Strongly nonfinitely based monoids

Mikhail Volkov

joint work with Sergey Gusev and Olga Sapir

PALS, March 21st, 2024

The Finite Basis Problem

The Finite Basis Problem (FBP) underlies the research reported in this talk.

The Finite Basis Problem

Given an interesting structure $\mathcal{S}$ (a set with a bunch of operations on it), determine whether or not the identities of S admit a finite basis, that is, a finite list of identities that infer every identity holding in $\mathcal{S}$.

If the identities of $\mathcal{S}$ admit a finite basis, we say that $\mathcal{S}$ is finitely based; otherwise $\mathcal{S}$ is said to be nonfinitely based. *FB or not FB, that is the question...*

The Finite Basis Problem for Finite Structures

Even a finite structure can be nonfinitely based. The smallest example is a 3-element structure of the form $(S,\cdot)$ known as Murskiĭ's groupoid.
Recently, another example of a nonfinitely based 3-element structure has been discovered by Marcel Jackson, Miaomiao Ren, and Xianzhong Zhao (Nonfinitely based ai-semirings with finitely based semigroup reducts, J. Algebra 611 (2022), 211–245): a 3-element semiring with idempotent and commutative addition and commutative multiplication.
A classical example is the Brandt monoid $\mathcal{B}_2^1$ formed by the following six 2×2-matrices:

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

the operation being the usual matrix multiplication. (Peter Perkins, Bases for equational theories of semigroups, J. Algebra 11 (1969), 298–314.)

Thus, here we see a very transparent, very natural, and very finite structure whose identities cannot be axiomatized by finite means.

Tarski's Finite Basis Problem

In the early 1960's, Tarski suggested to study the FBP for finite structures as a decision problem. Indeed, since any finite structure $\mathcal{S}$ 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 $\mathcal{S}$, would determine whether or not $\mathcal{S}$ is finitely based.

Tarski's Finite Basis Problem

Is there an algorithm that when given an effective description of a finite structure $\mathcal{S}$ decides whether $\mathcal{S}$ is finitely based or not?

This fundamental question was answered in the negative by Ralph McKenzie (Tarski's finite basis problem is undecidable, Int. J. Algebra Comput. 6 (1996), 49–104), even for finite structures with a single operation!

I think it is a good news for people involved in studying the FBP: since no mechanical procedure exists, you should be more clever than your computer to get an answer!

Finite Basis Problem for Finite Semigroups

The restriction of Tarski's Finite Basis Problem to the class of finite semigroups still remains open.

In fact, finite semigroups are unique with respect to the FBP: as already mentioned, for finite groupoids, the problem is known to be undecidable; on the other hand, finite groups, finite associative and Lie rings, finite lattices are all known to be finitely based whence Tarski's Finite Basis Problem restricted to these algebras is trivially decidable.

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.

In 1999–2000, I wrote a survey in which I tried to analyze the methods that existed at that time and to outline possible directions for further advances:

- The finite basis problem for finite semigroups: a survey, in P. Smith, E. Giraldes, P. Martins (eds.), Semigroups, World Scientific, 2000, 244–279;
- The finite basis problem for finite semigroups, Sci. Math. Jpn. 53 (2001), 171–199.

Inherently Nonfinitely Based Semigroups

The class of all semigroups satisfying all identities from a given set Σ is 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 $\mathcal{S}$, the variety defined by all identities of $\mathcal{S}$ is denoted by $\operatorname{var}\mathcal{S}$ and called the variety generated by $\mathcal{S}$.

A variety is finitely generated if it can be generated by a finite semigroup.

A variety is locally finite if all its finitely generated members are finite.

It is known that every finitely generated variety is locally finite.

A finite semigroup $\mathcal{S}$ is inherently nonfinitely based (INFB) if S does not belong to any finitely based locally finite variety.

Hence, if $\mathcal{S}$ is INFB and $\mathcal{T}$ is a finite semigroup such that $\mathcal{S} \in \operatorname{var}\mathcal{T}$, then $\mathcal{T}$ is nonfinitely based.

Thus, to prove that a given finite semigroup $\mathcal{T}$ is nonfinitely based, it suffices to exhibit an inherently nonfinitely based semigroup in the variety $\operatorname{var}\mathcal{T}$ — a powerful and easy-to-use method provided one has some supply of INFB semigroups. In fact, a priori it is not obvious that INFB semigroups exist.

A Bit of History

The term INFB was suggested by Peter Perkins (Basic questions for general algebras, Algebra Universalis 19 (1984), 16–23).

The very first example of an INFB finite algebra (in fact, a 3-element groupoid) was exhibited by Vadim Murskiĭ (On the number of k-element algebras with one binary operation without a finite basis of identities, Problemy Kibernet. 35 (1979), 5–27 [Russian]).

The question of whether or not the 6-element Brandt monoid $\mathcal{B}_2^1$ is INFB was put forward and promoted by the late George McNulty at the end of the 1970s.

Mark Sapir proved that $\mathcal{B}_2^1$ is INFB (Problems of Burnside type and the finite basis property in varieties of semigroups, Izv. Akad. Nauk SSSR, Ser. Mat. 51 (1987), 319–340 [Russian; Engl. translation Math. USSR–Izv. 30 (1987), 295–314]).

This was a consequence of a combinatorial characterization of INFB semigroups.

Later Sapir gave an algorithmically efficient structural description of INFB semigroups (Inherently nonfinitely based finite semigroups, Mat. Sb. 133, no.2 (1987), 154–166 [Russian; Engl. translation Math. USSR–Sb. 61 (1988), 155–166]).

Sapir's Combinatorial Description

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

A word u over an alphabet A avoids a word v over an alphabet X if u does not contain the value of v under any substitution $X \to A^+$.

A word w is unavoidable if over every finite alphabet there are only finitely many words that avoid w.

Theorem (Sapir, 1987)

A finite semigroup $\mathcal{S}$ is INFB iff every unavoidable word is an isoterm for $\mathcal{S}$.

One can make this result more concrete by using the Zimin words:

$$Z_1 := x_1, Z_2 := x_1x_2x_1, \ldots, Z_n := Z_{n-1}x_nZ_{n-1}, \ldots .$$

An equivalent formulation

A finite semigroup $\mathcal{S}$ is INFB iff every Zimin word is an isoterm for $\mathcal{S}$.

A Problem

Now let me quote from my 2000 survey:

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 $\mathcal{S}$ strongly nonfinitely based if $\mathcal{S}$ 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?

After 20+ years, we managed to answer this question in the affirmative.

Our First Example

Let $[m]$ stand for the chain $1 < 2 < \cdots < m$.
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 $[m]$.
A partial transformation α is order preserving if $i \leq j$ implies $i\alpha \leq j\alpha$ for all $i, j \in \operatorname{dom}\alpha$, and extensive if $i \leq i\alpha$ for every $i \in \operatorname{dom}\alpha$.
The set of all partial injections of $[m]$ that are extensive and order preserving forms a monoid that we denote by $\mathcal{IC}_m$ 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 refers to the cardinality of the monoid: $|\mathcal{IC}_m|$ is the $(m+1)$-th Catalan number. In particular, $|\mathcal{IC}_4|$ is the 5th Catalan number 42 aka the Answer to the Ultimate Question of Life, The Universe, and Everything.

Theorem

The i-Catalan monoid $\mathcal{IC}_4$ is strongly nonfinitely based.

Catalan Monoids

The Catalan monoid, denoted $\mathcal{C}_m$, is the monoid of all extensive and order preserving full transformations of $[m]$.
The cardinality of $\mathcal{C}_m$ is the mth Catalan number so that $|C_5| = 42$.

It can be shown that $\mathcal{IC}_m \in \operatorname{var} \mathcal{C}_{m+1}$ for all m (non-trivial!). Hence, we have

Corollary

The Catalan monoid $\mathcal{C}_5$ is strongly nonfinitely based.

There are a plethora of applications of these results to the FBP for finite semigroups, both explaining previously known facts from a unified viewpoint and solving the FBP for many as yet unexplored classes of finite semigroups.

- O.B. Sapir, M.V. Volkov, Catalan monoids inherently nonfinitely based relative to finite R-trivial semigroups, J. Algebra 633 (2023), 138–171;
- S.V. Gusev, O.B. Sapir, M.V. Volkov, Strongly nonfinitely based monoids, submitted, see also `https://arxiv.org/abs/2301.12426`.

skip example

Example: 0-Hecke Monoids

Coxeter groups are classical objects in algebra, geometry and combinatorics. A symmetric matrix $CD = (m_{ij})_{n\times n}$ with entries in $\mathbb{N} \cup \{+\infty\}$ is called a Coxeter matrix if $m_{ii} = 1$ for all i and $m_{ij} \geq 2$ for all $i \neq j$. Example: $\left(\begin{smallmatrix} 1 & 4 & 2 \\ 4 & 1 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right)$.

Such a matrix is represented as the Coxeter diagram with vertices $1, 2, \ldots, n$ that has the edge i ——— j if and only if $m_{ij} \geq 3$; in addition, if $m_{ij} > 3$, the edge is labeled m_{ij}. Our example matrix is depicted as the Coxeter diagram

B_3: • $\underset{4}{\text{———}}$ • ——— •.

If $CD = (m_{ij})_{n\times n}$ is a Coxeter matrix, then the Coxeter group $W(CD)$ is the group generated by $s_1, s_2, \ldots, s_n$ subject to the relations

$$(s_i s_j)^{m_{ij}} = 1 \text{ for all } i, j = 1, 2, \ldots, n \text{ such that } m_{ij} \neq +\infty. \qquad (\star)$$

Since $m_{ii} = 1$, the relations $(\star)$ for $i = j$ mean $s_i^2 = 1$, that is, each generator s_i is an involution. Using this, one can rewrite the relations $(\star)$ for $i \neq j$ as

$$\underbrace{s_i s_j \cdots}_{m_{ij} \text{ factors}} = \underbrace{s_j s_i \cdots}_{m_{ij} \text{ factors}}.$$

Example: 0-Hecke Monoids, continued

The 0-Hecke monoid of the group $W(CD)$ is the monoid $H_0(CD)$ generated by $s_1, s_2, \ldots, s_n$ subject to the relations

$$\underbrace{s_i s_j \cdots}_{m_{ij} \text{ factors}} = \underbrace{s_j s_i \cdots}_{m_{ij} \text{ factors}} .$$

for all $i \neq j$ such that $m_{ij} \neq +\infty$ and $s_i^2 = s_i$ for all $i = 1, 2, \ldots, n$.
Thus, one passes from $W(CD)$ to $H_0(CD)$ by merely converting each involution s_i into an idempotent with the same name.
Even though the 0-Hecke monoid of a Coxeter group radically differs from the group as an algebraic object, the monoid and the group share many combinatorial features. The reason for this is that the elements $W(CD)$ and $H_0(CD)$ can be shown to be representable as the same reduced words in the generators $s_1, s_2, \ldots, s_n$, albeit with different multiplication rules. In particular, the Coxeter group $W(CD)$ is finite if and only if so is its 0-Hecke monoid $H_0(CD)$.

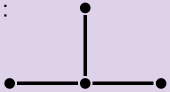
A complete classification of Coxeter diagrams giving rise to finite Coxeter groups is known (H. S. M. Coxeter, The complete enumeration of finite groups of the form $r_i^2 = (r_i r_j)^{k_{ij}} = 1$, J. London Math. Soc. s1-10 (1935), 21–25).

Example: 0-Hecke Monoids, finished

Our approach solves the FFB for almost all finite 0-Hecke monoids.

Theorem

A finite 0-Hecke monoid is nonfinitely based whenever a connected component of its Coxeter diagram has at least four vertices and is not D_4:



In view of the proposition and some earlier results by Olga Sapir, the FBP remains open for only six finite 0-Hecke monoids with connected Coxeter diagrams; the corresponding diagrams are I_4: • ——4—— •, I_5: • ——5—— •,

A_3: • ———— • ———— •, B_3, H_3: • ——5—— • ———— •, and D_4.

The sizes of these six monoids are 8, 10, 24, 48, 120, and 192.

What We Know and What We Don't

Combinatorial characterization? – only one way so far.
If a letter x occurs exactly once in a word u, then x is linear in u. A word u is sparse if every two occurrences of the same letter in u sandwich some linear letter.

Theorem

A finite monoid is strongly nonfinitely based if all sparse words are isoterms for it.

Similar to Sapir's combinatorial characterization of inherently nonfinitely based semigroups, this statement can be made more concrete. This time the Thue–Morse words come into the play:

$$T_1 := xy,\ T_2 := xyyx, \dots, T_n := T_{n-1}\overline{T_{n-1}}, \dots \text{ where } \overline{x} := y,\ \overline{y} := x.$$

The sparsification ST_n is obtained by inserting 2^n new linear variables at all even positions of T_n. Say, for $T_3 = xyyxyxxy$, one gets $ST_3 = h_0xyh_1yxh_2yxh_3xyh_4$.

An equivalent formulation

A finite monoid is strongly nonfinitely based if all words ST_n are isoterms for it.

What We Know and What We Don't, continued

We do not yet know if the converse is true. (My guess: it isn't.)

Structural description? — no idea so far!
We do not even now whether every strongly nonfinitely based semigroup contains a strongly nonfinitely based submonoid — for the INFB case this holds and is an important part of Sapir's structural description of INFB semigroups.

Algorithm? — no idea so far!

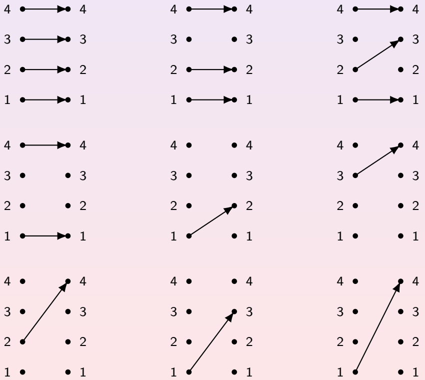
The main obstacle: we have no equational characterization of finitely generated semigroup varieties (the INFB case is based on Sapir's equational characterization of locally finite semigroup varieties). jump to summary

Now we pass to

Further Examples

Recall that our first example of a strongly nonfinitely based semigroup was the 42-element monoid $\mathcal{IC}_4$ of all partial extensive order preserving injections of $[4]$.

Looking for smaller examples, we have found a 10-element submonoid of $\mathcal{IC}_4$ which retains the property of being strongly nonfinitely based. The non-empty injections from the submonoid are displayed here:



Further Examples, continued

Our second example of a strongly nonfinitely based semigroup was the 42-element monoid $\mathcal{C}_5$ of all extensive and order preserving full transformations of $[5]$. It also contains a similar (but different) 10-element strongly nonfinitely based submonoid.

The smallest so far strongly nonfinitely based semigroup we know is the 9-element monoid obtained by adjoining 1 to the semigroup with the following Cayley table:

	0	a	e	b	a^2	ab	ea	eb
0	0	0	0	0	0	0	0	0
a	0	a^2	e	ab	a^2	eb	ea	eb
e	0	ea	e	eb	a^2	0	ea	eb
b	0	b	b	0	b	0	b	0
a^2	0	a^2	e	eb	a^2	eb	ea	eb
ab	0	ab	ab	0	ab	0	ab	0
ea	0	a^2	e	0	a^2	eb	ea	eb
eb	0	eb	eb	0	eb	0	eb	0

This monoid does not belong to the variety $\operatorname{var}\mathcal{IC}_4$. Thus, the 9-element example is smaller in size but not with respect to the varietal inclusion.

Minimal Examples?

The smallest inherently nonfinitely based semigroups are the 6-element Brandt monoid $\mathcal{B}_2^1$ and another 6-element monoid known as $\mathcal{A}_2^1$. There exist exactly two further nonfinitely based 6-element semigroups (Edmond W. H. Lee, Wen Tin Zhang, Finite basis problem for semigroups of order six, LMS J. Comput. Math. 18 (2015), 1–129), but they are not strongly nonfinitely based.

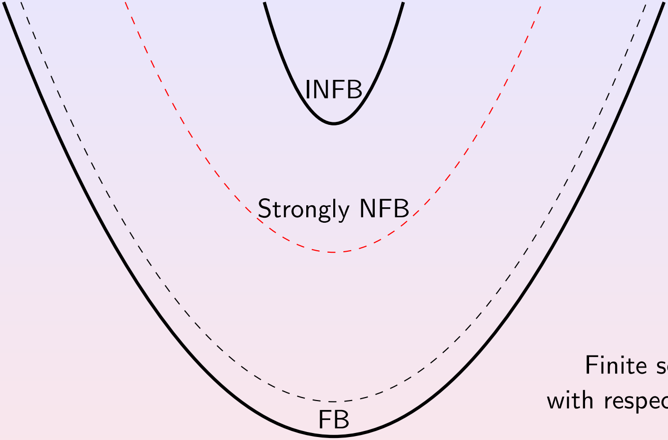
Hence, the minimum size of a strongly but not inherently nonfinitely based semigroup is either 7, or 8, or 9.

Using Mace4, we checked that

- no monoid of size 8 or less can be shown to be strongly but not inherently nonfinitely based via our present method;
- no monoid in the variety $\mathrm{var}\,\mathcal{IC}_4$ of size 9 or less can be shown to be strongly but not inherently nonfinitely based via our present method.

To compare: inherently nonfinitely based semigroups minimal with respect to the varietal inclusion were fully classified by Marcel Jackson (Small inherently nonfinitely based finite semigroups. Semigroup Forum 64 (2002) 297–324).

Summary



Finite semigroups
with respect to the FBP

'We can only see a short distance ahead, but we can see plenty there that needs to be done' (Alan Turing, Computing machine and intelligence, Mind 59 (1950), 433–460).