J. Comb. Algebra (Online first)
DOI 10.4171/JCA/100

Published by EMS Press

Strongly nonfinitely based monoids

Sergey V. Gusev, Olga B. Sapir, and Mikhail V. Volkov

Abstract. We show that the 42-element monoid of all partial order preserving and extensive injections on the 4-element chain is not contained in any variety generated by a finitely based finite semigroup.

1. General background: Identities and the Finite Basis Problem

The idea of an *identity* or a *law* is very basic and is arguably one of the very first abstract ideas that students come across when they start learning mathematics. We mean laws like the *commutative law of addition*:

A sum isn't changed at rearrangement of its addends.

At the end of high school, a student is aware (or, at least, is supposed to be aware) of a good dozen of laws:

- the commutative and associative laws of addition,
- the commutative and associative laws of multiplication,
- the distributive law of multiplication over addition,
- the difference of two squares identity,
- the Pythagorean trigonometric identity,

etc, etc. Moreover, the student may feel (though probably cannot explain) the difference between 'primary' identities such as

$$ab = ba \tag{1}$$

and

$$(ab)c = a(bc) \tag{2}$$

Mathematics Subject Classification 2020: 20M07.
Keywords: variety, finite basis problem, inherently nonfinitely based semigroup, strongly nonfinitely based semigroup, Catalan monoid.

and ‘secondary’ ones such as, for instance,

$$(ab)^2 = a^2b^2. \tag{3}$$

‘Primary’ laws such as (1) or (2) are *intrinsic* properties of objects (say, numbers) we multiply and of the way the multiplication is defined, whereas ‘secondary’ identities can be *formally inferred* from ‘primary’ ones, without knowing which objects are multiplied and how the multiplication is defined. Here is a simple example of such a formal inference:

$$\begin{aligned}
(ab)^2 &= (ab)(ab) && \text{by the definition of squaring}\\
&= a(ba)b && \text{by the law (2)}\\
&= a(ab)b && \text{by the law (1)}\\
&= (aa)(bb) && \text{by the law (2)}\\
&= a^2b^2 && \text{by the definition of squaring.}
\end{aligned}$$

Thus, (3) is a formal corollary of (2) and (1) and holds whenever and wherever the two laws hold. That is why, when extending the set of natural numbers (positive integers) to the set of integers, and then to the set of rationals, and then to the set of reals, and then to the set of complex numbers, we have to take care of preserving (2) and (1) in the sense that it has to be proved that the laws persist under each of these extensions. In contrast, there is no need to bother with ‘secondary’ identities like (3) as their formal proofs carry over.

A big part of algebra in fact deals with inferring some useful ‘secondary identities’ from some ‘primary’ laws. Identities to be inferred may be quite complicated, and the inference itself may be highly nontrivial. Think, for instance, of the product rule for a determinant:

$$\det AB = \det A \det B. \tag{4}$$

It looks quite innocent due to convenient notation, but the reader certainly realizes that in fact (4) constitutes a powerful identity whose explicit form is rather bulky already for matrices of a modest size. Indeed, if, say, $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} x & y \\ z & t \end{smallmatrix}\right)$, then (4) amounts to the identity

$$(ax + bz)(cy + dt) - (ay + bd)(cx + dz) = (ad - bc)(xt - yz),$$

and even imagining the explicit form of (4) for 3×3-matrices is painful, to say nothing of actually writing it down.

However, one can observe that usually only a few ‘primary’ laws are invoked in the course of the inference even if it is cumbersome. For instance, to deduce the identity (4), one needs only the very basic laws, namely, the commutative and associative

laws of addition and multiplication, the distributive law of multiplication over addition, and the existence of subtraction (that is expressed by the law $a = (a - b) + b$). This observation leads to the idea of composing a *complete* list of 'primary' laws that would allow one to infer *every* possible identity. Such a list is called an *identity basis*. It should be mentioned that even though this usage of the word 'basis' is quite common, its meaning here differs from the standard meaning of this term in linear algebra since no independence assumptions are made: the only requirement for a collection of identities Σ to form an identity basis is that every identity should be deducible from Σ!

Of course, in order to speak about an identity basis, one has to specify which identities are under consideration. In this paper, we deal with the simplest nontrivial case of a single *binary* operation. The attribute 'binary' means that the operation involves two operands, like addition and multiplication of numbers do. Thus, a binary operation on a nonempty set S is merely a map $S \times S \to S$.

The principal question on which studies of identity bases are focused is known as the *Finite Basis Problem* (FBP, for short). For the purpose of this paper, the FBP may be formulated as follows.

The Finite Basis Problem. Given a structure $(S, \cdot)$ where $\cdot$ is a binary operation on a set S, determine whether or not the identities of $(S, \cdot)$ have a finite basis.

The FBP is natural by itself, but it 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 to classical number-theoretic conjectures such as the Twin Prime, Goldbach, existence of odd perfect numbers and the infinitude of even perfect numbers—it has been shown by Peter Perkins [15] that each of these conjectures is equivalent to the FBP for a structure of the form $(S, \cdot)$.

We say that a structure $(S, \cdot)$ is *finitely based* if the answer to the FBP for $(S, \cdot)$ is positive, that is, if the identities of $(S, \cdot)$ have a finite basis. Otherwise, $(S, \cdot)$ is called *nonfinitely based*.

Even a *finite* structure of the form $(S, \cdot)$ can be nonfinitely based. The smallest example is a 3-element structure known as Murskiĭ's groupoid [12]. However, arguably, the most striking example (known as the 6-element *Brandt monoid* B_2^1) is formed by the following six 2×2-matrices:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{5}$$

the operation being the usual matrix multiplication. The example is due to Perkins [14]. Thus, here we see a very transparent, very natural, and very finite structure whose identities cannot be axiomatized by finite means.

In the 1960s, Alfred Tarski [21] suggested to study the FBP for finite structures as a *decision problem*. Indeed, since any finite structure is an object that can be given in a constructive way, one can ask for an algorithm which when presented with an effective description of the structure, would determine whether or not it is finitely based.

Tarski's Finite Basis Problem. Is there an algorithm that when given an effective description of a finite structure decides whether it is finitely based or not?

This fundamental question was answered in the negative by Ralph McKenzie [11] who showed that no algorithm can decide the FBP for finite structures of the form $(S, \cdot)$. Thus, no mechanical procedure for answering to the FBP exists in general, and one should be more clever than a computer to get an answer!

2. The Finite Basis Problem for semigroups and our contribution

In this paper, we deal with the FBP for *semigroups*, that is, structures of the form $(S, \cdot)$ satisfying the associative law (2). Perkins's example cited in Section 1 revealed that finite semigroups can be nonfinitely based. Moreover, it turns out that semigroups are the only 'classical' algebras for which finite nonfinitely based objects can exist: finite groups [13], finite associative and Lie rings [3, 8, 9], finite lattices [10] are all finitely based. Therefore studying finite semigroups from the viewpoint of the FBP has become a hot area in which many neat results have been achieved and several powerful methods have been developed, see the survey [22] for an overview. The present paper develops a novel approach to the Finite Basis Problem for finite semigroups initiated in [19] and solves one of the problems posed in [22]. As an application, we answer a question left open in [23].

In order to describe our contribution in precise way, we proceed with introducing a few notions and setting up our notation. The basic concepts we need come from equational logic; see, e.g., [4, Chapter II]. For the reader's convenience, we present them here in a form adapted to the use in this paper, that is, specialized to semigroups. When doing so, we closely follow [19, Section 1].

A (*semigroup*) *word* is a finite sequence of symbols, called *variables*. Sometimes we employ the *empty word*, that is, the empty sequence. Whenever words under consideration are allowed to be empty, we always say it explicitly.

We denote words by lowercase boldface letters. If $\mathbf{w} = x_1 \cdots x_k$, where $x_1, \ldots, x_k$ are variables, possibly with repeats, then the set $\{x_1, \ldots, x_k\}$ is denoted by $\mathrm{alph}(\mathbf{w})$. If $\mathbf{w}$ is empty, then $\mathrm{alph}(\mathbf{w}) = \varnothing$.

Words are multiplied by concatenation, that is, for any words $\mathbf{w}$, $\mathbf{w}'$, the sequence $\mathbf{w}\mathbf{w}'$ is obtained by appending the sequence $\mathbf{w}'$ to the sequence $\mathbf{w}$.

Any map $\varphi\colon \mathrm{alph}(\mathbf{w}) \to S$, where S is a semigroup, is called a *substitution*. The *value* $\varphi(\mathbf{w})$ of $\mathbf{w}$ under φ is the element of S that results from substituting $\varphi(x)$ for each variable $x \in \mathrm{alph}(\mathbf{w})$ and computing the product in S.

A (*semigroup*) *identity* is a pair of words written as a formal equality. From now on, we use the sign $\approx$ when writing identities (so that a pair $(\mathbf{w}, \mathbf{w}')$, say, is written as $\mathbf{w} \approx \mathbf{w}'$), saving the standard sign $=$ for 'genuine' equalities. A semigroup S *satisfies* $\mathbf{w} \approx \mathbf{w}'$ (or $\mathbf{w} \approx \mathbf{w}'$ *holds* in S) if $\varphi(\mathbf{w}) = \varphi(\mathbf{w}')$ for every substitution $\varphi\colon \mathrm{alph}(\mathbf{w}\mathbf{w}') \to S$, that is, substitutions of elements from S for the variables occurring in $\mathbf{w}$ or $\mathbf{w}'$ yield equal values to these words.

In Section 1 we mentioned formal inference of identities. For semigroup identities, the inference rules are very transparent as they amount to substituting a word for each occurrence of a variable in an identity, multiplying an identity through on the right or the left by a word, and using symmetry and transitivity of equality. Birkhoff's completeness theorem of equational logic [4, Theorem 14.17] gives a clear semantic meaning to formal inference: an identity $\mathbf{w} \approx \mathbf{w}'$ can be inferred from a set Σ of identities if and only if every semigroup satisfying all identities in Σ satisfies the identity $\mathbf{w} \approx \mathbf{w}'$ as well. In this situation, we say that an identity $\mathbf{w} \approx \mathbf{w}'$ *follows* from Σ or that Σ *implies* $\mathbf{w} \approx \mathbf{w}'$.

As defined in Section 1, a semigroup S is *finitely based* if it possesses a finite identity basis and *nonfinitely based* otherwise. We mentioned at the start of this section that the FBP restricted to finite semigroups becomes nontrivial; moreover, its algorithmic version, that is, Tarski's Finite Basis Problem restricted to semigroups, remains open so far.

The class of all semigroups satisfying all identities from a given set Σ is called the *variety defined by* Σ. A variety is *finitely based* if it can be defined by a finite set of identities; otherwise it is *nonfinitely based*. Given a semigroup S, the variety defined by the set of all identities S satisfies is denoted by $\mathrm{var}\, S$ and called the *variety generated by* S. A variety is called *finitely generated* if it can be generated by a finite semigroup.

A variety is *locally finite* if each of its finitely generated members is finite. A finite semigroup is called *inherently nonfinitely based* if it is not contained in any finitely based locally finite variety. The very first example of an inherently nonfinitely based semigroup was discovered by Mark Sapir [17], who proved that the 6-element Brandt monoid B_2^1 is inherently nonfinitely based. In [16] he gave a structural characterization of all inherently nonfinitely based semigroups, which, in particular, led to an algorithm to recognize whether or not a given finite semigroup is inherently nonfinitely based. (This sharply contrasts McKenzie's result [11] that no such algorithm exists for general finite structures.)

It is easy to see that the satisfaction of an identity is inherited by forming direct products and taking *divisors* (that is, homomorphic images of subsemigroups)

of semigroups so that each variety is closed under these two operators. In fact, this closure property characterizes varieties (the HSP-theorem; see [4, Theorem 11.9]). An easy byproduct of the proof of the HSP-theorem (see [4, Theorem 10.16]) is that every finitely generated variety is locally finite. By the definition, a semigroup and the variety it generates are simultaneously finitely or nonfinitely based. Hence, to prove that a given finite semigroup S is nonfinitely based, it suffices to exhibit an inherently nonfinitely based semigroup in the variety $\operatorname{var} S$. This argument, combined with Sapir's characterization of all inherently nonfinitely based semigroups, has become one of the most powerful and easy-to-use methods in studying the FBP for finite semigroups.

Now let us quote from the survey [22].

> If one focuses on the finite basis problem for finite semigroups (like we do in this survey), then the notion of an inherently nonfinitely based semigroup appears to be rather abundant. Why should we care about locally finite varieties which are not finitely generated when we are only interested in finitely generated ones? This question leads us to introduce the following notion: call a finite semigroup S *strongly nonfinitely based* if S cannot be a member of any finitely based finitely generated variety. Clearly, every inherently nonfinitely based finite semigroup is strongly nonfinitely based, and the question if the converse is true is another intriguing open problem:
>
> **Problem 4.4.** Is there a strongly nonfinitely based finite semigroup which is not inherently nonfinitely based?

In this paper, we answer the question asked in [22, Problem 4.4] in the affirmative. Our example is the 42-element semigroup IC_4 from [19] where it was shown to have a weaker property. We recall the definition of the semigroup IC_4 and one of its features in Section 3 and then prove our main result in Section 4. Section 5 presents an application.

3. Preliminaries

Following [19, Section 2], we introduce the semigroup IC_4 as a member of a family of transformation monoids.

Let $[m]$ stand for the set of the first m positive integers ordered in the usual way: $1 < 2 < \cdots < m$. By a *partial transformation* of $[m]$ we mean an arbitrary map α from a subset of $[m]$ (called the *domain* of α and denoted $\operatorname{dom}\alpha$) to $[m]$. We write partial transformations on the right of their arguments. A partial transformation α is *order preserving* if $i \le j$ implies $i\alpha \le j\alpha$ for all $i, j \in \operatorname{dom}\alpha$, and *extensive* if $i \le i\alpha$ for every $i \in \operatorname{dom}\alpha$. Clearly, if two transformations have either of the properties of

being injective, order preserving, or extensive, then so does their product, and the identity transformation enjoys all three properties. Hence, the set of all partial injections of $[m]$ that are extensive and order preserving forms a monoid[1] that we denote by IC_m and call the mth *i-Catalan monoid*. Both 'I' in the notation and 'i' in the name mean 'injective'; the 'Catalan' part of the name again refers to the cardinality of the monoid: $|IC_m|$ is the $(m+1)$th Catalan number. In particular, $|IC_4|$ is the fifth Catalan number 42 a.k.a. the Answer to the Ultimate Question of Life, The Universe, and Everything; see [1].

The key property of the monoid IC_4 for this paper involves two combinatorial notions, which we now recall.

Let $\mathbf{u}$ be a word and x a variable in $\mathrm{alph}(\mathbf{u})$. If x occurs exactly once in $\mathbf{u}$, then the variable is called *linear* in $\mathbf{u}$. If x occurs more than once in $\mathbf{u}$, then we say that the variable is *repeated* in $\mathbf{u}$. A word $\mathbf{u}$ is called *sparse* if every two occurrences of a repeated variable in $\mathbf{u}$ sandwich some linear variable.

Given a semigroup S, a word $\mathbf{u}$ is called an *isoterm for S* if the only word $\mathbf{v}$ such that S satisfies the identity $\mathbf{u} \approx \mathbf{v}$ is the word $\mathbf{u}$ itself.

Lemma 1 ([19, Lemma 3.4]). *Every sparse word is an isoterm for the monoid IC_4.*

We also need some properties of a class of finite semigroups defined in terms of the Green relation $\mathscr{D}$. Recall that for a semigroup S, the notation S^1 stands for the least monoid containing S, that is, $S^1 := S$ if S has an identity element and $S^1 := S \cup \{1\}$ if S has no identity element; in the latter case the multiplication in S is extended to S^1 in a unique way such that the fresh symbol 1 becomes the identity element in S^1. James Alexander Green (cf. [5]) introduced five equivalence relations on every semigroup S which are collectively referred to as *Green's relations*. Of those five relations, we need the following four:

$$
\begin{aligned}
x \mathscr{R} y &\Leftrightarrow xS^1 = yS^1, && \text{i.e., } x \text{ and } y \text{ generate the same right ideal;}\\
x \mathscr{L} y &\Leftrightarrow S^1x = S^1y, && \text{i.e., } x \text{ and } y \text{ generate the same left ideal;}\\
x \mathscr{J} y &\Leftrightarrow S^1xS^1 = S^1yS^1, && \text{i.e., } x \text{ and } y \text{ generate the same ideal;}\\
x \mathscr{D} y &\Leftrightarrow (\exists z \in S)\, x \mathscr{R} z \wedge z \mathscr{L} y, && \text{i.e., } \mathscr{D} = \mathscr{R}\mathscr{L}.
\end{aligned}
$$

In addition, we write $x \leq_{\mathscr{J}} y$ if $x \in S^1yS^1$.

An element e of a semigroup S is called an *idempotent* if $e^2 = e$. We let **DS** stand for the class of all finite semigroups in which every $\mathscr{D}$-class containing an idempotent is a subsemigroup. It is well known (and easy to verify) that **DS** is a *pseudovariety*,

[1] Recall that a *monoid* is a semigroup with an identity element.

that is, a class of finite semigroups closed under forming finite direct products and taking divisors.

The following proposition summarizes the features of semigroups in **DS** that we employ. They can all be found (or readily follow from some results) in either Jorge Almeida's monograph [2], where the pseudovariety **DS** is comprehensively studied in Chapter 8, or Lev Shevrin's memoir [20], where Section 3 treats a semigroup class whose finite members exactly constitute **DS**.

Proposition 2. *Let S be a semigroup in* **DS**.

(a) *Every $\mathscr{D}$-class of S containing an idempotent is a union of its subgroups.*

(b) *If $\mathbf{u}, \mathbf{v}$ are words with $\mathrm{alph}(\mathbf{u}) = \mathrm{alph}(\mathbf{v})$, then for any substitution $\varphi\colon \mathrm{alph}(\mathbf{u}) \to S$ such that $\varphi(\mathbf{u})$ is an idempotent, $\varphi(\mathbf{u}) \leq_{\mathscr{J}} \varphi(\mathbf{v})$.*

(c) *If $e \leq_{\mathscr{J}} a$ and $e \leq_{\mathscr{J}} b$ for some idempotent $e \in S$ and some $a, b \in S$, then $aeb \mathscr{D} e$.*

Proof. Claim (a) is contained in [20, Theorem 3]; see conditions (4a) or (4c) there.

For (b), we use condition (1b) in [20, Theorem 3]. It provides a homomorphism ψ from S onto a commutative semigroup of idempotents such that for every idempotent e and every element a in S, the equality $\psi(e) = \psi(a)$ implies $e \leq_{\mathscr{J}} a$. (In terminology of [20], this fact is expressed by saying that S is a semilattice of Archimedean semigroups.) The condition $\mathrm{alph}(\mathbf{u}) = \mathrm{alph}(\mathbf{v})$ readily implies

$$\psi(\varphi(\mathbf{u})) = \psi(\varphi(\mathbf{v})) = \prod_{x \in \mathrm{alph}(\mathbf{u})} \psi(\varphi(x))$$

due to commutativity and idempotency of the semigroup $\psi(S)$. Hence, $\varphi(\mathbf{u}) \leq_{\mathscr{J}} \varphi(\mathbf{v})$.

Claim (c) follows from [2, Lemma 8.1.4] combined with the observation that $\mathscr{D} = \mathscr{J}$ on every finite semigroup [5, Theorem 3]. ∎

The proof of the next lemma closely follows the proof pattern of [2, Lemma 8.1.9], but is included for the sake of completeness.

Lemma 3. *Let $S \in$ **DS** and $k := |S|!$. Then for every word $\mathbf{u}$ that can be decomposed as $\mathbf{u} = \mathbf{u}_0\mathbf{u}_1\cdots\mathbf{u}_n$ with $n > |S|$ and $\mathrm{alph}(\mathbf{u}_0) = \mathrm{alph}(\mathbf{u}_1) = \cdots = \mathrm{alph}(\mathbf{u}_n)$, the identity $\mathbf{u} \approx \mathbf{u}^{k+1}$ holds in S.*

Proof. Let $\mathbf{w}_i := \mathbf{u}_0\mathbf{u}_1\cdots\mathbf{u}_i$. Take an arbitrary substitution $\varphi\colon \mathrm{alph}(\mathbf{u}) \to S$. For brevity, let $u_i := \varphi(\mathbf{u}_i)$ and $w_i := \varphi(\mathbf{w}_i)$. The n elements $w_0, w_1, \ldots, w_{n-1}$ cannot be all distinct, and so there exist indices p, q with $0 \leq p < q < n$ such that $w_p = w_q$. Hence,

$$w_q = w_p u_{p+1} u_{p+2} \cdots u_q = w_q u_{p+1} u_{p+2} \cdots u_q,$$

from which we deduce the equality

$$w_q = w_q(u_{p+1}u_{p+2}\cdots u_q)^k. \tag{6}$$

It is known (and easy to verify) that the kth power of any element of S is an idempotent. Since $\mathrm{alph}(\mathbf{w}_q) = \mathrm{alph}(\mathbf{u}_i) = \mathrm{alph}(\mathbf{u}_{q+1}\mathbf{u}_{q+2}\cdots\mathbf{u}_n)$ for $i = 0, 1, \ldots, n$, Proposition 2 (b) implies that

$$(u_{p+1}u_{p+2}\cdots u_q)^k \leq_{\mathscr{J}} w_q \quad \text{and} \quad (u_{p+1}u_{p+2}\cdots u_q)^k \leq_{\mathscr{J}} u_{q+1}u_{q+2}\cdots u_n.$$

Then by Proposition 2 (c) the element

$$w_n = w_q u_{q+1}u_{q+2}\cdots u_n \overset{(6)}{=} w_q(u_{p+1}u_{p+2}\cdots u_q)^k u_{q+1}u_{q+2}\cdots u_n$$

and the idempotent $(u_{p+1}u_{p+2}\cdots u_q)^k$ lie in the same $\mathscr{D}$-class. By Proposition 2 (a) the $\mathscr{D}$-class of the element w_n is a union of its subgroups. Thus, w_n belongs to a subgroup of S. Then the idempotent w_n^k is the identity element of this subgroup, and $w_n^{k+1} = w_n$. Consequently, we have

$$\varphi(\mathbf{u}) = w_n = w_n^{k+1} = \varphi(\mathbf{u}^{k+1}).$$

Since the substitution $\varphi\colon \mathrm{alph}(\mathbf{u}) \to S$ is arbitrary, S satisfies the identity $\mathbf{u} \approx \mathbf{u}^{k+1}$, as claimed. ∎

By B_2 we denote the subsemigroup of the Brandt monoid B_2^1 consisting of the five nonidentity matrices in (5). The following characterization of finite semigroups beyond **DS** occurs as [2, Exercise 8.1.6]; the solution to this exercise follows from [20, Theorem 3].

Lemma 4. *A finite semigroup S does not belong to the pseudovariety* **DS** *if and only if $S \times S$ has the semigroup B_2 as a divisor.*

For each idempotent e of a semigroup S, the set $eSe := \{ese \mid s \in S\}$ is a subsemigroup in which e serves as an identity element. We call eSe the *local submonoid of S at e*. By **LDS** we denote the class of all finite semigroups all of whose local submonoids lie in **DS**. The class **LDS** also forms a pseudovariety; see [2, Section 5.2]. We need the following corollary of Lemma 4.

Corollary 5. *A finite semigroup S does not belong to the pseudovariety* **LDS** *if and only if $S \times S$ has the monoid B_2^1 as a divisor.*

Proof. For the 'if' part, observe that $B_2^1 \notin$ **LDS**. Indeed, B_2^1 is a local submonoid of itself, and the four matrix units in (5) form a $\mathscr{D}$-class that contains an idempotent matrix but is not closed under matrix multiplication. Now the claim follows from **LDS** being closed under forming finite direct products and taking divisors.

For the 'only if' part, take an arbitrary finite semigroup $S \notin \mathbf{LDS}$. Then for some idempotent $e \in S$, the local submonoid eSe does not belong to the pseudovariety $\mathbf{DS}$. By Lemma 4 we conclude that the monoid $T := eSe \times eSe$ has the semigroup B_2 as a divisor. Consider a subsemigroup U of T such that there exists an onto homomorphism $\varphi\colon U \to B_2$. The identity element $f := (e, e)$ of T cannot belong to U since otherwise its image $\varphi(f)$ would be an identity element in B_2, and B_2 has no identity element. The union $U' = U \cup \{f\}$ is a subsemigroup of T. We extend the homomorphism φ to an onto map $\varphi'\colon U' \to B_2^1$, letting $\varphi'(f) := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Clearly, φ' is a homomorphism whence the monoid B_2^1 as a divisor of T which is a submonoid of $S \times S$. ∎

4. Main result

The paper [19] has promoted the idea of relativizing the property of being inherently nonfinitely based (first suggested in [7] in the context of quasivarieties). If $\mathbf{C}$ is a class of semigroups, a semigroup T is called *inherently nonfinitely based relative to* $\mathbf{C}$ if every semigroup $S \in \mathbf{C}$ such that $T \in \operatorname{var} S$ is nonfinitely based. Specializing $\mathbf{C}$, one gets various concepts that occur in the literature. For instance, the property of being inherently nonfinitely based as considered by Mark Sapir in [16, 17] arises when $\mathbf{C}$ consists of all semigroups that generate locally finite varieties. If $\mathbf{C}$ is the class of all finite semigroups, one gets the property of being strongly nonfinitely based discussed in Section 2.

Theorem 3.1 in [19] shows that the i-Catalan monoid IC_4 is inherently nonfinitely based relative to the class of all finite semigroups in which Green's relation $\mathscr{R}$ is trivial (that is, coincides with the equality relation). We strengthen this result in Theorem 7 below, but first we provide a sufficient condition on a class of semigroups, under which IC_4 is inherently nonfinitely based relative to this class.

We fix a countably infinite set $\mathfrak{A}$ of variables. Denote by $\mathfrak{A}^+$ the set of all words whose variables lie in $\mathfrak{A}$ and let $\mathfrak{A}^*$ be $\mathfrak{A}^+$ with the empty word added. We assume that all words that we encounter below come from $\mathfrak{A}^*$.

Let $\mathbf{w}$ be a word. For $X \subseteq \operatorname{alph}(\mathbf{w})$, we denote by $\mathbf{w}(X)$ the word obtained from $\mathbf{w}$ by removing all occurrences of variables from $\operatorname{alph}(\mathbf{w}) \setminus X$. An occurrence of a word $\mathbf{u}$ in a word $\mathbf{w}$ as a *factor* is any decomposition of the form $\mathbf{w} = \mathbf{v}'\mathbf{u}\mathbf{v}''$, where the words $\mathbf{v}', \mathbf{v}''$ may be empty. If such a decomposition of $\mathbf{w}$ is unique, then we say that the factor $\mathbf{u}$ occurs in $\mathbf{w}$ once; otherwise, $\mathbf{u}$ occurs in $\mathbf{w}$ more than once.

Proposition 6. *Suppose that* $\mathbf{C}$ *is a class of semigroups and for each semigroup* $S \in \mathbf{C}$ *such that the* i*-Catalan monoid* IC_4 *belongs to the variety* $\operatorname{var} S$*, there exist an infinite*

sequence $\{\mathbf{u}_n \approx \mathbf{v}_n\}$ *of identities holding in* S *and an infinite sequence* $\{X_n\}$ *of sets of variables such that*

(P0) $\mathbf{u}_n(X_n) \neq \mathbf{v}_n(X_n)$*;*

(P1) *for all variables* y, z*, not necessarily distinct, the word* yz *occurs in* $\mathbf{u}_n(X_n)$ *as a factor at most once;*

(P2) *for every variable* z*, there are at least* n *pairwise distinct variables between any two occurrences of* z *in* $\mathbf{u}_n(X_n)$.

Then the i*-Catalan monoid* IC_4 *is inherently nonfinitely based relative to the class* $\mathbf{C}$.

Proof. We have to verify that each semigroup $S \in \mathbf{C}$ such that $IC_4 \in \operatorname{var} S$ is nonfinitely based. For this, it suffices to exhibit a property θ of words such that

(i) the word $\mathbf{u}_n$ has the property θ, while the word $\mathbf{v}_n$ does not have the property θ;

(ii) for an arbitrary identity $\mathbf{u}_n \approx \mathbf{u}$ of S such that the word $\mathbf{u}$ has the property θ, an application of any identity of S in less than $n-2$ variables to the word $\mathbf{u}$ preserves the property θ.

Indeed, a standard syntactic argument (see [22, Section 4] or [18, Fact 2.1]) then implies that for each n, the identity $\mathbf{u}_n \approx \mathbf{v}_n$ cannot be inferred from identities in less than $n-2$ variables holding in S. Therefore, no finite set of identities holding in S can infer all identities of this semigroup.

We show that the following property θ is relevant: a word $\mathbf{w}$ has θ if $\mathbf{w}(X_n) = \mathbf{u}_n(X_n)$. Evidently, (i) holds by the property (P0). It remains to verify (ii) provided that $IC_4 \in \operatorname{var} S$.

Let $\mathbf{u}_n \approx \mathbf{u}$ be an identity of S such that $\mathbf{u}(X_n) = \mathbf{u}_n(X_n)$. We need to establish that if a word $\mathbf{v}$ is obtained from $\mathbf{u}$ by an application of some identity $\mathbf{s} \approx \mathbf{t}$ of S in less than $n-2$ variables, then $\mathbf{v}(X_n) = \mathbf{u}_n(X_n)$. Obtaining $\mathbf{v}$ from $\mathbf{u}$ by an application of $\mathbf{s} \approx \mathbf{t}$ means that $\mathbf{u} = \mathbf{c}\,\varphi(\mathbf{s})\,\mathbf{d}$ and $\mathbf{v} = \mathbf{c}\,\varphi(\mathbf{t})\,\mathbf{d}$ for some $\mathbf{c}, \mathbf{d} \in \mathfrak{A}^*$ and some substitution $\varphi\colon \operatorname{alph}(\mathbf{st}) \to \mathfrak{A}^+$.

Take two variables $c, d \notin \operatorname{alph}(\mathbf{st})$. The identity $\mathbf{s} \approx \mathbf{t}$ implies each of the identities $c\,\mathbf{s}\,d \approx c\,\mathbf{t}\,d$, $c\,\mathbf{s} \approx c\,\mathbf{t}$, and $\mathbf{s}\,d \approx \mathbf{t}\,d$. If the words $\mathbf{c}$ and $\mathbf{d}$ are nonempty, then $\mathbf{u} = \psi(c\,\mathbf{s}\,d)$ and $\mathbf{v} = \psi(c\,\mathbf{t}\,d)$, where $\psi\colon \operatorname{alph}(\mathbf{st}\,cd) \to \mathfrak{A}^+$ is the substitution given by $\psi(c) := \mathbf{c}$, $\psi(d) := \mathbf{d}$ and $\psi(x) := \varphi(x)$ for each $x \in \operatorname{alph}(\mathbf{st})$. Similarly, if one of the words $\mathbf{c}$ and $\mathbf{d}$ is empty while the other is not, then the words $\mathbf{u}$ and $\mathbf{v}$ are images of either the words $c\,\mathbf{s}$ and respectively $c\,\mathbf{t}$ or the words $\mathbf{s}\,d$ and respectively $\mathbf{t}\,d$ under a suitable substitution. It follows that we may assume without any loss that $\mathbf{u} = \varphi(\mathbf{s})$ and $\mathbf{v} = \varphi(\mathbf{t})$, and $\mathbf{s} \approx \mathbf{t}$ is an identity of S in less than n variables.

Let

$$Y_n := \{z \in \operatorname{alph}(\mathbf{st}) \mid \operatorname{alph}(\varphi(z)) \cap X_n \neq \varnothing\}.$$

Let us verify that the word $\mathbf{s}(Y_n)$ is sparse. Indeed, for every repeated variable y of $\mathbf{s}$, the word $\varphi(y)$ occurs as a factor in $\mathbf{u}$ more than once. In view of the property (P1), we see that for each variable $y \in Y_n$ repeated in $\mathbf{s}$, the word $\varphi(y)(X_n)$ must be a single variable $x \in X_n$, say. Now choose two occurrences of ${}_1y$ and ${}_2y$ of y in $\mathbf{s}$ and let ${}_1x$ and ${}_2x$ be the corresponding occurrences of x in $\mathbf{u}$. By the property (P2) there are at least n pairwise distinct variables from X_n between ${}_1x$ and ${}_2x$ in $\mathbf{u}$. Since $|\operatorname{alph}(\mathbf{s})| < n$ and Y_n is the set of all variables whose images under φ contain variables from X_n, there must be a variable $t \in Y_n \cap \operatorname{alph}(\mathbf{s})$ such that $\varphi(t)$ involves at least two variables in X_n. In view of the property (P1), the variable t must be linear in $\mathbf{s}$. Therefore, the word $\mathbf{s}(Y_n)$ is sparse.

Since $IC_4 \in \operatorname{var} S$, the identity $\mathbf{s} \approx \mathbf{t}$ holds in IC_4. As IC_4 is a monoid, so does the identity $\mathbf{s}(Y_n) \approx \mathbf{t}(Y_n)$ since removing all occurrences of variables from $\operatorname{alph}(\mathbf{st}) \setminus Y_n$ has the same effect as substituting the identity element of IC_4 for these variables. By Lemma 1 every sparse word is an isoterm for the i-Catalan monoid IC_4. It follows that $\mathbf{t}(Y_n) = \mathbf{s}(Y_n)$. Hence, $\mathbf{v}(X_n) = \mathbf{u}(X_n) = \mathbf{u}_n(X_n)$, as required. ∎

Theorem 7. *The i-Catalan monoid IC_4 is inherently nonfinitely based relative to the pseudovariety* **LDS**.

Proof. Take any $S \in \mathbf{LDS}$ such that the variety $\operatorname{var} S$ contains IC_4; we have to prove that S is nonfinitely based.

Let $k = |S|!$; then the kth power of any element of S is an idempotent. In view of Proposition 6, it suffices to find an infinite sequence $\{\mathbf{u}_n \approx \mathbf{v}_n\}$ of identities holding in S and an infinite sequence $\{X_n\}$ of sets of variables such that the properties (P0), (P1) and (P2) hold. We will show that the following are relevant:

$$\mathbf{u}_n := \prod_{i=0}^{n} \mathbf{a}_n[\pi^i]\mathbf{b}_n[\pi^i], \quad \mathbf{v}_n := \mathbf{u}_n^{k+1},$$
$$X_n := \{x_0, y_0, z_0, x_1, y_1, z_1, \ldots, x_n, y_n, z_n\},$$

where π denotes the cyclic permutation $(01\cdots n)$ of the set $\{0, 1, \ldots, n\}$, and

$$\mathbf{a}_n[\tau] := x^k x_{0\tau} x^k y_1 x^k x_{1\tau} x^k y_2 x^k x_{2\tau} x^k \cdots x^k y_n x^k x_{n\tau} x^k,$$
$$\mathbf{b}_n[\tau] := x^k z_{0\tau} x^k y_1 x^k z_{1\tau} x^k y_2 x^k z_{2\tau} x^k \cdots x^k y_n x^k z_{n\tau} x^k$$

for any permutation τ of $\{0, 1, \ldots, n\}$.

By the definitions of the identities $\mathbf{u}_n \approx \mathbf{v}_n$ and the sets X_n, the properties (P0), (P1) and (P2) hold for each n. Since

$$\operatorname{alph}(\mathbf{a}_n[\pi^0]\mathbf{b}_n[\pi^0]) = \cdots = \operatorname{alph}(\mathbf{a}_n[\pi^i]\mathbf{b}_n[\pi^i]) = \cdots = \operatorname{alph}(\mathbf{a}_n[\pi^n]\mathbf{b}_n[\pi^n]),$$

Lemma 3 implies that every local submonoid of S satisfies the identity $\mathbf{u}_n(X_n) \approx \mathbf{v}_n(X_n)$ for all $n > |S|$. Since the kth power of any element of S is an idempotent, this implies that the identity $\mathbf{u}_n \approx \mathbf{v}_n$ holds in S. Theorem 7 is proved. ∎

Now it easy to deduce our main result. Recall that a semigroup is said to be strongly nonfinitely based if it is inherently nonfinitely based relative to the class of all finite semigroups.

Theorem 8. *The i-Catalan monoid IC_4 is strongly nonfinitely based.*

Proof. Take any finite semigroup S such that the variety $\operatorname{var} S$ contains IC_4; we have to prove that S is nonfinitely based. If $S \in \mathbf{LDS}$, this follows from Theorem 7. If $S \notin \mathbf{LDS}$, then Corollary 5 implies that the variety $\operatorname{var} S$ contains the 6-element Brandt monoid B_2^1. Since B_2^1 is inherently nonfinitely based [17, Corollary 6.1], we conclude that S is nonfinitely based is this case as well. ∎

It readily follows from the structural characterization of inherently nonfinitely based semigroups [16, Theorem 1] that such a semigroup must have a nonsingleton $\mathscr{D}$-class. Since all $\mathscr{D}$-classes of the i-Catalan monoid IC_4 are singletons, we conclude that IC_4 is not inherently nonfinitely based. Thus, Theorem 8 provides an example of a strongly nonfinitely based semigroup which is not inherently nonfinitely based, answering the question from [22] quoted in Section 2.

Remark 9. Reviewing the proofs of Proposition 6 and Theorems 7 and 8, one sees that all our arguments rely on only two properties of IC_4: that IC_4 is a monoid and that every sparse word is an isoterm for IC_4. Therefore, any monoid for which every sparse word is an isoterm is strongly nonfinitely based. Using this, the first-named author has constructed a strongly nonfinitely based monoid with only 9 elements which is not inherently nonfinitely based. This result will be published separately.

Remark 10. We point out a subtle yet important difference between the concept of being inherently nonfinitely based as considered in [16, 17] and that of being strongly nonfinitely based. The difference comes from the fact that the local finiteness of a variety is inherited by its subvarieties while the property of being finitely generated is not. Therefore, if a semigroup S is not contained in any finitely based locally finite semigroup variety, then S is contained in no finitely based locally finite variety $\mathbf{V}$ of groupoids—otherwise, the intersection of $\mathbf{V}$ with the variety of all semigroups would be a finitely based locally finite variety of semigroups containing S. Thus, when we speak about inherently nonfinitely based semigroups, it is unnecessary to specify within which class we work. In contrast, when we speak about strongly nonfinitely based semigroups, we should distinguish between the 'absolute' case and the case when we work within the class of all semigroups. In the present paper we

have only proved that every finitely generated *semigroup* variety containing the monoid IC_4 is nonfinitely based. This does not exclude the possibility that some finitely based finitely generated *groupoid* variety contains IC_4. The question of whether or not there exists a semigroup which, being not inherently nonfinitely based, is strongly nonfinitely based relative to the class of all finite groupoids still remains open.

For a more detailed discussion of the property of being strongly nonfinitely based in a broader universal-algebraic context, we refer the reader to [6, Section 1.1].

5. An application

Theorem 8 can be applied to prove the absence of a finite identity basis for many finite semigroups for which the FBP remained open so far. Here we restrict ourselves to just one application, resolving a question left open in [23].

Let $T_n(q)$ stand for the semigroup of all upper triangular $n \times n$-matrices over the finite field with q elements. In [23], it was shown that the semigroup $T_n(q)$ is inherently infinitely based if and only if $q > 2$ and $n > 3$. Thus, semigroups of upper triangular matrices over the 2-element field turn out to be not inherently nonfinitely based, but the question of whether or not they are finitely based remained unsolved for 20 years, with the only exception of the 8-element semigroup $T_2(2)$ that was proved to be finitely based in [24]. Now we are in a position to answer the question for all $n > 3$; the case $n = 3$ still remains open.

Theorem 11. *For each $n > 3$, the semigroup $T_n(2)$ of all upper triangular $n \times n$-matrices over the 2-element field is (strongly) nonfinitely based.*

Proof. Due to Theorem 8, it suffices to show that for each $n > 3$, the variety $\operatorname{var} T_n(2)$ contains the i-Catalan monoid IC_4. In fact, we construct an embedding $IC_4 \to T_4(2)$; since $T_4(2)$ naturally embeds into $T_n(2)$ for all $n \geq 4$, the claim will follow.

Recall that the monoid IC_4 consists of all extensive and order preserving partial injections of the chain $1 < 2 < 3 < 4$ into itself. Given any such partial injection α, we define a 4×4-matrix $A := (a_{ij})$ over the 2-element field by setting

$$a_{ij} := \begin{cases} 1 & \text{if } i\alpha = j, \\ 0 & \text{otherwise.} \end{cases}$$

Since α is extensive, $i\alpha = j$ implies $i \leq j$ whence the matrix A is upper triangular. Clearly, the map $\alpha \mapsto A$ is one-to-one, and it is easy to verify that the map is a homomorphism, using the fact that the image of IC_4 consists of row-monomial matrices so that one never adds two 1s when multiplying such matrices. ∎

Acknowledgements. The authors thank the anonymous referees for their feedback and helpful suggestions.

Funding. S. V. Gusev and M. V. Volkov were supported by the Ministry of Science and Higher Education of the Russian Federation, project FEUZ-2023-0022.

References

[1] D. Adams, *The hitchhiker's guide to the galaxy*. Pan Books, London, 1979

[2] J. Almeida, *Finite semigroups and universal algebra*. Ser. Algebra 3, World Scientific, River Edge, NJ, 1994 Zbl 0844.20039 MR 1331143

[3] Yu. A. Bakhturin and A. Yu. Ol'shanskiĭ, Identical relations in finite Lie rings. *Mat. Sb. (N.S.)* **96(138)** (1975), no. 4, 543–559 (Russian), English translation: *Mathematics of the USSR–Sbornik* **25** (1975), 507–523 Zbl 0332.17005 MR 0374219

[4] S. Burris and H. P. Sankappanavar, *A course in universal algebra*. Grad. Texts in Math. 78, Springer, New York-Berlin, 1981 Zbl 0478.08001 MR 0648287

[5] J. A. Green, On the structure of semigroups. *Ann. of Math. (2)* **54** (1951), 163–172 Zbl 0043.25601 MR 0042380

[6] M. Jackson and G. F. McNulty, The equational complexity of Lyndon's algebra. *Algebra Universalis* **65** (2011), no. 3, 243–262 Zbl 1230.08002 MR 2793398

[7] M. Jackson and M. Volkov, Relatively inherently nonfinitely q-based semigroups. *Trans. Amer. Math. Soc.* **361** (2009), no. 4, 2181–2206 Zbl 1171.08003 MR 2465833

[8] R. L. Kruse, Identities satisfied by a finite ring. *J. Algebra* **26** (1973), 298–318 Zbl 0276.16014 MR 0325678

[9] I. V. L'vov, Varieties of associative rings. I. *Algebra i Logika* **12** (1973), 269–297 (Russian), English translation: *Algebra and Logic* **12** (1973), 150–167 Zbl 0288.16008 MR 0389973

[10] R. McKenzie, Equational bases for lattice theories. *Math. Scand.* **27** (1970), 24–38 Zbl 0307.08001 MR 0274353

[11] R. McKenzie, Tarski's finite basis problem is undecidable. *Internat. J. Algebra Comput.* **6** (1996), no. 1, 49–104 Zbl 0844.08011 MR 1371734

[12] V. L. Murskiĭ, The existence in the three-valued logic of a closed class with a finite basis having no finite complete system of identities. *Dokl. Akad. Nauk SSSR* **163** (1965), 815–818 (Russian), English translation: *Soviet Math. Dokl.* **6** (1965), 1020–1024 Zbl 0154.25506 MR 0186539

[13] S. Oates and M. B. Powell, Identical relations in finite groups. *J. Algebra* **1** (1964), 11–39 Zbl 0121.27202 MR 0161904

[14] P. Perkins, Bases for equational theories of semigroups. *J. Algebra* **11** (1969), 298–314 Zbl 0186.03401 MR 0233911

[15] P. Perkins, Finite axiomatizability for equational theories of computable groupoids. *J. Symbolic Logic* **54** (1989), no. 3, 1018–1022 Zbl 0695.03023 MR 1011189

[16] M. V. Sapir, Inherently non-finitely based finite semigroups. *Mat. Sb. (N.S.)* **133(175)** (1987), no. 2, 154–166 (Russian), English translation: *Mathematics of the USSR–Sb.* **61** (1988), 155–166 Zbl 0655.20045 MR 0905002

[17] M. V. Sapir, Problems of Burnside type and the finite basis property in varieties of semigroups. *Izv. Akad. Nauk SSSR Ser. Mat.* **51** (1987), no. 2, 319–340 (Russian), English translation: *Mathematics of the USSR–Izv.* **30** (1988), 295–314 Zbl 0646.20047 MR 0897000

[18] O. B. Sapir, Non-finitely based monoids. *Semigroup Forum* **90** (2015), no. 3, 557–586 Zbl 1346.20075 MR 3345943

[19] O. B. Sapir and M. V. Volkov, Catalan monoids inherently nonfinitely based relative to finite $\mathscr{R}$-trivial semigroups. *J. Algebra* **633** (2023), 138–171 Zbl 1521.20124 MR 4610784

[20] L. N. Shevrin, On the theory of epigroups. I. *Mat. Sb.* **185** (1994), no. 8, 129–160 (Russian), English translation: *Russian Acad. Sci. Sb. Math.* **82** (1995), no. 2, 485–512 Zbl 0839.20073 MR 1302627

[21] A. Tarski, Equational logic and equational theories of algebras. In *Contributions to Math. Logic (Colloquium, Hannover, 1966)*, pp. 275–288, North-Holland, Amsterdam, 1968 Zbl 0209.01402 MR 0237410

[22] M. V. Volkov, The finite basis problem for finite semigroups. *Sci. Math. Jpn.* **53** (2001), no. 1, 171–199 Zbl 0990.20039 MR 1821612

[23] M. V. Volkov and I. A. Gol'dberg, Identities of semigroups of triangular matrices over finite fields. *Mat. Zametki* **73** (2003), no. 4, 502–510 (Russian), English translation: *Math. Notes* **73** (2003), no. 4, 474–481 Zbl 1064.20056 MR 1991897

[24] W. T. Zhang, J. R. Li, and Y. F. Luo, On the variety generated by the monoid of triangular 2×2 matrices over a two-element field. *Bull. Aust. Math. Soc.* **86** (2012), no. 1, 64–77 Zbl 1260.20076 MR 2960228

Received 16 January 2024; revised 11 June 2024.

Sergey V. Gusev
Institute of Natural Sciences and Mathematics, Ural Federal University, Lenina 51, 620000 Yekaterinburg, Russia; sergey.gusev@urfu.ru, sergey.gusb@gmail.com

Olga B. Sapir
Department of Mathematics, Ben-Gurion University of the Negev, P.O. Box 653, 8410501 Beersheba, Israel; olga.sapir@gmail.com

Mikhail V. Volkov
Institute of Natural Sciences and Mathematics, Ural Federal University, Lenina 51, 620000 Yekaterinburg, Russia; m.v.volkov@urfu.ru, mishavolkov@gmail.com