THE MANY VARIETIES OF VARIETY THEORY

Amaldev Manuel

School of Mathematics & Computer Science,
Indian Institute of Technology GOA

January 7, 2026

PLAN

FINITE MONOIDS

- A semigroup is a set with an associative binary operation. It is a monoid if the operation has an identity.
- An element x of the semigroup is idempotent if $x^2 = x$.
- Some examples.
    - $\{x_i \mid i \in I\}$ such that $x_i x_j = x_j$ for all $i, j \in I$. (Right-zero semigroup).
    - $I \times J$ with $(i,j) \cdot (i',j') = (i,j')$ (Rectangular band).
- In a finite semigroup every element x has a unique idempotent power denoted by x^ω. (Not necessarily true for infinite semigroups; consider the set of positive integers with addition.)

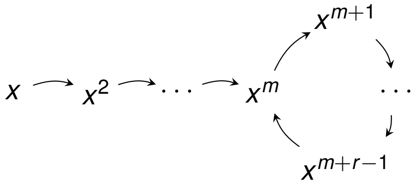

Figure. The lasso of powers of x in a finite semigroup

VARIETIES OF MONOIDS

- A class of monoids form a *(monoid) variety* if it is closed under Homomorphic images (quotients), Submonids and direct Products.
- Examples
  - Commutative monoids ($xy = yx$)
  - Bands or Idempotent Monoids ($x^2 = x$)
  - Semilattices or Idempotent & Commutative Monoids ($x^2 = x, xy = yx$).
  - Right-Zero semigroups ($xy = y$).
  - Rectangular bands ($x^2 = x, xyz = xz$ or equivalently $x^2 = x, xyx = x$).
- The class of all groups do not form a monoid variety: $(\mathbb{Z}, +)$ is a group but its submonoid $(\mathbb{N}, +)$ is not!.
- Varieties = Equationally Definable Classes (Birkhoff's HSP Theorem).
- It is a strengthening of Łoś–Tarski Preservation Theorem from model theory (A first-order sentence φ is preserved under substructures iff φ is equivalent to a universal sentence).

PSEUDOVARIETIES OF MONOIDS

- Let S and T be semigroups. *S divides T* if S is a quotient of a subsemigroup of T.
- Simplified definition: Varieties are classes closed under direct products and division.

Definition 1.1 (Pseudovariety)

A class of finite monoids constitute a pseudovariety if it is closed under finite direct products, quotients and submonoids (or equivalently finite direct products and division).

- Examples
  - the class of all finite monoids (**Fin**)
  - $\mathbf{V} \cap \mathbf{Fin}$ for each variety $\mathbf{V}$.
  - The class of all finite groups $\mathbf{G}$ is a pseudovariety ($x^\omega y = yx^\omega = y$ or $x^\omega = 1$ for short).
- Inverse and regular semigroups do not form pseudovarieties.

APERIODIC MONOIDS

- A semigroup/monoid is group-free if the only groups dividing it are trivial.
- Denote by **A** the class of all group-free monoids.
- **A** is a pseudo-variety (but not a variety) defined by the equation $x^\omega x = xx^\omega = x^\omega$ or $x^{\omega+1} = x^\omega$ for short.
- Pseudo-varities = Classes defined by profinite identities (terms built using variables, concatenation and ω-power) [Reiterman's Theorem].
- For example $(xy)^\omega x = y(xy)^\omega = (xy)^\omega$ defines the class of $\mathcal{J}$-trivial monoids.

RECOGNISABILITY BY MONOIDS

- Let Σ be a finite alphabet.
- A language $L \subseteq \Sigma^*$ is *recognised* by a finite monoid M if there is a morphism $h : \Sigma^* \to M$ such that $L = h^{-1}(P)$ for some $P \subseteq M$.
- Fix the alphabet $\Sigma = \{a, b\}$.
  - The trivial monoid recognises the languages $\emptyset, \Sigma^*$.
  - $a^* = \{\varepsilon, a, aa, \dots\}$ is recognised by $\langle 0, 1 \mid 0 \cdot 1 = 1 \cdot 0 = 0 \cdot 0 = 0 \rangle$.
  - $a^*b^* = \{\varepsilon, a, b, aa, ab, bb, \dots\}$ is recognised by $\{1, a, b, ab, 0\}$ with $aa = a, bb = b, ba = 0$.

MORE EXAMPLES

- Right-Zero Semigroup $\longleftrightarrow$ Union of languages of the form $\Sigma^* a$ for $a \in \Sigma$.
- Semilattice (idempotent & commutative) $\longleftrightarrow$ Boolean combination of languages of the form $\Sigma^* a \Sigma^*$ where $a \in \Sigma$.
- Nilpotent $\longleftrightarrow$ Finite or Cofinite Language.

RECOGNISABE LANGUAGES

- A language is *recognisable* if it is recognised by some finite monoid.
- Rational Subsets of the free monoid is the smallest family containing all finite subsets and closed under union ($X \cup Y$), concatenation ($X \cdot Y$) and Kleene closure ($X^* = X^0 \cup X^1 \cup \cdots$).
- Recognisable languages = Rational subsets [Kleene's Theorem]
- Recognisable languages are closed under
  - Boolean operations (if L_1 and L_2 are recognisable so are $L_1 \cap L_2$, $L_1 \cup L_2$ and $\overline{L_1} = \Sigma^* \setminus L_1$),
  - inverse images under morphisms between free monoids (if $L \subseteq B^*$ is recognisable and $\varphi : A^* \to B^*$ is a morphism then $\varphi^{-1}(L)$ is recognisable), and
  - *quotient by words* (if L is recognisable and $u, v \in \Sigma^*$ then $u^{-1}Lv^{-1} = w \in \Sigma^* \mid uwv \in L$ is recoginsable.

Syntactic Monoids

Definition 2.1

The syntactic congruence of $L \subseteq \Sigma^*$ is the relation $\sim$ on Σ^* as follows. For $u, v \in \Sigma^*$,

$$u \sim v \qquad \text{if} \qquad xuy \in L \Longleftrightarrow xvy \in L \text{ for all } x, y \in \Sigma^*.$$

Syntactic monoid of L is the quotient $\Sigma^* / \sim$.

The syntactic morphism factors through every morphism recognising L. Moreover, it can be computed!

$$\begin{tikzcd}
A^* \arrow[r, "\varphi"] \arrow[dr, "\eta_L"'] & M \arrow[d, dashed, "\exists!\,\psi"] \\
 & \mathsf{Synt}(L)
\end{tikzcd}$$

Figure. Universal Property of Syntactic Monoids

SYNTACTIC MONOIDS

- Not all monoids are syntactic monoids.
- For instance $M = \{1, x, y, z\}$ such that $ab = b$ for all $a \in M$ and $b \in \{x, y, z\}$.
- But every monoid M divides the product of syntactic monoids of languages recognised by M.

EILENBERG'S VARIETY THEOREM

Definition 3.1

A variety of recognisable languages is a map $\mathcal{V}$ each finite alphabet Σ to a set of recognisable languages $\mathcal{V}\Sigma^*$ such that

1. *$\mathcal{V}\Sigma^*$ is closed under Boolean operations,*
2. *if $L \in \mathcal{V}\Sigma^*$ and $\varphi : \Gamma^* \to \Sigma^*$ is a morphism then $\varphi^{-1}(L) \in \mathcal{V}\Gamma^*$, and*
3. *if $L \in \mathcal{V}\Sigma^*$ and $u, v \in \Sigma^*$ then $u^{-1}Lv^{-1} \in \mathcal{V}\Sigma^*$.*

Theorem 1 (Eilenberg c. 70)

Pseudovarieties of finite monoids and varieties of recognisable languages are in one-to-one correspondence.

Proof Idea.

Define the following maps.

$\mathbf{V} \mapsto \mathcal{V}$ where $\mathcal{V} : \Sigma \mapsto \{L \subseteq \Sigma^* \mid L \text{ is recognised by some } M \in \mathbf{V}\}$.

$\mathcal{V} \mapsto \mathbf{V}$ where $\mathbf{V}$ = pseudovariety generated by syntactic monoids of languages in $\mathcal{V}$.

Show that the maps are mutual inverses. □

CHARACTERISATION OF **A**

An expression is star-free if it does not use Kleene-star.

$$a^* b^* = \overline{\Sigma^* \, ba \, \Sigma^*} = \overline{\overline{\emptyset} \, ba \, \overline{\emptyset}}$$

First-order logic over words is the logic where the variables quantifies over positions and the predicates are of the form $a(x), x < y$ and $x = y$.

$$\forall x \forall y (a(x) \wedge b(y) \to x < y).$$

Theorem 2 (Schützenberger-McNaughton-Papert Theorem)

The following are equivalent for a language L.

1. *L is recognised by an aperiodic monoid.*
2. *$\mathrm{Syn}(L)$ is aperiodic.*
3. *L is star-free.*
4. *L is definable in first-order logic.*

The theorem gives an algorithm to check if a language is star-free or FO-definable!

1. Compute the syntactic monoid.
2. Check if it is aperiodic.

CHARACTERISATION OF **DA**

If **V** is a pseudovariety then **DV** is the pseudovariety of all monoids whose regular $\mathcal{D}$-classes are in **V**.

An expression is *unambiguous* if there is a unique way to parse any word given by the expression.

$$\Sigma^* a \Sigma^* b^* \Sigma^* = (\Sigma \setminus \{a\})^* \, a \, (\Sigma \setminus \{b\})^* \, b \, \Sigma^*$$

A monomial is an expression of the form $A_0^* a_1 A_1^* \cdots a_n A_n^*$ where each $a_i \in \Sigma$ and each $A_i \subseteq \Sigma$.

Theorem 3 (Therién-Wilke)

The following are equivalent for a language L.

1. *L is recognised by a* **DA***-monoid.*
2. $\mathrm{Syn}(L)$ *is in* **DA***.*
3. *L is a disjoint union of unambiguous monomials.*
4. *L is definable in two-variable first-order logic.*

VARIANTS OF VARIETY THEOREM

- Variant of Variety theorem exists for recognisable sets of
    - ω-words
    - biinfinite words
    - scattered words (countable linear orders with no dense suborder)
    - countable linear orders
    - functions ($\Sigma^* \mapsto \mathbb{N} \cup \{\infty\}$)
    - words with an involution
    - $\ldots$
- In each case decidable characterisation leads to algorithms.

EXAMPLE: LANGUAGES WITH AN INVOLUTION

- An operation $\star : S \to S$ is an involution on the semigroup on S is $(a^\star)^\star = a$ and $(ab)^\star = b^\star a^\star$.
- A pseudovariety of $\star$-semigroups is a collection of $\star$-semigroups closed under finite direct products, $\star$-subsemigroups and quotients.
- An involutary alphabet $(A, \dagger)$ is an alphabet A with an involution on A (a function that is its own inverse).
- A variety of involutary languages is a class of languages closed under Boolean operations, involution, quotients by words and inverse morphisms.

Theorem 4 (M., Nevatia)

Pseudovarieties of $\star$-semigroups and varieties of involutary languages are in one-to-one correspondence.

Theorem 5 (M., Nevatia)

A language is definable in first-order logic with the neighbour relation iff it is recognised by a locally hermitian semidirect product of an aperiodic & commutative semigroup and a locally trivial semigroup.

EXAMPLE: COUNTABLE LINEAR ORDERS

We define $\circ$-**DA** to be the subclass of $\circ$-monoids that satisfy the following equations.

1. $(xyz)^* y (xyz)^* = (xyz)^*$
2. $x^* = x^\omega x^{\omega *}$
3. $\{x_1, \ldots, x_k\}^\eta = (x_1 \cdots x_k)^{\omega^*} (x_1 \cdots x_k)^\omega$.

Theorem 6 (M., Sreejith)

The following are equivalent for a set of labelled-countable linear orders L.

1. *L is definable in* FO^2.
2. *L is recognised by a* $\circ$-**DA** *algebra.*

CONCLUSION

- Variety theory is a rich avenue for interesting mathematical results that have computational consequences, and, more importantly
- there is plenty of room here!