Matrix identities involving multiplication and transposition

Karl Auinger   Igor Dolinka   Mikhail V. Volkov

Abstract

We study matrix identities involving multiplication and unary operations such as transposition or Moore-Penrose inversion. We prove that in many cases such identities admit no finite basis.

Keywords. matrix transposition, symplectic transpose, Moore-Penrose inverse, matrix law, identity basis, finite basis problem

Background and motivation

Matrices[1] and matrix operations constitute basic tools for many branches of mathematics. Important properties of matrix operations are often expressed in form of *laws* or *identities* such as the associative law for multiplication of matrices. Studying matrix identities that involve multiplication and addition is a classic research direction that was motivated by several important problems in geometry and algebra (see [1] for a survey of the origins of the theory) and that has eventually led to the profound and beautiful theory of PI-rings [14, 21, 23, 47]. Matrix identities involving along with multiplication and addition also certain involution operations (such as taking the usual or symplectic transpose of a matrix) have attracted much attention as well, see, for instance, [9, 10, 15, 20, 47].

If one aims to classify matrix identities of a certain type, then a natural approach is to look for a collection of 'basic' identities such that all other identities would follow

K. Auinger: Fakultät für Mathematik, Universität Wien, Nordbergstrasse 15, A-1090 Wien, Austria; e-mail: karl.auinger@univie.ac.at

I. Dolinka: Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia; e-mail: dockie@dmi.uns.ac.rs

M. V. Volkov: Faculty of Mathematics and Mechanics, Ural State University, Lenina 51, 620083 Ekaterinburg, Russia; e-mail: mikhail.volkov@usu.ru

[1]In this paper, the word 'matrix' always means a square matrix of finite size; also, in order to avoid trivialities, we always assume that the size is at least two.

from these basic identities. Such a collection is usually referred to as an *identity basis* or simply a *basis*. For instance, all identities of matrices over an infinite field involving only multiplication are known to follow from the associative law, see [22, Lemma 2]. Thus, the associative law forms a basis of such 'multiplicative' identities. For identities involving both multiplication and addition, an explicit basis is known for 2×2-matrices (except the case of an infinite ground field of characteristic 2), see [8, 13, 30, 36, 46], and for 3×3- and 4×4-matrices over a finite field, see [17, 18]. However, for matrices of arbitrary size over a finite field or a field of characteristic 0 the powerful results by Kruse-L'vov [31, 34] and Kemer [24, 25] ensure at least the existence of a finite identity basis for such identities.

In contrast, multiplicative identities of matrices over a finite field admit no finite basis. This rather surprising fact was proved in the mid-1980s by the third author [53, Proposition 3] and Sapir [49, Corollary 6.2]. It is worth noting that methods used in [53] and [49] were very different but each of them sufficed to cover multiplicative identities of matrices of every fixed size over every finite field.

In the present paper we study matrix identities involving multiplication and one or two natural one-place operations such as taking various transposes or Moore-Penrose inversion. For this we first have to adapt the methods of [53] and [49]. We present the corresponding results in Section 2 while Section 1 collects necessary preliminaries. Applications to the finite basis problem for matrix identities are presented in Section 3. Both methods of Section 2 are used here, and it turns out that they in some sense complement one another since, in contrast to the case of purely multiplicative identities, none of the methods alone suffices to cover, say, identities of matrices of every size over every finite field involving multiplication and the usual transposition of matrices.

Our main results may be a summarized as follows.

Theorem. *Each of following sets of matrix identities admits no finite identity basis:*

- *the identities of $n \times n$-matrices over a finite field involving multiplication and usual transposition;*

- *the identities of $2n \times 2n$-matrices over a finite field involving multiplication and symplectic transposition;*

- *the identities of 2×2-matrices over the field of complex numbers involving either multiplication and Moore-Penrose inversion or multiplication, Moore-Penrose inversion and Hermitian conjugation;*

- *the identities of Boolean $n \times n$-matrices involving multiplication and transposition.*

We mention in passing that tools developed in Section 2 admit many further applications that will be published in a separate paper.

As far as the theory of matrices is concerned, we use fairly standard concepts of linear algebra, see [38]. Our proofs however involve some notions of equational logic and semigroup theory which are briefly recalled in the next section.

1 Preliminaries

1.1 Vocabulary of equational logic

The concepts of an identity and an identity basis are intuitively clear. Nevertheless, any precise reasoning about these concepts requires a formal framework, especially when one aims at 'negative' results as we do in this paper. Such a framework, provided by equational logic, is concisely presented, e.g., in [5, Chapter II]. For the reader's convenience, we briefly overview the basic vocabulary of equational logic in a form adapted to the use in this paper. Readers familiar with equational logic may, of course, skip this overview.

A non-empty set A endowed with operations $f_1 : \underbrace{A \times A \times \cdots \times A}_{n_1} \to A$, $f_2 : \underbrace{A \times A \times \cdots \times A}_{n_2} \to A$, ... is called an *algebraic structure of type* $(n_1, n_2, \dots)$ *with the carrier A*. Algebraic structures considered in this paper are mostly of types $(2, 1)$ or $(2, 1, 1)$ so that they have one binary operation and one or two unary operations. Since binary operations involved are always assumed to satisfy the associative law, our structures are *semigroups* equipped with one or two additional unary operations; for brevity, we shall call such structures *unary semigroups*.

We formally define notions related to unary semigroup identities only for the case of one unary operation leaving the straightforward modification for the case of two unary operations to the reader. Given a countably infinite set X, we define the set $T(X)$ of all *unary semigroup terms* over X as follows:

- every $x \in X$ is a unary semigroup term and so is x^*;

- if u and v are unary semigroup terms, then so is their concatenation uv;

- if u is a unary semigroup term, then so is $(u)^*$.

The algebraic structure $\mathcal{T}(X) = \langle T(X), \cdot, {}^* \rangle$ of type $(2, 1)$ whose binary operation $\cdot$ is concatenation and whose unary operation is $u \mapsto (u)^*$ is called the *free unary semigroup* over X. By a *unary semigroup identity* over X we mean a formal expression $u = v$ where $u, v \in T(X)$. A unary semigroup $\mathcal{S} = \langle S, \cdot, {}^* \rangle$ *satisfies* the identity $u = v$ if the equality $\varphi(u) = \varphi(v)$ holds in $\mathcal{S}$ under all possible homomorphisms $\varphi : \mathcal{T}(X) \to \mathcal{S}$. Given $\mathcal{S}$, we denote by $\mathrm{Eq}\,\mathcal{S}$ the set of all unary semigroup identities it satisfies.

Given any collection Σ of unary semigroup identities, we say that an identity $u = v$ *follows* from Σ or that Σ *implies* $u = v$ if every unary semigroup satisfying all identities of Σ satisfies the identity $u = v$ as well. Birkhoff's completeness theorem of equational logic (see [5, Theorem 14.17]) shows that this notion (which we have given a semantic definition) can be captured by a very transparent set of inference rules. These rules in fact formalize the most natural things one does with identities: substitution of a term for a variable, application of operations to identities (such as, say, multiplying an identity through on the right by the same term) and using symmetry and transitivity of equality. We need not going into more detail here because the completeness theorem is not utilized in this paper.

Given a unary semigroup $\mathcal{S}$, an *identity basis* for $\mathcal{S}$ is any set $\Sigma \subseteq \operatorname{Eq} \mathcal{S}$ such that every identity of $\operatorname{Eq} \mathcal{S}$ follows from Σ. A unary semigroup $\mathcal{S}$ is said to be *finitely based* if it possesses a finite identity basis; otherwise $\mathcal{S}$ is called *nonfinitely based*.

The class of all unary semigroups satisfying all identities from a given set Σ of unary semigroup identities is called the *variety defined by* Σ. A variety is is said to be *finitely based* if it can be defined by a finite set of identities; otherwise it is called *nonfinitely based*.

It is easy to see that the satisfaction of an identity is inherited by forming direct products, taking unary subsemigroups and homomorphic images so that each variety is closed under these operators. In fact, varieties can be characterized by this closure property (the HSP-theorem, see [5, Theorem 11.9]). Given a unary semigroup $\mathcal{S}$, the variety defined by $\operatorname{Eq} \mathcal{S}$ is the *variety generated by* $\mathcal{S}$; we denote this variety by $\operatorname{var} \mathcal{S}$. From the HSP-theorem it follows that every member of $\operatorname{var} \mathcal{S}$ is a homomorphic image of a unary subsemigroup of a direct product of several copies of $\mathcal{S}$. Observe also that a unary semigroup and the variety it generates are simultaneously finitely or nonfinitely based.

A variety $\mathbf{V}$ is said to be *locally finite* if every finitely generated member of $\mathbf{V}$ is finite. A finite unary semigroup is called *inherently nonfinitely based* if it is not contained in any finitely based locally finite variety. Since the variety generated by a finite unary semigroup is locally finite (this is an easy consequence of the HSP-theorem, see [5, Theorem 10.16]), the property of being inherently nonfinitely based implies the property of being nonfinitely based; in fact, the former property is much stronger.

1.2   Vocabulary of semigroup theory

Since the algebraic structures we deal with are semigroups with some additional operation(s), we freely use the standard terminology and notation of semigroup theory, mostly following the early chapters of the textbook [6]. It should be noted, however, that our presentation is to a reasonable extent self-contained and does not require any specific semigroup-theoretic background.

In general, considering a unary semigroup $\mathcal{S} = \langle S, \cdot, {}^* \rangle$, we do not assume any additional identities involving the unary operation *. If the identities $(xy)^* = y^* x^*$ and $(x^*)^* = x$ happen to hold in $\mathcal{S}$, in other words, if the unary operation $x \mapsto x^*$ is an involutory anti-automorphism of the semigroup $\langle S, \cdot \rangle$, we call $\mathcal{S}$ an *involutory semigroup*. If, in addition, the identity $x = xx^* x$ holds, $\mathcal{S}$ is said to be a *regular $*$-semigroup*. Each group, subject to its inverse operation $x \mapsto x^{-1}$, is an involutory semigroup, even a regular $*$-semigroup; throughout the paper, any group is considered as a unary semigroup with respect to this inverse unary operation.

A wealth of examples of involutory semigroups and regular $*$-semigroups can be obtained via the following 'unary' version of the well known Rees matrix construction[2]. Let $\mathcal{G} = \langle G, \cdot, {}^{-1} \rangle$ be a group, 0 a symbol beyond G, and I a non-empty set. We formally set $0^{-1} = 0$. Given an $I \times I$-matrix $P = (p_{ij})$ over $G \cup \{0\}$ such that $p_{ij} = p_{ji}^{-1}$ for all $i, j \in I$, we define a multiplication $\cdot$ and a unary operation * on the set $(I \times G \times I) \cup \{0\}$

---

[2]See [6, Section 3.1] for a description of the construction in the plain semigroup case and [19, Section 2] for a detailed analysis of its unary version.

by the following rules:

$$a \cdot 0 = 0 \cdot a = 0 \ \text{ for all } a \in (I \times G \times I) \cup \{0\},$$

$$(i, g, j) \cdot (k, h, \ell) = \begin{cases} (i, gp_{jk}h, \ell) & \text{if } p_{jk} \neq 0, \\ 0 & \text{if } p_{jk} = 0; \end{cases}$$

$$(i, g, j)^* = (j, g^{-1}, i), \ 0^* = 0.$$

It can be easily checked that $\langle (I \times G \times I) \cup \{0\}, \cdot, ^* \rangle$ becomes an involutory semigroup; it will be a regular $*$-semigroup precisely when $p_{ii} = e$ (the identity element of the group $\mathcal{G}$) for all $i \in I$. We denote this unary semigroup by $\mathcal{M}^0(I, \mathcal{G}, I; P)$ and call it the *unary Rees matrix semigroup over $\mathcal{G}$ with the sandwich matrix P*. If the involved group $\mathcal{G}$ happens to be the trivial group $\mathcal{E} = \{e\}$, then we shall ignore the group entry and represent the non-zero elements of such a Rees matrix semigroup by the pairs (i, j) with $i, j \in I$.

In this paper, the 10-element unary Rees matrix semigroup over the trivial group $\mathcal{E} = \{e\}$ with the sandwich matrix

$$\begin{pmatrix} e & e & e \\ e & e & 0 \\ e & 0 & e \end{pmatrix}$$

plays a key role; we denote this semigroup by $\mathcal{K}_3$. Thus, subject to the convention mentioned above, $\mathcal{K}_3$ consists of the nine pairs (i, j), $i, j \in \{1, 2, 3\}$, and the element 0, and the operations restricted to its non-zero elements can be described as follows:

$$(i, j) \cdot (k, \ell) = \begin{cases} (i, \ell) & \text{if } (j, k) \neq (2, 3), (3, 2), \\ 0 & \text{otherwise;} \end{cases} \tag{1.1}$$

$$(i, j)^* = (j, i).$$

Another unary semigroup that will be quite useful in the sequel is the *free involutory semigroup $\mathcal{FI}(X)$* on a given alphabet X. It can be constructed as follows. Let $\overline{X} = \{x^* \mid x \in X\}$ be a disjoint copy of X and define $(x^*)^* = x$ for all $x^* \in \overline{X}$. Then $\mathcal{FI}(X)$ is the free semigroup $(X \cup \overline{X})^+$ endowed with an involution * defined by

$$(x_1 \cdots x_m)^* = x_m^* \cdots x_1^*$$

for all $x_1, \ldots, x_m \in X \cup \overline{X}$. See [19, Section 3] for more details on $\mathcal{FI}(X)$.

We will refer to elements of $\mathcal{FI}(X)$ as *involutory words over X* while elements of the free semigroup X^+ will be referred to as (plain semigroup) words over X.

1.3 A property of matrices of rank 1

Given a field $\mathcal{K} = \langle K, +, \cdot \rangle$, we denote the set of all $n \times n$-matrices over $\mathcal{K}$ by $\mathrm{M}_n(\mathcal{K})$. As mentioned in the introduction, in order to avoid trivialities, we always assume that $n \geq 2$.

We conclude our preparations with registering a simple property of rank 1 matrices. This property is, of course, known, but we do provide a proof for the sake of completeness.

Lemma 1.1. *If a matrix $A \in \mathrm{M}_n(\mathcal{K})$ has rank 1, then $A^2BA = ABA^2$ for any matrix $B \in \mathrm{M}_n(\mathcal{K})$.*

Proof. Consider the natural action of A on the vector space K^n. The null-space $N(A) = \{x \in K^n \mid xA = 0\}$ has dimension $n - 1$ whence the generalized eigenspace of A corresponding to the eigenvalue 0 coincides with either K^n or $N(A)$. In the former case $A^2 = 0$ and $A^2BA = 0 = ABA^2$ for any matrix B. In the latter case K^n decomposes into the direct sum of $N(A)$ and the range $R(A) = \{xA \mid x \in K^n\}$, see [38, Section 5.10]. Then $R(A)$ is in fact the (generalized) eigenspace of A corresponding to a non-zero eigenvalue $\alpha \in K$ and the minimal polynomial of A is $x(x - \alpha)$. Thus, A satisfies the equation $A^2 - \alpha A = 0$, whence $A^2BA = \alpha ABA = AB\alpha A = ABA^2$, as required. $\square$

Let $\mathrm{L}_n(\mathcal{K})$ denote the set of all $n \times n$-matrices of rank at most 1 over $\mathcal{K}$. Adding the identity matrix to $\mathrm{L}_n(\mathcal{K})$ we get a set which we denote by $\mathrm{L}_n^1(\mathcal{K})$. Clearly, it is closed under matrix multiplication. From Lemma 1.1 we immediately obtain

Corollary 1.2. *The semigroup $\langle \mathrm{L}_n^1(\mathcal{K}), \cdot \rangle$ satisfies the identity*

$$x^2yx = xyx^2. \tag{1.2}$$

Observe that every group satisfying (1.2) is abelian.

2   Tools

2.1   A unary version of the critical semigroup method

Here we present a 'unary' modification of the approach used in [53]. According to the classification proposed in the survey [54], this approach is referred to as the *critical semigroup method*.

The formulation of the corresponding result involves two simple operators on unary semigroup varieties. For any unary semigroup $\mathcal{S} = \langle S, \cdot, {}^* \rangle$ we denote by $\mathrm{H}(\mathcal{S})$ the unary subsemigroup of $\mathcal{S}$ which is generated by all elements of the form xx^*, where $x \in S$. We call $\mathrm{H}(\mathcal{S})$ the *Hermitian subsemigroup* of $\mathcal{S}$. For any variety $\mathbf{V}$ of unary semigroups, let $\mathrm{H}(\mathbf{V})$ be the subvariety of $\mathbf{V}$ generated by all Hermitian subsemigroups of members of $\mathbf{V}$. Likewise, given a positive integer n, let $\mathrm{P}_n(\mathcal{S})$ be the unary subsemigroup of $\mathcal{S}$ which is generated by all elements of the form x^n, where $x \in S$, and let $\mathrm{P}_n(\mathbf{V})$ be the subvariety of $\mathbf{V}$ generated by all subsemigroups $\mathrm{P}_n(\mathcal{S})$, where $\mathcal{S} \in \mathbf{V}$.

The following easy observation will be useful in the sequel as it helps calculating the effect of the operators H and P_n.

Lemma 2.1. $\mathrm{H}(\mathsf{var}\, \mathcal{S}) = \mathsf{var}\, \mathrm{H}(\mathcal{S})$ *and* $\mathrm{P}_n(\mathsf{var}\, \mathcal{S}) = \mathsf{var}\, \mathrm{P}_n(\mathcal{S})$ *for every unary semigroup* $\mathcal{S}$ *and for each* n.

Proof. The non-trivial part of the first claim is the inclusion $\mathrm{H}(\mathsf{var}\, \mathcal{S}) \subseteq \mathsf{var}\, \mathrm{H}(\mathcal{S})$. Let $\mathcal{T} \in \mathsf{var}\, \mathcal{S}$, then $\mathcal{T}$ is a homomorphic image of a unary subsemigroup $\mathcal{U}$ of a direct product of several copies of $\mathcal{S}$. But then $\mathrm{H}(\mathcal{T})$ is a homomorphic image of $\mathrm{H}(\mathcal{U})$. As is easy to see, $\mathrm{H}(\mathcal{U})$ is a unary subsemigroup of a direct product of several copies of $\mathrm{H}(\mathcal{S})$. Thus $\mathrm{H}(\mathcal{T}) \in \mathsf{var}\, \mathrm{H}(\mathcal{S})$. Since this holds for an arbitrary $\mathcal{T} \in \mathsf{var}\, \mathcal{S}$, we conclude that $\mathrm{H}(\mathsf{var}\, \mathcal{S}) \subseteq \mathsf{var}\, \mathrm{H}(\mathcal{S})$. The second assertion can be treated in a completely similar way. $\square$

We are now ready to state the main result of this subsection.

Theorem 2.2. *Let* $\mathbf{V}$ *be any unary semigroup variety such that* $\mathcal{K}_3 \in \mathbf{V}$. *If either*

- *there exists a group* $\mathcal{G}$ *such that* $\mathcal{G} \in \mathbf{V}$ *but* $\mathcal{G} \notin \mathrm{H}(\mathbf{V})$

or

- *there exist a positive integer* d *and a group* $\mathcal{G}$ *of exponent dividing* d *such that* $\mathcal{G} \in \mathbf{V}$ *but* $\mathcal{G} \notin \mathrm{P}_d(\mathbf{V})$,

then $\mathbf{V}$ *has no finite basis of identities.*

Proof. Assume first that there exists a group $\mathcal{G} \in \mathbf{V}$ for which $\mathcal{G} \notin \mathrm{H}(\mathbf{V})$.

1. First we recall the basic idea of 'the critical semigroup method' in the unary setting. Suppose that $\mathbf{V}$ is finitely based. If Σ is a finite identity basis of the variety $\mathbf{V}$ then there exists a positive integer ℓ such that all identities from Σ depend on at most ℓ letters. Therefore identities from Σ hold in a unary semigroup $\mathcal{S}$ whenever all ℓ-generated unary subsemigroups of $\mathcal{S}$ satisfy Σ. In other words, $\mathcal{S}$ belongs to $\mathbf{V}$ whenever all of its ℓ-generated unary subsemigroups are in $\mathbf{V}$. We see that in order to prove our theorem it is sufficient to construct, for any given positive integer k, a unary semigroup $\mathcal{T}_k \notin \mathbf{V}$ for which all k-generated unary subsemigroups of $\mathcal{T}_k$ belong to $\mathbf{V}$.

2. Fix an identity $u(x_1, \ldots, x_m) = v(x_1, \ldots, x_m)$ that holds in $\mathrm{H}(\mathbf{V})$ but fails in the group $\mathcal{G} = \langle G, \cdot, ^{-1} \rangle$. The latter means that, for some $g_1, \ldots, g_m \in G$, substitution of g_i for x_i yields

$$u(g_1, \ldots, g_m) \neq v(g_1, \ldots, g_m). \tag{2.1}$$

Now, for each positive integer k, let $n = \max\{4, 2k+1\}$, $I = \{1, \ldots, nm\}$ and consider the unary Rees matrix semigroup $\mathcal{T}_k = \mathcal{M}^0(I, \mathcal{G}, I; P_k)$ over the group $\mathcal{G}$ with the sandwich matrix

$$P_k = \begin{pmatrix}
M_n(g_1) & E_n & O_n & O_n & \cdots & O_n & E_n^T \\
E_n^T & M_n(g_2) & E_n & O_n & \cdots & O_n & O_n \\
O_n & E_n^T & M_n(g_3) & E_n & \cdots & O_n & O_n \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
O_n & O_n & O_n & O_n & \cdots & E_n & O_n \\
O_n & O_n & O_n & O_n & \cdots & M_n(g_{m-1}) & E_n \\
E_n & O_n & O_n & O_n & \cdots & E_n^T & M_n(g_m)
\end{pmatrix},$$

where O_n is the zero $n \times n$-matrix, E_n is the $n \times n$-matrix having e (the identity of $\mathcal{G}$) in the position $(n, 1)$ and 0 in all other positions, E_n^T is the transpose of E_n, and $M_n(g)$ denotes the $n \times n$-matrix of the form

$$M_n(g) = \begin{pmatrix}
e & g & 0 & \cdots & 0 & 0 & e \\
g^{-1} & e & e & \cdots & 0 & 0 & 0 \\
0 & e & e & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & e & e & 0 \\
0 & 0 & 0 & \cdots & e & e & e \\
e & 0 & 0 & \cdots & 0 & e & e
\end{pmatrix}.$$

(This construction is in a sense a combination of those of the first and the third authors' papers [3] and [53].) We aim to prove that $\mathcal{T}_k$ enjoys the two properties needed, namely, it does not belong to $\mathbf{V}$, but each k-generated unary subsemigroup of $\mathcal{T}_k$ lies in $\mathbf{V}$.

3. In order to prove that $\mathcal{T}_k \notin \mathbf{V}$, we construct an identity that holds in $\mathbf{V}$, but fails in $\mathcal{T}_k$. Consider the following m terms in mn letters $x_1, \ldots, x_{mn}$ (the square brackets in these terms serve only to improve readability):

$$w_1 = [x_1 x_1^* \cdots x_n x_n^*][(x_{n+1} \cdots x_{2n})(x_{n+1} \cdots x_{2n})^* x_{2n} x_{2n}^*] \cdots$$
$$[(x_{(m-1)n+1} \cdots x_{mn})(x_{(m-1)n+1} \cdots x_{mn})^* x_{mn} x_{mn}^*],$$
$$w_2 = [(x_1 \cdots x_n)(x_1 \cdots x_n)^* x_n x_n^*][x_{n+1} x_{n+1}^* \cdots x_{2n} x_{2n}^*] \times$$
$$[(x_{2n+1} \cdots x_{3n})(x_{2n+1} \cdots x_{3n})^* x_{3n} x_{3n}^*] \cdots$$
$$[(x_{(m-1)n+1} \cdots x_{mn})(x_{(m-1)n+1} \cdots x_{mn})^* x_{mn} x_{mn}^*],$$
$$w_3 = [(x_1 \cdots x_n)(x_1 \cdots x_n)^* x_n x_n^*][(x_{n+1} \cdots x_{2n})(x_{n+1} \cdots x_{2n})^* x_{2n} x_{2n}^*] \times$$
$$[x_{2n+1} x_{2n+1}^* \cdots x_{3n} x_{3n}^*] \cdots [(x_{(m-1)n+1} \ldots x_{mn})(x_{(m-1)n+1} \ldots x_{mn})^* x_{mn} x_{mn}^*],$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$w_m = [(x_1 \cdots x_n)(x_1 \cdots x_n)^* x_n x_n^*][(x_{n+1} \cdots x_{2n})(x_{n+1} \cdots x_{2n})^* x_{2n} x_{2n}^*] \cdots$$
$$[(x_{(m-2)n+1} \cdots x_{(m-1)n})(x_{(m-2)n+1} \cdots x_{(m-1)n})^* x_{(m-1)n} x_{(m-1)n}^*] \times$$
$$[x_{(m-1)n+1} x_{(m-1)n+1}^* \cdots x_{mn} x_{mn}^*].$$

Substituting w_i for x_i in u respectively v, we get the identity

$$u(w_1, \ldots, w_m) = v(w_1, \ldots, w_m) \tag{2.2}$$

which holds in the variety $\mathbf{V}$. Indeed, if we take any $\mathcal{S} \in \mathbf{V}$, then, since $ss^* \in \mathrm{H}(\mathcal{S})$ for any s in $\mathcal{S}$, all the values of w_i belong to the Hermitian subsemigroup $\mathrm{H}(\mathcal{S})$ of $\mathcal{S}$. This subsemigroup, however, lies in $\mathrm{H}(\mathbf{V})$, and therefore, satisfies the identity $u = v$.

Now we shall show that (2.2) fails in $\mathcal{T}_k$. Indeed, substituting $(i, e, i) \in \mathcal{T}_k$ for x_i, we calculate that the value of every term of the form

$$(x_{(j-1)n+1} \cdots x_{jn})(x_{(j-1)n+1} \cdots x_{jn})^* x_{jn} x_{jn}^*$$

is equal to $((j-1)n+1, e, jn)$ while the value of each term of the form

$$x_{(j-1)n+1} x_{(j-1)n+1}^* \cdots x_{jn} x_{jn}^*$$

is equal to $((j-1)n+1, g_j, jn)$. Hence the value of w_j is just $(1, g_j, mn)$. Therefore, under this substitution, the left hand part of (2.2) takes the value $(s, u(g_1, \ldots, g_m), t)$ for suitable $s, t \in \{1, mn\}$ while the value of the right hand part of (2.2) is $(s', v(g_1, \ldots, g_m), t')$ (again for suitable $s', t' \in \{1, mn\}$). In view of the inequality (2.1), these elements do not coincide in $\mathcal{T}_k$.

4. It remains to prove that each k-generated unary subsemigroup of $\mathcal{T}_k$ lies in $\mathbf{V}$. For every m-tuple $(\lambda_1, \ldots, \lambda_m)$ of positive integers satisfying

$$1 \leq \lambda_1 \leq n < \lambda_2 \leq 2n < \lambda_3 \leq \ldots (m-1)n < \lambda_m \leq mn, \tag{2.3}$$

consider the unary subsemigroup $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$ of $\mathcal{T}_k$ consisting of 0 and all triples (i, g, j) such that $g \in G$ and $i, j \notin \{\lambda_1, \ldots, \lambda_m\}$. Using that $2k < n$ according to our choice of n, one concludes that any given k elements of $\mathcal{T}_k$ must be contained in

$\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$ for suitable $\lambda_1, \ldots, \lambda_m$. Thus it is sufficient to prove that each semigroup of the form $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$ belongs to the variety $\mathbf{V}$.

Let us fix positive integers $\lambda_1, \ldots, \lambda_m$ satisfying (2.3). When multiplying triples from $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$, the $\lambda_1^{\text{th}}, \ldots, \lambda_m^{\text{th}}$ rows and columns of the sandwich matrix P_k are never involved. Therefore we can identify $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$ with the unary Rees matrix semigroup $\mathcal{M}^0(I', \mathcal{G}, I'; P_k')$ over the group $\mathcal{G}$ where $I' = I \setminus \{\lambda_1, \ldots, \lambda_m\}$ and the sandwich matrix $P_k' = P_k(\lambda_1, \ldots, \lambda_m)$ is obtained from P_k by deleting its $\lambda_1^{\text{th}}, \ldots, \lambda_m^{\text{th}}$ rows and columns. Note that by (2.3) exactly one row and one column of each block $M_n(g_i)$ is deleted.

Now we transform the matrix $P_k(\lambda_1, \ldots, \lambda_m)$ as follows. For each i such that $(i-1)n + 2 < \lambda_i$, we multiply successively

$$
\begin{aligned}
&\text{the row } ((i-1)n+2) \text{ by } g_i \text{ from the left and} \\
&\text{the column } ((i-1)n+2) \text{ by } g_i^{-1} \text{ from the right;} \\
&\text{the row } ((i-1)n+3) \text{ by } g_i \text{ from the left and} \\
&\text{the column } ((i-1)n+3) \text{ by } g_i^{-1} \text{ from the right;} \qquad (2.4) \\
&\dotfill \\
&\text{the row } (\lambda_i - 1) \text{ by } g_i \text{ from the left and} \\
&\text{the column } (\lambda_i - 1) \text{ by } g_i^{-1} \text{ from the right.}
\end{aligned}
$$

In order to help the reader to understand the effect of the transformations (2.4), we illustrate their action on the block obtained from $M_n(g_i)$ by removing the λ_i^{th} row and column in the following scheme in which λ_i has been chosen to be equal to $(i-1)n + 5$. (The transformations have no effect beyond $M_n(g_i)$ because all the rows and columns of $P_k(\lambda_1, \ldots, \lambda_m)$ involved in (2.4) have non-zero entries only within $M_n(g_i)$.)

<table>
<tr><td align="center">The block obtained from $M_n(g_i)$ by erasing
the $((i-1)n+5)^{\text{th}}$ row and column</td><td align="center">After the first
transformation</td></tr>
</table>

$$
\begin{pmatrix}
e & g_i & 0 & 0 & 0 & \cdots & 0 & e \\
g_i^{-1} & e & e & 0 & 0 & \cdots & 0 & 0 \\
0 & e & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & e & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & e & e \\
e & 0 & 0 & 0 & 0 & \cdots & e & e
\end{pmatrix}
\qquad
\begin{pmatrix}
e & e & 0 & 0 & 0 & \cdots & 0 & e \\
e & e & g_i & 0 & 0 & \cdots & 0 & 0 \\
0 & g_i^{-1} & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & e & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & e & e \\
e & 0 & 0 & 0 & 0 & \cdots & e & e
\end{pmatrix}
$$

<table>
<tr><td align="center">After the second
transformation</td><td align="center">After the third
transformation</td></tr>
</table>

$$
\begin{pmatrix}
e & e & 0 & 0 & 0 & \cdots & 0 & e \\
e & e & e & 0 & 0 & \cdots & 0 & 0 \\
0 & e & e & g_i & 0 & \cdots & 0 & 0 \\
0 & 0 & g_i^{-1} & e & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & e & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & e & e \\
e & 0 & 0 & 0 & 0 & \cdots & e & e
\end{pmatrix}
\qquad
\begin{pmatrix}
e & e & 0 & 0 & 0 & \cdots & 0 & e \\
e & e & e & 0 & 0 & \cdots & 0 & 0 \\
0 & e & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & e & e & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & e & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & e & e \\
e & 0 & 0 & 0 & 0 & \cdots & e & e
\end{pmatrix}
$$

Now it should be clear that also in general the transformations (2.4) result in a matrix Q_k all of whose non-zero entries are equal to e. On the other hand, it is known (see, e.g., [3, Proposition 6.2]) that the transformations (2.4) of the sandwich matrix do not change the unary semigroup $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$; in other words $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m)$ is isomorphic to the $I' \times I'$ unary Rees matrix semigroup $\mathcal{R}_k$ over $\mathcal{G}$ with the sandwich matrix Q_k. Let $\mathcal{U}_k$ be the $I' \times I'$ unary Rees matrix semigroup over the trivial group $\mathcal{E}$ with the sandwich matrix Q_k. It is easy to check that the mapping $\mathcal{G} \times \mathcal{U}_k \to \mathcal{R}_k$ defined by

$$(g, (i, j)) \mapsto (i, g, j), \quad (g, 0) \mapsto 0$$

for all $g \in G$, $i, j \in I'$, is a unary semigroup homomorphism onto $\mathcal{R}_k$. Now we note that $\mathcal{G} \in \mathbf{V}$ and $\mathcal{U}_k$ belongs to the variety generated by $\mathcal{K}_3$ (see [3, Theorem 5.2]). This yields $\mathcal{T}_k(\lambda_1, \ldots, \lambda_m) \cong \mathcal{R}_k \in \mathbf{V}$.

The case when there exist a positive integer d and a group $\mathcal{G}$ of exponent dividing d such that $\mathcal{G} \in \mathbf{V}$ but $\mathcal{G} \notin \mathrm{P}_d(\mathbf{V})$ can be treated in a very similar way. The construction of the critical semigroups remains the same, and the only modification to be made in the rest of the proof is to replace the terms w_i above by the following plain semigroup words:

$w_1 = [x_1^d \cdots x_n^d][x_{n+1} \cdots x_{2n}]^d \cdots [x_{(m-1)n+1} \cdots x_{mn}]^d,$
$w_2 = [x_1 \cdots x_n]^d[x_{n+1}^d \cdots x_{2n}^d] \cdots [x_{(m-1)n+1} \cdots x_{mn}]^d,$
$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$
$w_m = [x_1 \cdots x_n]^d[x_{n+1} \cdots x_{2n}]^d \cdots [x_{(m-1)n+1}^d \cdots x_{mn}^d].$

(These words already have been used in the plain semigroup case by the third author [53].)

$\square$

2.2 A unary version of the method of inherently nonfinitely based semigroups

Here we prove a sufficient condition for an involutory semigroup to be inherently nonfinitely based and exhibit two concrete examples of involutory semigroups satisfying this condition. These examples will be essential in our applications in Section 3.

Let $x_1, x_2, \ldots, x_n, \ldots$ be a sequence of letters. The sequence $\{Z_n\}_{n=1,2,\ldots}$ of *Zimin words* is defined inductively by $Z_1 = x_1$, $Z_{n+1} = Z_n x_{n+1} Z_n$. We say that an involutory word v is an *involutory isoterm for a unary semigroup* $\mathcal{S}$ if the only involutory word v' such that $\mathcal{S}$ satisfies the involutory semigroup identity $v = v'$ is the word v itself.

Theorem 2.3. *Let $\mathcal{S}$ be a finite involutory semigroup. If all Zimin words are involutory isoterms for $\mathcal{S}$, then $\mathcal{S}$ is inherently nonfinitely based.*

Proof. Arguing by contradiction, suppose that $\mathcal{S}$ belongs to a finitely based locally finite variety $\mathbf{V}$. If Σ is a finite identity basis of $\mathbf{V}$, then there exists a positive integer ℓ such that all identities from Σ depend on at most ℓ letters. Clearly, all identities in Σ hold in $\mathcal{S}$. In the following, our aim will be to construct, for any given positive integer k, an infinite, finitely generated involutory semigroup $\mathcal{T}_k$ which satisfies all identities in at most k variables that hold in $\mathcal{S}$. In particular, $\mathcal{T}_\ell$ will satisfy all identities from Σ. This yields a contradiction, as then we must conclude that $\mathcal{T}_\ell \in \mathbf{V}$, which is impossible by the local finiteness of $\mathbf{V}$.

We shall employ a construction invented by Sapir [49], see also his lecture notes [51]. We fix k and let $r = 6k + 2$. Consider the $r^2 \times r$-matrix M shown in the left hand part of Fig. 1. All odd columns of M are identical and equal to the transpose of the

$$
M = \begin{pmatrix}
1 & 1 & \cdots & 1 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & r & \cdots & 1 & r \\
2 & 1 & \cdots & 2 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
2 & r & \cdots & 2 & r \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r & 1 & \cdots & r & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r & r & \cdots & r & r
\end{pmatrix}
\qquad
M_A = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1r-1} & a_{1r} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{11} & a_{r2} & \cdots & a_{1r-1} & a_{rr} \\
a_{21} & a_{12} & \cdots & a_{2r-1} & a_{1r} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{21} & a_{r2} & \cdots & a_{2r-1} & a_{rr} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{r1} & a_{12} & \cdots & a_{rr-1} & a_{1r} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{r1} & a_{r2} & \cdots & a_{rr-1} & a_{rr}
\end{pmatrix}
$$

Figure 1: The matrices M and M_A

row $(1, 1, \ldots, 1, 2, 2, \ldots, 2, \ldots, r, r, \ldots, r)$ where each number occurs r times. All even columns of M are identical and equal to the transpose of the row $(1, 2, \ldots, r, \ldots, 1, 2, \ldots, r)$ in which the block $1, 2, \ldots, r$ occurs r times.

Now consider the alphabet $A = \{a_{ij} \mid 1 \leq i, j \leq r\}$ of cardinality r^2. We convert the matrix M to the matrix M_A (shown in the right hand part of Fig. 1) by replacing numbers by letters according to the following rule: whenever the number i occurs in the column j of M, we substitute it with the letter a_{ij} to get the corresponding entry in M_A.

Let v_t be the word in the tth row of the matrix M_A. Consider the endomorphism $\gamma : A^+ \to A^+$ defined by

$$\gamma(a_{ij}) = v_{(i-1)r+j}.$$

Let V_k be the set of all factors of the words in the sequence $\{\gamma^m(a_{11})\}_{m=1,2,\ldots}$ and let 0 be a symbol beyond V_k. We define a multiplication $\cdot$ on the set $V_k \cup \{0\}$ as follows:

$$
u \cdot v = \begin{cases}
uv & \text{if } u, v, uv \in V_k, \\
0 & \text{otherwise.}
\end{cases}
$$

Clearly, $\langle V_k \cup \{0\}, \cdot \rangle$ becomes a semigroup which we denote by $\mathcal{V}_k^0$. Using this semigroup, we can conveniently reformulate two major combinatorial results by Sapir:

Proposition 2.4. [51, Proposition 2.1] *Let $X_k = \{x_1, \ldots, x_k\}$ and $w \in X_k^+$. Assume that there exists a homomorphism $\varphi : X_k^+ \to \mathcal{V}_k^0$ for which $\varphi(w) \neq 0$. Then there is an endomorphism $\psi : X_k^+ \to X_k^+$ such that the word $\psi(w)$ appears as a factor in the Zimin word Z_k.*

Proposition 2.5. [51, Lemma 4.14] *Let $X_k = \{x_1, \ldots, x_k\}$ and $w, w' \in X_k^+$. Assume that there exists a homomorphism $\varphi : X_k^+ \to \mathcal{V}_k^0$ for which $\varphi(w) \neq \varphi(w')$. Then the identity $w = w'$ implies a non-trivial semigroup identity of the form $Z_{k+1} = z$.*

Now let $\overline{\mathcal{V}}_k^0$ denote the semigroup anti-isomorphic to $\mathcal{V}_k^0$; we shall use the notation $x \mapsto x^*$ for the mutual anti-isomorphisms between $\mathcal{V}_k^0$ and $\overline{\mathcal{V}}_k^0$ in both directions and denote $\{v^* \mid v \in V_k\}$ by $\overline{V}_k$. Let

$$\mathcal{T}_k = \langle V_k \cup \overline{V}_k \cup \{0\}, \cdot, ^* \rangle$$

be the 0-direct union of $\mathcal{V}_k^0$ and $\overline{\mathcal{V}}_k^0$; this means that we identify 0 with 0^*, preserve the multiplication in both $\mathcal{V}_k^0$ and $\overline{\mathcal{V}}_k^0$, and set $u \cdot v^* = u^* \cdot v = 0$ for all $u, v \in V_k$. This is the unary semigroup we need.

It is clear that $\mathcal{T}_k$ is infinite and is generated (as a unary semigroup) by the set A which is finite. It remains to verify that $\mathcal{T}_k$ satisfies every identity in at most k variables that holds in our initial unary semigroup $\mathcal{S}$. So, let $p, q \in \mathcal{FI}(X_k)$ and suppose that the identity $p = q$ holds in $\mathcal{S}$ but fails in $\mathcal{T}_k$. Then there exists a unary semigroup homomorphism $\varphi : \mathcal{FI}(X_k) \to \mathcal{T}_k$ for which $\varphi(p) \neq \varphi(q)$. Hence, at least one of the elements $\varphi(p)$ and $\varphi(q)$ is not equal to 0; (without loss of generality) assume that $\varphi(p) \neq 0$. Then we may also assume $\varphi(p) \in V_k$; otherwise we may consider the identity $p^* = q^*$ instead of $p = q$. Since $\varphi(p) \neq 0$, there is no letter $x \in X_k$ such that p contains both x and x^*. Now we define a substitution $\sigma : \mathcal{FI}(X_k) \to \mathcal{FI}(X_k)$ as follows:

$$\sigma(x) = \begin{cases} x^* & \text{if } p \text{ contains } x^*, \\ x & \text{otherwise.} \end{cases}$$

Then $\sigma(p)$ does not contain any starred letter, thus being a plain word in X_k^+. Since σ^2 is the identity mapping, we have $\varphi(p) = (\varphi\sigma)(\sigma(p))$, and $\varphi\sigma$ maps X_k^+ into $\mathcal{V}_k^0$. Now we consider two cases.

Case 1: $\sigma(q)$ contains a starred letter. We apply Proposition 2.4 to the plain word $\sigma(p)$ and the semigroup homomorphism $X_k^+ \to \mathcal{V}_k^0$ obtained by restricting $\varphi\sigma$ to X_k^+. We conclude that there is an endomorphism ψ of X_k^+ such that the word $\psi(\sigma(p))$ appears as a factor in the Zimin word Z_k. Thus, $Z_k = z'\psi(\sigma(p))z''$ for some z', z'' (that may be empty). The endomorphism ψ extends in a natural way to an endomorphism of the free involutory semigroup $\mathcal{FI}(X_k)$ and there is no harm in denoting the extension by ψ as well. The identity $p = q$ implies the identity

$$z'\psi(\sigma(p))z'' = z'\psi(\sigma(q))z''. \tag{2.5}$$

The left hand side of (2.5) is Z_k and the identity is not trivial because its right hand side involves a starred letter. Since $p = q$ holds in our initial semigroup $\mathcal{S}$, so does (2.5). But this contradicts the assumption that all Zimin words are involutory isoterms for $\mathcal{S}$.

Case 2: $\sigma(q)$ contains no starred letter. In this case $\sigma(q)$ is a plain word in X_k^+, and we are in a position to apply Proposition 2.5 to the semigroup identity $\sigma(p) = \sigma(q)$ and the semigroup homomorphism $X_k^+ \to \mathcal{V}_k^0$ obtained by restricting $\varphi\sigma$ to X_k^+. We conclude that $\sigma(p) = \sigma(q)$ implies a non-trivial semigroup identity $Z_{k+1} = z$. Therefore the identity $p = q$ implies $Z_{k+1} = z$, and we again get a contradiction. $\square$

Before passing to concrete examples of inherently nonfinitely based involutory semigroups, we formulate a corollary of our proof of Theorem 2.3 which will be useful for

our subsequent paper(s). Proposition 2.4 easily implies that no word in the sequence $\{\gamma^m(a_{11})\}_{m=1,2,...}$ has any square (that is, a word of the form ww) as a factor. Hence, the semigroup $\mathcal{V}_k^0$ satisfies the identity[3] $x^2 = 0$. This identity is clearly inherited by the involutory semigroup $\mathcal{T}_k$ which by its construction satisfies also the identity $xx^* = 0$. Since $\mathcal{T}_k$ is finitely generated and infinite, we arrive at the following conclusion:

Corollary 2.6. *If a variety* $\mathbf{V}$ *of involutory semigroups satisfies no non-trivial identity of the form* $Z_{k+1} = z$ *and all members of* $\mathbf{V}$ *satisfying the identities* $xx^* = x^2 = 0$ *are locally finite, then* $\mathbf{V}$ *is not finitely based.*

This result is parallel to [49, Proposition 3] in the plain semigroup case.

Now consider the *twisted Brandt monoid* $\mathcal{T}\mathcal{B}_2^1 = \langle B_2^1, \cdot, * \rangle$, where B_2^1 is the set of the following six 2×2-matrices:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

the binary operation $\cdot$ is the usual matrix multiplication and the unary operation $*$ fixes the matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and swaps each of the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

with the other one.

Corollary 2.7. *The twisted Brandt monoid* $\mathcal{T}\mathcal{B}_2^1$ *is inherently nonfinitely based.*

Proof. By Theorem 2.3 we only have to show that $\mathcal{T}\mathcal{B}_2^1$ satisfies no non-trivial involutory semigroup identity of the form $Z_n = z$. If z is a plain semigroup word, we can refer to [49, Lemma 3.7] which shows that the semigroup $\langle B_2^1, \cdot \rangle$ does not satisfy any non-trivial **semigroup** identity of the form $Z_n = z$. If we suppose that the involutory word z contains a starred letter, we can substitute the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} \right)$ for all letters occurring in Z_n and z. Since this matrix is idempotent, the value of the word Z_n under this substitution equals $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} \right)$. On the other hand, z evaluates to a product involving the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} \right)^* = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} \right)$, and it is easy to see that such a product is equal to either $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ or $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix} \right)$. Thus, the identity $Z_n = z$ cannot hold in $\mathcal{T}\mathcal{B}_2^1$ in this case as well. $\qquad\square$

An equivalent way to define $\mathcal{T}\mathcal{B}_2^1$ is to consider the 5-element unary Rees matrix semigroup over the trivial group $\mathcal{E} = \{e\}$ with the sandwich matrix

$$\begin{pmatrix} 0 & e \\ e & 0 \end{pmatrix}$$

---

[3] Strictly speaking the following expression is not an identity as defined before (since 0 is not a term) but rather an abbreviation for the identities $x^2 y = x^2 = yx^2$. However, referring to such abbreviations as identities is a standard convention which we adopt.

and then to adjoin to this unary Rees matrix semigroup an identity element. For convenience and later use we note that $\mathcal{TB}_2^1$ can thus be realized as the set

$$\{(1,1),(1,2),(2,1),(2,2),0,1\}$$

endowed with the operations

$$(i,j)\cdot(k,\ell) = \begin{cases} (i,\ell) & \text{if } (j,k)\in\{(1,2),(2,1)\}, \\ 0 & \text{otherwise}; \end{cases} \tag{2.6}$$

$$1\cdot x = x = x\cdot 1,\ 0\cdot x = 0 = x\cdot 0 \text{ for all } x;$$

$$(i,j)^* = (j,i),\ 1^* = 1,\ 0^* = 0.$$

Suppose that $\mathcal{S}$ is a finite unary semigroup for which the variety var $\mathcal{S}$ contains an inherently nonfinitely based finite unary semigroup. Immediately from the definition it follows that $\mathcal{S}$ is also inherently nonfinitely based. This observation is useful, in particular, for the justification of our second example of an involutory inherently nonfinitely based semigroup. This is a 'twisted version' $\mathcal{TA}_2^1$ of another 6-element semigroup that often shows up under the name A_2^1 in the theory of semigroup varieties. The unary semigroup $\mathcal{TA}_2^1$ is formed by the 6 matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

under the usual matrix multiplication and the unary operation that swaps each of the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

with the other one and fixes all other matrices. Alternatively, $\mathcal{TA}_2^1$ is obtained from the 5-element unary Rees matrix semigroup $\mathcal{A}_2$ over $\mathcal{E} = \{e\}$ with the sandwich matrix

$$\begin{pmatrix} 0 & e \\ e & e \end{pmatrix} \tag{2.7}$$

by adjoining an identity element. Again, for later use, we note that $\mathcal{TA}_2^1$ can be realized as the set

$$\{(1,1),(1,2),(2,1),(2,2),0,1\}$$

endowed with the operations

$$(i,j)\cdot(k,\ell) = \begin{cases} (i,\ell) & \text{if } (j,k)\neq(1,1) \\ 0 & \text{if } (j,k)=(1,1); \end{cases} \tag{2.8}$$

$$1\cdot x = x = x\cdot 1,\ 0\cdot x = 0 = x\cdot 0 \text{ for all } x;$$

$$(i,j)^* = (j,i),\ 1^* = 1,\ 0^* = 0.$$

Corollary 2.8. *The involutory semigroup $\mathcal{TA}_2^1$ is inherently nonfinitely based.*

Proof. We represent $\mathcal{T}\mathcal{A}_2^1$ as in (2.8) and $\mathcal{T}\mathcal{B}_2^1$ as in (2.6) and consider the direct square $\mathcal{T}\mathcal{A}_2^1 \times \mathcal{T}\mathcal{A}_2^1$. It is then easy to check that the twisted Brandt monoid $\mathcal{T}\mathcal{B}_2^1$ is a homomorphic image of the unary subsemigroup of $\mathcal{T}\mathcal{A}_2^1 \times \mathcal{T}\mathcal{A}_2^1$ generated by the pairs $(1, 1)$, $\big((1, 1), (2, 2)\big)$ and $\big((2, 2), (1, 1)\big)$. Thus, $\mathcal{T}\mathcal{B}_2^1$ belongs to $\mathsf{var}\, \mathcal{T}\mathcal{A}_2^1$. Since by Corollary 2.7 $\mathcal{T}\mathcal{B}_2^1$ is inherently nonfinitely based, so is $\mathcal{T}\mathcal{A}_2^1$. □

Sapir [49, Proposition 7] has shown that a (plain) finite semigroup $\mathcal{S}$ is inherently non-finitely based **if and only if** all Zimin words are isoterms for $\mathcal{S}$, that is, $\mathcal{S}$ satisfies no non-trivial semigroup identity of the form $Z_n = z$. Our Theorem 2.3 models the 'if' part of this statement but we do not know whether or not the 'only if' part transfers to the involutory environment. Some partial results in this direction have been recently obtained by the second author [11]. Here we present yet another special result which however suffices for our purposes.

Proposition 2.9. *Let* $\mathcal{S} = \langle S, \cdot, {}^* \rangle$ *be a finite involutory semigroup and suppose that there exists an involutory word* $\omega(x)$ *in one variable* x *such that* $\mathcal{S}$ *satisfies the identity* $x = x\omega(x)x$. *Then* $\mathcal{S}$ *is not inherently nonfinitely based.*

Proof. We have to construct a finite set of identities that defines a locally finite variety of involutory semigroups containing $\mathcal{S}$. The crucial step towards this goal consists in verifying that the identity $x = x\omega(x)x$ allows one to express right divisibility in terms of equational logic. This being done, we shall be in a position to closely follow powerful arguments developed by Margolis and Sapir in [37].

We say that elements $a, b \in S$ *divide each other on the right* and write $a \mathscr{R} b$ if either $a = b$ or there exist $s, t \in S$ such that $a = bs$ and $b = at$. We say that b *strictly divides* a and write $a <_{\mathscr{R}} b$ if $a = bs$ for some $s \in S$ but $b \neq a$ and $b \neq at$ for any $t \in S$. Clearly, $\mathscr{R}$ is an equivalence relation (known as the *right Green relation* in semigroup theory) and $<_{\mathscr{R}}$ is transitive and anti-reflexive.

Since $\mathcal{S}$ satisfies the identity $x = x\omega(x)x$, we have $a \mathscr{R} a\omega(a)$ for each element $a \in S$. (Indeed, $a = a\omega(a) \cdot a$ and $a\omega(a) = a \cdot \omega(a)$.) Thus, for $a, b \in S$, we have $a \mathscr{R} b$ if and only if $a\omega(a) \mathscr{R} b\omega(b)$. Since $a\omega(a)$ and $b\omega(b)$ are idempotents, that latter condition is equivalent to the two equalities $a\omega(a) \cdot b\omega(b) = b\omega(b)$ and $b\omega(b) \cdot a\omega(a) = a\omega(a)$. In particular, for $u, v \in S$ we have $uv \mathscr{R} u$ if and only if $uv\omega(uv) \cdot u\omega(u) = u\omega(u)$ (since the second equality $u\omega(u) \cdot uv\omega(uv) = uv\omega(uv)$ is always true).

Let Z_n' be the word obtained from the Zimin word Z_n by deleting the last letter (which is x_1), that is, $Z_n' x_1 = Z_n$. Further let h denote the length of the longest possible chain of the form

$$s_1 <_{\mathscr{R}} s_2 <_{\mathscr{R}} \cdots <_{\mathscr{R}} s_k.$$

Set $n = h+1$; Lemma 7 in [37] shows that under every evaluation of the letters $x_1, \ldots, x_n$ in $\mathcal{S}$, the values of the words Z_n' and Z_n divide each other on the right. As explained in the previous paragraph, this can be restated as saying that the values of the terms $Z_n\omega(Z_n) \cdot Z_n'\omega(Z_n')$ and $Z_n'\omega(Z_n')$ coincide under every evaluation of $x_1, \ldots, x_n$ in $\mathcal{S}$, that is, $\mathcal{S}$ satisfies the identity

$$Z_n\omega(Z_n) \cdot Z_n'\omega(Z_n') = Z_n'\omega(Z_n'). \tag{2.9}$$

On the other hand, in each involutory semigroup $\mathcal{T} = \langle T, \cdot, {}^* \rangle$ which satisfies $x = x\omega(x)x$ and (2.9), the values of the words Z_n' and Z_n under every evaluation of $x_1, \ldots, x_n$ necessarily divide each other on the right. This implies that such $\mathcal{T}$ satisfies the implication

$$xZ_n = yZ_n \rightarrow xZ_n' = yZ_n'. \tag{2.10}$$

Indeed, suppose that under some evaluation φ of the letters $x, y, x_1, \ldots, x_n$ in $\mathcal{T}$, the words xZ_n and yZ_n happen to take the same value, that is, $\varphi(xZ_n) = \varphi(yZ_n)$. Since $\varphi(Z_n) \mathscr{R}$ $\varphi(Z_n')$, there exists $t \in T$ such that $\varphi(Z_n)t = \varphi(Z_n')$. Hence

$$\varphi(xZ_n') = \varphi(xZ_n)t = \varphi(yZ_n)t = \varphi(yZ_n'),$$

that is, the words xZ_n' and yZ_n' also take a common value under φ.

Lemma 8 in [37] shows that a finitely generated semigroup satisfying (2.10) is finite if and only if it satisfies the identity

$$x^k = x^{k+\ell} \tag{2.11}$$

for some $k, \ell \geq 1$ and has only locally finite subgroups. We note that an involutory semigroup $\mathcal{T} = \langle T, \cdot, {}^* \rangle$ is finitely generated if and only if so is the semigroup $\langle T, \cdot \rangle$. An identity of the form (2.11) definitely holds in $\mathcal{S}$ since $\mathcal{S}$ is finite. Hence it suffices to find a finite number of identities which hold in $\mathcal{S}$ and which force each (involutory) semigroup to have only locally finite subgroups.

We can proceed as at the end of [37]. Let $\mathcal{G}$ be the direct product of all maximal subgroups of $\mathcal{S}$. By the Oates-Powell theorem [41], see also [40, §5.2], the locally finite variety $\mathsf{var}\,\mathcal{G}$ generated by the finite group $\mathcal{G}$ can be defined by a single identity $v(x_1, \ldots, x_m) = 1$. The left hand side v of this identity can be assumed to contain no occurrence of the inversion ${}^{-1}$, that is, we may assume that v is a plain semigroup word in the letters $x_1, \ldots, x_m$. Now let $\mathcal{F} = \mathcal{F}(x_1, \ldots, x_m)$ be the m-generated relatively free semigroup in the (locally finite) semigroup variety generated by the semigroup $\langle S, \cdot \rangle$. The semigroup $\mathcal{F}$ is finite and hence has a least ideal; this ideal is known to be a union of (isomorphic) subgroups (Sushkevich's theorem, see [6]). Let $\mathcal{H}$ be any of these subgroups and let e be the identity element of $\mathcal{H}$. We denote by $u(x_1, \ldots, x_m)$ a word whose value in $\mathcal{F}$ is e. Since $e^2 = e$, we see that $\mathcal{F}$ (and therefore $\mathcal{S}$) satisfies the identity

$$u = u^2. \tag{2.12}$$

For every element $g \in \mathcal{F}$, the product ege belongs to $\mathcal{H}$. As observed in [28], the group $\mathcal{H}$ belongs to the variety $\mathsf{var}\,\mathcal{G}$. Consequently, $\mathcal{F}$ (and therefore $\mathcal{S}$) satisfies the identity

$$v(ux_1u, \ldots, ux_mu) = u. \tag{2.13}$$

Note that both sides of (2.13) are plain semigroup words in the letters $x_1, \ldots, x_m$.

Now consider the variety $\mathbf{V}$ of involutory semigroups defined by the identity $x = x\omega(x)x$, an identity of the form (2.11) holding in $\mathcal{S}$, and the identities (2.9), (2.12), and (2.13). Since by the construction $\mathcal{S}$ satisfies all the listed identities, $\mathcal{S}$ is a member of $\mathbf{V}$. Let $\mathcal{T} = \langle T, \cdot, {}^* \rangle$ be any finitely generated member of $\mathbf{V}$; then, as already mentioned, the semigroup $\langle T, \cdot \rangle$ is also finitely generated. The first and the third identity ensure that the

semigroup $\langle T, \cdot \rangle$ satisfies the implication (2.10), and therefore, it is finite provided that all its subgroups are locally finite. Each group that satisfies the identities (2.12) and (2.13) satisfies the identity $v(x_1, \ldots, x_m) = 1$, whence this group belongs to var $\mathcal{G}$ and so is locally finite. Altogether, $\mathcal{T}$ is finite. Thus, $\mathbf{V}$ is locally finite and finitely based, and the proposition is proved. $\qquad\square$

Proposition 2.9 implies in particular that no finite regular $*$-semigroup can be inherently nonfinitely based as one can use x^* in the role of the term $\omega(x)$. In particular, the unary semigroup $\langle B_2^1, \cdot, {}^T \rangle$, where the unary operation is the usual matrix transposition, is not inherently nonfinitely based (this fact was first discovered by Sapir, see [50]), even though it is not finitely based [29].

3 Applications

3.1 Matrix semigroups with Moore-Penrose inverse

Certainly, the most common unary operation for matrices is transposition. However, it is convenient for us to start with analyzing matrix semigroups with Moore-Penrose inverse because this analysis will help us in considering semigroups with transposition.

We first recall the notion of Moore-Penrose inverse. This has been discovered by Moore [39] and independently by Penrose [42] for complex matrices, but has turned out to be a fruitful concept in a more general setting—see [4] for a comprehensive treatment.

The following results were obtained by Drazin [12].

Proposition 3.1. [12, Proposition 1] *Let $\mathcal{S}$ be an involutory semigroup. Then, for any given $a \in \mathcal{S}$, the four equations*

$$axa = a, \; xax = x, \; (ax)^* = ax, \; (xa)^* = xa \tag{3.1}$$

have at most one common solution $x \in \mathcal{S}$.

For an element a of an involutory semigroup $\mathcal{S}$, we denote by $a^\dagger$ the unique common solution x of the equations (3.1), provided it exists, and call $a^\dagger$ the *Moore-Penrose inverse* of a.

Recall that an element $a \in \mathcal{S}$ is said to be *regular*, if there is an $x \in \mathcal{S}$ such that $axa = a$. Concerning existence of the Moore-Penrose inverse, we have the following

Proposition 3.2. [12, Proposition 2] *Let $\mathcal{S}$ be an involutory semigroup satisfying the implication*

$$x^*x = x^*y = y^*x = y^*y \to x = y. \tag{3.2}$$

Then for an arbitrary $a \in \mathcal{S}$, the Moore-Penrose inverse $a^\dagger$ exists if and only if a^*a and aa^* are regular elements.

Let $\langle R, +, \cdot \rangle$ be a ring. An *involution of the ring* is an involution $x \mapsto x^*$ of the semigroup $\langle R, \cdot \rangle$ satisfying in addition the identity $(x+y)^* = x^*+y^*$. For ring involutions, the implication (3.2) is easily seen to be equivalent to

$$x^*x = 0 \to x = 0. \tag{3.3}$$

Now suppose that $\mathcal{K} = \langle K, +, \cdot \rangle$ is a field that admits an involution $x \mapsto \overline{x}$. Then the matrix ring $\mathrm{M}_n(\mathcal{K})$ has an involution that naturally arises from the involution of $\mathcal{K}$, namely $(a_{ij}) \mapsto (a_{ij})^* := (\overline{a_{ij}})^T$. This involution of $\mathrm{M}_n(\mathcal{K})$ in general does not satisfy the implication (3.3). However, it does satisfy (3.3) if and only if the equation

$$x_1\overline{x_1} + x_2\overline{x_2} + \cdots + x_n\overline{x_n} = 0 \tag{3.4}$$

admits only the trivial solution $(x_1, \ldots, x_n) = (0, \ldots, 0)$ in K^n. Since all elements of $\mathrm{M}_n(\mathcal{K})$ are regular, this means that the Moore-Penrose inverse exists — subject to the involution $(a_{ij}) \mapsto (a_{ij})^* = (\overline{a_{ij}})^T$ — whenever (3.4) admits only the trivial solution. (The classical Moore-Penrose inverse is thereby obtained by putting $\mathcal{K} = \mathbb{C}$, the field of complex numbers, endowed with the usual complex conjugation $z \mapsto \overline{z}$.) On the other hand, it is easy to see that the condition that (3.4) has only the trivial solution is necessary: if $(a_1, \ldots, a_n)$ were a non-trivial solution to (3.4), then the matrix formed by n identical rows $(a_1, \ldots, a_n)$ would have no Moore-Penrose inverse.

The proof of the main result of this subsection requires an explicit calculation of the Moore-Penrose inverses of certain rank 1 matrices. Thus, we present a simple method for such a calculation. For a row vector $a = (a_1, \ldots, a_n) \in K^n$, where $\mathcal{K} = \langle K, +, \cdot \rangle$ is a field with an involution $x \mapsto \overline{x}$, let a^* denote the column vector $(\overline{a_1}, \ldots, \overline{a_n})^T$. It is easy to see that any $n \times n$-matrix A of rank 1 over $\mathcal{K}$ can be represented as $A = b^*c$ for some non-zero row vectors $b, c \in K^n$. Provided that (3.4) admits only the trivial solution in K^n, one gets $A^\dagger$ as follows:

$$A^\dagger = c^*(cc^*)^{-1}(bb^*)^{-1}b. \tag{3.5}$$

Here bb^* and cc^* are non-zero elements of $\mathcal{K}$ whence their inverses in $\mathcal{K}$ exist. In order to justify (3.5), it suffices to check that the right hand side of (3.5) satisfies the simultaneous equations (3.1) with the matrix A in the role of a, and this is straightforward. Note that formula (3.5) immediately shows that $A^\dagger$ is a scalar multiple of $A^* = c^*b$, namely

$$A^\dagger = \frac{1}{cc^* \cdot bb^*} A^*. \tag{3.6}$$

So, we can formulate one of the highlights of the section — a result that reveals an unexpected feature of a rather classical and well studied object.

Theorem 3.3. *Let $\mathcal{K} = \langle K, +, \cdot \rangle$ be a field having an involution $x \mapsto \overline{x}$ for which the equation $x\overline{x} + y\overline{y} = 0$ has only the trivial solution $(x, y) = (0, 0)$ in K^2. Then the unary semigroup $\langle \mathrm{M}_2(\mathcal{K}), \cdot, ^\dagger \rangle$ of all 2×2-matrices over $\mathcal{K}$ endowed with Moore-Penrose inversion † — subject to the involution $(a_{ij}) \mapsto (a_{ij})^* = (\overline{a_{ij}})^T$ — has no finite basis of identities.*

Proof. Set $\mathcal{S} = \langle \mathrm{M}_2(\mathcal{K}), \cdot, ^\dagger \rangle$. By Theorem 2.2 and Lemma 2.1 it is sufficient to show that

1) $\mathcal{K}_3 \in \mathsf{var}\,\mathcal{S}$,

2) there exists a group $\mathcal{G} \in \mathsf{var}\,\mathcal{S}$ such that $\mathcal{G} \notin \mathsf{var}\,\mathrm{H}(\mathcal{S})$.

In order to prove 1), consider the following sets of rank 1 matrices in $M_2(\mathcal{K})$:

$$
H_{11} = \left\{ \begin{pmatrix} x & x \\ x & x \end{pmatrix} \right\}, \quad
H_{12} = \left\{ \begin{pmatrix} x & 0 \\ x & 0 \end{pmatrix} \right\}, \quad
H_{13} = \left\{ \begin{pmatrix} 0 & x \\ 0 & x \end{pmatrix} \right\},
$$

$$
H_{21} = \left\{ \begin{pmatrix} x & x \\ 0 & 0 \end{pmatrix} \right\}, \quad
H_{22} = \left\{ \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} \right\}, \quad
H_{23} = \left\{ \begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix} \right\}, \quad (3.7)
$$

$$
H_{31} = \left\{ \begin{pmatrix} 0 & 0 \\ x & x \end{pmatrix} \right\}, \quad
H_{32} = \left\{ \begin{pmatrix} 0 & 0 \\ x & 0 \end{pmatrix} \right\}, \quad
H_{33} = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & x \end{pmatrix} \right\},
$$

where in each case x runs over $K \setminus \{0\}$. Observe that $\mathcal{K}$ cannot be of characteristic 2, since the equation $x\overline{x} + y\overline{y} = 0$ has only the trivial solution in K^2. Taking this into account, a straightforward calculation shows that

$$
H_{ij} \cdot H_{k\ell} = \begin{cases} H_{i\ell} & \text{if } (j,k) \neq (2,3), (3,2), \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)
$$

Hence the set

$$
T = \bigcup_{1 \leq i,j \leq 3} H_{ij} \cup \{0\}
$$

is closed under multiplication so that this set forms a subsemigroup $\mathcal{T}$ of $\mathcal{S}$ and the partition $\mathcal{H}$ of T into the classes H_{ij} and $\{0\}$ is a congruence on $\mathcal{T}$. Equation (3.6) shows that

$$
H_{ij}^{\dagger} = H_{ji}. \quad (3.9)
$$

We see that T is closed under Moore-Penrose inversion and $\mathcal{H}$ respects †, thus is a congruence on the unary semigroup $\mathcal{T}' = \langle T, \cdot, {}^{\dagger} \rangle$. Now comparing (3.8) and (3.9) with the multiplication and inversion rules in $\mathcal{K}_3$ (see (1.1)), we conclude that $\mathcal{T}'/\mathcal{H}$ and $\mathcal{K}_3$ are isomorphic as unary semigroups. Hence $\mathcal{K}_3$ is in var $\mathcal{S}$.

For 2) we merely let $\mathrm{GL}_2(\mathcal{K})$, the group of all invertible 2×2-matrices over $\mathcal{K}$, play the role of $\mathcal{G}$. Since Moore-Penrose inversion on $\mathrm{GL}_2(\mathcal{K})$ coincides with usual matrix inversion, we observe that $\mathrm{GL}_2(\mathcal{K})$ is a unary subsemigroup of $\mathcal{S}$. Moreover, since $AA^{\dagger}$ is the identity matrix for every invertible matrix A, we conclude that, with the exception of the identity matrix, the Hermitian subsemigroup $\mathrm{H}(\mathcal{S})$ contains only matrices of rank 1, that is, $\mathrm{H}(\mathcal{S}) \subseteq \mathrm{L}_2^1(\mathcal{K})$, the set of all matrices of rank at most 1 with the identity matrix adjoined. By Corollary 1.2 the semigroup $\langle \mathrm{L}_2^1(\mathcal{K}), \cdot \rangle$ satisfies the identity $x^2 y x = x y x^2$. Consequently, each group in var $\mathrm{H}(\mathcal{S})$ is abelian, while the group $\mathrm{GL}_2(\mathcal{K})$ is non-abelian. Thus, $\mathrm{GL}_2(\mathcal{K})$ is contained in var $\mathcal{S}$ but is not contained in var $\mathrm{H}(\mathcal{S})$, as required. $\qquad \square$

Remark 3.1. Apart from any subfield of $\mathbb{C}$ closed under complex conjugation, Theorem 3.3 applies, for instance, to finite fields $\mathcal{K} = \langle K, +, \cdot \rangle$ for which $|K| \equiv 3 \pmod 4$, endowed with the trivial involution $x \mapsto \overline{x} = x$; the latter follows from the fact that the equation $x^2 + 1 = 0$ admits no solution in $\mathcal{K}$ if and only if $|K| \equiv 3 \pmod 4$ (cf. [33, Theorem 3.75]). Moreover, by slightly changing the arguments one can show an analogous result for $\mathcal{K}$ being any skew-field of quaternions closed under conjugation.

The reader may ask whether or not the restriction on the size of matrices is essential in Theorem 3.3. For some fields, it definitely is. For instance, for finite fields with the

trivial involution $x \mapsto \overline{x} = x$, no extension of Theorem 3.3 to $n \times n$-matrices with $n > 2$ is possible simply because the Moore-Penrose inverse is only a partial operation in this case. Indeed, it is a well known corollary of the Chevalley-Warning theorem (see, e.g., [52, Corollary 2 in §1.2]) that the equation $x_1^2 + \cdots + x_n^2 = 0$ (that is (3.4) with the trivial involution) admits a non-trivial solution in any finite field whenever $n > 2$.

The situation is somewhat more complicated for subfields of $\mathbb{C}$. Theorem 2.2 does not apply here because of the following obstacle. It is well known (see, for example, [35, p. 101]) that the two matrices

$$\zeta = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \text{ and } \eta = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \tag{3.10}$$

generate a free subgroup of $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot, ^{-1} \rangle$. On the other hand, it is easy to verify that for any subfield $\mathcal{K}$ of $\mathbb{C}$ closed under complex conjugation, the mapping $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot, ^{-1} \rangle \rightarrow \langle \mathrm{M}_3(\mathcal{K}), \cdot, ^\dagger \rangle$ defined by $A \mapsto \begin{pmatrix} A & 0 \\ 0\ 0 & 0 \end{pmatrix}$ is an embedding of unary semigroups. Thus, for $n > 2$, the unary semigroup $\langle \mathrm{Sing}_n(\mathcal{K}), \cdot, ^\dagger \rangle$ of singular $n \times n$-matrices contains a free non-abelian group, whence every group belongs to the unary semigroup variety generated by $\langle \mathrm{Sing}_n(\mathcal{K}), \cdot, ^\dagger \rangle$. Now we observe that $\mathrm{Sing}_n(\mathcal{K})$ is contained in the Hermitian subsemigroup of $\langle \mathrm{M}_n(\mathcal{K}), \cdot, ^\dagger \rangle$. Indeed, it was proved in [16] (see also [2] for a recent elementary proof) that the semigroup $\langle \mathrm{Sing}_n(\mathcal{K}), \cdot \rangle$ is generated by idempotent matrices. For an arbitrary idempotent matrix $A \in \mathrm{M}_n(\mathcal{K})$, let

$$N(A) = \{x \in K^n \mid xA = 0\} \text{ and } F(A) = \{x \in K^n \mid xA = x\}$$

be the null-space and the fixed-point-space of A, respectively. Now consider two matrices of orthogonal projectors: P_1, the matrix of the orthogonal projector to the space $F(A)$, and P_2, the matrix of the orthogonal projector to the space $N(A)^\perp$. As any orthogonal projector matrix P satisfies $P = P^2 = P^\dagger$, both $P_1 = P_1 P_1^\dagger$ and $P_2 = P_2 P_2^\dagger$ belong to the Hermitian subsemigroup $\mathrm{H}(\mathrm{M}_n(\mathcal{K}))$, but then A also belongs to $\mathrm{H}(\mathrm{M}_n(\mathcal{K}))$ since $A = (P_1 P_2)^\dagger$, see [38, Exercise 5.15.9a]. Thus, $\mathrm{Sing}_n(\mathcal{K}) \subseteq \mathrm{H}(\mathrm{M}_n(\mathcal{K}))$, whence no group $\mathcal{G}$ can satisfy the condition of Theorem 2.2.

However, the fact that Theorem 2.2 cannot be applied to, say, the unary semigroup $\langle \mathrm{M}_3(\mathbb{C}), \cdot, ^\dagger \rangle$ does not yet mean that the identities of this semigroup are finitely based. We thus have the following open question.

Problem 3.1. Is the unary semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, ^\dagger \rangle$ not finitely based for each subfield $\mathcal{K}$ of $\mathbb{C}$ closed under complex conjugation and for all $n > 2$?

It is known that $\langle \mathrm{M}_n(\mathcal{K}), \cdot, ^\dagger \rangle$ satisfies rather involved identities (see [7] for an example), so the conjecture that these identities admit no finite basis looks quite natural. In connection with Problem 3.1, we also observe that the proofs of Theorem 3.3 and Corollary 1.2 readily yield the following:

Remark 3.2. For each conjugation-closed subfield $\mathcal{K}$ of $\mathbb{C}$ and for all $n > 2$, the unary semigroup $\langle \mathrm{L}_n(\mathcal{K}) \cup \mathcal{G}, \cdot, ^\dagger \rangle$ consisting of all matrices of rank at most 1 and all matrices from some non-abelian subgroup $\mathcal{G}$ of $\mathrm{GL}_n(\mathcal{K})$ has no finite identity basis.

Another natural related structure is the semigroup $M_n(\mathcal{K})$ endowed with **both** unary operations $\dagger$ and $*$. Recall that the Moore-Penrose inverse is in fact defined in terms of identities involving both these operations—namely, Proposition 3.2 implies that the algebraic structure $\langle M_n(\mathcal{K}), \cdot, \dagger, * \rangle$ satisfies the identities

$$xx^\dagger x = x, \; x^\dagger x x^\dagger = x^\dagger, \; (xx^\dagger)^* = xx^\dagger, \; (x^\dagger x)^* = x^\dagger x \tag{3.11}$$

and these identities uniquely determine the operation $A \mapsto A^\dagger$. This might have provoked one to conjecture that the identities (3.11) together with the identities $(xy)^* = y^*x^*$ and $(x^*)^* = x$ form a basis for $\mathrm{Eq}\langle M_n(\mathcal{K}), \cdot, \dagger, * \rangle$. However our techniques show that this is not the case at least for $n = 2$.

Theorem 3.4. *Let $\mathcal{K}$ be a field as in Theorem 3.3; then $\langle M_2(\mathcal{K}), \cdot, \dagger, * \rangle$ is not finitely based as an algebraic structure of type $(2, 1, 1)$.*

Proof. The characteristic of $\mathcal{K}$ is not 2 whence the group

$$\mathcal{G} = \{ A \in \mathrm{GL}_2(\mathcal{K}) \mid A^\dagger = A^* \}$$

is non-abelian. Indeed, on the prime subfield of $\mathcal{K}$, the involution $x \mapsto \overline{x}$ is the identity automorphism; so, for matrices over the prime subfield, conjugation $*$ coincides with transposition, and thus, for example, $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right)$ and $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix} \right)$ are two non-commuting members of $\mathcal{G}$. Set $\mathcal{A} = \langle M_2(\mathcal{K}), \cdot, \dagger, * \rangle$; the algebraic structure $\langle G, \cdot, \dagger, * \rangle$, that is, the group $\mathcal{G}$ with inversion taken twice as unary operation, belongs to $\mathrm{var}\,\mathcal{A}$. Now, as in the proof of Theorem 3.3, consider the set

$$T = \bigcup_{1 \leq i,j \leq 3} H_{ij} \cup \{0\}$$

where H_{ij} are defined via (3.7). Obviously, $H_{ij}^* = H_{ji}$ whence $\mathcal{T} = \langle T, \cdot, \dagger, * \rangle$ is a substructure of $\mathcal{A}$ and the partition $\mathscr{H}$ of T into the classes H_{ij} and $\{0\}$ is a congruence on this substructure. The quotient $\mathcal{T}/\mathscr{H}$ is then isomorphic to the semigroup $\mathcal{K}_3$ endowed twice with its unary operation. We conclude that $\mathcal{K}_3$ treated this way also belongs to $\mathrm{var}\,\mathcal{A}$. By Corollary 1.2 the identity $x^2yx = xyx^2$ holds in $\mathrm{H}(\mathcal{A})$ (by which we mean the substructure of $\mathcal{A}$ generated by all elements of the form $AA^\dagger$). Now construct the semigroups $\mathcal{T}_k$ (by use of the identity $x^2yx = xyx^2$) as in Step 2 in the proof of Theorem 2.2 and endow each of them twice with its unary operation. The arguments in Steps 3 and 4 in the proof then show that $\mathcal{T}_k$ does not belong to $\mathrm{var}\,\mathcal{A}$ while each k-generated substructure of $\mathcal{T}_k$ does belong to $\mathrm{var}\,\mathcal{A}$. Thus, $\mathcal{T}_k$ can play the role of critical structures for $\mathrm{var}\,\mathcal{A}$ whence the desired conclusion follows by the reasoning as in Step 1 in the proof of Theorem 2.2. $\square$

Also in this setting, our result gives rise to a natural question.

Problem 3.2. Is the algebraic structure $\langle M_n(\mathcal{K}), \cdot, \dagger, * \rangle$ of type $(2, 1, 1)$ not finitely based for each subfield $\mathcal{K}$ of $\mathbb{C}$ closed under complex conjugation and for all $n > 2$?

Here an observation similar to Remark 3.2 can be stated: for each conjugation-closed subfield $\mathcal{K} \subseteq \mathbb{C}$ and for all $n > 2$, the algebraic structure $\langle \mathrm{L}_n(\mathcal{K}) \cup \mathrm{GL}_n(\mathcal{K}), \cdot, \dagger, * \rangle$ consisting of all matrices of rank at most 1 and all invertible matrices has no finite identity basis.

3.2 Matrix semigroups with transposition over infinite fields

Here we show that the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ is finitely based for any infinite field $\mathcal{K}$. More precisely, we verify that all identities holding in $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ follow from the associativity and the involution laws $(xy)^T = y^T x^T$, $(x^T)^T = x$. This is an involutory analogue of a result from [22] mentioned in the introduction; the proof given there does not immediately show the intended analogue, but the ideas below are inspired by the arguments in [22].

Let us start with an auxiliary construction and consider first, for an arbitrary field $\mathcal{K}$, the set $\mathrm{M}_2(\mathcal{K}[x])$ of all 2×2-matrices over the polynomial ring $\mathcal{K}[x]$. A matrix

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

with p_{ij} non-zero members of $\mathcal{K}[x]$ is said to be *ascending* if

$$\deg(p_{11}) < \min\{\deg(p_{12}), \deg(p_{21})\}, \ \max\{\deg(p_{12}), \deg(p_{21})\} < \deg(p_{22})$$

and *descending* if

$$\deg(p_{11}) > \max\{\deg(p_{12}), \deg(p_{21})\}, \ \min\{\deg(p_{12}), \deg(p_{21})\} > \deg(p_{22}.)$$

Let Asc and Desc stand for the sets of all ascending and descending matrices respectively. It is straightforward to see that Asc and Desc are disjoint and both of them are closed under multiplication and transposition. Set

$$A = \begin{pmatrix} 1 & 0 \\ x^2 & x \end{pmatrix};$$

by induction one obtains that

$$A^n = \begin{pmatrix} 1 & 0 \\ p_{n+1} & p_n \end{pmatrix}$$

for polynomials p_n and p_{n+1} with $\deg(p_{n+1}) = n + 1$ and $\deg(p_n) = n$. Further, set $B = A^T$; then for any $n, m \geq 1$ one gets

$$A^n B^m = \begin{pmatrix} 1 & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$$

where $\deg(s_{12}) = m + 1$, $\deg(s_{21}) = n + 1$ and $\deg(s_{22}) = m + n + 2$, while

$$B^m A^n = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$$

where $\deg(r_{11}) = m + n + 2$, $\deg(r_{12}) = \deg(r_{21}) = m + n + 1$ and $\deg(r_{22}) = m + n$. In particular, $A^n B^m \in \text{Asc}$ while $B^m A^n \in \text{Desc}$.

Lemma 3.5. *For all $m, n \geq 1$, the sets $\text{Asc}\, A^n$ and $\text{Desc}\, B^m$ are disjoint.*

Proof. Let a, b, c, d, p, q be non-zero polynomials with

$$\deg(a) < \min\{\deg(b), \deg(c)\}, \ \max\{\deg(b), \deg(c)\} < \deg(d) \ \text{ and } \ \deg(p) < \deg(q).$$

Then $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a typical matrix in Asc, $\left(\begin{smallmatrix} d & b \\ c & a \end{smallmatrix}\right)$ is such in Desc, A^n is of the form $\left(\begin{smallmatrix} 1 & 0 \\ q & p \end{smallmatrix}\right)$ and B^m is of the form $\left(\begin{smallmatrix} 1 & q \\ 0 & p \end{smallmatrix}\right)$. Now

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ q & p \end{pmatrix} = \begin{pmatrix} a + bq & bp \\ c + dq & dp \end{pmatrix}.$$

Since $\deg(a) < \deg(b)$ and $\deg(p) < \deg(q)$ we obtain

$$\deg(a + bq) = \deg(bq) > \deg(bp).$$

In particular, for any $C = (c_{ij}) \in \operatorname{Asc} A^n$ we get

$$\deg(c_{11}) > \deg(c_{12}) \tag{3.12}$$

On the other hand,

$$\begin{pmatrix} d & b \\ c & a \end{pmatrix} \begin{pmatrix} 1 & q \\ 0 & p \end{pmatrix} = \begin{pmatrix} d & dq + bp \\ c & cq + ap \end{pmatrix}.$$

Here we have that

$$\deg(d) < \deg(dq) = \deg(dq + bp).$$

Again, this shows that for any $D = (d_{ij}) \in \operatorname{Desc} B^m$,

$$\deg(d_{11}) < \deg(d_{12}). \tag{3.13}$$

Conditions (3.12) and (3.13) immediately imply that $\operatorname{Asc} A^n \cap \operatorname{Desc} B^m = \varnothing$. $\qquad\square$

We are able to draw our first conclusion.

Proposition 3.6. *Let* $u = u(a, b), v = v(a, b) \in \{a, b\}^+$ *be two distinct words,* $A = \left(\begin{smallmatrix} 1 & 0 \\ x^2 & x \end{smallmatrix}\right) \in \mathrm{M}_2(\mathcal{K}[x])$ *and* $B = A^T$. *Then* $u(A, B) \neq v(A, B)$.

Proof. Suppose that the equality $u(A, B) = v(A, B)$ holds. The matrices A and B are invertible as matrices over the field of rational functions over $\mathcal{K}$ whence we can cancel the longest common prefix and the longest common suffix of the words u and v obtaining new words u' and v' that still fulfil $u'(A, B) = v'(A, B)$. Observe that none of the words u' and v' are empty because otherwise A and B would be invertible over the polynomial ring $\mathcal{K}[x]$ which is not true. Thus, we may assume that u and v start and end with different symbols and this means that the words are of either of the following two forms:

$$u = a^{n_1} b^{m_1} \cdots a^{n_s} b^{m_s} \ \text{ and } \ v = b^{\ell_1} a^{k_1} \cdots b^{\ell_t} a^{k_t} \tag{3.14}$$

$$u = a^{n_1} b^{m_1} \cdots a^{n_{s-1}} b^{m_{s-1}} a^{n_s} \ \text{ and } \ v = b^{\ell_1} a^{k_1} \cdots b^{\ell_{t-1}} a^{k_{t-1}} b^{k_t} \tag{3.15}$$

with $s, t \geq 1$ and all $n_i, m_i, k_i, \ell_i \geq 1$. In case (3.14), $u(A, B) \in \operatorname{Asc}$ and $v(A, B) \in \operatorname{Desc}$ while in case (3.15) $u(A, B) \in \operatorname{Asc} A^{n_s}$ and $v(A, B) \in \operatorname{Desc} B^{k_t}$. In any case, using Lemma 3.5 we obtain $u(A, B) \neq v(A, B)$, a contradiction. $\qquad\square$

Theorem 3.7. *For each infinite field $\mathcal{K}$, all the identities of the involutory semigroup $\langle M_n(\mathcal{K}), \cdot, {}^T \rangle$ follow from the laws $(xy)z = x(yz)$, $(xy)^T = y^T x^T$, $(x^T)^T = x$.*

Proof. If we call identities that follow from the associativity and the involution laws *trivial*, our claim amounts to saying that $\langle M_n(\mathcal{K}), \cdot, {}^T \rangle$ does not satisfy any non-trivial involutory identity. We show this already holds for the involutory semigroup $\mathcal{GL}_2(\mathcal{K}) = \langle GL_2(\mathcal{K}), \cdot, {}^T \rangle$ that obviously embeds into $\langle M_n(\mathcal{K}), \cdot, {}^T \rangle$ for each n.

It is known (and easy to verify) that the free involutory semigroup on one generator z, say, contains as a unary subsemigroup a free involutory semigroup on countably many generators, namely, $\mathcal{FI}(Z)$ where

$$Z = \{ zz^*z, \ z(z^*)^2z, \ \ldots, \ z(z^*)^nz, \ \ldots \}.$$

Therefore we only need verifying that $\mathcal{GL}_2(\mathcal{K})$ satisfies no non-trivial involutory identity in one letter z. Such an identity can be written as $u(z, z^T) = v(z, z^T)$ with u and v being two distinct words. Proposition 3.6 implies that $u(A, B) \neq v(A, B)$ in $M_2(\mathcal{K}[x])$. Setting

$$u(A, B) = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \ \text{ and } \ v(A, B) = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix},$$

we see that for some indices i and j the polynomials u_{ij} and v_{ij} are distinct whence $u_{ij} - v_{ij}$ is a non-zero polynomial. Now take any $\lambda \in \mathcal{K}$ and set $z(\lambda) = \left(\begin{smallmatrix} 1 & 0 \\ \lambda^2 & \lambda \end{smallmatrix} \right)$. If the equality

$$u(z(\lambda), z(\lambda)^T) = v(z(\lambda), z(\lambda)^T) \tag{3.16}$$

holds then λ must be a root of the non-zero polynomial $u_{ij} - v_{ij}$ whence (3.16) can hold only for finitely many elements λ of $\mathcal{K}$. Since $\mathcal{K}$ is infinite, equality (3.16) fails for all but finitely many λ, and so, in particular, the identity $u(z, z^T) = v(z, z^T)$ fails in $\mathcal{GL}_2(\mathcal{K})$. $\square$

Similarly, it can be shown that the involutory semigroup $\langle M_n(R), \cdot, {}^* \rangle$ is finitely based for each subring $R \subseteq \mathbb{C}$ closed under complex conjugation—here * stands for the complex-conjugate transposition $(a_{ij})^* = (\overline{a_{ij}})^T$. Indeed, we have already mentioned that the two matrices ζ and η in (3.10) generate a free subgroup of $\langle SL_2(\mathbb{Z}), \cdot, {}^{-1} \rangle$ and hence a free subsemigroup of $\langle SL_2(\mathbb{Z}), \cdot \rangle$. But $\eta = \zeta^*$ whence the involutory subsemigroup in $\langle SL_2(\mathbb{Z}), \cdot, {}^* \rangle$ generated by ζ is isomorphic to the free involutory semigroup $\mathcal{FI}(\{\zeta\})$.

We note that Theorem 2.2 and Corollary 1.2 prove the non-existence of a finite identity bases for the unary subsemigroup of $\langle M_n(\mathbb{C}), \cdot, {}^* \rangle$ [respectively $\langle M_n(\mathbb{R}), \cdot, {}^T \rangle$] that consists of all matrices of rank at most 1 together with all unitary [respectively all orthogonal] matrices.

3.3 Matrix semigroups with transposition over finite fields

Now we demonstrate that the case of finite fields is completely opposite to the one of infinite fields with respect to the finite basis problem for matrix semigroups with transposition. We start with considering 2×2-matrices. Here Theorem 2.2 solves the finite basis problem in the negative for the involutory semigroup $\langle M_2(\mathcal{K}), \cdot, {}^T \rangle$ for each finite field $\mathcal{K}$

except $\mathcal{K} = \mathbb{F}_2$, the 2-element field. (The exception is due to the fact that the involutory semigroup $\langle \mathrm{M}_2(\mathbb{F}_2), \cdot, {}^T \rangle$ satisfies the identity

$$(xx^*)^3(yy^*)^3 = (yy^*)^3(xx^*)^3$$

which does not hold in $\mathcal{K}_3$; consequently, $\mathcal{K}_3$ is not in $\mathsf{var}\langle \mathrm{M}_2(\mathbb{F}_2), \cdot, {}^T \rangle$ and Theorem 2.2 does not apply here.) In the following theorem, we shall demonstrate the application of Theorem 2.2 only in the case when $\mathcal{K}$ has odd characteristic. With some additional effort we could include also the case when the characteristic of $\mathcal{K}$ is 2 and $|K| \geq 4$. We shall omit this since that case will be covered by a different kind of proof later.

Theorem 3.8. *For each finite field $\mathcal{K}$ of odd characteristic, the involutory semigroup $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ has no finite identity basis.*

Proof. Let $\mathcal{S} = \langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$. As in the proof of Theorem 3.3 one shows that $\mathcal{K}_3$ is in $\mathsf{var}\,\mathcal{S}$. Furthermore, let d be the exponent of the group $\mathrm{GL}_2(\mathcal{K})$. By Corollary 1.2, each group in $\mathrm{P}_d(\mathsf{var}\,\mathcal{S}) = \mathsf{var}\,\mathrm{P}_d(\mathcal{S})$ satisfies the identity $x^2yx = xyx^2$ and therefore is abelian. On the other hand, as in the proof of Theorem 3.4, the group $\mathcal{G} = \{A \in \mathrm{GL}_2(\mathcal{K}) \mid A^T = A^{-1}\}$ is in $\mathsf{var}\,\mathcal{S}$ but is non-abelian. Thus, Theorem 2.2 applies. $\qquad\square$

The next theorem contains the even characteristic case and proves, in fact, a stronger assertion.

Theorem 3.9. *For each finite field $\mathcal{K} = \langle K, +, \cdot \rangle$ with $|K| \not\equiv 3 \pmod 4$, the involutory semigroup $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ is inherently nonfinitely based.*

Proof. As mentioned in Remark 3.1, there exists $x \in K$ for which $1 + x^2 = 0$. Now consider the following matrices:

$$H_{11} = \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}, \; H_{12} = \begin{pmatrix} 1 & 0 \\ x & 0 \end{pmatrix}, \; H_{21} = \begin{pmatrix} 1 & x \\ 0 & 0 \end{pmatrix}, \; H_{22} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; O = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then the set $M = \{H_{11}, H_{12}, H_{21}, H_{22}, I, O\}$ is closed under multiplication and transposition, hence $\mathcal{M} = \langle M, \cdot, {}^T \rangle$ is an involutory subsemigroup of $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$. The mapping $\mathcal{TA}_2^1 \to \mathcal{M}$ given by

$$(i, j) \mapsto H_{ij}, \; 0 \mapsto O, \; 1 \mapsto I$$

is an isomorphism of involutory semigroups. The result now follows from Corollary 2.8. $\qquad\square$

The case of matrix semigroups of size greater than 2 is similar.

Theorem 3.10. *For each finite field $\mathcal{K}$, the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ with $n \geq 3$ is inherently nonfinitely based.*

Proof. It follows from the Chevalley-Warning theorem [52, Corollary 2 in §1.2] that there exist $x, y \in K$ satisfying $1 + x^2 + y^2 = 0$. Now consider the following matrices:

$$H_{11} = \begin{pmatrix} 1 & x & y \\ x & x^2 & xy \\ y & xy & y^2 \end{pmatrix}, \quad H_{12} = \begin{pmatrix} 1 & 0 & 0 \\ x & 0 & 0 \\ y & 0 & 0 \end{pmatrix}, \quad H_{21} = \begin{pmatrix} 1 & x & y \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$H_{22} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad O = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Again, the set $M = \{H_{11}, H_{12}, H_{21}, H_{22}, I, O\}$ is closed under multiplication and transposition, and as in the previous proof, $\mathcal{M} = \langle M, \cdot, {}^T \rangle$ forms an involutory subsemigroup of $\langle \mathrm{M}_3(\mathcal{K}), \cdot, {}^T \rangle$ that is isomorphic with $\mathcal{TA}_2^1$. Hence $\langle \mathrm{M}_3(\mathcal{K}), \cdot, {}^T \rangle$ is inherently nonfinitely based. The assertion for $\mathrm{M}_n(\mathcal{K})$ for $n \geq 3$ now follows in an obvious way. $\qquad\square$

Remark 3.3. The statements of Theorems 3.8, 3.9, 3.10 remain valid if the unary operation $A \mapsto A^T$ is replaced with an operation of the form $A \mapsto A^{\sigma T}$ for any automorphism σ of $\mathcal{K}$, where $(a_{ij})^{\sigma T} := (a_{ij}^\sigma)^T$.

We are ready to prove the main result of this subsection. With the exception of the 'only if' part of item (2), this is a summary of Theorems 3.8, 3.9, and 3.10.

Theorem 3.11. *Let $\mathcal{K} = \langle K, +, \cdot \rangle$ be a finite field. Then*

1. *the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ is not finitely based;*

2. *the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ is inherently nonfinitely based if and only if either $n \geq 3$ or $n = 2$ and $|K| \not\equiv 3 \pmod 4$.*

Proof. The only assertion of this theorem not covered by our previous results is that $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ is **not** inherently nonfinitely based if $|K| \equiv 3 \pmod 4$. We employ Proposition 2.9 to prove this.

Recall that the condition $|K| \equiv 3 \pmod 4$ corresponds precisely to the case when each matrix A in $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ admits a Moore-Penrose inverse $A^\dagger$ (Remark 3.1). Let A be a matrix of rank 1; by (3.6) there exists a scalar $\alpha \in \mathcal{K} \setminus \{0\}$ such that $\alpha A^\dagger = A^T$. Let $r = |K| - 1$; then $\alpha^r = 1$. Since the multiplicative subgroup of $\mathcal{K}$ is a cyclic subgroup of $\mathrm{GL}_2(\mathcal{K})$, the number r divides the exponent d of $\mathrm{GL}_2(\mathcal{K})$ whence $\alpha^d = 1$. Consequently,

$$A(A^T A)^d = A(\alpha A^\dagger A)^d = \alpha^d A(A^\dagger A)^d = A.$$

If $A \in \mathrm{GL}_2(\mathcal{K})$, we also have $A = A(A^T A)^d$ because $(A^T A)^d$ is the identity matrix; clearly, the equality $A = A(A^T A)^d$ holds also for the case when A is the zero matrix. Summarizing, we conclude that the identity $x = x(x^T x)^d$ holds in the involutory semigroup $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$. Setting $\omega(x) := x^T (x x^T)^{d-1}$, we see that $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ satisfies the identity $x = x \omega(x) x$, as required by Proposition 2.9. $\qquad\square$

Remark 3.4. It is known [49, Corollary 6.2] that the matrix semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ is inherently nonfinitely based (as a plain semigroup) for every finite field $\mathcal{K}$. Thus, the involutory semigroups $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ over finite fields $\mathcal{K}$ such that $|K| \equiv 3 \pmod 4$ provide a natural series of unary semigroups whose equational properties essentially differ from the equational properties of their semigroup reducts.

3.4 Matrix semigroups with symplectic transpose

For a $2m \times 2m$-matrix

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with A, B, C, D being $m \times m$-matrices over any field $\mathcal{K}$, the *symplectic transpose* X^S is defined by

$$X^S = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix},$$

see, e.g., [45, (5.1.1)]. The symplectic transpose is an involution of $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot \rangle$ whose properties essentially differ from those of the usual transposition. In fact, every involution of the semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ that fixes the scalar matrices is in a certain sense similar to either the usual transposition or the symplectic transpose[4].

The definition of the symplectic transpose resembles that of the involution in the twisted Brandt monoid $\mathcal{TB}_2^1$ (defined in terms of 2×2-matrices) and leads to the following application.

Theorem 3.12. *The involutory semigroup $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$ is inherently nonfinitely based for each $m \geq 1$ and each finite field $\mathcal{K} = \langle K, +, \cdot \rangle$.*

Proof. Consider the following sets of $2m \times 2m$-matrices:

$$H_{11} = \left\{ \pm \begin{pmatrix} O_m & I_m \\ O_m & O_m \end{pmatrix} \right\}, \ H_{12} = \left\{ \pm \begin{pmatrix} I_m & O_m \\ O_m & O_m \end{pmatrix} \right\},$$

$$H_{21} = \left\{ \pm \begin{pmatrix} O_m & O_m \\ O_m & I_m \end{pmatrix} \right\}, \ H_{22} = \left\{ \pm \begin{pmatrix} O_m & O_m \\ I_m & O_m \end{pmatrix} \right\}$$

where for any positive integer k, we denote be I_k, respectively, O_k the identity, respectively, zero $k \times k$-matrix. Let

$$T = \bigcup_{1 \leq i,j \leq 2} H_{ij} \cup \{O_{2m}, I_{2m}\}.$$

The set T is closed under multiplication and symplectic transposition whence $\langle T, \cdot, {}^S \rangle$ forms an involutory subsemigroup of $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$. On the other hand, the mapping

$$H_{ij} \mapsto (i, j), \ I_{2m} \mapsto 1, \ O_{2m} \mapsto 0$$

is a homomorphism of $\langle T, \cdot, {}^S \rangle$ onto $\mathcal{TB}_2^1$. Altogether, the twisted Brandt monoid $\mathcal{TB}_2^1$ is a homomorphic image of an involutory subsemigroup of $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$. $\square$

---

[4] We do not want to formalize this phrase in general because its meaning actually depends on some intrinsic properties of the field $\mathcal{K}$. In the simplest case, when $\mathcal{K}$ is algebraically closed of characteristic $\neq 2$, the similarity takes the strongest form: every involutory semigroup of the form $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^* \rangle$ such that * fixes the scalar matrices is isomorphic to either $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ or $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^S \rangle$ (in the latter case n should be even). This is well known for involutions that respect the addition of matrices (see, e.g., [44, Corollary 14.2]) but it easily follows from a classic result by Khalezov [26] that every involution of the semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ automatically preserves addition.

Remark 3.5. It is easy to see that the involutory semigroup $\langle \mathrm{M}_m(\mathcal{K}), \cdot, {}^T \rangle$ embeds into $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$ via the mapping $A \mapsto \begin{pmatrix} A & O_m \\ O_m & A \end{pmatrix}$. Therefore, if $m > 1$ and $\mathcal{K}$ is an infinite field, then $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$ satisfies no non-trivial involutory identity and is finitely based (see Subsection 3.2). The 2×2-matrices over an infinite field fulfill non-trivial identities involving multiplication and the symplectic transpose, for instance, $xx^S y = yxx^S$ or $xx^S = x^S x$. In fact, we have verified that these two identities together with the associativity and the involution laws $(xy)^S = y^S x^S$, $(x^S)^S = x$ form an identity basis for $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^S \rangle$ with $\mathcal{K}$ infinite. (The proof of this will be published separately.) Thus, the involutory semigroup $\langle \mathrm{M}_{2m}(\mathcal{K}), \cdot, {}^S \rangle$ is finitely based if and only if $\mathcal{K}$ is an infinite field.

3.5   Boolean matrices

Recall that a *Boolean* matrix is a matrix with entries 0 and 1 only. The multiplication of such matrices is as usual, except that addition and multiplication of the entries is defined as: $a + b = \max\{a, b\}$ and $a \cdot b = \min\{a, b\}$. Let B_n denote the set of all Boolean $n \times n$-matrices. It is well known that the semigroup $\langle B_n, \cdot \rangle$ is essentially the same as the semigroup of all binary relations on an n-element set subject to the usual composition of binary relations. The operation ^{T} of forming the matrix transpose then corresponds to the operation of forming the dual binary relation.

Theorem 3.13. *The involutory semigroup $\mathcal{B}_n = \langle B_n, \cdot, {}^T \rangle$ of all Boolean $n \times n$-matrices endowed with transposition is inherently nonfinitely based.*

Proof. Consider the Boolean matrices

$$B_{11} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \ B_{12} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \ B_{21} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ B_{22} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

$$O = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \ I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The set $M = \{B_{11}, B_{12}, B_{21}, B_{22}, O, I\}$ is closed under multiplication and transposition whence $\mathcal{M} = \langle M, \cdot, {}^T \rangle$ is an involutory subsemigroup of $\mathcal{B}_2$. The mapping $\mathcal{TB}_2^1 \to \mathcal{M}$ given by

$$(i, j) \mapsto B_{ij}, \ 0 \mapsto O, \ 1 \mapsto I$$

is an isomorphism of involutory semigroups. By Corollary 2.7 $\mathcal{M}$ is inherently nonfinitely based whence so is $\mathcal{B}_2$. Since $\mathcal{B}_2$ can be embedded as an involutory semigroup into $\mathcal{B}_n$ for each n, the result follows. $\square$

Remark 3.6. Our proof of Theorem 3.13 also applies to some important involutory subsemigroups of $\mathcal{B}_n$. In order to introduce an interesting instance, recall that Boolean $n \times n$-matrices are in a 1-1 correspondence with bipartite directed graphs whose parts are of size n: the bipartite graph of a matrix $A = (a_{ij})$ has the row set and the column set of A as its parts and has an edge from the ith row to the jth column if and only if $a_{ij} = 1$. If the graph of A admits a perfect matching (i.e. a set of edges so that every vertex is incident to precisely one of them), A is said to be a *Hall matrix* (the name suggested in [27]

is, of course, inspired by Hall's marriage theorem). It is easy to see that the collection HB_n of all Hall $n \times n$-matrices is closed under multiplication and transposition. Since all the matrices $B_{11}, B_{12}, B_{21}, B_{22}, O, I$ from the above proof are Hall matrices, we readily conclude that the involutory semigroup $\langle HB_n, \cdot, {}^T \rangle$ is inherently nonfinitely based.

Remark 3.7. We can unify Theorem 3.13 and some results in Subsection 3.3 by considering matrices over **semirings**. A *semiring* is an algebraic structure $\mathcal{L} = \langle L, +, \cdot \rangle$ of type $(2,2)$ such that $\langle L, + \rangle$ is a commutative semigroup, $\langle L, \cdot \rangle$ is a semigroup and multiplication distributes over addition. From the proofs of Theorems 3.9 and 3.13 we see that the involutory matrix semigroup $\langle M_n(\mathcal{L}), \cdot, {}^T \rangle$ over a finite semiring is inherently nonfinitely based whenever the semiring $\mathcal{L}$ has a zero 0 (that is, a neutral element for $\langle L, + \rangle$ which is at the same time an absorbing element for $\langle L, \cdot \rangle$) and satisfies one of the following two conditions:

1. there exist (not necessarily distinct) elements $e, x \neq 0$ such that $e^2 = e$, $ex = xe = x$, $e + x^2 = 0$;

2. there exists an element $e \neq 0$ such that $e^2 = e = e + e$.

We have already met an infinite series of semirings satisfying (1): it consists of the finite fields $\mathcal{K} = \langle K, +, \cdot \rangle$ with $|K| \not\equiv 3 \pmod 4$. It should be noted that semirings satisfying (2) are even more plentiful: for example, finite distributive lattices as well as the power semirings of finite semigroups (with the subset union as addition and the subset product as multiplication) fall in this class.

A Boolean matrix $A = (a_{ij})$ is said to be *upper triangular* if $a_{ij} = 0$ whenever $i > j$. Let BT_n stand for the set of all Boolean upper triangular $n \times n$-matrices. The semigroups $\langle BT_n, \cdot \rangle$ play an important role in the theory of formal languages, see [43]; their identities have been studied by the third author and Goldberg in [56]. Observe that this semigroup admits quite a natural unary operation: the reflection with respect to the secondary diagonal (the diagonal from the top right to the bottom left corner). We denote by A^D the result of applying this operation to the matrix A. It is easy to see that the operation $A \mapsto A^D$ is in fact an involution; this follows, for instance, from the fact that $A^D = J A^T J$ where J is the Boolean matrix with 1s in the secondary diagonal and 0s elsewhere.

Theorem 3.14. *For each integer $n \geq 3$, the involutory semigroup $\mathcal{BT}_n = \langle BT_n, \cdot, {}^D \rangle$ of all Boolean upper triangular $n \times n$-matrices endowed with the reflection with respect to the secondary diagonal is inherently nonfinitely based.*

Proof. Consider the involutory submonoid $\mathcal{M}$ in $\mathcal{BT}_n$ generated by the following two Boolean matrices:

$$X = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Clearly, for each matrix (m_{ij}) in this submonoid one has $m_{11} = m_{nn} = 1$, whence the set of all matrices (m_{ij}) such that $m_{1n} = 1$ forms an ideal in $\mathcal{M}$. We denote this ideal by $\mathcal{N}$. A straightforward calculation shows that, besides X, Y, and the identity matrix I, only the two matrices

$$XY = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \quad \text{and} \quad YX = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

belong to $\mathcal{M} \setminus \mathcal{N}$. This allows one to organize the following bijection between $\mathcal{M} \setminus \mathcal{N}$ and the set of non-zero matrices in $\mathcal{TA}_2^1$:

$$I \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \quad X \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right), \quad Y \mapsto \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right), \quad XY \mapsto \left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right), \quad YX \mapsto \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right).$$

One easily checks that extending this bijection to $\mathcal{M}$ by sending all elements from $\mathcal{N}$ to $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ yields an involutory semigroup homomorphism from $\mathcal{M}$ onto $\mathcal{TA}_2^1$. Thus, $\mathcal{TA}_2^1$ as a homomorphic image of an involutory subsemigroup in $\mathcal{BT}_n$ belongs to the involutory semigroup variety generated by $\mathcal{BT}_n$. Corollary 2.8 implies that $\mathcal{BT}_n$ is inherently non-finitely based. $\qquad\square$

Remark 3.8. In [56] it was shown that the semigroup $\langle BT_n, \cdot \rangle$ is inherently nonfinitely based for $n \geq 4$. However, the construction used there does not imply the same fact for the involutory case. Our proof of Theorem 3.14 follows a different construction suggested for the plain semigroup case by Li and Luo [32]. Li and Luo have also verified that the semigroup $\langle BT_2, \cdot \rangle$ is finitely based. We do not know whether or not the involutory semigroup $\mathcal{BT}_2$ is finitely based.

Other interesting involutory semigroups of Boolean matrices include the semigroup $\mathcal{BR}_n = \langle BR_n, \cdot, ^T \rangle$ of all Boolean $n \times n$-matrices with 1s on the main diagonal (such matrices correspond to reflexive binary relations) and the semigroup $\mathcal{BU}_n = \langle BU_n, \cdot, ^D \rangle$ of all Boolean upper triangular $n \times n$-matrices with 1s on the main diagonal. (The unary operation is the usual transpose in the former case and the reflection with respect to the secondary diagonal in the latter one.) Our present methods do not yet suffice to handle the finite basis problem for these unary semigroups so we just formulate

Problem 3.3. Are the involutory semigroups $\mathcal{BR}_n$ and $\mathcal{BU}_n$ finitely based for $n \geq 3$?

The involutory semigroups $\mathcal{BR}_2$ and $\mathcal{BU}_2$ are easily seen to be finitely based. The finite basis problem for the plain semigroups $\langle BR_n, \cdot \rangle$ and $\langle BU_n, \cdot \rangle$ has been solved by the third author [55].

References

[1] Amitsur, S. A.: Polynomial identities. Israel J. Math. **19**, 183–199 (1974)

[2] Araújo, J., Mitchell, J. D.: An elementary proof that every singular matrix is a product of idempotent matrices. Amer. Math. Monthly **112**, 641–645 (2005)

[3] Auinger, K.: Strict regular ∗-semigroups. In: Proceedings of the Conference on Semigroups with Applications, Howie, J. M., Munn, W. D., Weinert, H.-J. (eds.), pp. 190–204, World Scientific, Singapore (1992)

[4] Ben-Israel, A., Greville, Th.: Generalized Inverses: Theory and Applications. Springer-Verlag, Berlin–Heidelberg–New York (2003)

[5] Burris, S., Sankappanavar, H. P.: A Course in Universal Algebra. Springer-Verlag, Berlin–Heidelberg–New York (1981)

[6] Clifford, A. H., Preston, G. B.: The Algebraic Theory of Semigroups, Vol. I. Amer. Math. Soc., Providence (1961)

[7] Cline, R. E.: Note on the generalized inverse of the product of matrices. SIAM Review **6**, 57–58 (1964)

[8] Colombo, J., Koshlukov, P.: Central polynomials in the matrix algebra of order two. Linear Algebra Appl. **377**, 53–67 (2004)

[9] D'Amour, A., Racine, M.: *-Polynomial identites of matrices with the transpose involution: the low degrees. Trans. Amer. Math. Soc. **351**, 5089–5106 (1999)

[10] D'Amour, A., Racine, M.: *-Polynomial identites of matrices with the symplectic involution: the low degrees. Comm. Algebra **32**, 895-918 (2004)

[11] Dolinka, I.: On identities of finite involution semigroups. Semigroup Forum **80**, 105–120 (2010)

[12] Drazin, M. P.: Regular semigroups with involution. In: Proceedings of the Symposium on Regular Semigroups, pp. 29–46, Northern Illinois University, De Kalb (1979)

[13] Drensky, V. S.: A minimal basis of identities for a second-order matrix algebra over a field of characteristic 0. Algebra i Logika **20**, 282–290 (1981) [Russian; English transl. Algebra and Logic **20**, 188–194 (1982)]

[14] Drensky, V. S., Formanek, E.: Polynomial Identity Rings. Birkhäuser, Basel (2004)

[15] Drensky, V. S., Giambruno, A.: On the *-polynomial identities of minimal degree for matrices with involution. Boll. Unione Math. Ital. Ser. A **9**, 471–482 (1995)

[16] Erdos, J. A.: On products of idempotent matrices. Glasgow Math. J. **8**, 118–122 (1967)

[17] Genov, G. K.: A basis of identities of the algebra of third-order matrices over a finite field. Algebra i Logika **20**, 385–388 (1981) [Russian; English transl. Algebra and Logic **20**, 241–257 (1982)]

[18] Genov, G. K., Siderov, P. N.: A basis of the identities of the fourth order matrix algebra over a finite field. I, II. Serdica **8**, 313–323, 351–366 (1982) [Russian]

[19] Gerhard, J. A., Petrich, M.: Free involutorial completely simple semigroups. Canad. J. Math. **37**, 271–295 (1985)

[20] Giambruno, A.: On *-polynomial identities for $n \times n$-matrices. J. Algebra **133**, 433-438 (1990)

[21] Giambruno, A., Zaicev, M.: Polynomial Identities and Asymptotic Methods. Amer. Math. Soc., Providence (2005)

[22] Golubchik, I. Z., Mikhalev, A. V.: A note on varieties of semiprime rings with semigroup identities. J. Algebra **54**, 42–45 (1978)

[23] Kanel-Belov, A., Rowen, L. H.: Computational Aspects of Polynomial Identities. A K Peters Ltd., Wellesley (2005)

[24] Kemer, A. R.: The finite basis property of identities of associative algebras. Algebra i Logika **26**, 597–641 (1987) [Russian; English transl. Algebra and Logic **26**, 362–397 (1987)]

[25] Kemer, A. R.: Ideals of Identities of Associative Algebras. Amer. Math. Soc., Providence (1991)

[26] Khalezov, E. A.: Isomorphisms of matrix semigroups. Ivanov. Gos. Ped. Inst., Uchenye Zap., Fiz.-Mat. Nauki **5**, 42–56 (1954) [Russian]

[27] Kim, K. H.: The semigroups of Hall relations. Semigroup Forum **9**, 253–260 (1974)

[28] Kim, K. H., Roush, F.: On groups in varieties of semigroups. Semigroup Forum **16**, 201–202 (1978)

[29] Kleiman, E. I.: Bases of identities of varieties of inverse semigroups. Sib. Mat. Zh. **20**, 760–777 (1979) [Russian; English transl. Sib. Math. J. **20**, 530–543 (1979)]

[30] Koshlukov, P.: Basis of the identities of the matrix algebra of order two over a field of characteristic $p \neq 2$. J. Algebra **241**, 410–434 (2001)

[31] Kruse, R. L.: Identities satisfied by a finite ring. J. Algebra **26**, 298–318 (1973)

[32] Li, J. R., Luo, Y. F.: On the finite basis problem for the monoids of triangular boolean matrices. Algebra Univers. (to appear)

[33] Lidl, R., Niederreiter, H.: Finite Fields. Addison-Wesley, Cambridge (1997)

[34] L'vov, I. V.: Varieties of associative rings. I. Algebra i Logika **12**, 269–297 (1973) [Russian; English transl. Algebra and Logic **12**, 150–167 (1973)]

[35] Magnus, W., Karras, A., Solitar, D.: Combinatorial Group Theory. Wiley, New York–London–Singapore (1966)

[36] Maltsev, Yu. N., Kuzmin, E. N.: A basis for the identities of the algebra of second-order matrices over a finite field. Algebra i Logika **17**, 28–32 (1978) [Russian; English transl. Algebra and Logic **17**, 18–21 (1978)]

[37] Margolis, S. W., Sapir, M. V.: Quasi-identities of finite semigroups and symbolic dynamics. Israel J. Math. **92**, 317–331 (1995)

[38] Meyer, C. D.: Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia (2000)

[39] Moore, E. H.: On the reciprocal of the general algebraic matrix. Bull. Amer. Math. Soc. **26**, 394–395 (1920)

[40] Neumann, H.: Varieties of Groups. Springer-Verlag, Berlin–Heidelberg–New York (1967)

[41] Oates, S., Powell, M. B.: Identical relations in finite groups. J. Algebra **1**, 11–39 (1964)

[42] Penrose, R.: A generalized inverse for matrices. Proc. Cambridge Phil. Soc. **51**, 406–413 (1955)

[43] Pin, J.-E., Straubing, H.: Monoids of upper triangular matrices. In: Semigroups. Structure and Universal Algebraic Problems, Pollák, Gy., Schwarz, Št., Steinfeld, O. (eds.), Colloq. Math. Soc. János Bolyai **39**, pp. 259–272, North-Holland, Amsterdam–Oxford–New York (1985)

[44] Procesi, C.: The invariant theory of $n \times n$ matrices. Adv. Math. **19**, 306–381 (1976)

[45] Procesi, C.: Lie Groups: an Approach through Invariants and Representations. Springer-Verlag, Berlin–Heidelberg–New York (2006)

[46] Razmyslov, Yu. P.: Finite basing of the identities of a matrix algebra of second order over a field of characteristic zero. Algebra i Logika **12**, 83–113 (1973) [Russian; English transl. Algebra and Logic **12**, 47–63 (1973)]

[47] Rowen, L. H.: Polynomial Identities in Ring Theory. Academic Press, New York–London (1980)

[48] Sapir, M. V.: Inherently nonfinitely based finite semigroups. Mat. Sb. **133**, 154–166 (1987) [Russian; English transl. Math. USSR-Sb. **61**, 155–166 (1988)]

[49] Sapir, M. V.: Problems of Burnside type and the finite basis property in varieties of semigroups. Izv. Akad. Nauk SSSR, Ser. Mat. **51**, 319–340 (1987) [Russian; English transl. Math. USSR-Izv. **30**, 295–314 (1987)]

[50] Sapir, M. V.: Identities of finite inverse semigroups. Internat. J. Algebra Comput. **3**, 115–124 (1993)

[51] Sapir, M. V.: Combinatorics on Words with Applications. IBP-Litp 1995/32: Rapport de Recherche Litp, Université Paris 7 (1995). Available online under
`http://www.math.vanderbilt.edu/~msapir/ftp/course/course.pdf`

[52] Serre, J.-P.: Cours d'Arithmetique. Presses Universitaires de France, Paris (1980)

[53] Volkov, M. V.: On finite basedness of semigroup varieties. Mat. Zametki **45**, 12–23 (1989) [Russian; English transl. Math. Notes **45**, 187–194 (1989)]

[54] Volkov, M. V.: The finite basis problem for finite semigroups. Sci. Math. Jpn. **53**, 171–199 (2001)

[55] Volkov, M. V.: Reflexive relations, extensive transformations and piecewise testable languages of a given height. Internat. J. Algebra Comput. **14**, 817–827 (2004)

[56] Volkov, M. V., Goldberg, I. A.: The finite basis problem for monoids of triangular boolean matrices. In: Algebraic Systems, Formal Languages, and Conventional and Unconventional Computation Theory, Surikaisekikenkyusho Kokyuroku **1366**, pp. 205–214, Research Institute for Mathematical Sciences, Kyoto University, Kyoto (2004)