THE FINITE BASIS PROBLEM FOR SEMIGROUPS OF MATRICES WITH NATURAL UNARY OPERATIONS

K. AUINGER, I. DOLINKA, AND M. V. VOLKOV

ABSTRACT. We study equational theories of matrices in the language involving multiplication and natural unary operations such as transposition or Moore-Penrose inversion. We prove that in many cases such theories are not finitely axiomatizable.

1. BACKGROUND AND MOTIVATION

A fundamental and widely studied question connected with an algebraic structure is whether its equational theory is finitely axiomatizable. This question is often referred to as the *finite basis problem* because structures for which it is answered in the affirmative are usually said to be *finitely based*. Being very natural by itself, the finite basis problem has also revealed a number of interesting and unexpected relations to many issues of theoretical and practical importance ranging from feasible algorithms for membership in certain classes of formal languages (see [1]) to classical number-theoretic conjectures (such as the Twin Prime, Goldbach, existence of odd perfect numbers and the infinitude of even perfect numbers: see [21] where it is shown that each of these conjectures is equivalent to the finite basability of a particular groupoid).

Since matrices play a distinguished role in mathematics, algebraic structures whose carriers are sets of matrices form important and popular objects of study. In particular, the finite basis problem has been exhaustively investigated for the semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ of all $n \times n$-matrices[1] over a field $\mathcal{K}$. Corresponding results can be summarized in the following statement.

Theorem 1.1. *Let $\mathcal{K}$ be a field, $n \geq 2$. The semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ is finitely based if and only if $\mathcal{K}$ is infinite.*

The "if" part of Theorem 1.1 is a straightforward corollary of the general fact that the semigroup of all $n \times n$-matrices over an infinite field satisfies only trivial identities, that is, consequences of the associative law, see [10, Lemma 2]. The "only if" part, that is, the claim that the semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$ over a finite field is non-finitely based, was proved in the mid-1980s by the third author [29, Proposition 3] and Sapir [24, Corollary 6.2].

[1] In order to avoid boring trivialities, we always assume $n \geq 2$ when speaking about $n \times n$-matrices.

For a forthcoming discussion it is worth noting that methods used in [29] and [24] were rather different but each of them sufficed to cover all finite semigroups of the form $\langle \mathrm{M}_n(\mathcal{K}), \cdot \rangle$. Moreover, Sapir's approach implied that all these semigroups are even inherently non-finitely based, where a finite semigroup is called *inherently non-finitely based* if it is not contained in any finitely based locally finite variety.

Matrices admit several fairly natural unary operations such as transposition, for instance. Here is the starting point of the present paper: we shall treat the finite basis problem for **unary semigroups of matrices** where a *unary semigroup* is merely a semigroup equipped with some additional unary operation. For this we first have to adapt to the unary environment the methods of [29] and [24]. We present the corresponding results in Section 2 while Section 3 collects several applications to the finite basis problem for unary semigroups of matrices. Both methods of Section 2 are used here, and it turns out that they in some sense complement each other since, in contrast to the plain semigroup case, none of them alone suffice to cover, say, all finite unary semigroups of the form $\langle \mathrm{M}_n(\mathcal{K}), \cdot, ^T \rangle$ where the unary operation $A \mapsto A^T$ is the usual transposition of matrices. For this particular case our main result is the following.

Theorem 1.2. *Let $\mathcal{K} = \langle K, +, \cdot \rangle$ be a finite field and $n \geq 2$. 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 non-finitely based if and only if either $n \geq 3$ or $n = 2$ and $|K| \not\equiv 3 \pmod 4$.*

Other types of unary operations with matrices studied in this paper include, for instance, Moore-Penrose inversion and symplectic transposition. We also consider unary semigroups of Boolean matrices under transposition.

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

We assume the reader's acquaintance with basic concepts of equational logic and the theory of varieties such as the HSP-theorem, see, e.g., [5, Chapter II]. As far as semigroup notions are concerned, we adopt the standard terminology and notation from [6]. It should be noted, however, that the presentation is to a reasonable extent self-contained so that most of the material should be accessible to readers without any specific semigroup-theoretic background.

2. Tools

2.1. **Preliminaries.** As mentioned, by a *unary semigroup* we mean an algebraic structure $\mathcal{S} = \langle S, \cdot, ^* \rangle$ of type $(2, 1)$ such that the binary operation $\cdot$ is associative, i. e. $\langle S, \cdot \rangle$ is a semigroup. In general, 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 (see [6, Section 3.1] for a description of the construction in the plain semigroup case). 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 $(I \times G \times I) \cup \{0\}$ by the following rules:

$$a \cdot 0 = 0 \cdot a = 0 \quad \text{for all } a \in (I \times G \times I) \cup \{0\},$$
$$(i, g, j) \cdot (k, h, \ell) = \begin{cases} (i, gp_{jk}h, \ell) & \text{if } p_{jk} \neq 0, \\ 0 & \text{if } p_{jk} = 0; \end{cases}$$
$$(i, g, j)^* = (j, g^{-1}, i),\ 0^* = 0.$$

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 usually 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:

$$(i, j) \cdot (k, \ell) = \begin{cases} (i, \ell) & \text{if } (j, k) \neq (2, 3), (3, 2), \\ 0 & \text{otherwise}; \end{cases} \tag{2.1}$$
$$(i, j)^* = (j, i).$$

Another unary semigroup that will be quite useful in the sequel is the *free involutory semigroup* $\mathrm{FI}(X)$ 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 $\mathrm{FI}(X)$ 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 $x_1, \ldots, x_m \in X \cup \overline{X}$. We will refer to elements of $\mathrm{FI}(X)$ as to *involutory words over* X while elements of the free semigroup X^+ will be referred to as (plain semigroup) words over X.

2.2. **A unary version of the critical semigroup method.** Here we present a 'unary' modification of the approach used in [29]. According to a classification proposed in [30], 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 $\mathcal{S} \in \mathbf{V}$.

Denote by $\mathsf{var}\,\mathcal{S}$ the variety generated by a given unary semigroup $\mathcal{S}$. 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* $n \in \mathbb{N}$.

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. □

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 $u(x_1, \dots, x_m) = v(x_1, \dots, x_m)$ that holds in $\mathrm{H}(\mathbf{V})$ but fails in the group $\mathcal{G}$. The latter means that, for some $g_1, \dots, g_m \in G$, substitution of g_i for x_i yields

$$u(g_1, \dots, g_m) \neq v(g_1, \dots, g_m). \tag{2.2}$$

Now, for each positive integer k, let $n = \max\{4, 2k+1\}$, $I = \{1, \dots, nm\}$ 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 [29].) We are going 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 $x_1, \dots, x_{mn}$:

$$\begin{aligned}
w_1 = {} & [x_1x_1^*\cdots x_nx_n^*][(x_{n+1}\cdots x_{2n})(x_{n+1}\cdots x_{2n})^*x_{2n}x_{2n}^*]\cdots \\
& [(x_{(m-1)n+1}\cdots x_{mn})(x_{(m-1)n+1}\cdots x_{mn})^*x_{mn}x_{mn}^*], \\
w_2 = {} & [(x_1\cdots x_n)(x_1\cdots x_n)^*x_nx_n^*][x_{n+1}x_{n+1}^*\cdots x_{2n}x_{2n}^*]\times \\
& [(x_{2n+1}\cdots x_{3n})(x_{2n+1}\cdots x_{3n})^*x_{3n}x_{3n}^*]\cdots \\
& [(x_{(m-1)n+1}\cdots x_{mn})(x_{(m-1)n+1}\cdots x_{mn})^*x_{mn}x_{mn}^*], \\
w_3 = {} & [(x_1\cdots x_n)(x_1\cdots x_n)^*x_nx_n^*][(x_{n+1}\cdots x_{2n})(x_{n+1}\cdots x_{2n})^*x_{2n}x_{2n}^*]\times \\
& [x_{2n+1}x_{2n+1}^*\cdots x_{3n}x_{3n}^*]\cdots[(x_{(m-1)n+1}\ldots x_{mn})(x_{(m-1)n+1}\ldots x_{mn})^*x_{mn}x_{mn}^*], \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
w_m = {} & [(x_1\cdots x_n)(x_1\cdots x_n)^*x_nx_n^*][(x_{n+1}\cdots x_{2n})(x_{n+1}\cdots x_{2n})^*x_{2n}x_{2n}^*]\ldots \\
& [(x_{(m-2)n+1}\cdots x_{(m-1)n})(x_{(m-2)n+1}\cdots x_{(m-1)n})^*x_{(m-1)n}x_{(m-1)n}^*]\times \\
& [x_{(m-1)n+1}x_{(m-1)n+1}^*\cdots x_{mn}x_{mn}^*].
\end{aligned}$$

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

$$u(w_1,\ldots,w_m) = v(w_1,\ldots,w_m) \tag{2.3}$$

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 (2.3) 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 (2.3) 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 (2.3) is $(s',v(g_1,\ldots,g_m),t')$ (again for suitable $s',t'\in\{1,mn\}$). In view of the inequality (2.2), 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

$$1\le\lambda_1\le n<\lambda_2\le 2n<\lambda_3\le\ldots(m-1)n<\lambda_m\le mn, \tag{2.4}$$

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 $g\in\mathcal{G}$ 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 (2.4). When multiplying triples from $\mathcal{T}_k(\lambda_1,\ldots,\lambda_m)$, the $\lambda_1^{\mathrm{th}},\ldots,\lambda_m^{\mathrm{th}}$ 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 $\lambda_1^{\text{th}},\dots,\lambda_m^{\text{th}}$ rows and columns. Note that by (2.4) exactly one row and one column of each block $M_n(g_i)$ is deleted.

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

$$\begin{array}{l}\text{the row } ((i-1)n+2) \text{ by } g_i \text{ from the left and}\\ \text{the column } ((i-1)n+2) \text{ by } g_i^{-1} \text{ from the right;}\\ \text{the row } ((i-1)n+3) \text{ by } g_i \text{ from the left and}\\ \text{the column } ((i-1)n+3) \text{ by } g_i^{-1} \text{ from the right;}\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ \text{the row } (\lambda_i-1) \text{ by } g_i \text{ from the left and}\\ \text{the column } (\lambda_i-1) \text{ by } g_i^{-1} \text{ from the right.}\end{array} \tag{2.5}$$

In order to help the reader to understand the effect of the transformations (2.5), 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 $P_k(\lambda_1,\dots,\lambda_m)$ involved in (2.5) have non-zero entries only within $M_n(g_i)$.)

The block obtained from $M_n(g_i)$ by erasing the $((i-1)n+5)^{\text{th}}$ row and column

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

After the first transformation

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

After the second transformation

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

After the third transformation

$$\begin{pmatrix} e & e & 0 & 0 & 0 & \cdots & 0 & e\\ e & e & e & 0 & 0 & \cdots & 0 & 0\\ 0 & e & e & e & 0 & \cdots & 0 & 0\\ 0 & 0 & e & e & 0 & \cdots & 0 & 0\\ 0 & 0 & 0 & 0 & e & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & 0 & \cdots & e & e\\ e & 0 & 0 & 0 & 0 & \cdots & e & e \end{pmatrix}$$

Now it should be clear that also in general the transformations (2.5) 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 (2.5) of the sandwich matrix do not change the unary semigroup $\mathcal{T}_k(\lambda_1,\dots,\lambda_m)$; 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

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

for all $g \in \mathcal{G}$, $i, j \in I'$, 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:

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

(These words already have been used in the plain semigroup case by the third author [29].) □

2.3. **A unary version of the method of inherently non-finitely based semigroups.** Recall that a finite algebraic structure of type τ is called *inherently non-finitely based* if it is not contained in any finitely based locally finite variety of type τ. Here we prove a sufficient condition for an involutory semigroup to be inherently non-finitely based and exhibit two concrete examples of involutory semigroups satisfying this condition. These examples will be essentially 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 non-finitely 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, $\mathcal{T}_k$ 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 [24], see also his lecture notes [27]. We fix k and let $r = 6k + 2$. Consider the $r^2 \times r$-matrix M shown in Fig. 1 on the left. All odd columns of M are identical and equal

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

FIGURE 1. The matrices M and M_A

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

Now consider the alphabet $A = \{a_{ij} \mid 1 \le i, j \le r\}$ of cardinality r^2. We convert the matrix M to the matrix M_A (shown in Fig. 1 on the right) 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 t^{th} 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 $\{\gamma^m(a_{11})\}_{m=1,2,\dots}$ 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. [27, Proposition 2.1] *Let $X_k = \{x_1, \dots, x_k\}$ 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. [27, Lemma 4.14] *Let $X_k = \{x_1, \dots, x_k\}$ 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 $p, q \in \mathrm{FI}(X_k)$ and suppose that the identity $p = q$ holds in $\mathcal{S}$ but fails in $\mathcal{T}_k$. Then there exists a unary semigroup homomorphism $\varphi : \mathrm{FI}(X_k) \to \mathcal{T}_k$ 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 $\sigma : \mathrm{FI}(X_k) \to \mathrm{FI}(X_k)$ 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 $\mathrm{FI}(X_k)$ and there is no harm in denoting the extension by ψ as well. The identity $p = q$ implies the identity

$$z'\psi(\sigma(p))z'' = z'\psi(\sigma(q))z''. \tag{2.6}$$

The left hand side of (2.6) 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 (2.6). 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. $\square$

As a concrete example of an inherently non-finitely based involutory semigroup, consider the *twisted Brandt monoid* $\mathcal{TB}_2^1 = \langle B_2^1, \cdot, ^* \rangle$, 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 2.6. *The twisted Brandt monoid $\mathcal{TB}_2^1$ is inherently non-finitely based.*

Proof. By Theorem 2.3 we only have to show that $\mathcal{TB}_2^1$ satisfies no non-trivial involutory semigroup identity of the form $Z_n = z$. If z is a plain semigroup word, we can refer to [24, 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 $\mathcal{TB}_2^1$ in this case as well. $\square$

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

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

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

$$(i,j)\cdot(k,\ell)=\begin{cases}(i,\ell) & \text{if } (j,k)\in\{(1,2),(2,1)\},\\ 0 & \text{otherwise;}\end{cases} \tag{2.7}$$
$$1\cdot x=x=x\cdot 1,\ 0\cdot x=0=x\cdot 0 \text{ for all } x;$$
$$(i,j)^*=(j,i),\ 1^*=1,\ 0^*=0.$$

Suppose that $\mathcal{A}$ is a finite algebraic structure for which the variety $\mathsf{var}\,\mathcal{A}$ contains an inherently non-finitely based algebraic structure. Immediately from the definition it follows that $\mathcal{A}$ is also inherently non-finitely based. This observation is useful, in particular, for the justification of our second example of an involutory inherently non-finitely 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

$$\begin{pmatrix}0&e\\e&e\end{pmatrix} \tag{2.8}$$

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

$$(i,j)\cdot(k,\ell)=\begin{cases}(i,\ell) & \text{if } (j,k)\neq(1,1)\\ 0 & \text{if } (j,k)=(1,1);\end{cases} \tag{2.9}$$
$$1\cdot x=x=x\cdot 1,\ 0\cdot x=0=x\cdot 0 \text{ for all } x;$$
$$(i,j)^*=(j,i),\ 1^*=1,\ 0^*=0.$$

Corollary 2.7. *The involutory monoid $\mathcal{TA}_2^1$ is inherently non-finitely based.*

Proof. We represent $\mathcal{TA}_2^1$ as in (2.9) and $\mathcal{TB}_2^1$ as in (2.7) and consider the direct square $\mathcal{TA}_2^1\times\mathcal{TA}_2^1$. It is then easy to check that the twisted Brandt monoid $\mathcal{TB}_2^1$ is a homomorphic image of the unary subsemigroup of $\mathcal{TA}_2^1\times\mathcal{TA}_2^1$ generated by the pairs $(1,1)$, $\big((1,1),(2,2)\big)$ and $\big((2,2),(1,1)\big)$. Thus, $\mathcal{TB}_2^1$ belongs to $\mathsf{var}\,\mathcal{TA}_2^1$. Since by Corollary 2.6 $\mathcal{TB}_2^1$ is inherently non-finitely based, so is $\mathcal{TA}_2^1$. $\square$

Sapir [24, 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 [7]. Here we present yet another special result which however suffices for our purposes.

Proposition 2.8. *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 non-finitely based.*

Proof. This is a consequence of some deep facts obtained by Margolis and Sapir in [15] and we shall follow their main arguments. The reason that these arguments apply here is that in $\mathcal{S}$, the Green relation $\mathscr{R}$ can be expressed in terms of equational logic. This ensures that we can construct a finite set of identities that defines a locally finite variety of involutory semigroups containing $\mathcal{S}$.

First of all, we recall the definition of the relation $\mathscr{R}$. As usual, for a semigroup $a \in S$, the set $\{as \mid s \in S\} \cup \{a\}$ (that is, the right ideal generated by a) is denoted by aS^1. Now we say that elements $a, b \in S$ are *$\mathscr{R}$-related* if $aS^1 = bS^1$.

Since $\mathcal{S}$ satisfies the identity $x = x\omega(x)x$, we have $a \mathrel{\mathscr{R}} a\omega(a)$ for each element $a \in S$. Thus, for $a, b \in S$, we have $a \mathrel{\mathscr{R}} b$ if and only if $a\omega(a) \mathrel{\mathscr{R}} b\omega(b)$. 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 $uv \mathrel{\mathscr{R}} u$ 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 *$\mathscr{R}$-height* of $\mathcal{S}$, that is, h is the length of the longest possible $\mathscr{R}$-chain

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

where $a <_{\mathscr{R}} b$ means that $aS^1 \subseteq bS^1$ but $bS^1 \nsubseteq aS^1$. Set $n = h+1$; Lemma 7 in [15] shows that $\mathcal{S}$ satisfies $Z_n' \mathrel{\mathscr{R}} Z_n$ whence it satisfies the identity

$$Z_n\omega(Z_n) \cdot Z_n'\omega(Z_n') = Z_n'\omega(Z_n'). \tag{2.10}$$

On the other hand, each involutory semigroup $\mathcal{T}$ which satisfies $x = x\omega(x)x$ and (2.10) necessarily satisfies $Z_n' \mathrel{\mathscr{R}} Z_n$, hence (its semigroup reduct) belongs to the quasivariety $\mathbf{Q}_n$ of semigroups defined by the quasi-identity

$$xZ_n = yZ_n \to xZ_n' = yZ_n'.$$

Lemma 8 in [15] then shows that a finitely generated semigroup $\mathcal{T} \in \mathbf{Q}_n$ is finite if and only if it is periodic and all subgroups of $\mathcal{T}$ are locally finite. We note that an involutory semigroup is finitely generated if and only if so is its semigroup reduct. Hence it suffices to find a finite number of identities which hold in $\mathcal{S}$ and which force each (involutory) semigroup to be periodic and to have only locally finite subgroups.

We can proceed as at the end of [15]: first, since $\mathcal{S}$ is finite, it satisfies the identity $x^k = x^{k+\ell}$ for some $k, \ell \geq 1$. This identity definitely forces any (involutory) semigroup to be periodic. Next, let $\mathcal{G}$ be the direct product of all maximal subgroups of $\mathcal{S}$. By the Oates-Powell theorem [19], see also [18, §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 negative letters, that is, 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 reduct of $\mathcal{S}$. Let e be an idempotent in the minimal ideal of $\mathcal{F}$ and let $u(x_1, \ldots, x_m)$ be a word whose value in $\mathcal{F}$ is e. It follows that $\mathcal{F}$ (and therefore $\mathcal{S}$) satisfies the identity

$$u = u^2. \tag{2.11}$$

For every element $g \in \mathcal{F}$, the product ege belongs to the maximal subgroup of $\mathcal{F}$ with idempotent e. Consequently, $\mathcal{F}$ (and therefore $\mathcal{S}$) satisfies the identity

$$v(ux_1u, \ldots, ux_mu) = u. \tag{2.12}$$

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

Now consider the variety of involutory semigroups defined by the identities $x = x\omega(x)x$, $x^k = x^{k+\ell}$ together with (2.10), (2.11) and (2.12). By construction, $\mathcal{S}$ is a member of that variety. Let $\mathcal{T}$ be any finitely generated member; then, as already mentioned, the semigroup reduct of $\mathcal{T}$ is also finitely generated. The first and the third identity ensure that the semigroup reduct of $\mathcal{T}$ belongs to the quasivariety $\mathbf{Q}_n$, and therefore it is finite provided that it is periodic and all its subgroups are locally finite. Periodicity is, of course, guaranteed by the second identity. Finally, each group $\mathcal{H}$ that satisfies the identities (2.11) and (2.12) satisfies the identity $v(x_1, \ldots, x_m) = 1$, whence $\mathcal{H}$ belongs to $\mathsf{var}\,\mathcal{G}$ and so is locally finite. Altogether, $\mathcal{T}$ is finite and the proposition is proved. □

Proposition 2.8 implies in particular that no finite regular $*$-semigroup can be inherently non-finitely 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 non-finitely based (the fact first discovered by Sapir, see [26]), even though it is not finitely based [12].

3. Applications

3.1. **A property of matrices of rank 1.** Recall that, for a field $\mathcal{K}$, we denote the set of all $n \times n$-matrices over $\mathcal{K}$ by $\mathrm{M}_n(\mathcal{K})$. We start 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 3.1. *If $A \in \mathrm{M}_n(\mathcal{K})$, $n \geq 2$, has rank 1, then $A^nBA^{n-1} = A^{n-1}BA^n$ for any matrix $B \in \mathrm{M}_n(\mathcal{K})$.*

Proof. By the Cayley-Hamilton theorem, A satisfies the equation

$$A^n - \mathrm{tr}(A)A^{n-1} = 0,$$

where $\mathrm{tr}(A)$ stands for the trace of the matrix A. Hence

$$A^nBA^{n-1} = \mathrm{tr}(A)A^{n-1}BA^{n-1} = A^{n-1}B\,\mathrm{tr}(A)A^{n-1} = A^{n-1}BA^n,$$

as required. □

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 obtain the set which we denote by By $\mathrm{L}^1_n(\mathcal{K})$. Clearly, it is closed under matrix multiplication. From Lemma 3.1 we immediately obtain

Corollary 3.2. *For any field $\mathcal{K}$ and $n \geq 2$, the semigroup $\langle \mathrm{L}^1_n(\mathcal{K}), \cdot \rangle$ satisfies the identity*

$$x^nyx^{n-1} = x^{n-1}yx^n. \tag{3.1}$$

Observe that every group satisfying (3.1) is abelian.

3.2. **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 [17] and independently by Penrose [20] 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 [8].

Proposition 3.3. [8, Proposition 1] *Let $\mathcal{S}$ be an involutory semigroup. Then, for any given $a \in \mathcal{S}$, the four equations*

$$axa = a,\ xax = x,\ (ax)^* = ax,\ (xa)^* = xa \tag{3.2}$$

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 (3.2), 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 3.4. [8, Proposition 2] *Let $\mathcal{S}$ be an involutory semigroup satisfying the quasi-identity*

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

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 quasi-identity (3.3) is easily seen to be equivalent to

$$x^*x = 0 \to x = 0. \tag{3.4}$$

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 quasi-identity (3.4). However, it does satisfy (3.4) if and only if the equation

$$x_1\overline{x_1} + x_2\overline{x_2} + \cdots + x_n\overline{x_n} = 0 \tag{3.5}$$

admits only the trivial solution $(x_1, \dots, x_n) = (0, \dots, 0)$ 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 (3.5) 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 (3.5) has only the trivial solution is necessary: if $(a_1, \dots, a_n)$ were a non-trivial solution to (3.5), then the matrix formed by n identical rows $(a_1, \dots, a_n)$ 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 $a = (a_1, \dots, a_n) \in K^n$, where $\mathcal{K} = \langle K, +, \cdot \rangle$ is a field with an involution $x \mapsto \overline{x}$, let a^* denote the column vector $(\overline{a_1}, \dots, \overline{a_n})^T$. 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 (3.5) admits only the trivial solution in K^n, one gets $A^\dagger$ as follows:

$$A^\dagger = c^*(cc^*)^{-1}(bb^*)^{-1}b. \tag{3.6}$$

Here bb^* and cc^* are non-zero elements of $\mathcal{K}$ whence their inverses in $\mathcal{K}$ exist. In order to justify (3.6), it suffices to check that the right hand side of (3.6) satisfies the simultaneous equations (3.2) with the matrix A in the role of

a, and this is straightforward. Note that formula (3.6) immediately shows that $A^\dagger$ is a scalar multiple of $A^* = c^*b$, namely

$$A^\dagger = \frac{1}{cc^* \cdot bb^*} A^*. \tag{3.7}$$

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 3.5. *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 $\mathrm{M}_2(\mathcal{K})$:

$$\begin{aligned}
H_{11} &= \left\{\begin{pmatrix} x & x \\ x & x \end{pmatrix}\right\}, & H_{12} &= \left\{\begin{pmatrix} x & 0 \\ x & 0 \end{pmatrix}\right\}, & H_{13} &= \left\{\begin{pmatrix} 0 & x \\ 0 & x \end{pmatrix}\right\}, \\
H_{21} &= \left\{\begin{pmatrix} x & x \\ 0 & 0 \end{pmatrix}\right\}, & H_{22} &= \left\{\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}\right\}, & H_{23} &= \left\{\begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix}\right\}, \\
H_{31} &= \left\{\begin{pmatrix} 0 & 0 \\ x & x \end{pmatrix}\right\}, & H_{32} &= \left\{\begin{pmatrix} 0 & 0 \\ x & 0 \end{pmatrix}\right\}, & H_{33} &= \left\{\begin{pmatrix} 0 & 0 \\ 0 & x \end{pmatrix}\right\},
\end{aligned} \tag{3.8}$$

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

$$H_{ij} \cdot H_{k\ell} = \begin{cases} H_{i\ell} & \text{if } (j,k) \neq (2,3), (3,2), \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

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 $\mathscr{H}$ of T into the classes H_{ij} and $\{0\}$ is a congruence on $\mathcal{T}$. Equation (3.7) shows that

$$H_{ij}^\dagger = H_{ji}. \tag{3.10}$$

We see that T is closed under Moore-Penrose inversion and $\mathscr{H}$ respects †, thus is a congruence on the unary semigroup $\mathcal{T}' = \langle T, \cdot, {}^\dagger \rangle$. Now comparing (3.9) and (3.10) with the multiplication and inversion rules in $\mathcal{K}_3$ (see (2.1)), we conclude that $\mathcal{T}'/\mathscr{H}$ and $\mathcal{K}_3$ are isomorphic as unary semigroups. Hence $\mathcal{K}_3$ is in $\mathsf{var}\,\mathcal{S}$.

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 unary semigroup of all matrices of rank at most 1 with the identity matrix adjoined. By Corollary 3.2 the monoid $\mathrm{L}_2^1(\mathcal{K})$ satisfies the identity $x^2yx = xyx^2$. Consequently, each group in $\mathsf{var}\,\mathrm{H}(\mathcal{S})$ is abelian, while the group $\mathrm{GL}_2(\mathcal{K})$ is non-abelian. Thus, $\mathrm{GL}_2(\mathcal{K})$ is contained in $\mathsf{var}\,\mathcal{S}$ but is not contained in $\mathsf{var}\,\mathrm{H}(\mathcal{S})$, as required. □

Remark 3.1. Apart from any subfield of $\mathbb{C}$ closed under complex conjugation, Theorem 3.5 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. [13, 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 3.5. For some fields, it definitely is. For instance, for finite fields with the trivial involution $x \mapsto \overline{x} = x$, no extension of Theorem 3.5 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., [28, Corollary 2 in §1.2]) that the equation $x_1^2+\cdots+x_n^2 = 0$ (that is (3.5) 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, [14, p. 101]) that the two matrices

$$\zeta = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad \eta = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \tag{3.11}$$

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 $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot, {}^{-1}\rangle \to \langle \mathrm{M}_3(\mathcal{K}), \cdot, {}^\dagger\rangle$ defined by $A \mapsto \begin{pmatrix} A & \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} \\ \begin{smallmatrix} 0 & 0 \end{smallmatrix} & 0 \end{pmatrix}$ 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 the unary monoid $\langle \mathrm{Sing}_n^1(\mathcal{K}), \cdot, {}^\dagger\rangle$. Now we observe that $\mathrm{Sing}_n^1(\mathcal{K})$ is contained in (actually, coincides with) the Hermitian subsemigroup of $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^\dagger\rangle$. Indeed, it was proved in [9] (see also [2] for a recent elementary proof) that the semigroup $\langle \mathrm{Sing}_n^1(\mathcal{K}), \cdot\rangle$ 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 $P_1 = P_1P_1^\dagger$ and $P_2 = P_2P_2^\dagger$ 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 $A = (P_1P_2)^\dagger$, see [16, Exercise 5.15.9a]. Thus, $\mathrm{Sing}_n^1(\mathcal{K}) \subseteq \mathrm{H}(\mathrm{M}_n(\mathcal{K}))$, 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 $n > 2$?

In connection with this problem, we observe that the proofs of Theorem 3.5 and Corollary 3.2 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 $\mathrm{M}_n(\mathcal{K})$ endowed with **both** unary operations † and *. Here our techniques produce a similar result.

Theorem 3.6. *Let $\mathcal{K}$ be a field as in Theorem* 3.5*; then $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^\dagger, {}^*\rangle$ 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 $\mathcal{A} = \langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^\dagger, {}^*\rangle$; the algebraic structure $\langle G, \cdot, {}^\dagger, {}^*\rangle$, that is, the group $\mathcal{G}$ with inversion taken twice as unary operation, belongs to $\mathsf{var}\,\mathcal{A}$. Now, as in the proof of Theorem 3.5, consider the set

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

where H_{ij} are defined via (3.8). 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 $\mathsf{var}\,\mathcal{A}$. By Corollary 3.2 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 $\mathsf{var}\,\mathcal{A}$ while each k-generated substructure of $\mathcal{T}_k$ does belong to $\mathsf{var}\,\mathcal{A}$. Thus, $\mathcal{T}_k$ can play the role of critical structures for $\mathsf{var}\,\mathcal{A}$ 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 $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^\dagger, {}^* \rangle$ 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.

3.3. **Matrix semigroups with transposition.** First of all, wee observe that, for each field $\mathcal{K}$ of characteristic 0, the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ is finitely based. Indeed, we have already mentioned that the two matrices ζ and η in (3.11) generate a free subgroup of $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot, {}^{-1} \rangle$ and hence a free subsemigroup of $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot \rangle$. But $\eta = \zeta^T$ whence the unary subsemigroup in $\langle \mathrm{SL}_2(\mathbb{Z}), \cdot, {}^T \rangle$ generated by ζ is isomorphic to the free monogenic involutory semigroup $\mathrm{FI}(\{\zeta\})$. The latter semigroup is known to contain as a unary subsemigroup a free involutory semigroup on countably many generators, namely, $\mathrm{FI}(Z)$ where

$$Z = \{\zeta\zeta^T\zeta,\ \zeta(\zeta^T)^2\zeta,\ \ldots,\ \zeta(\zeta^T)^n\zeta,\ \ldots\}.$$

Hence all identities holding in $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T \rangle$ with $n \geq 2$ and $\mathcal{K}$ of characteristic 0 follow from the associativity and the involution laws $(xy)^T = y^T x^T$, $(x^T)^T = x$. Similarly, $\langle \mathrm{M}_n(R), \cdot, {}^* \rangle$ is finitely based for each $n \geq 2$ and each subring $R \subseteq \mathbb{C}$ closed under complex conjugation—here * stands for the complex-conjugate transposition $(a_{ij})^* = (\overline{a_{ij}})^T$.

We note that Theorem 2.2 and Corollary 3.2 prove the non-existence of a finite identity bases for the unary subsemigroup of $\langle \mathrm{M}_n(\mathbb{C}), \cdot, {}^* \rangle$ [respectively $\langle \mathrm{M}_n(\mathbb{R}), \cdot, {}^T \rangle$] that consists of all matrices of rank at most 1 together with all unitary [respectively all orthogonal] matrices.

For the case of finite fields, one can show that Theorem 2.2 solves the finite basis problem in the negative for the involutory semigroup $\langle \mathrm{M}_2(\mathcal{K}), \cdot, {}^T \rangle$ for each finite field $\mathcal{K}$ except $\mathcal{K} = \mathbb{F}_2$, the 2-element field. Indeed, it can be verified 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 3.7. *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 3.5 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 3.2, each group in $\mathrm{P}_d(\mathsf{var}\,\mathcal{S}) = \mathsf{var}\,\mathrm{P}_d(\mathcal{S})$ satisfies the law $x^2yx = xyx^2$ and therefore is abelian. On the other hand, as in the proof of Theorem 3.6, 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. $\square$

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

Theorem 3.8. *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 non-finitely 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 2.7. $\square$

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

Theorem 3.9. *For each $n \geq 3$ and each finite field $\mathcal{K}$, the involutory semigroup $\langle \mathrm{M}_n(\mathcal{K}), \cdot, {}^T\rangle$ is inherently non-finitely based.*

Proof. It follows from the Chevalley-Warning theorem [28, Corollary 2 in §1.2] that there exist $x, y \in K$ satisfying $1 + x^2 + y^2 = 0$. Now consider the following matrices:

$$H_{11}=\begin{pmatrix}1&x&y\\x&x^2&xy\\y&xy&y^2\end{pmatrix},\ H_{12}=\begin{pmatrix}1&0&0\\x&0&0\\y&0&0\end{pmatrix},\ H_{21}=\begin{pmatrix}1&x&y\\0&0&0\\0&0&0\end{pmatrix},$$
$$H_{22}=\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix},\ I=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\ O=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}.$$

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 non-finitely based. The assertion for $\mathrm{M}_n(\mathcal{K})$ for $n\geq 3$ now follows in an obvious way. □

Remark 3.3. The statements of Theorems 3.7, 3.8, 3.9 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 result announced in Section 1 as Theorem 1.2 (for the reader's convenience we reproduce this result here as Theorem 3.10). With the exception of the 'only if' part of item (2), this is a summary of Theorems 3.7, 3.8, and 3.9.

Theorem 3.10. *Let $n\geq 2$ and $\mathcal{K}=\langle K,+,\cdot\rangle$ 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 non-finitely based if and only if either $n\geq 3$ or $n=2$ and $|K|\not\equiv 3 \pmod 4$.*

It remains to prove that $\langle \mathrm{M}_2(\mathcal{K}),\cdot,{}^T\rangle$ is **not** inherently non-finitely based if $|K|\equiv 3 \pmod 4$. The assertion follows from Proposition 2.8.

Proof of Theorem 3.10. Suppose that $|K|\equiv 3 \pmod 4$. This is precisely 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 (3.7) 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,

$$A(A^TA)^d=A(\alpha A^\dagger A)^d=\alpha^dA(A^\dagger A)^d=A.$$

If $A\in\mathrm{GL}_2(\mathcal{K})$, we also have $A=A(A^TA)^d$ because $(A^TA)^d$ is the identity matrix; clearly, the equality $A=A(A^TA)^d$ holds also for the case when A is the zero matrix. Summarizing, we conclude that the identity $x=x(x^Tx)^d$ holds in the involutory semigroup $\langle \mathrm{M}_2(\mathcal{K}),\cdot,{}^T\rangle$. Setting $\omega(x):=x^T(xx^T)^{d-1}$, we see that $\langle \mathrm{M}_2(\mathcal{K}),\cdot,{}^T\rangle$ satisfies the identity $x=x\omega(x)x$, as required by Proposition 2.8. □

Remark 3.4. It is known [24, Corollary 6.2] that the matrix semigroup $\langle \mathrm{M}_n(\mathcal{K}),\cdot\rangle$ is inherently non-finitely based (as a plain semigroup) for every

$n \geq 2$ and 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 $2n \times 2n$-matrix

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

with A, B, C, D being $n \times n$-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. [22, (5.1.1)]. This definition is similar to 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 3.11. *The involutory semigroup $\langle \mathrm{M}_{2n}(\mathcal{K}), \cdot, ^S \rangle$ is inherently nonfinitely based for each $n \geq 1$ and each finite field $\mathcal{K} = \langle K, +, \cdot \rangle$.*

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

$$H_{11} = \left\{ \pm \begin{pmatrix} O_n & I_n \\ O_n & O_n \end{pmatrix} \right\}, \ H_{12} = \left\{ \pm \begin{pmatrix} I_n & O_n \\ O_n & O_n \end{pmatrix} \right\},$$
$$H_{21} = \left\{ \pm \begin{pmatrix} O_n & O_n \\ O_n & I_n \end{pmatrix} \right\}, \ H_{22} = \left\{ \pm \begin{pmatrix} O_n & O_n \\ I_n & O_n \end{pmatrix} \right\}$$

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

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

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

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

is a homomorphism of $\mathcal{T}$ onto $\mathcal{TB}_2^1$. Altogether, the twisted Brandt monoid $\mathcal{TB}_2^1$ divides $\langle \mathrm{M}_{2n}(\mathcal{K}), \cdot, ^S \rangle$. □

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 3.12. *For each integer $n \geq 2$, the involutory semigroup $\mathcal{B}_n = \langle B_n, \cdot, {}^T \rangle$ of all Boolean $n \times n$-matrices endowed with transposition is inherently non-finitely 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 2.6 $\mathcal{M}$ is inherently non-finitely 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. □

Remark 3.5. We can unify Theorem 3.12 and some results in Subsection 2.3 by considering matrix semigroups with transposition 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 3.8 and 3.12 we see that the involutory matrix semigroup $\langle \mathrm{M}_n(\mathcal{L}), \cdot, {}^T \rangle$ over a finite semiring is inherently non-finitely based whenever $n \geq 2$ and 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 either 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.

Acknowledgement. The second author was supported by Grant No. 144011 of the Ministry of Science and Technological Development of the Republic of Serbia. The third author acknowledges support from the Federal Education Agency of Russia, grant 2.1.1/3537, and from the Russian Foundation for Basic Research, grant 09-01-12142.

References

[1] J. Almeida, *Finite Semigroups and Universal Algebra*, World Scientific, Singapore, 1995.

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

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

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

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

[6] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups I*, Amer. Math. Soc., Providence, 1961.

[7] I. Dolinka, *On identities of finite involution semigroups*, Semigroup Forum, to appear.

[8] M. P. Drazin, *Regular semigroups with involution*, pp.29–46 in: Proceedings of the Symposium on Regular semigroups, Northern Illinois University, De Kalb, 1979.

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

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

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

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

[13] R. Lidl and H. Niederreiter, *Finite Fields*, Addison-Wesley, Cambridge, 1997.

[14] W. Magnus, A. Karras and D. Solitar, *Combinatorial Group Theory*, Wiley, New York–London–Singapore, 1966.

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

[16] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.

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

[18] H. Neumann, *Varieties of Groups*, Springer-Verlag, Berlin–Heidelberg–New York, 1967.

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

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

[21] P. Perkins, *Finite axiomatizability for equational theories of computable groupoids*, J. Symbolic Logic **54** (1989), 1018–1022.

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

[23] M. V. Sapir, *Inherently non-finitely based finite semigroups*, Mat. Sb. **133**, no.2 (1987), 154–166 [Russian; English transl. Math. USSR-Sb. **61** (1988), 155–166].

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

[25] M. V. Sapir, *On Cross semigroup varieties and related questions*, Semigroup Forum **42** (1991), 345–364.

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

[27] M. V. Sapir, *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).

[28] J.-P. Serre, *Cours d'Arithmetique*, Presses Universitaires de France, Paris, 1980.

[29] M. V. Volkov, *On finite basedness of semigroup varieties*, Mat. Zametki **45**, no.3 (1989), 12–23 [Russian; English transl. Math. Notes **45** (1989), 187–194].

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

Fakultät für Mathematik, Universität Wien, Nordbergstrasse 15, A-1090 Wien, Austria

E-mail address: karl.auinger@univie.ac.at

Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

E-mail address: dosz@eunet.rs

Faculty of Mathematics and Mechanics, Ural State University, Lenina 51, 620083 Ekaterinburg, Russia

E-mail address: mikhail.volkov@usu.ru