2 \\ c_1e_2 - e_1c_2 & c_1b_2 - b_1c_2 \end{pmatrix}. \tag{2.1}$$

For later use, record the following straightforward but useful lemma.

Lemma 2.1. *Let $\mathcal{A} = (A_1, A_2) \in V_2(R)$ and A_1 be a non-scalar matrix. If A_1 is upper triangular, then $[A_1, A_2] = 0$ if and only if A_2 is also upper triangular and*

$$b_1e_2 - e_1b_2 = 0. \tag{2.2}$$

Similarly, let A_1 be diagonal non-scalar. Then $[A_1, A_2] = 0$ if and only if A_2 is also diagonal.

Proof. Suppose $[A_1, A_2] = 0$ with $c_1 = 0$. Then we have $b_1e_2 - e_1b_2 = b_1c_2 = e_1c_2 = 0$, using (2.1). Since A_1 is non-scalar either e_1 or b_1 is non-zero. In an integral domain, this implies $b_1e_2 - e_1b_2 = c_2 = 0$. The other statement is similar. □

This implies the following well known result. Note that a 2×2 matrix is non-scalar if and only if it is nonderogatory.

Proposition 2.2. *Let $\mathcal{A}$ be a commutative matrix sequence (i.e, all terms pairwise commute) of finite or infinite length over an integral domain R. Then $\mathcal{A}$ is triangularizable if and only if one of its non-scalar terms is triangularizable. Similarly, $\mathcal{A}$ is diagonalizable if and only if one of its non-scalar terms is diagonalizable.*

As a consequence, the proof of Proposition 1.9 is complete after the following. Let us denote by $\mathbb{A}$ the algebra generated over $\mathbb{F}$, the field of fractions of R, by the terms of $\mathcal{A}$. We use a well known result: $\mathcal{A}$ is commutative if and only if the dimension of $\mathbb{A}$ is $\leqslant 2$.

Proposition 2.3. *Let $n \geqslant 2$ and $\mathcal{A} \in V_n(R)$ be a non-commutative matrix sequence. Then $\mathcal{A}$ is triangularizable if and only if all subsequences of $\mathcal{A}$ of length $\leqslant 3$ are triangularizable.*

Proof. If $\mathcal{A}$ is triangularizable, it is clear that all subsequences of $\mathcal{A}$ are also triangularizable. Conversely, since $\mathcal{A}$ is non-commutative, we can assume without loss of generality, that $[A_1, A_2] \neq 0$, and that, after a suitable conjugation, both A_1 and A_2 are upper triangular. By hypothesis, all triples of the form (A_1, A_2, A_k), $k = 3, \ldots, n$ are triangularizable. Then, the algebra $\mathbb{A}_k$ generated by $\mathcal{A}_k = (A_1, A_2, A_k)$ equals the one generated by (A_1, A_2), since one is a subset of the other but both are three dimensional over $\mathbb{F}$. Indeed, if $\mathbb{A}_k$ was of dimension $\leqslant 2$, $\mathcal{A}_k$ would be commutative, and if $\mathbb{A}_k$ was four dimensional, $\mathcal{A}_k$ would not be triangularizable. Therefore, for all $j = 3, \ldots, n$, $A_j = p_j(A_1, A_2)$ for some polynomial p_j with coefficients in $\mathbb{F}$ and thus, A_j is also upper triangular, for all j. □

Recall also the following.

Lemma 2.4. *Let $\mathcal{A} = (A_1, A_2) \in V_2(\mathbb{F})$ be commutative and A_1 be a non-scalar matrix. Then $A_2 = p(A_1)$ for some degree 1 polynomial $p(x) \in \mathbb{F}[x]$.*

Proof. From Eq. (2.1), the conditions $[A_1, A_2] = 0$ can be written as $Mu = 0$, where

$$M = \begin{pmatrix} 0 & c_1 & -e_1 \\ -c_1 & 0 & b_1 \\ e_1 & -b_1 & 0 \end{pmatrix}, \quad u = (b_2, e_2, c_2). \tag{2.3}$$

Note that M has rank exactly 2, since A_1 is non-scalar. So, the vector $u = (b_2, e_2, c_2)$ is in the nullspace of M, which is generated by $(b_1, e_1, c_1) \neq 0$. Thus, there is an $\alpha \in \mathbb{F}$ such that $(b_2, e_2, c_2) = \alpha(b_1, e_1, c_1)$. Then, $A_2 = (d_2 - \alpha d_1)I + \alpha A_1$ is, explicitly, the required polynomial (I denotes the identity 2×2 matrix). □

It is clear that a single matrix is triangularizable over an algebraically closed field, but not necessarily so over a general integral domain or field. In Appendix A we include a short account of the conditions for triangularization of a single matrix in $V_1(R)$. The following notion will play a central role. If $\mathcal{A}$ is a subsequence of $\mathcal{B}$, we will write $\mathcal{A} \subseteq \mathcal{B}$.

Definition 2.5. A matrix sequence with at least one non-scalar term $\mathcal{A} = (A_1, \dots) \in V_n(R)$ is called *reduced* if there are no commuting pairs among its terms, that is, $[A_j, A_k] \neq 0$, for all $1 \leqslant j < k \leqslant n$. We say that $\mathcal{A}$ is a *reduction* of $\mathcal{B}$ if $\mathcal{A}$ is reduced and is obtained from $\mathcal{B}$ by deleting some of its terms. Finally, we say that $\mathcal{A}$ is a *maximal reduction* of $\mathcal{B}$, and that l is its *reduced length*, if $\mathcal{A}$ is a reduction of $\mathcal{B}$ of length l, and any subsequence $\mathcal{A}' \subseteq \mathcal{B}$ with length $> l$ is not reduced.

Over a field, by Lemma 2.4, a reduced sequence $\mathcal{A}$ is one where no term is a polynomial function of another (so all terms generate an algebra of dimension 2, and no two terms generate the same subalgebra of the full matrix algebra). It is clear that any two maximal reductions have the same length. Note also that any subsequence of a reduced sequence is also reduced. The following facts show that important properties like existence of a triangularization are captured by any maximal reduction of a matrix sequence.

Proposition 2.6. *Let $\mathcal{A} = (A_1, \dots, A_n) \in V_n(R)$ and let A_{n+1} commute with at least one non-scalar term of $\mathcal{A}$. Then $\mathcal{A}$ is triangularizable if and only if $\mathcal{A}' := (A_1, \dots, A_n, A_{n+1})$ is triangularizable.*

Proof. Naturally if $\mathcal{A}'$ is triangularizable, $\mathcal{A}$ is also. For the converse, without loss of generality assume $[A_1, A_{n+1}] = 0$ with A_1 non-scalar and $\mathcal{A}$ in upper triangular form. Then, by Lemma 2.1, A_{n+1} is also upper triangular, so that $\mathcal{A}'$ is triangularizable. $\square$

Corollary 2.7. *If $\mathcal{A}$ and $\mathcal{A}'$ are arbitrary sequences with a common maximal reduction (of length $\geqslant 1$), then either they are both triangularizable or both not triangularizable.*

Proof. Let $\mathcal{B}$ be such a common maximal reduction. Then $\mathcal{A}$ and $\mathcal{A}'$ are obtained from $\mathcal{B}$ by adding scalar matrices or matrices that commute with some of the non-scalar terms of $\mathcal{B}$. So, if $\mathcal{B}$ is triangularizable, both $\mathcal{A}$ and $\mathcal{A}'$ are triangularizable by repeatedly applying Proposition 2.6. The case when $\mathcal{B}$ is not triangularizable is analogous. $\square$

2.2. Necessary conditions for triangularization via invariants

We continue to work over an integral domain R. Define the following important $GL_2(R)$-invariant functions. For a matrix $A \in V_1$, let δ_A denote the discriminant of its characteristic polynomial, that is $\delta_A = \mathrm{tr}^2 A - 4 \det A$.

Definition 2.8. Let $\tau, \sigma : V_2(R) \to R$ and $\Delta : V_3(R) \to R$ be defined by

$$\begin{aligned}
\tau(A, B) &:= 2\mathrm{tr}(AB) - \mathrm{tr}(A)\mathrm{tr}(B),\\
\sigma(A, B) &:= \det(AB - BA),\\
\Delta(A, B, C) &:= (\mathrm{tr}(ABC - CBA))^2 .
\end{aligned}$$

When a matrix sequence is written as $\mathcal{A} = (A_1, \dots, A_n) \in V_n(R)$ we will also use

$$\begin{aligned}
\tau_{jk} &= \tau_{jk}(\mathcal{A}) := \tau(A_j, A_k),\\
\sigma_{jk} &= \sigma_{jk}(\mathcal{A}) := \sigma(A_j, A_k),\\
\Delta_{jkl} &= \Delta_{jkl}(\mathcal{A}) := \Delta(A_j, A_k, A_l)
\end{aligned}$$

for any indices $j,k,l \in \{1,\ldots,n\}$. By simple computations , we can express these functions in terms of b_j, c_j, e_j as follows:

$$\tau_{jk} = e_j e_k + 2b_j c_k + 2c_j b_k,$$

$$\sigma_{jk} = (b_j e_k - e_j b_k)(c_j e_k - e_j c_k) - (b_j c_k - c_j b_k)^2, \tag{2.4}$$

$$\Delta_{jkl} = \begin{vmatrix} b_j & b_k & b_l \\ e_j & e_k & e_l \\ c_j & c_k & c_l \end{vmatrix}^2. \tag{2.5}$$

Note that τ,σ and Δ are symmetric under permutation of any matrices/indices, but σ and Δ vanish when 2 matrices/indices coincide. Since the above expressions do not depend explicitly on the variables a_j or d_j but only on the difference $e_j = a_j - d_j$, the functions τ,σ and Δ are invariant under translation of any argument by a scalar matrix, that is, for any matrices A, B and scalar matrices λ,μ, we have $\tau(A+\lambda, B+\mu) = \tau(A,B)$ and similarly for σ and Δ.

Remark 2.9. Note that these are essentially the same functions used in Drensky's theorem 1.8 [3]. They also coincide with the functions used in [5], up to a constant factor. There are some interesting relations between these invariants which are obtained from simple calculations. In particular, we have

$$\tau(A,A) = \delta_A = \operatorname{tr}^2 A - 4\det A,$$

$$\sigma(A,B) = \operatorname{tr}(A[A,B]B) = \frac{1}{4}\left(\tau(A,A)\tau(B,B) - \tau(A,B)^2\right), \tag{2.6}$$

$$\Delta(A,B,C) = -\frac{1}{4}\begin{vmatrix} \tau(A,A) & \tau(A,B) & \tau(A,C) \\ \tau(B,A) & \tau(B,B) & \tau(B,C) \\ \tau(C,A) & \tau(C,B) & \tau(C,C) \end{vmatrix}$$

for all matrices A,B,C over R, in agreement with Eq. (1.4).

The following is a simple necessary condition for triangularization.

Proposition 2.10. *Let $\mathcal{A} \in V_n$ be a triangularizable sequence. Then $\sigma(A,B)$ and $\Delta(A,B,C)$ vanish for all terms A,B,C of $\mathcal{A}$.*

Proof. Since σ and Δ are G invariant, we can assume that $\mathcal{A}$ is upper triangular. By direct computation, $\sigma(A,B) = \det([A,B]) = 0$, and $\Delta(A,B,C) = (\operatorname{tr}(ABC - CBA))^2 = 0$. $\square$

Note that the vanishing of all $\sigma_{jk} = \sigma(A_j, A_k)$ is not sufficient for $\mathcal{A}$ to be triangularizable, as the following important example shows. We adopt the usual convention that blank matrix entries stand for zero entries (in this case $0 \in R$).

Example 2.11. Let $\mathcal{A} = (A_1, A_2, A_3) \in V_3$ have the form

$$A_1 = \begin{pmatrix} a_1 & \\ & d_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & b_2 \\ & d_2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} a_3 & \\ c_3 & d_3 \end{pmatrix},$$

for some $a_1,\ldots,d_3 \in R$. Then $\sigma_{12} = \sigma_{13} = 0$ and $\sigma_{23} = -b_2c_3(e_2e_3 + b_2c_3)$. Assume that $e_2e_3 + b_2c_3 = 0$ and that $e_1b_2c_3 \neq 0$, so that all σ_{jk} vanish, neither A_2 or A_3 are diagonal, and (since these assumptions imply $e_2e_3 \neq 0$) all three matrices have distinct eigenvalues. So, in this case, all subsequences of length $\leqslant 2$ are triangularizable, but the next Proposition will show that $\mathcal{A}$ is not triangularizable.

Proposition 2.12. *As in Example 2.11, let $\mathcal{A} = (A_1, A_2, A_3) \in V_3$ be a triple of the form*

$$A_1 = \begin{pmatrix} a_1 & \\ & d_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & b_2 \\ & d_2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} a_3 & \\ c_3 & d_3 \end{pmatrix}, \tag{2.7}$$

with $e_2e_3 \neq 0$. Then, the following are equivalent.

(i) $\mathcal{A}$ *is reduced,* (ii) $\mathcal{A}$ *is not triangularizable,* (iii) $\Delta_{123} \neq 0$ *(i.e, $e_1b_2c_3 \neq 0$).*

Proof. (i) or (ii) imply (iii): If $e_1b_2c_3 = 0$ at least one of the factors is zero. In each case, A_1 is a scalar, $\mathcal{A}$ is lower triangular, or $\mathcal{A}$ is upper triangular, respectively, so $\mathcal{A}$ is triangularizable and is not reduced, since A_1 commutes with one or both of the other terms. (iii) implies (ii) and (i): Suppose that $e_1b_2c_3 \neq 0$. Then, none of the three numbers is zero. Let g be the $SL_2(R)$ matrix with columns (x, y) and (z, w). Then

$$\begin{aligned} gA_1g^{-1} &= \begin{pmatrix} * & -xze_1 \\ ywe_1 & * \end{pmatrix}, \\ gA_2g^{-1} &= \begin{pmatrix} * & x(xb_2 - ze_2) \\ y(we_2 - yb_2) & * \end{pmatrix}, \\ gA_3g^{-1} &= \begin{pmatrix} * & -z(xe_3 + zc_3) \\ w(ye_3 + wc_3) & * \end{pmatrix} \end{aligned} \tag{2.8}$$

from which it follows that there is no $g \in G$ that will make $g \cdot \mathcal{A}$ upper or lower triangular, so $\mathcal{A}$ is not triangularizable. Also, by Lemma 2.1, none of the 3 commutators between the pairs will vanish, so $\mathcal{A}$ is reduced. $\square$

2.3. Numerical criteria for triangularization

A simple necessary and sufficient numerical condition for triangularization of a pair of 2×2 matrices over an algebraically closed field was given in the article [6], which also describes the similarity classes of pairs of $m \times m$ matrices in great generality.

Proposition 2.13 [6]. *A pair $(A_1, A_2) \in V_2(\overline{\mathbb{F}})$ is triangularizable if and only if $\sigma_{12} = \det[A_1, A_2] = 0$.*

Note that Friedland writes the condition $\sigma_{12} = 0$ in a different, but equivalent form (see Eq. (2.6) in Remark 2.9). The generalization to integral domains is as follows.

Theorem 2.14. *A pair $\mathcal{A} = (A_1, A_2) \in V_2(R)$ is triangularizable if and only if both A_1 and A_2 are triangularizable and $\sigma_{12} = \det[A_1, A_2] = 0$.*

Proof. The conditions are clearly necessary. For the converse, let us suppose that both A_1 and A_2 are triangularizable and $\det[A_1, A_2] = 0$. Then, as the determinant is $GL_2(R)$ invariant, we can assume A_1 upper triangular ($c_1 = 0$). If $[A_1, A_2] = 0$ the pair (A_1, A_2) is triangularizable by Corollary 2.2, so we can assume that $[A_1, A_2] \neq 0$. Eq. (2.1) shows that

$$0 = \det[A_1, A_2] = -b_1^2c_2^2 + e_1c_2(e_1b_2 - b_1e_2) = -c_2(b_1^2c_2 + e_1b_1e_2 - e_1^2b_2). \tag{2.9}$$

If $c_2 = 0$, $\mathcal{A}$ is triangularizable. If not, $c_2 \neq 0$ and we distinguish four cases. (i) If $e_1 = 0$, then $b_1 \neq 0$ (as A_1 is non-scalar) which makes Eq. (2.9) impossible to solve. (ii) If $b_1 = 0$, then $e_1 \neq 0$, and Eq. (2.9) implies $b_2 = 0$ and $\mathcal{A}$ is lower triangular. (iii) Suppose now $e_1b_2 = b_1e_2$. Then $0 = \det[A_1, A_2] = -c_2^2b_1^2$ and so $b_1 = 0$ which reduces to the previous case.

Finally, consider the general case (iv) with $\delta_{12} = b_1e_2 - e_1b_2 \neq 0$ and non-zero b_1 and e_1. So, we are assuming $c_2e_2b_2 \neq 0$. From Eq. (2.9), the quadratic equation $Q_2(x, y) \equiv c_2x^2 - e_2xy - b_2y^2 = 0$ associated to A_2 (see Appendix A) has a non-trivial solution: $(b_1, -e_1) \in R^2$. Suppose that A_2 is nondegenerate ($\delta_A \neq 0$). Then, by Lemma A.2, there are z', w' in $\mathbb{F}$, the field of fractions of R, such that

$w'b_1 + z'e_1 \neq 0$ and the eigenvectors of A_2 are multiples of $(b_1, -e_1)$ and of (z', w'). So, we choose an eigenvector of A_2 of the form $(z', w') \in R^2$ colinear with $(z', w') \in \mathbb{F}^2$. Moreover, by Proposition A.1, there are $\alpha, \beta \in \mathbb{F}^* = \mathbb{F} \setminus \{0\}$ so that the eigenvectors $(x, y) = \alpha(b_1, -e_1)$ and $(z, w) = \beta(z', w')$ verify either $xR + yR = R$ or $zR + wR = R$. If the first alternative holds, let $xq - yp = 1$ for some $p, q \in R$, (note that (p, q) is not necessarily an eigenvector of A_2) and put:

$$g = \begin{pmatrix} x & p \\ y & q \end{pmatrix}.$$

Then, conjugating by g^{-1} gives

$$g^{-1}A_1 g = \begin{pmatrix} * & * \\ -e_1 yx - b_1 y^2 & * \end{pmatrix} = \begin{pmatrix} * & * \\ \alpha(y^2 x - xy^2) & * \end{pmatrix} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$$
$$g^{-1}A_2 g = \begin{pmatrix} * & * \\ c_2 x^2 - e_2 xy - b_2 y^2 & * \end{pmatrix} = \begin{pmatrix} * & * \\ \alpha^2 Q_2(b_1, -e_1) & * \end{pmatrix} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix},$$

so $\mathcal{A}$ is again triangularizable. If the second alternative holds, we do the same interchanging the roles of (x, y) and (z, w). Finally, suppose that A_2 is degenerate ($\delta_A = 0$). Then, there is only one eigenvector of A_2 and the solutions to $Q_2(x, y) = 0$ form a single line through the origin in $\mathbb{F}^2$, so all solutions $(x, y) \in R^2$ are multiples of $(b_1, -e_1) \in R^2$. Since A_2 is triangularizable, by Proposition A.1, we can choose (x, y) so that $xR + yR = R$ and we proceed as before. $\square$

Example 2.15. Over the integral domain $R = \mathbb{C}[u, v]$, consider the pair

$$A_1 = \begin{pmatrix} -v & u \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} uv & u^2 \\ 2v^2 & 0 \end{pmatrix}.$$

Then we have A_1 upper triangular and $\sigma_{12} = 0$ as can be checked, but A_2 is not triangularizable over $\mathbb{C}[u, v]$, as no eigenvector $(x, y) \in R^2$ satisfies $xR + yR = R$. So, $\sigma_{12} = 0$ and A_1 is triangularizable but not the pair (A_1, A_2).

We finally arrive to Theorem 1.11, which is a converse to Proposition 2.10.

Theorem 2.16. *A sequence $\mathcal{A} = (A_1, \ldots, A_n) \in V_n(R)$ is triangularizable if and only if each A_j is triangularizable and $\sigma_{jk} = \Delta_{jkl} = 0$ for all $1 \leqslant j, k, l \leqslant n$.*

Proof. Consider $\mathcal{A} = (A_1, \ldots, A_n)$ reduced with $\sigma_{jk} = \Delta_{jkl} = 0$ for all $1 \leqslant j, k, l \leqslant n$. By Theorem 2.14 the conditions $\sigma_{jk} = 0$ and A_j triangularizable mean that all subsequences of $\mathcal{A}$ of length $\leqslant 2$ are triangularizable. So, after a similarity that puts A_1 and A_2 in upper triangular form, we can assume (A_1, A_2, A_3) to be in the form

$$A_1 = \begin{pmatrix} a_1 & b_1 \\ & d_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & b_2 \\ & d_2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} a_3 & b_3 \\ c_3 & d_3 \end{pmatrix}.$$

Since $\mathcal{A}$ is reduced, by hypothesis $\delta_{12} = b_1 e_2 - e_1 b_2 \neq 0$. From Eq. (2.5), we see that $0 = \Delta_{123} = (b_1 e_2 - e_1 b_2)^2 c_3^2$, so $c_3 = 0$ which means that (A_1, A_2, A_3) is triangularizable. Repeating the argument for all triples (A_j, A_k, A_l) we see that all subsequences of $\mathcal{A}$ of length $\leqslant 3$ are triangularizable. So $\mathcal{A}$ is triangularizable by Proposition 1.9. Finally, if $\mathcal{A}$ is not reduced, we consider the above argument for any maximal reduction $\mathcal{B}$, triangularize $\mathcal{B}$, and apply Corollary 2.7. $\square$

2.4. Improved criterion for triangularization

We now prove an inductive property of reduced sequences that allow us to improve the criterion of Theorem 2.16.

Theorem 2.17. *Let $n \geqslant 4$ and $\mathcal{A} = (A_1, \ldots, A_{n-1}) \in V_{n-1}$ be a reduced triangularizable sequence. If $\sigma(A_j, A_n) = 0$ for some matrix A_n, and all $j = 1, \ldots, n-1$ then $(A_1, \ldots, A_n)$ is also triangularizable.*

Proof. We can suppose that $(A_1, \ldots, A_{n-1})$ has been conjugated so that it is already an upper triangular matrix sequence. So all the σ_{jk} vanish, for indices j, k between 1 and $n-1$ (by Proposition 2.10). To reach a contradiction, assume that $\mathcal{A}' = (A_1, \ldots, A_n)$ is not triangularizable so that

$$A_n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$$

is non-scalar with $c_n \neq 0$. Since $\mathcal{A}$ is reduced, none of the A_j can be scalar. We can also assume that A_n does not commute with some A_j otherwise $\mathcal{A}'$ would be triangularizable by Proposition 2.6. This means that $\delta_{jn} := b_j e_n - e_j b_n \neq 0$, for $j = 1, \ldots, n-1$, by Lemma 2.1. Using Formula (2.4), the $n-1$ conditions $\sigma_{jn} = 0, j = 1, \ldots, n-1$ can be written as (because $c_n \neq 0$)

$$b_j^2 c_n + b_j e_j e_n - e_j^2 b_n = 0 \quad \text{for } j = 1, \ldots, n-1.$$

Since A_n is non-scalar, we are looking for a non-zero solution $u = (b_n, e_n, c_n) \in R^3$ to the matrix equation $Bu = 0$ where

$$B = \begin{pmatrix} -e_1^2 & e_1 b_1 & b_1^2 \\ \vdots & \vdots & \vdots \\ -e_{n-1}^2 & e_{n-1} b_{n-1} & b_{n-1}^2 \end{pmatrix}.$$

A simple computation shows that every 3×3 minor of B is of the form $\pm\delta_{jk}\delta_{kl}\delta_{lj}$. Since all these minors are non-zero by hypothesis, there is no non-zero solution $u \in R^3$, and we have a contradiction. Hence $c_n = 0$ and $\mathcal{A}'$ is triangularizable. $\square$

From two finite matrix sequences $\mathcal{A} = (A_1, \ldots, A_{n_1})$ and $\mathcal{B} = (B_1, \ldots, B_{n_2})$ one can form their concatenation $\mathcal{A} \cup \mathcal{B} := (A_1, \ldots, A_{n_1}, B_1, \ldots, B_{n_2})$. The following corollary may be called the *concatenation principle* for triangularizable sequences.

Corollary 2.18. *If $\mathcal{A} \in V_{n_1}$ and $\mathcal{B} \in V_{n_2}$ are triangularizable matrix sequences and they have a common reduction of length $\geqslant 3$, then their concatenation $\mathcal{A} \cup \mathcal{B} \in V_{n_1+n_2}$ is also triangularizable.*

Proof. Let $\mathcal{C}$ be such a common reduction, which we can assume to have length 3. $\mathcal{C} = (C_1, C_2, C_3)$ is obviously triangularizable and $\sigma(C_j, A_k) = \sigma(C_j, B_k) = 0$ for all possible indices, since both $\mathcal{A}$ and $\mathcal{B}$ are triangularizable. So the Proposition above applies. $\square$

To prove Theorems 1.12 and 1.13, we will need the following easy fact.

Lemma 2.19. *Let $n \geqslant 2$. A reduced triangularizable sequence of length n as at least $n-1$ diagonalizable terms.*

Proof. We can assume that $\mathcal{A}$ is already is upper triangular form, and let A_1 and A_2 be two non diagonalizable terms of $\mathcal{A}$. Then, they are of the form

$$A_1 = \begin{pmatrix} a_1 & b_1 \\ & a_1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_2 & b_2 \\ & a_2 \end{pmatrix}$$

for some $a_1, a_2, b_1, b_2 \in R$, $b_1, b_2 \neq 0$. Thus, they commute, so $\mathcal{A}$ is not reduced. $\square$

Proposition 2.20. *Let $\mathcal{A} = (A_1, A_2, A_3, A_4)$ be a reduced quadruple whose terms A_j are all triangularizable. Then $\mathcal{A}$ is triangularizable if and only if $\sigma_{jk} = 0$ for all $j, k \in \{1, 2, 3, 4\}$.*

Proof. One direction is a consequence of Proposition 2.10. Conversely, let $\mathcal{A}$ be reduced with $\sigma_{jk} = 0$. By Theorem 2.14 we may assume that A_1 and A_2 are upper triangular, so that $c_1 = c_2 = 0$. From Lemma 2.19, we can also assume that A_1 is diagonalizable so that $b_1 = 0$ and $e_1 \neq 0$. Then, reducibility implies that $b_2 \neq 0$. In order to obtain a contradiction, suppose that $c_3c_4 \neq 0$. Then we have

$$\begin{aligned}
0 &= \sigma_{13} = e_1^2 b_3 c_3, \\
0 &= \sigma_{14} = e_1^2 b_4 c_4, \\
0 &= \sigma_{23} = -b_2c_3(b_2c_3 + e_2e_3), \\
0 &= \sigma_{24} = -c_4(b_2^2c_4 + b_2e_2e_4 - e_2^2b_4), \\
0 &= \sigma_{34} = -b_4(c_3^2b_4 + c_3e_3e_4 - e_3^2c_4).
\end{aligned}$$

This implies $b_3 = 0$, $b_4 = 0$, $b_2c_3 + e_2e_3 = 0$ and $b_2c_4 + e_2e_4 = 0$. The last two equations imply that (e_2, b_2) is a nontrivial solution of the matrix equation

$$\begin{pmatrix} e_3 & -c_3 \\ e_4 & -c_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This implies $e_3c_4 - c_3e_4 = 0$, which together with $b_3 = b_4 = 0$ contradict the reducibility of $\mathcal{A}$. So, $c_3c_4 = 0$ and either $\Delta_{123} = 0$ or $\Delta_{124} = 0$, by Eq. (2.5). Assuming, without loss of generality, that $\Delta_{123} = 0$, (A_1, A_2, A_3) is triangularizable by Theorem 2.16, and the result follows from Proposition 2.12. $\square$

Now, we are ready to prove Theorems 1.12 and 1.13.

Proof of Theorem 1.12. Suppose $\mathcal{A}$ is reduced of length $l \geqslant 4$ and assume all subsequences of $\mathcal{A}$ of length $\leqslant 2$ are triangularizable. Then, by Proposition 2.20 all quadruples of $\mathcal{A}$ are triangularizable, so that $\mathcal{A}$ is triangularizable by Proposition 1.9. $\square$

Proof of Theorem 1.13. Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a matrix sequence with maximal reduction $\mathcal{A}' = (A_1, \ldots, A_l)$ of length l. Then $\mathcal{A}$ is triangularizable if and only $\mathcal{A}'$ is, by Corollary 2.7, showing (i) for $l \leqslant 3$. If $l = 4$, (A_1, A_2, A_3, A_4) is triangularizable by Proposition 2.20. Then, we are in the hypothesis of Theorem 2.17, which shows that we can apply induction to conclude that $\mathcal{A}'$ is triangularizable for all $l \geqslant 4$. $\square$

Remark 2.21. To summarize and relate our results with McCoy's Theorem, we have shown that over an algebraically closed field and under the simple condition that $\mathcal{A}$ has reduced length $\neq 3$, $\mathcal{A}$ is triangularizable if and only if the commutators $[A_j, A_k]$ are nilpotent, for $j = 1, 2, 3$ and $k > j$. In particular, the monomials $p(x)$ in Theorem 1.3 are unnecessary (for reduced length $\neq 3$).

3. Simultaneous similarity

3.1. Similarity and subtriples

Again, let $G = GL_2(R)$. We now prove Proposition 1.10 with the help of the following easy lemma.

Lemma 3.1. *Let $[A_1, A_2] \neq 0$ and $g \in G$. If $g \cdot (A_1, A_2) = (A_1, A_2)$ then g is a scalar.*

Proof. Let g have columns (x, y) and (z, w) with $xw - yz$ invertible. The conditions $A_jg = gA_j$, for $j = 1, 2$ can be written as $M_ju = 0$, where

$$M_j = \begin{pmatrix} 0 & c_j & -e_j \\ -c_j & 0 & b_j \\ e_j & -b_j & 0 \end{pmatrix}, \quad u = (z, x - w, y), \; j = 1, 2.$$

So, the vector $(z, x - w, y) \in R^3$ lies in the intersection of the nullspace of the M_j, each being generated by the non-zero vector (b_j, e_j, c_j), $j = 1, 2$, (each M_j has rank exactly 2, as A_1 and A_2 are non-scalars). So, $(z, x - w, y)$ is zero unless (b_1, e_1, c_1) and (b_2, e_2, c_2) are colinear. But this is not the case by Eq. (2.1) since $[A_1, A_2] \neq 0$. Thus, $z = y = 0$ and $x = w$, so that g is scalar as wanted. $\square$

Proof of Proposition 1.10. If $\mathcal{A}$ and $\mathcal{B}$ are similar, then $\mathcal{A}_J \sim \mathcal{B}_J$ for any index set $J = (j_1, \ldots, j_l)$ with $l \leqslant n$ and $1 \leqslant j_1 < \cdots < j_l \leqslant n$. Conversely, let $\mathcal{A}_J \sim \mathcal{B}_J$ for index sets J of cardinality $\leqslant 3$. We divide the proof into three cases. (i) Suppose $\mathcal{A}$ is scalar. Since scalar matrices are invariant under conjugation, $A_j \sim B_j$ implies that $A_j = B_j$, so $\mathcal{A} = \mathcal{B}$. (ii) Suppose now that $\mathcal{A}$ is non-scalar, but commutative. Then some term of $\mathcal{A}$, say A_1 is non-scalar. Since $A_1 \sim B_1$, B_1 is also non-scalar, and after a conjugation, we can assume that $B_1 = A_1$. Moreover, $(A_j, A_k) \sim (B_j, B_k)$ and $[A_j, A_k] = 0$ implies that $[B_j, B_k] = 0$. As a consequence, $\mathcal{B}$ is also commutative and non-scalar. Since A_1 is a non-scalar 2×2 matrix, it is a nonderogatory matrix, so that every matrix commuting with A_1 is a polynomial in A_1 with coefficients in $\mathbb{F}$, the field of fractions of R. Therefore $\mathcal{A} = (A_1, p_2(A_1), \ldots, p_n(A_1))$ for some polynomials $p_j(x)$, $j = 2, \ldots, n$. Since pairs are similar, let $g_j \in G$, $j = 2, \ldots, n$, be such that $(B_1, B_j) = (A_1, B_j) = (g_j A_1 g_j^{-1}, g_j A_j g_j^{-1})$. Then $B_j = g_j A_j g_j^{-1} = g_j\, p_j(A_1) g_j^{-1} = p_j(g_j A_1 g_j^{-1}) = p_j(A_1)$, for all $j = 2, \ldots, n$. So $\mathcal{B} = (B_1, \ldots, B_n) = (A_1, p_2(A_1), \ldots, p_n(A_1)) = \mathcal{A}$. (iii) Finally, let $\mathcal{A}$ be non-commutative. Then, some pair does not commute, say $[A_1, A_2] \neq 0$. Since all pairs are similar we may, after a suitable conjugation, assume that $(A_1, A_2) = (B_1, B_2)$. Since all triples are similar, let $g_j \in G$, $j = 3, \ldots, n$, be such that $g_j \cdot (A_1, A_2, A_j) = (A_1, A_2, B_j)$. Then, since $g_j \cdot (A_1, A_2) = (A_1, A_2)$ Lemma 3.1 implies that g_j is scalar for all $j = 3, \ldots, n$, so $B_j = g_j \cdot A_j = A_j$ and $\mathcal{B} = \mathcal{A}$. $\square$

3.2. Similarity for semisimple sequences

We now work over a field $\mathbb{F}$. Recall the following definition.

Definition 3.2. A matrix sequence $\mathcal{A} \in V_n$ is called *semisimple* if its G-orbit (for the conjugation action (1.1)) is Zariski closed in V_n.

This notion of semisimplicity generalizes that of a single matrix. In the context of geometric invariant theory, semisimplicity can be translated into more algebraic terms as follows. Recall that, in the general situation of a general affine (algebraic) reductive group K acting on an affine variety V (both over $\mathbb{F}$) one defines the affine quotient variety $V /\!\!/ K$ as the maximal spectrum of the ring of invariant functions on V, $\mathrm{Spm}(\mathbb{F}[V]^K)$, which comes equipped with a projection

$$q : V \to V /\!\!/ K$$

induced from the canonical inclusion of algebras $\mathbb{F}[V]^K \subset \mathbb{F}[V]$ (see, for example, [12] or [11]). The set of closed orbits is in bijective correspondence with geometric points of the quotient $V /\!\!/ K$. Recall also that a vector $x \in V$ is said to be *stable* if the corresponding 'orbit map'

$$\psi_x : K \to V, \quad g \mapsto g \cdot x$$

is proper. It is easy to see that $x \in V$ is stable if and only if the K-orbit of x is closed and its stabilizer K_x is finite. Another useful criterion for stability is the *Hilbert–Mumford numerical criterion*, which is stated in terms of nontrivial homomorphisms $\phi : \mathbb{F}^* \to K$, called one parameter subgroups (1PS) of K ($\mathbb{F}^* := \mathbb{F} \setminus \{0\}$). To any such ϕ and to a point $x \in V$ one associates the composition $\phi_x := \psi_x \circ \phi : \mathbb{F}^* \to V$. If ϕ_x can be extended to a morphism $\overline{\phi_x} : \mathbb{F} \to V$, we say that $\lim_{\lambda \to 0} \phi_x(\lambda)$ exists and equals $\overline{\phi_x}(0)$.

Theorem 3.3 (Hilbert–Mumford, see [12]). *A point $x \in V$ is not stable if and only if there is a one parameter subgroup ϕ of K, such that $\overline{\phi_x}(0)$ exists.*

Let us return to our example of the conjugation action (1.1) of $G = GL_2(\mathbb{F})$ on $V_n = V_n(\mathbb{F})$, and let $\mathcal{S} \subset V_n = V_n(\mathbb{F})$ denote the subset of semisimple sequences. Then $\mathcal{S}/G$, being the set of closed orbits, is in bijection with $V_n /\!\!/ G = \mathrm{Spm}(\mathbb{F}[V_n]^G)$. As described in Section 1, Drensky's theorem (Theorem 1.8) realized the algebra of invariants as a quotient $\mathbb{F}[V_n]^G = \mathbb{F}[\mathbf{x}]/I$ where I is the ideal of relations (see Eq. (1.4)) and

$$\mathbf{x} = (t_1, \ldots, t_n, t_{11}, \ldots, t_{nn}, u_{12}, \ldots, u_{n-1,n}, s_{123}, \ldots, s_{n-2,n-1,n}) \in \mathbb{F}^N$$

is the list of generators $\left(N = 2n + \binom{n}{2} + \binom{n}{3}\right)$. Dualizing the sequence

$$\mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{x}]/I = \mathbb{F}[V_n]^G \subset \mathbb{F}[V_n],$$

we obtain:

$$V_n \to V_n /\!\!/ G = \mathcal{S}/G \subset \mathbb{F}^N.$$

By standard arguments, the last inclusion is precisely the map $\Phi([\mathcal{A}]) = \mathbf{x}$, that sends a G orbit to its values on the generators, so we conclude the following.

Proposition 3.4. *For a field of characteristic zero, the map $\Phi : \mathcal{S}/G \to \mathbb{F}^N$ is injective.*

To obtain an analogous map for non-semisimple sequences, we first characterize these ones as triangularizable but not diagonalizable sequences.

Note that, for the action of G on V_n there are no stable points, since the scalar nonzero matrices stabilize any sequence $\mathcal{A} \in V_n$. This is not a problem, since the same orbit space is obtained with the conjugation action of the affine reductive group $G_1 = SL_2(\mathbb{F})$ (determinant one matrices in $GL_2(\mathbb{F})$) on V_n which has generically $\mathbb{Z}_2$ stabilizers:

$$V_n /\!\!/ G = V_n /\!\!/ G_1.$$

Note that any diagonal matrix sequence $\mathcal{A} \in D := \{\mathcal{A} \in V_n : b = c = 0\}$ has the subgroup

$$H = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \subset G_1, \quad \lambda \in \mathbb{F}^* \tag{3.1}$$

contained in its stabilizer.

Proposition 3.5. *A 2×2 matrix sequence is stable (for the $SL_2(\mathbb{F})$ conjugation action) if and only if it is not triangularizable. A 2×2 matrix sequence is semisimple if and only if it is either stable or diagonalizable.*

Proof. This follows from general results (see [1]). For completeness, we include a proof of this particular case, using the Hilbert–Mumford criterion (Theorem 3.3). For the first statement, we can assume that $\mathcal{A}$ is upper triangular. Then, a simple computation shows that the closure of the orbit of $\mathcal{A}$ under the subgroup $H \subset G_1 = SL_2(\mathbb{F})$ (Eq. 3.1) intersects D. So, either $\mathcal{A}$ is in D (and it is commutative and semisimple) and its stabilizer contains D, or $\mathcal{A}$ is not commutative, $\mathcal{A} \notin D$ so $G \cdot \mathcal{A} = G_1 \cdot \mathcal{A}$ is not closed. In either case, $\mathcal{A}$ is not stable. Conversely, let $\mathcal{A}$ be not stable for the action of G_1. By elementary representation theory, any one parameter subgroup of G_1 is conjugated to

$$\lambda \mapsto \phi_n(\lambda) = \begin{pmatrix} \lambda^n & 0 \\ 0 & \lambda^{-n} \end{pmatrix}, \quad n \in \{1, 2, \ldots\}.$$

In other words, any 1PS can be written as $\phi = g^{-1}\phi_n g$, for some $g \in G_1$ and some ϕ_n so,

$$\lim_{\lambda \to 0} \phi_{\mathcal{A}}(\lambda) = \lim_{\lambda \to 0} \phi(\lambda) \cdot \mathcal{A} = g^{-1} \lim_{\lambda \to 0} \phi_{\mathrm{n}}(\lambda) \cdot (g \cdot \mathcal{A}). \tag{3.2}$$

Writing $g \cdot \mathcal{A}$ as

$$g \cdot \mathcal{A} = \begin{pmatrix} a(g) & b(g) \\ c(g) & d(g) \end{pmatrix},$$

we obtain

$$\phi_n(\lambda) \cdot (g \cdot \mathcal{A}) = \begin{pmatrix} a(g) & b(g)\lambda^{2n} \\ c(g)\lambda^{-2n} & d(g) \end{pmatrix}.$$

By the Hilbert–Mumford criterion, the limit (3.2) exists for some 1PS, so we must have $c(g) = 0$, for some $g \in G_1$. This means that $g \cdot \mathcal{A}$ is upper triangular, hence not irreducible. The second statement is analogous. □

3.3. Similarity for non-commutative triangularizable sequences

To obtain a numerical similarity criterion for non-semisimple sequences, we are reduced, by Proposition 3.5, to the case of triangularizable (non-diagonalizable) sequences. Here, we consider the non-commutative triangularizable case.

Let $\mathsf{U} \subset V_n$ be the affine variety of upper triangular matrix sequences and let $\mathsf{K} \subset \mathsf{U}$ be the subset of commutative sequences. For $n \geqslant 2$ and $\mathcal{A} \in \mathsf{U}$, let $P_{\mathcal{A}}$ denote the following $2 \times n$ matrix, and $\delta_{jk} = \delta_{jk}(\mathcal{A})$ the corresponding 2×2 minors

$$P_{\mathcal{A}} = \begin{pmatrix} e_1 & \cdots & e_n \\ b_1 & \cdots & b_n \end{pmatrix}, \quad \delta_{jk} = b_j e_k - e_j b_k,\ j,k \in \{1,\ldots,n\}.$$

According to Lemma 2.1, K is characterized as consisting of sequences $\mathcal{A}$ such that the rank of $P_{\mathcal{A}}$ is $\leqslant 1$ (it is 0 when all terms in $\mathcal{A}$ are scalars). Hence, when $\mathcal{A} \in \mathsf{U} \setminus \mathsf{K}$, $P_{\mathcal{A}}$ has rank 2 and defines an element $\pi_{\mathcal{A}}$ of the Grassmanian $\mathbb{G}(2,n)$ of 2-planes in $\mathbb{F}^n$. It is easy to see that the action of $G = GL_2(\mathbb{F})$, restricted to $\mathsf{U} \setminus \mathsf{K}$, preserves the plane $\pi_{\mathcal{A}}$:

Lemma 3.6. *Let $g \in G$. If $\mathcal{A}$ and $\mathcal{A}' = g \cdot \mathcal{A}$ are both in $\mathsf{U} \setminus \mathsf{K}$, then $e = e'$ and $\pi_{\mathcal{A}} = \pi_{\mathcal{A}'}$.*

Proof. We can assume that $\mathcal{A}' = g \cdot \mathcal{A}$ for some $g \in SL_2(\mathbb{F})$ and compute, for $\mathcal{A} \in \mathsf{U} \setminus \mathsf{K}$,

$$g \cdot \mathcal{A} = \begin{pmatrix} x & z \\ y & w \end{pmatrix}\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}\begin{pmatrix} w & -z \\ -y & x \end{pmatrix} = \begin{pmatrix} * & * \\ y(ew - by) & * \end{pmatrix}.$$

Since $P_{\mathcal{A}}$ has rank 2, $ew \neq by$ as vectors in $\mathbb{F}^n$. Thus, in order that $g \cdot \mathcal{A}$ be again in $\mathsf{U} \setminus \mathsf{K}$, we need $y = 0$, which simplifies the formula above to

$$g \cdot \mathcal{A} = \begin{pmatrix} x & z \\ 0 & w \end{pmatrix}\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}\begin{pmatrix} w & -z \\ 0 & x \end{pmatrix} = \begin{pmatrix} a & x(bx - ez) \\ 0 & d \end{pmatrix} \tag{3.3}$$

because $xw = \det g = 1$. Then, $e = e' = a - d$. Moreover, the upper right entry is $b' = x(bx - ez) = x^2 b - xze$, a linear combination of b and e. So $\pi_{\mathcal{A}} = \pi_{\mathcal{A}'}$, as asserted. □

The correspondence $\mathcal{A} \mapsto \pi_{\mathcal{A}}$ is therefore well defined on the G-orbits of non-commutative triangularizable sequences. To make this more precise, consider the following algebraic subvarieties of V_n. Let $\mathcal{U} = G \cdot \mathsf{U}$ be the variety of triangularizable matrix sequences and let $\mathcal{K} \subset \mathcal{U}$ be the subvariety of commutative sequences. Note that $\mathcal{U}$ is indeed irreducible, as the image in V_n of the (irreducible) algebraic variety $G \times \mathsf{U}$ under the morphism $(g, \mathcal{A}) \mapsto g \cdot \mathcal{A}$. The irreducibility of $\mathcal{K}$ was first noted in [7] (see [8] for a proof).

By the previous lemma, given a sequence $\mathcal{A} \in \mathcal{U} \setminus \mathcal{K}$, we can define $\pi_{\mathcal{A}} := \pi_{\mathcal{B}}$ using any matrix sequence $\mathcal{B} \in G \cdot \mathcal{A} \cap \mathsf{U}$, that is, $\mathcal{B}$ is upper triangular and similar to $\mathcal{A}$ (even though the matrix $P_{\mathcal{A}}$ is not defined). So, we can define a map $\psi : \mathcal{U}\setminus\mathcal{K} \to \mathbb{P}^{N-1}$, where $\mathbb{P}^{N-1}$ denotes the projective space over $\mathbb{F}$ of dimension $N - 1 = \binom{n}{2} - 1$, as the composition

$$\begin{array}{ccc} \mathcal{U}\setminus\mathcal{K} & & \\ \pi \downarrow & \searrow \psi & \\ \mathbb{G}(2,n) & \hookrightarrow & \mathbb{P}^{N-1}, \end{array}$$

where the bottom inclusion is the Plücker embedding of the Grassmanian $\mathbb{G}(2,n)$. In more concrete terms we have.

Lemma 3.7. *Let $\mathcal{A} \in \mathcal{U}\backslash\mathcal{K}$. Then $\psi(\mathcal{A}) = [\delta_{12} : \delta_{13} : \cdots : \delta_{n-1,n}]$, where $\delta_{jk} = \delta_{jk}(\mathcal{B})$, for any $\mathcal{B} \in G \cdot \mathcal{A} \cap \mathsf{U}$.*

Proof. The formula follows from the definition of the Plücker embedding. We just need to show that the map is well defined. The point $[\delta_{12} : \delta_{13} : \cdots : \delta_{n-1,n}]$ is in projective space since at least a pair of terms, say A_j and A_k do not commute, so that $\delta_{jk}(\mathcal{B}) = b_j e_k - e_j b_k$ is nonzero, by Lemma 2.1. On the other hand for a different sequence $\mathcal{B}' \in G \cdot \mathcal{A} \cap \mathsf{U}$, Eq. (3.3) and $e = e'$ (Lemma 3.6) imply

$$\delta_{jk}(\mathcal{B}') = b_j' e_k' - e_j' b_k' = x(b_j x - e_j z)e_k - e_j x(b_k x - e_k z) = x^2 \delta_{jk}(\mathcal{B}),$$

so the point $\psi(\mathcal{A}) \in \mathbb{P}^{N-1}$ is indeed independent of the choice of $\mathcal{B} \in G \cdot \mathcal{A} \cap \mathsf{U}$. $\square$

By Lemma 3.7, the quotients δ_{jk}/δ_{lm} (for $j \neq k$ and $l \neq m$) are well defined rational G-invariant functions on $\mathcal{U} \setminus \mathcal{K}$ (defined on the open dense complement of $\delta_{lm}^{-1}(0) \subset \mathcal{U} \setminus \mathcal{K}$), so they descend to the quotient space $(\mathcal{U} \setminus \mathcal{K})/G$. Note, however, that these are not quotients of regular (polynomial) invariants.

Define now the map $\Psi : (\mathcal{U} \setminus \mathcal{K})/G \to \mathbb{F}^n \times \mathbb{F}^{N+n} \times \mathbb{P}^{N-1}$, $N = \binom{n}{2}$, by

$$\Psi([\mathcal{A}]) = \left(\{t_j\}_{j=1,\dots,n},\ \{t_{jk}\}_{j,k=1,\dots,n},\ \psi\right),$$

where $[\mathcal{A}]$ denotes the conjugacy class of $\mathcal{A}$, $t_j = \operatorname{tr}(A_j)$, $t_{jk} = \operatorname{tr}(A_j A_k)$, and ψ was given by Lemma 3.7. The main result of this subsection is the following. For the proof, we use a standard consequence of the Noether–Deuring theorem; namely, if $\mathcal{A}$ and $\mathcal{A}'$ are elements in $V_{m,n}(\mathbb{F})$ and $g \in GL_m(\overline{\mathbb{F}})$ verifies $g \cdot \mathcal{A} = \mathcal{A}'$ (i.e, similarity over $\overline{\mathbb{F}}$) then they are similar over $\mathbb{F}$ (see e.g. [2, p. 200]).

Theorem 3.8. *Let $n \geqslant 2$. The map Ψ is two-to-one. More precisely, the rational invariants δ_{jk}/δ_{lm}, together with the regular invariants t_j and t_{jk}, distinguish G-orbits in $\mathcal{U} \setminus \mathcal{K}$, except for the identification $e \leftrightarrow -e$, when written in triangular form.*

Proof. Let both sequences $\mathcal{A}$ and $\mathcal{A}'$ be triangularizable and non-commutative, so there exist $\mathcal{B} \in G \cdot \mathcal{A} \cap (\mathsf{U} \setminus \mathsf{K})$ and $\mathcal{B}' \in G \cdot \mathcal{A}' \cap (\mathsf{U} \setminus \mathsf{K})$. Let us write $t_j = \operatorname{tr}(B_j) = \operatorname{tr}(A_j)$ etc., as before, and use primed letters to denote corresponding quantities for $\mathcal{A}'$ or $\mathcal{B}'$. To prove injectivity, suppose $\Psi([\mathcal{A}]) = \Psi([\mathcal{A}'])$ so that $t_j = t_j'$ and $t_{jk} = t_{jk}'$ for all appropriate indices. Then,

$$e_j e_k = 2t_{jk} - t_j t_k = 2t_{jk}' - t_j' t_k' = e_j' e_k'.$$

Since any triple $(x^2, xy, y^2) \in \mathbb{F}^3$ determines (x, y) up to sign, the equation above implies that $e' = \pm e$, as vectors in $\mathbb{F}^n$. Suppose $e' = e$. The hypothesis $\Psi([\mathcal{A}]) = \Psi([\mathcal{A}'])$ means also that $\psi(\mathcal{B}) = \psi(\mathcal{B}')$. Thus $\pi_{\mathcal{B}} = \pi_{\mathcal{B}'}$ as planes in $\mathbb{G}(2,n)$ due to the injectivity of the Plücker map. Therefore $b' = \alpha b + \beta e$ for some complex numbers $\alpha, \beta, \alpha \neq 0$. So, using the invertible

$$g = \begin{pmatrix} \sqrt{\alpha} & -\beta/\sqrt{\alpha} \\ & \sqrt{\alpha}^{-1} \end{pmatrix},$$

where $\sqrt{\alpha}$ is any square root of α in $\overline{\mathbb{F}}$, it is easy to see that $g \cdot \mathcal{B} = \mathcal{B}'$, so that $\mathcal{A}$ and $\mathcal{A}'$ are similar over $\overline{\mathbb{F}}$. So, $\mathcal{A}$ and $\mathcal{A}'$ are also similar over $\mathbb{F}$. Finally, with $e' = -e$ (note $e \neq 0$ by hypothesis), one can see that $\mathcal{B}'$ is not similar to $\mathcal{B}$, hence the result. $\square$

In view of Proposition 3.5, Theorem 3.8 and Proposition 3.4 give a solution to the invariants problem (ii) for non-commutative sequences. Theorem 1.14 provides a more efficient solution (and works for any characteristic $\neq 2$), and its proof will be given after exploring the canonical forms described in the next section.

Remark 3.9. Finding maps analogous to Φ or Ψ in the commutative case is more involved! See Friedland [6] for a discussion of the case of a commuting pair of 2×2 matrices. However, testing similarity of commutative 2×2 matrix sequences is a trivial task after reduction to triangular form, as recalled in Appendix B.

We end this section with the following conjecture on the generalization of Proposition 1.10 to $m \times m$ matrices. Since an irreducible sequence is a generic sequence and the conjugacy classes of triangularizable or block-triangularizable sequences seem to depend on less data (i.e, less regular or rational invariants) than the irreducible case, we propose the following problem. Define the "semisimple similarity number" $S(m)$ as

$$S(m) = \min\{k \in \mathbb{N} : \forall n \in \mathbb{N},\ \forall \mathcal{A}, \mathcal{B} \in \mathcal{S}_{n,m},\ \mathcal{A} \sim \mathcal{B} \Leftrightarrow \mathcal{A}_J \sim \mathcal{B}_J\ \forall J \text{ with } |J| \leqslant k\},$$

where $\mathcal{S}_{n,m} \subset V_{n,m}$ is the subset of semisimple sequences.

Conjecture. For arbitrary $m \times m$ sequences (not necessarily semisimple) $\mathcal{A} \sim \mathcal{B}$ if and only if $\mathcal{A}_J \sim \mathcal{B}_J$ for all index vectors J with length $\leqslant S(m)$.

Note that, by Procesi's Theorem, $S(m) \leqslant N(m)$. This conjecture is true for $m = 2$, by Proposition 1.10, since Proposition 2.12 gives $S(2) = 3$.

4. Canonical forms and reconstruction of sequences

4.1. Canonical forms

We now describe canonical forms for 2×2 matrix sequences over an algebraically closed field $\overline{\mathbb{F}}$. Informally, this means the indication, for each given matrix sequence $\mathcal{A} \in V_n$, of an element in its conjugacy class which has a simple form, and this same simple form should be used for the biggest possible set of sequences, hence the term 'canonical'. We will assume $n \geqslant 2$ since such canonical forms for $n = 1$ are provided by the well known Jordan decomposition. Also, for commutative sequences, the canonical forms are the same as for $n = 1$. This is recalled in Appendix B. For non-commutative sequences, it turns out that canonical forms can be divided into five cases.

By the preceding results, it is no surprise that we need to consider distinct canonical forms for the stable and for the reducible cases. We make the following choices.

Definition 4.1. We say that a stable (i.e, irreducible) matrix sequence $\mathcal{A} = (A_1, A_2, \ldots)$ is in *(stable) canonical form* if A_1 is in Jordan canonical form and $b_2 = 1$. We say that a triangularizable sequence $\mathcal{A}$ is in *(triangular) canonical form* if $\mathcal{A}$ is upper triangular, A_1 is diagonal and $b_2 = 1$.

Recall that, by Proposition 2.19 a triangularizable sequence of length $n \geqslant 2$ has at least $n - 1$ semisimple (or diagonalizable) terms. By contrast, a stable sequence can have all terms which are non-diagonalizable. To see this, just consider the family of matrices of the form

$$\begin{pmatrix} -\alpha\beta & \alpha^2 \\ -\beta^2 & \alpha\beta \end{pmatrix}$$

for some parameters $\alpha, \beta \in \mathbb{C}$ not both zero. Then, all these matrices are similar to the 2×2 Jordan block with zero diagonal (i.e, $\alpha = 1$, $\beta = 0$), but the σ of two of these is zero only when the vectors $(\alpha, \beta) \in \mathbb{C}^2$ are colinear. We have, however, the following fact.

Proposition 4.2. *Let $\mathcal{A}$ have reduced length $n = 2$ or $n \geqslant 4$. Then $\mathcal{A}$ is stable if and only if some $\sigma_{jk} \neq 0$.*

Proof. If $\mathcal{A}$ is not stable, then it is triangularizable, by Proposition 3.5, so that all $\sigma = 0$, by Proposition 2.10. Conversely, suppose $\mathcal{A}$ is stable with reduced length $n = 2$. Then $\sigma_{12} \neq 0$ because of Friedland's result (or Theorem 1.11). Finally, if $n \geqslant 4$, Proposition 2.20 implies that at least one $\sigma_{jk} \neq 0$. On the other hand, Example 2.11 shows that the statement is not true for $n = 3$. $\square$

Theorem 4.3. *All non-commutative matrix sequences can be put in canonical form. More precisely, after rearranging terms, any sequence is similar to a sequence with A_1, A_2 and A_3 as described by the following table.*

Stable case	*1.a*	$\sigma_{12} \neq 0, \delta_1 \neq 0$	A_1 *diagonal*, $b_2 = 1$
	1.b	$\sigma_{12} \neq 0, \delta_1 = \delta_2 = 0$	A_1 *Jordan block*, $b_2 = 1$
	1.c	$\sigma_{12} = \sigma_{13} = \sigma_{23} = 0, \Delta_{123} \neq 0$	A_1 *diagonal*, $b_2 = 1$
Triangularizable case	*2.a*	*all diagonalizable*	A_1 *diagonal*, $b_2 = 1$
	2.b	*1 non-diagonalizable* (A_2)	A_1 *diag.*, A_2 *Jordan bl.*

Proof. Let us show that only the five possibilities above occur, and may be put in the given forms, starting when $\mathcal{A}$ is stable. If $n \geqslant 4$ or $n = 2$, Proposition 4.2 implies that there is some $\sigma \neq 0$, so we can rearrange the terms of $\mathcal{A}$ so that we have $\sigma_{12} \neq 0$ (in particular, $[A_1, A_2] \neq 0$). If A_1 or A_2 is diagonalizable, we have possibility 1.a. Assuming A_1 diagonalizable, we may suppose that A_1 is already in diagonal form. Since A_1 is nonscalar $\delta_1 \neq 0$, and the stabilizer of A_1 are the diagonal invertible matrices $H \subset G$. Then $\sigma_{12} = e_1^2 b_2 c_2 \neq 0$ implies that both b_2 and c_2 are nonzero. Let $g = \mathrm{diag}(x, x^{-1})$ for some $x \in \overline{\mathbb{F}}^*$. A simple computation shows that $g^{-1}A_1 g$ is diagonal and $A_2' = g^{-1}A_2 g$ has $b_2' = x^{-2}b_2$. So, we can solve for x in order to have $b_2' = 1$, as claimed. If A_1 and A_2 are both not diagonalizable, we have case 1.b. In this case, we may suppose A_1 is already in Jordan form. Conjugating A_2 with a matrix of the form

$$g = \begin{pmatrix} 1 & z \\ & 1 \end{pmatrix},$$

a simple computation shows that we can solve for z in order to obtain $b_2 = 1$. If $n = 3$ and some $\sigma_{12} \neq 0$, we return to cases 1.a or 1.b. So it remains the case $n = 3$ and all $\sigma = 0$, which is 1.c. From Example 2.11 we see that necessarily $\Delta_{123} \neq 0$ (after an eventual rearrangement of terms), and a diagonal conjugation will achieve $b_2 = 1$. Finally, when $\mathcal{A}$ is triangularizable, so by Proposition 2.19 either none or one of the terms of $\mathcal{A}$ are non-semisimple, cases 2.a and 2.b respectively. The forms mentioned will be obtained by conjugation with a diagonal invertible matrix. $\square$

In the above table we have $[A_1, A_2] \neq 0$ in all cases, so if g stabilizes (A_1, A_2), then g is scalar, by Lemma 3.1. Therefore, we have a uniqueness statement for canonical forms.

Proposition 4.4. *Let $\mathcal{A}, \mathcal{B}$ be two sequences in canonical form with $[A_1, A_2] \neq 0$. Assume that $(B_1, B_2) = (A_1, A_2)$. Then, $\mathcal{A} \sim \mathcal{B}$ if and only if $\mathcal{A} = \mathcal{B}$.*

4.2. Reconstruction of sequences from invariants

We continue to work over an algebraically closed field $\overline{\mathbb{F}}$, and here we restrict to characteristic $\neq 2$. Let $\mathcal{S}'(\overline{\mathbb{F}})$ denote the subset of $V_n(\overline{\mathbb{F}})$ of semisimple sequences such that A_1 is diagonalizable and $[A_1, A_2] \neq 0$. Then $\mathcal{S}'(\mathbb{F})$ consists of irreducible (stable) sequences that can be put in the form 1.a.

Let $\overline{\Phi} : \mathcal{S}'(\overline{\mathbb{F}})/G(\overline{\mathbb{F}}) \to \overline{\mathbb{F}}^{4n-3}$ denote the map

$$\overline{\Phi}([\mathcal{A}]) := (t_1, t_{11}, t_2, t_{22}, t_{12}, \ldots, t_k, t_{1k}, t_{2k}, s_{12k} \ldots, t_n, t_{1n}, t_{2n}, s_{12n}) . \tag{4.1}$$

We now describe a process of (re)constructing a sequence from its values under the map $\overline{\Phi}$. Let $v \in \overline{\mathbb{F}}^{4n-3}$ be given. We want to find $\mathcal{A} \in \mathcal{S}'(\overline{\mathbb{F}})$ such that $\overline{\Phi}([\mathcal{A}]) = v$.

Let a_1 and d_1 (in $\overline{\mathbb{F}}$) be the roots of the polynomial $\lambda^2 - t_1\lambda + \frac{t_1^2 - t_{11}}{2} = 0$ (the characteristic polynomial of a diagonal matrix whose trace is t_1 and the trace of its square is t_{11}). If they are equal, there is no solution to our problem simply because an $\mathcal{A}$ satisfying $\overline{\Phi}([\mathcal{A}]) = v$ will have either A_1 non-diagonalizable or $[A_1, A_2] = 0$. So, with $e_1 = a_1 - b_1 \neq 0$, put $b_1 = c_1 = 0, b_2 = 1$ and

$$c_2 = \frac{t_{22} - a_2^2 - d_2^2}{2} = \frac{1}{4e_1^2}\begin{vmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{vmatrix} \neq 0,$$

$$\begin{pmatrix} a_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ a_1 & d_1 \end{pmatrix}^{-1} \begin{pmatrix} t_2 \\ t_{12} \end{pmatrix}. \tag{4.2}$$

Then, one easily checks that the pair $(A_1, A_2) \in V_2(\overline{\mathbb{F}})$ whose entries are $a, b, c, d \in \overline{\mathbb{F}}^2$ is in canonical form 1.a and satisfies $\overline{\Phi}([\mathcal{A}]) = (t_1, t_{11}, t_2, t_{22}, t_{12})$. Moreover, this pair is unique except for the choice of assigning to a_1 or d_1 one or the other root of the characteristic polynomial. The two possible choices are:

$$\left(\begin{pmatrix} a_1 & \\ & d_1 \end{pmatrix}, \begin{pmatrix} a_2 & 1 \\ c_2 & d_2 \end{pmatrix}\right), \quad \left(\begin{pmatrix} d_1 & \\ & a_1 \end{pmatrix}, \begin{pmatrix} d_2 & 1 \\ c_2 & a_2 \end{pmatrix}\right),$$

These pairs are similar (for e_1 and c_2 both non-zero), with similarity matrix

$$g = \begin{pmatrix} & 1/x \\ -x & \end{pmatrix}, \qquad x = \sqrt{-c_2}. \tag{4.3}$$

Moreover, up to a non-zero scalar multiple, this is the unique matrix sending one pair to the other. Let $\mathcal{A}^\vee := g \cdot \mathcal{A}$ denote the sequence obtained by acting with this matrix. Note that $[\mathcal{A}] = [\mathcal{A}^\vee]$ and if $\mathcal{A}$ is in canonical form, then so is $\mathcal{A}^\vee$. Now, let

$$\begin{pmatrix} a_k \\ b_k \\ c_k \\ d_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ a_1 & 0 & 0 & d_1 \\ a_2 & c_2 & 1 & d_2 \\ 0 & -c_2e_1 & e_1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} t_k \\ t_{1k} \\ t_{2k} \\ s_{12k} \end{pmatrix}, \quad k = 3, \ldots, n. \tag{4.4}$$

The determinant of this matrix is $-2e_1^2c_2$, so our hypothesis imply that it is indeed invertible. Hence the transformation above provides an isomorphism of vector spaces, from the variables (a_k, b_k, c_k, d_k) to the variables $(t_k, t_{1k}, t_{2k}, s_{12k})$, $k = 3, \ldots, n$.

Now, consider the reconstruction of triangularizable sequences. Let $\mathcal{U}'(\overline{\mathbb{F}})$ denote the subset of $V_n(\overline{\mathbb{F}})$ of triangularizable sequences such that A_1 is diagonalizable and $[A_1, A_2] \neq 0$ (so that $\mathcal{U}'(\overline{\mathbb{F}}) \subset \mathcal{U} \setminus \mathcal{K}$). Define the map $\overline{\Psi} : \mathcal{U}'(\overline{\mathbb{F}})/G \to \overline{\mathbb{F}}^{2n} \times \mathbb{P}^{n-2}$, where $\mathbb{P}^k$ denotes the projective space over $\overline{\mathbb{F}}$ of dimension k, by the formula

$$\overline{\Psi}([\mathcal{A}]) = \left(t_1, t_{11}, \ldots, t_k, t_{1k}, \ldots, t_n, t_{1n};\ \psi'\right),$$

where $\psi'([\mathcal{A}]) = [1 : \delta_{13} : \cdots : \delta_{1n}] \in \mathbb{P}^{n-2}$, $\delta_{jk} := b_je_k - e_jb_k$ assuming $\mathcal{A}$ is in upper triangular form. Let $w \in \overline{\mathbb{F}}^{2n} \times \mathbb{P}^{n-2}$ be given. Again, let a_1 and d_1 be the (distinct) roots of the characteristic polynomial $\lambda^2 - t_1\lambda + \frac{t_1^2 - t_{11}}{2} = 0$. Then, with $e_1 = a_1 - d_1 \neq 0$, put $b_1 = 0, b_2 = 1, c_j = 0, j = 1, \ldots, n$, and

$$\begin{pmatrix} a_k \\ d_k \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ a_1 & d_1 \end{pmatrix}^{-1} \begin{pmatrix} t_k \\ t_{1k} \end{pmatrix},$$

$$b_k = -\frac{\delta_{1k}}{e_1}, \quad k = 3, \ldots, n.$$

Then the matrix sequence $\mathcal{A}$ with the entries $a, b, c, d \in \overline{\mathbb{F}}$ is in triangular canonical form (2.a or 2.b) and satisfies $\overline{\Psi}([\mathcal{A}]) = w$, with $\psi'(\mathcal{A}) = [1 : \delta_{13} : \cdots : \delta_{1n}] \in \mathbb{P}^{n-2}$. The other solution $\mathcal{A}'$ is obtained by changing e to $-e$ (that is, exchanging a with d). We have.

Proposition 4.5. *The map* $\overline{\Phi} : \mathcal{S}'(\overline{\mathbb{F}})/G(\overline{\mathbb{F}}) \to \overline{\mathbb{F}}^{4n-3}$ *is injective, and* $\overline{\Psi} : \mathcal{U}'(\overline{\mathbb{F}})/G(\overline{\mathbb{F}}) \to \overline{\mathbb{F}}^{2n} \times \mathbb{P}^{n-2}$ *is two-to-one.*

Proof. Let $\mathcal{A}_c = (A_1, A_2, \ldots)$ and $\mathcal{B}_c = (B_1, B_2, \ldots)$ be canonical forms associated to $\mathcal{A}, \mathcal{B} \in \mathcal{S}'(\overline{\mathbb{F}})$ respectively. If $\overline{\Phi}([\mathcal{A}_c]) = \overline{\Phi}([\mathcal{B}_c])$ then either $(A_1, A_2) = (B_1, B_2)$ or $(A_1, A_2) = (B_1, B_2)^\vee$. If the first situation holds, then $\mathcal{A}_c = \mathcal{B}_c$ by the isomorphism given by Eq. (4.4). Hence, Proposition 4.4 shows that $\mathcal{A} \sim \mathcal{B}$. In the second alternative, just replace $\mathcal{B}_c$ with $\mathcal{B}_c^\vee$ (using g in Eq. (4.3)) and the conclusion is the same, showing that $\overline{\Phi}$ is injective. The case of $\overline{\Psi}$ is analogous, although in this case the two possibilities for a pair in triangular canonical form with given values of $(t_1, t_{11}, t_2, t_{22})$ are:

$$\left(\begin{pmatrix} a_1 & \\ & d_1 \end{pmatrix}, \begin{pmatrix} a_2 & 1 \\ & d_2 \end{pmatrix}\right), \quad \left(\begin{pmatrix} d_1 & \\ & a_1 \end{pmatrix}, \begin{pmatrix} d_2 & 1 \\ & a_2 \end{pmatrix}\right).$$

These pairs are not similar (note $e_1 \neq 0$), by Lemma 3.6, resulting in $\overline{\Psi}$ being 2-1. □

Now, the proof of Theorem 1.14 is easy.

Proof of Theorem 1.14. Let $\mathbb{F}$ be a field of characteristic $\neq 2$ and $\overline{\mathbb{F}}$ its algebraic closure. We have shown that $\overline{\Phi} : \mathcal{S}'(\overline{\mathbb{F}})/G(\overline{\mathbb{F}}) \to \overline{\mathbb{F}}^{4n-3}$ is injective. Then, the map $\Phi' : \mathcal{S}'(\mathbb{F})/G(\mathbb{F}) \to \mathbb{F}^{4n-3}$ defined in Eq. (1.6), having the same form of $\overline{\Phi}$, is just its restriction to $\mathcal{S}'(\mathbb{F})/G(\mathbb{F})$, being therefore injective as well. The inclusion $\mathcal{S}'(\mathbb{F})/G(\mathbb{F}) \subset \mathcal{S}'(\overline{\mathbb{F}})/G(\overline{\mathbb{F}})$ reflects the fact that two sequences in $\mathcal{S}'(\mathbb{F})$, similar over $G(\overline{\mathbb{F}})$, are also similar over $G(\mathbb{F})$, by the Noether–Deuring theorem (see e.g. [2, p. 200]). The case of $\overline{\Psi}$ is analogous. □

Remark 4.6. Finally, we argue that the other cases of semisimple non-commutative sequences in the table of Theorem 4.3 can also be reconstructed by a simple modification of the above procedure. In the case 1.b, we substitute the pair (A_1, A_2) by the pair $(A_1 - A_2, A_1 + A_2)$, and in the case 1.c (assuming, without loss of generality, that A_1 is diagonal non-scalar), we perform the substitution $(A_1, A_2, A_3) \mapsto (A_1, A_2 + A_3, A_2 - A_3)$. In both cases we end up with a matrix in $\mathcal{S}'$, that is, the first matrix is diagonalizable non-scalar and the first pair does not commute.

In conclusion, the only conjugacy classes that the maps Φ' and Ψ' fail to distinguish (not counting the involution $e \leftrightarrow -e$ in the Ψ' case, and after the reduction to $\mathcal{S}'$ described above) are the following types:

$$\begin{pmatrix} a & b \\ & a \end{pmatrix}, \quad \begin{pmatrix} a & \\ & a \end{pmatrix}, \quad a, b \in \mathbb{F}^n,$$

which have the same value under Φ' (note that these are not in the domain of Ψ', as they are commutative), regardless of $b \in \mathbb{F}^n$.

Acknowledgements

I would like to thank J. Dias da Silva whose interesting questions were at the genesis of this article, and R. Guralnick for calling my attention to related work in the literature and for providing simplified proofs of some statements in a previous version. I thank also my colleagues J. Mourão, J.P. Nunes and S. Lawton for many interesting and motivating conversations on this and related topics. Work partially supported by the CAMGSD, IST, Technical Univ. of Lisbon, and “Fundação para a Ciência e a Tecnologia” through the programs Praxis XXI, POCI/MAT/58549/2004 and FEDER. Some computations performed using CoCoA; document typeset using LyX.

Appendix A. Triangularization of singlets over R

In this Appendix, we treat the triangularization problem for a single 2×2 matrix over an integral domain R. We do not claim originality of Proposition A.1 and Lemma A.2 below, although the author was unable to find a suitable reference. Their exposition is mainly intended to provide self-contained proofs of Theorems 1.11 and 2.14.

When is a single matrix with entries in R triangularizable? Let us write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in V_1(R), \quad g = \begin{pmatrix} x & z \\ y & w \end{pmatrix} \in GL_2(R). \tag{A.1}$$

Since g is invertible, so is $\det g = xw - yz$, and conjugation of A by g^{-1} gives:

$$g^{-1} \cdot A = g^{-1} A\, g = \frac{1}{xw - yz} \begin{pmatrix} * & bw^2 + ezw - cz^2 \\ cx^2 - exy - by^2 & * \end{pmatrix}. \tag{A.2}$$

So, triangularizing A amounts to finding a solution $(x, y, z, w) \in R^4$ to one of the equations $cx^2 - exy - by^2 = 0$ or $bw^2 + ezw - cz^2 = 0$, such that $xw - yz$ is invertible in R. Note that both equations are given by the same quadratic form

$$Q(x, y) := cx^2 - exy - by^2 = \frac{1}{2}(x, y)Q_A(x, y)^T$$

associated to matrix

$$Q_A = \begin{pmatrix} 2c & -e \\ -e & -2b \end{pmatrix}.$$

The discriminant of Q is $\delta_Q := -\det Q_A = e^2 + 4bc = \mathrm{tr}^2 A - 4 \det A$, and coincides precisely with the discriminant δ_A of the characteristic polynomial of A.

Eq. (A.2) provides a necessary condition for triangularization: If A is triangularizable over R, then its eigenvalues lie in R. Indeed, if $g^{-1}Ag = T$ for some $g \in GL_2(R)$ and some upper triangular matrix T, $\delta_A = \delta_T = (\lambda_1 - \lambda_2)^2$ is a square in R, where $\lambda_1, \lambda_2 \in R$ are the diagonal elements of T, which are also the eigenvalues of A.

A necessary and sufficient condition is the following.

Proposition A.1. *Let $A \in V_1(R)$ be a 2×2 matrix over an integral domain R. A is triangularizable if and only if it has an eigenvector of the form $(x, y) \in R^2$, such that $xR + yR = R$ (in particular, the ideal (x, y) is principal).*

Proof. The equation $Ag = gT$, for some invertible $g \in G(R)$ and upper triangular T, means that the first column of g is an eigenvector (x, y) of A, so that $x, y \in R$ and there exist $w, z \in R$ (forming the second column of g) so that $xw - yz$ is a unit. In particular, $xR + yR = R$. Conversely, let A be as in (A.1), with an eigenvector $(x, y) \in R^2$ verifying $xR + yR = R$. From simple computations the eigenvalues of A (both in R) are $\lambda_1 = \frac{a+d+r}{2}$ and $\lambda_2 = \frac{a+d-r}{2}$, where r is a square root of $\delta_A = e^2 - 4bc$, and the respective eigenvectors are $v_1 = (\lambda_1 - d, c)$ and $v_2 = (\lambda_2 - d, c)$ (both in R^2). So, without loss of generality, let (x, y) be in the eigenspace of v_1: $(x, y) = \alpha(\lambda_1 - d, c) = \alpha\left(\frac{e+r}{2}, c\right)$, for some nonzero α in the field of fractions of R. By hypothesis, there exist $z, w \in R$ verifying $xw - yz = 1$. Thus, using the invertible matrix g with columns (x, y) and (z, w), we have that $g^{-1}Ag$ is upper triangular (with λ_1 and λ_2 in the main diagonal). Indeed, by Eq. (A.2) its lower left entry is $cx^2 - exy + -by^2 = \frac{\alpha^2}{4}c(r^2 - e^2 - 4bc) = 0$. $\square$

The following fact is used in the proof of Theorem 2.14.

Lemma A.2. *Fix nonzero elements x, y in a field $\mathbb{F}$. If $A \in V_1(\mathbb{F})$ is nondegenerate ($\delta_A \neq 0$) and verifies $cx^2 - exy - by^2 = 0$, then A is diagonalizable and one of its eigenvectors is (x, y). In particular, there is another eigenvector $(z, w) \in \mathbb{F}^2$ with $wx - yz \neq 0$.*

Proof. To satisfy $cx^2 - exy - by^2 = 0$, the triple (b, e, c) must be a linear combination of the vectors $(-x, y, 0)$ and $(0, x, y)$. So, the matrix A is given (uniquely up to the addition of a scalar) by the triple

$(b, e, c) = (-zx, zy + wx, wy)$, for some $z, w \in \mathbb{F}$. Then, a simple computation shows that the discriminant of A is a square in $\mathbb{F}$: $\delta_A = (wx - zy)^2$. So A is triangularizable over $\mathbb{F}$. Moreover, its eigenvalues are easily checked to be $\lambda_1 = wx$ and $\lambda_2 = yz$. Since, by hypothesis $\delta_A \neq 0$, the eigenvalues are distinct, so A is diagonalizable over $\mathbb{F}$. Also, the eigenvectors are multiples of (x, y) and (z, w), respectively. $\square$

Appendix B. Canonical forms and similarity of commuting sequences

For completeness, we include the following well known description of all canonical forms of commuting sequences over an algebraically closed field $\overline{\mathbb{F}}$.

Theorem B.1. *A matrix sequence of length n with coefficients in $\overline{\mathbb{F}}$ is commutative if and only if it is similar to a matrix sequence in one of the forms (called diagonal or triangular forms, respectively (both forms include the scalar sequences)):*

$$\mathcal{A} = \begin{pmatrix} a & \\ & d \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} a & b \\ & a \end{pmatrix}, \; a, d, b \in \overline{\mathbb{F}}^n.$$

Proof. The sufficiency is clear by Lemma 2.1 (condition (2.2) is satisfied because $b = 0$ (resp. $e = 0$) in the first (resp. second) case). Conversely, since $\mathcal{A}$ is commutative it is triangularizable, by Corollary 2.2. Let $k \geqslant 1$ be the smallest integer such that A_k is non-scalar. By an appropriate conjugation, one can assume that either A_k is diagonal or it is written as a single Jordan block (here we are using the assumption on $\overline{\mathbb{F}}$). Then lemma 2.1 implies that $\mathcal{A}$ is either in diagonal or in triangular form, as wanted. $\square$

The following consequence (a similarity test for commutative sequences) is clear.

Corollary B.2. *Let $\mathcal{A}$ and $\mathcal{A}'$ be commutative, both of the same form as described in the Theorem above. If $\mathcal{A}$ is in diagonal form, then $\mathcal{A} \sim \mathcal{A}'$ if and only if $a = a'$ and $d = d'$ or $a = d'$ and $d = a'$. If $\mathcal{B}$ is in triangular form then $\mathcal{B} \sim \mathcal{B}'$ if and only if $a = a'$ and $b = \lambda b'$ for some $\lambda \in \overline{\mathbb{F}} \setminus \{0\}$.*

References

[1] M. Artin, On Azumaya algebras and finite dimensional representations of rings, J. Algebra 11 (1969) 532–563.
[2] C.W. Curtis, I. Reiner, Representation Theory of Finite Groups and Associative Algebras, Pure and Applied Math., John Wiley & Sons, 1962.
[3] V. Drensky, Defining relations for the algebra of invariants of 2×2 matrices, Algebra Represent. Theory 6 (2003) 193–214.
[4] J. Dubnov, V. Ivanov, Sur l'abaissement du degré des polinômes en affineurs, C.R. (Doklady) Acad. Sci. URSS (N.S.) 41 (1943) 95–98; G. Higman, On a conjecture of Nagata, Proc. Cambridge Philos. Soc. 52 (1956) 1–4.
[5] C. Florentino, Invariants of 2×2 matrices, irreducible $SL(2, \mathbb{C})$ characters and the Magnus trace map, Geom. Dedicata 121 (2007) 167–186.
[6] S. Friedland, Simultaneous similarity of matrices, Adv. Math. 50 (1983) 189–265.
[7] M. Gerstenhaber, On dominance and varieties of commuting matrices, Ann. Math. 73 (1961) 324–348.
[8] R. Guralnick, A note on commuting pairs of matrices, Linear and Multilinear Algebra 31 (1992) 71–75.
[9] T. Laffey, Simultaneous reduction of sets of matrices under similarity, Linear Algebra Appl. 84 (1986) 123–138.
[10] N. McCoy, On the characteristic roots of matric polynomials, Bull. Amer. Math. Soc. 42 (1936) 592–600.
[11] S. Mukai, An Introduction to Invariants and Moduli, Cambridge UP, 2003.
[12] D. Mumford, J. Fogarty, F. Kirwan, Geometric Invariant Theory, third ed., Springer-Verlag, 1994.
[13] A. Paz, An application of the Cayley–Hamilton theorem to matrix polynomials in several variables, Linear and Multilinear Algebra 15 (1984) 161–170.
[14] C.J. Pappacena, An upper bound for the length of a finite-dimensional algebra, J. Algebra 197 (1997) 535–545.
[15] C. Procesi, The invariant theory of $n \times n$ matrices, Adv. Math. 19 (1976) 306–381.
[16] Yu. Razmyslov, Identities with trace in full matrix algebras over a field of characteristic zero, Izv. Akad. Nauk SSSR Ser. Mat. 38 (1974) 723–756.
[17] H. Radjavi, P. Rosenthal, Simultaneous Triangularization, Universitext, Springer-Verlag, 2000.