On Varieties of Automata Enriched with an Algebraic Structure

Ondřej Klíma

Department of Mathematics and Statistics
Masaryk University, Brno
Czech Republic

AFL 2014 — May 28

An Outline

I. Algebraic Theory of Regular Languages

Varieties of regular languages, syntactic monoid, Eilenberg correspondence and its generalizations.

II. Varieties of Automata

Minimal DFA, closure operators, Eilenberg type correspondence.

III. Automata Enriched with an Algebraic Structure

Ordered automata, meet-automata, DL-automata, Eilenberg type correspondence.

Thanks

The presented point of view is based on numerous interesting discussions on the topic with my colleagues Michal Kunc and Libor Polák.

I. Algebraic Theory of Regular Languages

Examples

- Goal of the study: effective characterizations of certain natural classes of regular languages.
- Typical result: a language belongs to a given class iff its syntactic monoid belongs to a certain class of monoids.

Examples

- Goal of the study: effective characterizations of certain natural classes of regular languages.
- Typical result: a language belongs to a given class iff its syntactic monoid belongs to a certain class of monoids.

Theorem (Schützenberger – 1966)

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Theorem (Simon — 1972)

A regular language L is piecewise testable if and only if the syntactic monoid of L is $\mathcal{J}$-trivial.

- General framework – Eilenberg correspondence.

Varieties of Languages

Definition

A variety of languages $\mathcal{V}$ associates to every finite alphabet A a class $\mathcal{V}(A)$ of regular languages over A in such a way that

- $\mathcal{V}(A)$ is closed under finite unions, finite intersections and complements (in particular $\emptyset, A^* \in \mathcal{V}(A)$),
- $\mathcal{V}(A)$ is closed under quotients, i.e.
  $L \in \mathcal{V}(A),\ u, v \in A^*$ implies
  $u^{-1}Lv^{-1} = \{ w \in A^* \mid uwv \in L \} \in \mathcal{V}(A)$,
- $\mathcal{V}$ is closed under preimages in morphisms, i.e.
  $f : B^* \to A^*,\ L \in \mathcal{V}(A)$ implies
  $f^{-1}(L) = \{ v \in B^* \mid f(v) \in L \} \in \mathcal{V}(B)$.

Pseudovarieties of Monoids

Definition

A pseudovariety of finite monoids is a class of finite monoids closed under submonoids, morphic images and products of finite families.

- For a regular language $L \subseteq A^*$ we define a relation $\sim_L$ on A^* by the rule $u \sim_L v$ iff u and v have the same contexts in L.

  Formally: $u \sim_L v \iff \{(p,q) \mid puq \in L\} = \{(p,q) \mid pvq \in L\}$.
- $\sim_L$ is the syntactic congruence of L and $A^*/\sim_L = M_L$ is the syntactic monoid of L.

The Eilenberg Correspondence

- For each pseudovariety of monoids $\mathbf{V}$ we denote $\alpha(\mathbf{V})$ the variety of regular languages given by

$$(\alpha(\mathbf{V}))(A) = \{L \subseteq A^* \mid M_L \in \mathbf{V}\} .$$

- For each variety of regular languages $\mathcal{L}$ we denote by $\beta(\mathcal{L})$ the pseudovariety of monoids generated by syntactic monoids M_L, where $L \in \mathcal{L}(A)$ for some alphabet A.

Theorem (Eilenberg – 1976)

The mappings α and β are mutually inverse isomorphisms between the lattice of all pseudovarieties of finite monoids and the lattice of all varieties of regular languages.

A Formal Definition of a DFA

Definition

A deterministic finite automaton over the alphabet A is a five-tuple $\mathcal{A} = (Q, A, \cdot, i, F)$, where

- Q is a nonempty set of states,
- $\cdot : Q \times A \to Q$ is a complete transition function, which can be extended to a mapping $\cdot : Q \times A^* \to Q$ by $q \cdot \lambda = q,\ q \cdot (ua) = (q \cdot u) \cdot a$,
- $i \in Q$ is the initial state,
- $F \subseteq Q$ is the set of final states.

The automaton $\mathcal{A}$ accepts a word $u \in A^*$ iff $i \cdot u \in F$.

The automaton $\mathcal{A}$ recognizes the language $L_{\mathcal{A}} = \{u \in A^* \mid i \cdot u \in F\}$.

A Relationship between DFAs and Monoids

If we have a DFA $\mathcal{A} = (Q, A, \cdot, i, F)$, then:

- Each word $u \in A^*$ performs the transformation $\tau_u : Q \to Q$ where $\tau_u(q) = q \cdot u$ for each $q \in Q$.
- The transition monoid of $\mathcal{A}$ is $(\{\tau_u \mid u \in A^*\}, \circ)$.
- The transition monoid of the minimal automaton of L is isomorphic to the syntactic monoid M_L of L.
- Every monoid can be transformed to a DFA.

Motivations for a Notion of a Variety of Automata

- Why monoids instead of automata?
  - An equational description of pseudovarieties of monoids by pseudoidentities.
  - Other algebraic constructions, e.g. products (semidirect, wreath, Mal'cev).

Motivations for a Notion of a Variety of Automata

- Why monoids instead of automata?
  - An equational description of pseudovarieties of monoids by pseudoidentities.
  - Other algebraic constructions, e.g. products (semidirect, wreath, Mal'cev).
- Why are we still interested in automata characterizations?
  - Usually, a regular language is given by an automaton. And computation of the syntactic monoid need not to be effective (can be exponentially larger).
  - Sometimes a “graph condition” on automata can be easier to test than an equational condition on monoids.

Generalizations of the Eilenberg Correspondence

Since not all natural classes of regular languages are varieties, one of the recent directions of the research in algebraic theory of regular languages is devoted to generalizations of the Eilenberg correspondence.

Generalizations of the Eilenberg Correspondence

Since not all natural classes of regular languages are varieties, one of the recent directions of the research in algebraic theory of regular languages is devoted to generalizations of the Eilenberg correspondence.

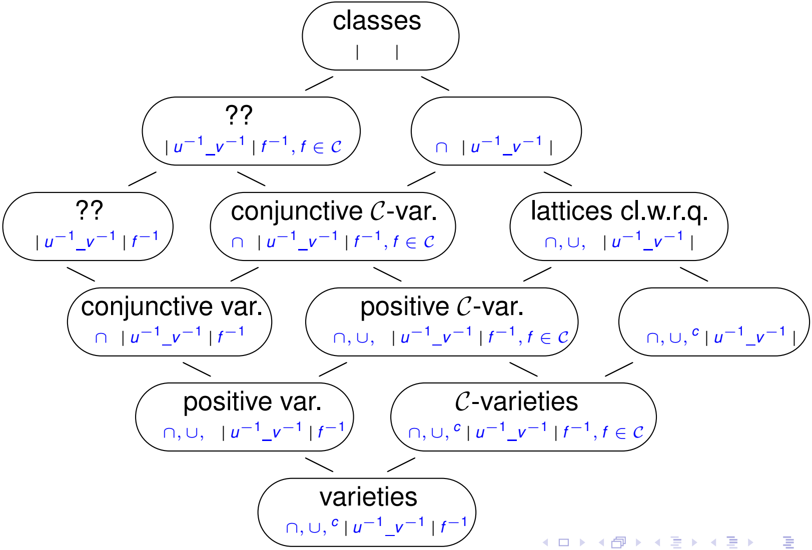
- Pin (1995): Positive varieties of regular languages — closure under complementation is not required.
  Algebraic counterparts are pseudovarieties of finite ordered monoids.
  (Syntactic monoid is implicitly ordered.)

Generalizations of the Eilenberg Correspondence

Since not all natural classes of regular languages are varieties, one of the recent directions of the research in algebraic theory of regular languages is devoted to generalizations of the Eilenberg correspondence.

- Pin (1995): Positive varieties of regular languages — closure under complementation is not required. Algebraic counterparts are pseudovarieties of finite ordered monoids.
  (Syntactic monoid is implicitly ordered.)
- Polák (1999): Conjunctive (and disjunctive) varieties.
- Straubing (2002): $\mathcal{C}$-varieties of languages.
- Ésik, Larsen (2003): literal varieties of languages.
- Gehrke, Grigorieff, Pin (2008): Lattices of regular languages.

Variants of Varieties of Regular Languages



II. Varieties of Automata

The Construction of a Minimal DFA by Brzozowski

- For a language $L \subseteq A^*$ and $u \in A^*$, we define a left quotient $u^{-1}L = \{ w \in A^* \mid uw \in L \}$.

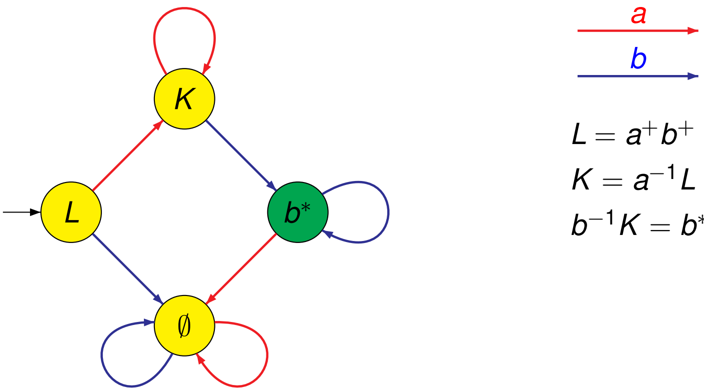
Definition

The canonical deterministic automaton of L is $\mathcal{D}_L = (D_L, A, \cdot, L, F)$, where

- $D_L = \{ u^{-1}L \mid u \in A^* \}$,
- $q \cdot a = a^{-1}q$, for each $q \in D_L$, $a \in A$,
- $q \in F$ iff $\lambda \in q$.

- Each state $q = u^{-1}L$ is formed by all words transforming the state q into a final state.
- The set of all such words can be considered in an arbitrary automaton for every state.

An Example of a Canonical Automaton



$$L = a^+ b^+$$

$$K = a^{-1}L = a^* b^+$$

$$b^{-1}K = b^*$$

A Future of a State in a DFA

- For a DFA $\mathcal{A} = (Q, A, \cdot, i, F)$ and an arbitrary state $q \in Q$, we consider $L_q = \{w \in A^* \mid q \cdot w \in F\}$.
  (Sometimes L_q is called the future of the state q.)
- If q is a reachable state, i.e. there is $u \in A^*$ such that $q = i \cdot u$, then

  $$w \in L_q \iff (i \cdot u) \cdot w \in F \iff uw \in L \iff w \in u^{-1}L\,.$$

  Hence $L_q = u^{-1}L$.
- Thus we can consider a mapping φ from the subautomaton $\mathcal{A}' = (Q', A, \cdot, i, F \cap Q')$ of $\mathcal{A}$ consisting of the reachable states $Q' = \{i \cdot u \mid u \in A^*\}$ onto $\mathcal{D}_L$.
- In particular, this proves that the canonical DFA $\mathcal{D}_L$ is a minimal automaton for L.

Right Quotients

- Another consequence: for every $u \in A^*$, the language $u^{-1}L$ is recognized by the same DFA as L with a different initial state.
- What about right quotients Lu^{-1}?

Right Quotients

- Another consequence: for every $u \in A^*$, the language $u^{-1}L$ is recognized by the same DFA as L with a different initial state.
- What about right quotients Lu^{-1}? We take another set of final states $F' = \{q \in Q \mid q \cdot u \in F\}$.
- Is the converse direction also true? Yes.

Right Quotients

- Another consequence: for every $u \in A^*$, the language $u^{-1}L$ is recognized by the same DFA as L with a different initial state.
- What about right quotients Lu^{-1}? We take another set of final states $F' = \{q \in Q \mid q \cdot u \in F\}$.
- Is the converse direction also true? Yes.

Lemma

Let $\mathcal{D}_L = (D_L, A, \cdot, L, F)$ be the canonical automaton of a language L and $F' \subseteq D_L$. Then $(D_L, A, \cdot, L, F')$ recognizes a language which can be expressed as a Boolean combination of languages of the form Lu^{-1}.

Automata without Initial and Final States

- Varieties of languages are closed under Boolean operations and quotients, therefore the choice of an initial state and final states in automata can be left free.
- When we talk about varieties of automata, by an automaton we mean the underlying labeled graph, i.e. a triple $(Q, A, \cdot)$.
- We say that $(Q, A, \cdot)$ recognizes a language L if there are $i \in Q$ and $F \subseteq Q$ such that $L = L_{\mathcal{A}}$ where $\mathcal{A} = (Q, A, \cdot, i, F)$.

A Product of Automata

- Since varieties of languages are closed under taking unions and intersections, it is convenient to include direct products of automata in our varieties of automata.
- Formally: let $(Q_j, A, \cdot_j)_{j\in I}$ be automata, I be a finite set. We define $(\prod_{j\in I} Q_j, A, \cdot)$, where $(q_j)_{j\in I} \cdot a = (q_j \cdot_j a)_{j\in I}$.
- If for each $j \in I$ a language L_j is recognized by $\mathcal{A}_j = (Q_j, A, \cdot_j, i_j, F_j)$, then $(\prod_{j\in I} Q_j, A, \cdot, (i_j)_{j\in I}, F)$
  - recognizes $\bigcap L_j$ if $F = \{(q_j)_{j\in I} \mid \forall j : q_j \in F_j\}$ or
  - recognizes $\bigcup L_j$ if $F = \{(q_j)_{j\in I} \mid \exists j : q_j \in F_j\}$.

A Disjoint Union of Automata

- Our motivation is an Eilenberg type correspondence.
- Therefore, we need to have classes of automata which are closed under all possible constructions which do not recognize more languages. The basic algebraic constructions follow.
- A disjoint union of automata: let $(Q_j, A, \cdot_j)_{j \in I}$ be automata, I be a finite set and Q_j be pairwise disjoint sets.
  We define $(Q, A, \cdot)$, where $Q = \bigcup_{j \in I} Q_j$, $q \cdot a = q \cdot_j a$ with $j \in I$ such that $q \in Q_j$.
- If L is recognized by a disjoint union of automata then it is recognized by some of them.

Morphisms of DFAs

- Let $(Q, A, \cdot)$ and $(P, A, \circ)$ be automata and $\varphi : Q \to P$ be a mapping. Then φ is called a morphism of automata if

$$\varphi(q \cdot a) = \varphi(q) \circ a \quad \text{for all} \quad a \in A, q \in Q\,.$$

- If there exists a surjective morphism of automata from $(Q, A, \cdot)$ to $(P, A, \circ)$, then we say that $(P, A, \circ)$ is a morphic image of $(Q, A, \cdot)$.
- If $(P, A, \circ)$ is a morphic image of $(Q, A, \cdot)$ and L is recognized by $(P, A, \circ)$, then L is recognized by $(Q, A, \cdot)$.

Trivial Automata and Subautomata

- Denote by $\mathcal{T}_n(A)$ an n-state trivial automaton, i.e $\mathcal{T}_n(A) = (I_n, A, \cdot)$, where $I_n = \{1, \ldots, n\}$ and $j \cdot a = j$ for all $j \in I_n$, $a \in A$.
- The disjoint union of automata $(Q_j, A, \cdot_j)_{j \in I_n}$ is a morphic image of the product $(\prod_{j \in I_n \cup \{0\}} Q_j, A, \cdot)$ where $(Q_0, A, \cdot_0) = \mathcal{T}_n(A)$.

Trivial Automata and Subautomata

- Denote by $\mathcal{T}_n(A)$ an n-state trivial automaton, i.e $\mathcal{T}_n(A) = (I_n, A, \cdot)$, where $I_n = \{1, \dots, n\}$ and $j \cdot a = j$ for all $j \in I_n$, $a \in A$.
- The disjoint union of automata $(Q_j, A, \cdot_j)_{j \in I_n}$ is a morphic image of the product $(\prod_{j \in I_n \cup \{0\}} Q_j, A, \cdot)$ where $(Q_0, A, \cdot_0) = \mathcal{T}_n(A)$.
- Let $(Q, A, \cdot)$ be an automaton and $P \subseteq Q$ be a non-empty subset. If $p \cdot a \in P$ for every $p \in P$, $a \in A$, then $(P, A, \cdot)$ is called s subautomaton of $(Q, A, \cdot)$.

Preimages in Morphisms

- Varieties of languages are closed under taking preimages.
- Let $f : B^* \to A^*$ be a morphism, $L \in \mathcal{V}(A)$ be recognized by an automaton $\mathcal{A} = (Q, A, \cdot, i, F)$.
- Then we construct $\mathcal{B} = (Q, B, \circ, i, F)$, where $q \circ b = q \cdot f(b)$ for every $q \in Q$, $b \in B$.
- The automaton $(Q, B, \circ, i, F)$ recognizes the language $f^{-1}(L)$.
- We say that $(P, B, \circ)$ is an f-subautomaton of $(Q, A, \cdot)$ if $P \subseteq Q$ and $q \circ b = q \cdot f(b)$ for every $q \in P$, $b \in B$.
- Taking $f = \mathrm{id} : A^* \to A^*$, we obtain the original notion of subautomaton.

Varieties of Automata

Definition

A variety of automata $\mathbb{V}$ associates to every finite alphabet A a class $\mathbb{V}(A)$ of automata (without initial and final states) over alphabet A in such a way that

- $\mathbb{V}(A) \neq \emptyset$ is closed under disjoint unions, finite direct products and morphic images,
- $\mathbb{V}$ is closed under f-subautomata.

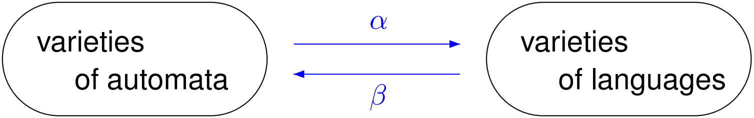
- If we define $\mathbb{T}(A) = \{\mathcal{T}_n(A) \mid n \in \mathbb{N}\}$ then the first condition can be written equivalently in the following way: $\mathbb{T}(A) \subseteq \mathbb{V}(A)$ and $\mathbb{V}(A)$ is closed under finite direct products and morphic images.
- In particular, the class of trivial automata $\mathbb{T}$ forms the smallest variety of automata.

An Eilenberg Type Correspondence

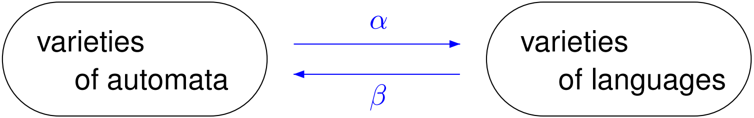
- For each variety of automata $\mathbb{V}$ we denote by $\alpha(\mathbb{V})$ the variety of regular languages given by

$$(\alpha(\mathbb{V}))(A) = \{L \subseteq A^* \mid \exists \mathcal{A} = (Q, A, \cdot, i, F) :$$
$$L = L_{\mathcal{A}} \wedge (Q, A, \cdot) \in \mathbb{V}(A)\}\,.$$

- For each variety of regular languages $\mathcal{L}$ we denote by $\beta(\mathcal{L})$ the variety of automata generated by all DFAs $\mathcal{D}_L$, where $L \in \mathcal{L}(A)$ for some alphabet A.



An Eilenberg Type Correspondence II



Theorem (Ésik and Ito, Chaubard, Pin and Straubing)

The mappings α and β are mutually inverse isomorphisms between the lattice of all varieties of automata and the lattice of all varieties of regular languages.

- A version for $\mathcal{C}$-varieties is obvious: we consider f-subautomata (etc.) just for $f \in \mathcal{C}$.
- Ésik and Ito were working with literal varieties (morphisms map letters to letters, i.e $f(B) \subseteq A$) and used disjoint union.
- Chaubard, Pin and Straubing called the automata $\mathcal{C}$-actions and used trivial automata.

An Examples – Acyclic Automata

- One of the conditions in Simon's characterization of piecewise testable languages is that a minimal DFA is acyclic.
- A content $\mathrm{c}(u)$ of a word $u \in A^*$ is the set of all letters occurring in u.
- We say that $(Q, A, \cdot)$ is a acyclic if for each $u \in A^*$ and $q \in Q$ we have

$$q \cdot u = q \implies (\forall a \in \mathrm{c}(u) : q \cdot a = q) .$$

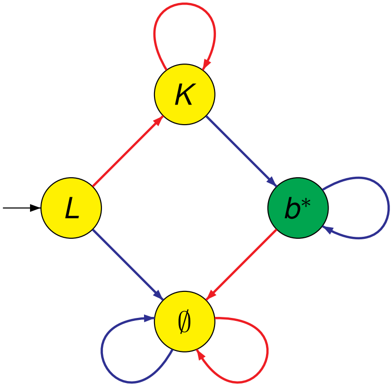
- The class of all acyclic automata is a variety.
- The corresponding variety of languages (well-known): (disjoint) unions of the languages of the form

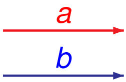
$$A_0^* a_1 A_1^* a_2 A_2^* \dots A_{n-1}^* a_n A_n^*, \quad \text{where} \quad a_i \notin A_{i-1} \subseteq A .$$

An Example – Piecewise Testable Languages

- In DLT'13 we gave an alternative condition for automata recognizing piecewise testable languages.
- We call an acyclic automaton $(Q, A, \cdot)$ locally confluent, if for each state $q \in Q$ and every pair of letters $a, b \in A$, there is a word $w \in \{a, b\}^*$ such that $(q \cdot a) \cdot w = (q \cdot b) \cdot w$.
- A stronger condition: an acyclic automaton $(Q, A, \cdot)$ is confluent, if for each state $q \in Q$ and every pair of words $u, v, \in \{a, b\}^*$, there is a word $w \in \{a, b\}^*$ such that $(q \cdot u) \cdot w = (q \cdot v) \cdot w$.
- Each acyclic automaton is confluent iff it is locally confluent.
- The class of all acyclic confluent automata is a variety which corresponds to the variety of piecewise testable languages.

An Example of a Piecewise Testable Language



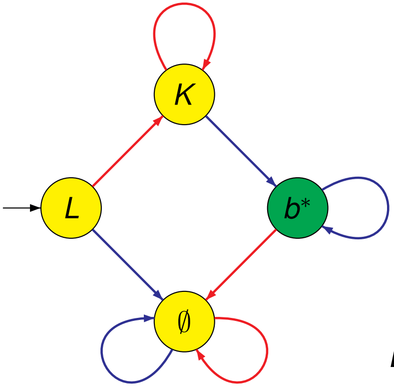


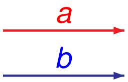
$L = a^+ b^+$

$K = a^{-1}L = a^* b^+$

$b^{-1}K = b^*$

An Example of a Piecewise Testable Language





$$L = a^+ b^+$$

$$K = a^{-1} L = a^* b^+$$

$$b^{-1} K = b^*$$

$$L = A^* a A^* b A * \cap (A^* b A^* a A^*)^c$$

The Eilenberg Correspondence – Remark

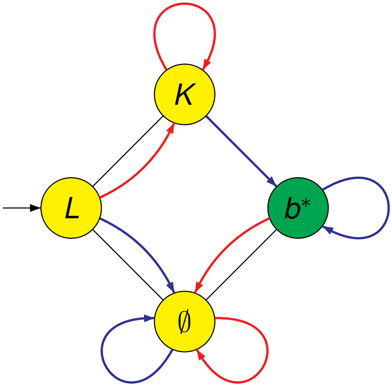
- One of the crucial step is to prove the following: an arbitrary automaton $(Q, A, \cdot)$ can be reconstruct from minimal automata of languages which are recognized by the automaton $(Q, A, \cdot)$.
- This is given in two steps.
  In the first step we fix an initial state $i \in Q$ and we reconstruct subautomaton $\{i \cdot u \mid u \in A^*\}$ from minimal automata of languages which are recognized when an initial state is i. Here a notion of a product of automata is used. (Also subautomata and morphic images.)
- In the second step, different initial states are considered and a notion of a disjoint union of automata is used. (A disjoint union of automata is a sum (co-product) from the categorical point of view.)

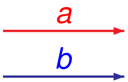
III. Automata Enriched with an Algebraic Structure

A Natural Ordering of the Canonical Automaton

- For a language $L \subseteq A^*$, we have defined a the canonical deterministic automaton: $\mathcal{D}_L = (D_L, A, \cdot, L, F)$, where
  - $D_L = \{ u^{-1}L \mid u \in A^* \}$,
  - $q \cdot a = a^{-1}q$, for each $q \in D_L$, $a \in A$,
  - $q \in F$ iff $\lambda \in q$.
- Therefore states are ordered by inclusion, which means that each minimal automaton is implicitly equipped with a partial order.
- The action by each letter a is an isotone mapping: for all states p, q such that $p \subseteq q$ we have $p \cdot a = a^{-1}p \subseteq a^{-1}q = q \cdot a$.
- The final states form an upward closed subset w.r.t. $\subseteq$.

An Example of an Ordered Automaton



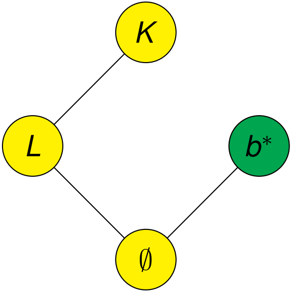


$L = a^+ b^+$

$K = a^{-1} L = a^* b^+$

$L \subseteq K$

An Example of an Ordered Automaton



$L = a^+ b^+$

$K = a^{-1}L = a^* b^+$

$L \subseteq K$

An Ordered Automaton

Definition

An ordered automaton over the alphabet A is a six-tuple $\mathcal{A} = (Q, A, \cdot, \leq, i, F)$, where

- $\mathcal{A} = (Q, A, \cdot, i, F)$ is a usual DFA;
- $\leq$ is a partial order;
- an action by every letter is an isotone mapping from the partial ordered set $(Q, \leq)$ to itself;
- F is an upward closed set, i.e. $p \leq q, p \in F \implies q \in F$.

An Ordered Automaton

Definition

An ordered automaton over the alphabet A is a six-tuple $\mathcal{A} = (Q, A, \cdot, \leq, i, F)$, where

- $\mathcal{A} = (Q, A, \cdot, i, F)$ is a usual DFA;
- $\leq$ is a partial order;
- an action by every letter is an isotone mapping from the partial ordered set $(Q, \leq)$ to itself;
- F is an upward closed set, i.e. $p \leq q, p \in F \implies q \in F$.

- For the subautomaton $\mathcal{A}' = (Q', A, \cdot, \leq, i, F \cap Q')$, with $Q' = \{i \cdot u \mid u \in A^*\}$, for states $p \leq q$ we have $L_p \subseteq L_q$.
  ($u \in L_p \implies p \cdot u \in F$ (upward closed) $\implies q \cdot u \in F \implies u \in L_q$.)
- Thus φ given by $\varphi(q) = L_q$ is, in fact, a morphism from $\mathcal{A}'$ onto the canonical ordered automaton of L.

A Transition Monoid of an Ordered Automaton

If we have an ordered automaton $(Q, A, \cdot, \leq)$, then:

- we have defined $\tau_u : Q \to Q$ transformation by a word $u \in A^*$.
- These transformations can be ordered:

$$\tau_u \leq \tau_v \iff \forall p \in Q : p \cdot \tau_u \leq p \cdot \tau_v .$$

- An ordered transition monoid.
- In particular, the ordered transition monoid of the canonical ordered automaton of L is isomorphic to the syntactic ordered monoid M_L of L.

Algebraic Constructions on Ordered Automata

- The definitions of a disjoint union of ordered automata and f-subautomata are obvious.
- In a product of ordered automata $(Q_j, A, \cdot_j, \leq_j)_{j\in I}$ the order $\leq$ is define in usual way: $(q_j)_{j\in I} \leq (p_j)_{j\in I}$ iff $\forall j : q_j \leq p_j$.
- In the definition of morphism of automata we add assumption that considered mapping is isotone.

Definition

A variety of ordered automata $\mathbb{V}$ associates to every finite alphabet A a class $\mathbb{V}(A)$ of ordered automata (without initial and final states) over alphabet A in such a way that

- $\mathbb{V}(A) \neq \emptyset$ is closed under disjoint union, finite direct products and morphic images,
- $\mathbb{V}$ is closed under f-subautomata.

An Eilenberg Type Correspondence

- If L is recognized by an ordered automaton $(Q, A, \cdot, \leq, i, F)$ then we can not take new finite states $Q \setminus F$ – this set is not an upward closed subset.
  Therefore $(Q, A, \cdot, \leq)$ does not recognize L^c.
- In a product of ordered automata $(\prod_{j \in I} Q_j, A, \cdot, \leq, (i_j)_{j \in I}, F)$, both subsets $\{(q_j)_{j \in I} \mid \forall j : q_j \in F_j\}$ and $\{(q_j)_{j \in I} \mid \exists j : q_j \in F_j\}$ are upward closed, therefore the product recognizes intersection and union.

An Eilenberg Type Correspondence

- If L is recognized by an ordered automaton $(Q, A, \cdot, \leq, i, F)$ then we can not take new finite states $Q \setminus F$ – this set is not an upward closed subset.
  Therefore $(Q, A, \cdot, \leq)$ does not recognize L^c.
- In a product of ordered automata $(\prod_{j \in I} Q_j, A, \cdot, \leq, (i_j)_{j \in I}, F)$, both subsets $\{(q_j)_{j \in I} \mid \forall j : q_j \in F_j\}$ and $\{(q_j)_{j \in I} \mid \exists j : q_j \in F_j\}$ are upward closed, therefore the product recognizes intersection and union.

Theorem (Pin)

There are mutually inverse isomorphisms between the lattice of all varieties of ordered automata and the lattice of all positive varieties of regular languages.

The Level $1/2$

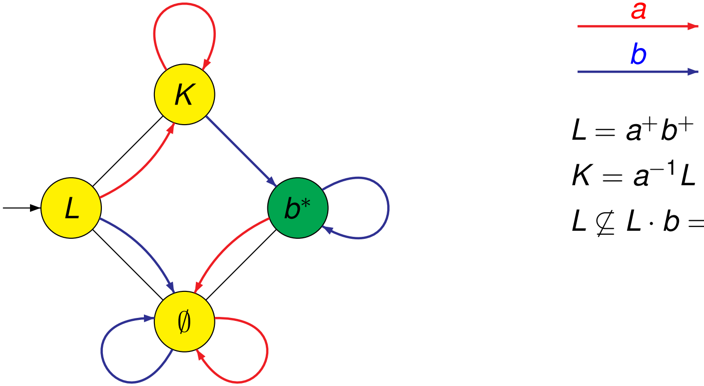
- Piecewise testable languages are Boolean combinations of languages of the form

$$A^* a_1 A^* a_2 A^* \dots A^* a_\ell A^*, \text{ where } a_1, \dots, a_\ell \in A,\ \ell \geq 0 .$$

- Piecewise testable languages form level 1 in Straubing-Thérien hierarchy.
- Level $1/2$ is formed just by finite unions of intersections of languages above.
- The corresponding variety of ordered automata is the class of all ordered automata where actions by letters are increasing mappings. I.e. ordered automata satisfying:

$$\forall q \in Q, a \in A : q \cdot a \geq q .$$

An Example of an Ordered Automaton outside 1/2



$L = a^+ b^+$

$K = a^{-1}L = a^* b^+$

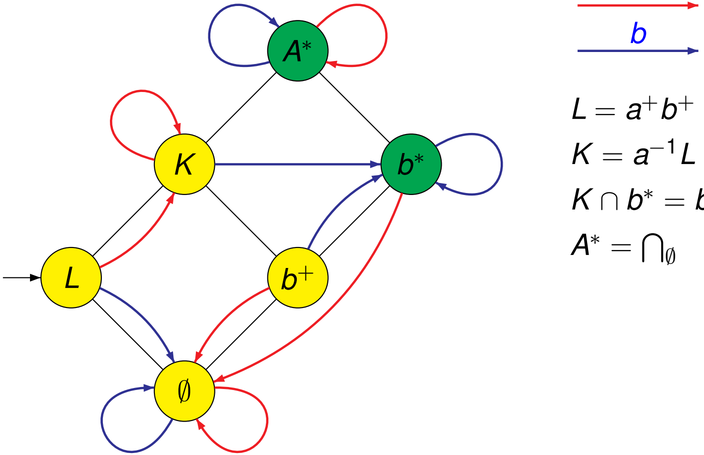
$L \not\subseteq L \cdot b = \emptyset$

III.2 Meet Automata

Intersections of Left Quotients

- For a language $L \subseteq A^*$ we extend the canonical automaton $(D_L, A, \cdot)$, where states are subsets of A^*.
- We can consider intersections of states: $U_L = \{\bigcap_{j\in I} K_j \mid I \text{ finite set }, K_j \in D_L\}$. If $I = \emptyset$ then we put $\bigcap_{j\in I} K_j = A^*$.
- The finite set U_L is equipped with the operation intersection $\cap$ and we can define $(\bigcap_{j\in I} K_j) \cdot a = \bigcap_{j\in I}(K_j \cdot a)$.
- We have the automaton $(U_L, A, \cdot)$ with semilattice operation $\cap$. Moreover, A^* is the largest element in the semilattice $(U_L, \cap)$ and it is an absorbing state in $(U_L, A, \cdot)$.
- Naturally $F = \{K \mid \lambda \in K\}$ is closed w.r.t. $\cap$ and F is upward closed, i.e. F is a filter.

An Example of a Meet Automaton



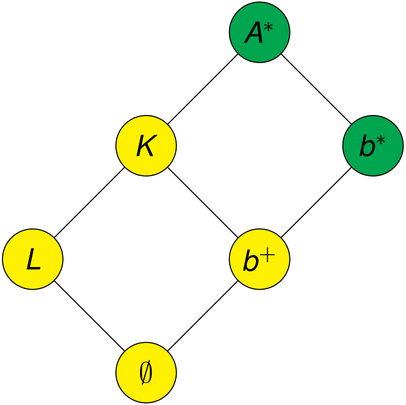
$L = a^+ b^+$

$K = a^{-1}L = a^* b^+$

$K \cap b^* = b^+$

$A^* = \bigcap_\emptyset$

An Example of a Meet Automaton



$$L = a^+ b^+$$
$$K = a^{-1}L = a^* b^+$$
$$K \cap b^* = b^+$$
$$A^* = \bigcap_\emptyset$$

Meet Automata

Definition

A structure $(Q, A, \cdot, \wedge, \top)$ is a meet automaton if

- $(Q, A, \cdot)$ is a DFA,
- $(Q, \wedge)$ is a semilattice with the largest element $\top$,
- actions by letters are endomorphisms of the semilattice $(Q, \wedge)$, i.e. $\forall p, q \in Q, a \in A : (p \wedge q) \cdot a = p \cdot a \wedge q \cdot a$
- $\top$ is an absorbing state.

This meet automaton recognizes a language L if there are $i, f \in Q$ such that $L = \{u \in A^* \mid i \cdot u \wedge f = f\}$.

Meet Automata

Definition

A structure $(Q, A, \cdot, \wedge, \top)$ is a meet automaton if

- $(Q, A, \cdot)$ is a DFA,
- $(Q, \wedge)$ is a semilattice with the largest element $\top$,
- actions by letters are endomorphisms of the semilattice $(Q, \wedge)$, i.e. $\forall p, q \in Q, a \in A : (p \wedge q) \cdot a = p \cdot a \wedge q \cdot a$
- $\top$ is an absorbing state.

This meet automaton recognizes a language L if there are $i, f \in Q$ such that $L = \{u \in A^* \mid i \cdot u \wedge f = f\}$.

Let $\mathcal{A}$ be a meet automaton recognizing a language L. The automaton $\mathcal{U}_L = (U_L, A, \cdot, \wedge, A^*)$, the canonical meet automaton of L, is a minimal meet automaton of L: there is a subautomaton $\mathcal{A}'$ of $\mathcal{A}$ and surjective morphism of meet automata $\varphi : \mathcal{A}' \to \mathcal{U}_L$.

Meet Automata – Differences

- We can consider transformations of a meet automata given by words again, but now we need to “intersect” them. Thus instead of a transition monoid we define a transition semiring. A transition semiring of the canonical meet automaton of L is isomorphic to the syntactic (idempotent) semirng of L.

Meet Automata – Differences

- We can consider transformations of a meet automata given by words again, but now we need to “intersect” them. Thus instead of a transition monoid we define a transition semiring. A transition semiring of the canonical meet automaton of L is isomorphic to the syntactic (idempotent) semirng of L.
- f-subautomata, morphic images – obvious modifications.
- A product is defined in the same way but, it does not recognize unions of languages now: if for each $j \in I$ we have a state f_j in $(Q_j, A, \cdot_j, \wedge_j, \top_j)$, then the set $X = ((q_j)_{j \in I} \mid \exists j : q_j \geq f_j\}$ is not a filter.

Varieties of Meet Automata

- A disjoint union of meet automata is not considered now, since it is not a meet automaton. The role of a disjoint union was a sum (from the categorical point of view) of automata. In the case of meet automata, this role can be played by the product of meet automata.

Definition

A variety of meet automata $\mathbb{V}$ associates to every finite alphabet A a class $\mathbb{V}(A)$ of meet automata over alphabet A in such a way that

- $\mathbb{V}(A) \neq \emptyset$ is closed under direct finite products and morphic images,
- $\mathbb{V}$ is closed under f-subautomata.

An Eilenberg Type Correspondence

Theorem (Klíma, Polák)

There are mutually inverse isomorphisms between the lattice of all varieties of meet automata and the lattice of all conjunctive varieties of regular languages.

Varieties of Meet Automata – An Example

Example

For each alphabet A, a meet automata $(Q, A, \cdot, \wedge, \top)$ belongs to $\mathbb{S}(A)$ if $\forall q \in Q, a \in A \;:\; q \cdot a = q \cdot a \wedge q$ and

$$\forall q \in Q, a, b \in A \;:\; q \cdot ab = q \cdot a \wedge q \cdot b \,. \quad (*)$$

Then $\mathbb{S}$ is a variety of meet automata and the corresponding conjunctive variety of languages $\mathcal{S}$ satisfies $\mathcal{S}(A) = \{B^* \mid B \subseteq A\} \cup \{\emptyset\}$.

- $\mathcal{S}$ is a conjunctive variety of languages.
- For all $B \subseteq A$, the canonical meet automaton of $L = B^*$ is $\mathcal{U}_L = (U_L, A, \cdot, \cap, A^*)$ and it has just three states: $B^*, A^*, \emptyset$. We see $B^* \cdot a \in \{B^*, \emptyset\}$ and $\mathcal{U}_L$ satisfies $(*)$.

An Example

- Let $(Q, A, \cdot, \wedge, \top)$ be a meet automaton satisfying

$$\forall q \in Q, a, b \in A \cup \{\lambda\} \ : \ q \cdot ab = q \cdot a \wedge q \cdot b \, . \quad (*)$$

  And we choose $i, f \in Q$.
- For $a, b, c \in A$ we have

$$(q \cdot a) \cdot bc = (q \cdot a) \cdot b \wedge (q \cdot a) \cdot c = q \cdot a \wedge q \cdot b \wedge q \cdot c \, .$$

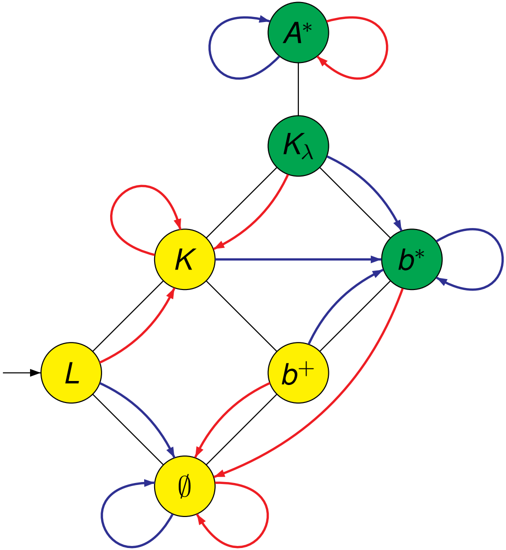
- In general $q \cdot a_1 \dots a_n = q \cdot a_1 \wedge \dots \wedge q \cdot a_n$ and (for $q = i$) we have $a_1 \dots a_n \in L \iff a_1 \in L, \dots, a_n \in L$.
- We can denote $B = L \cap A$ and we have $L = B^*$.
- Note that an equational characterization is that syntactic semiring satisfies the equality $xy = x \wedge y$.

III.3 DL-automata

The Canonical DL-Automaton of a Language

- For a language $L \subseteq A^*$ we extend the canonical meet automaton $(U_L, A, \cdot, \wedge, A^*)$ by unions of states:
  $W_L = \{\bigcup_{j \in I} M_j \mid I \text{ finite set }, M_j \in U_L\}$.
  If $I = \emptyset$ then we put $\bigcup_{j \in I} M_j = \emptyset$.
- The finite set W_L is equipped with the operations intersection $\cap$ and union $\cup$ (due to distributive laws).
  We can define $(\bigcup_{j \in I} M_j) \cdot a = \bigcup_{j \in I}(M_j \cdot a)$.
- We have the automaton $(W_L, A, \cdot)$ and a distributive lattice $(W_L, \cap, \cup)$. Moreover, A^* is the largest element, $\emptyset$ is the smallest element – both are absorbing states in $(W_L, A, \cdot)$.
- Naturally $F = \{M \mid \lambda \in M\}$ is closed w.r.t. $\cap$, upward closed, and $M_1 \cup M_2 \in F$ implies $M_1 \in F$ or $M_2 \in F$.
  I.e. F is an ultrafilter. In other words the intersection of all elements in F (the minimum in F) is join-irreducible.

An Example of a Canonical DL-Automaton



a

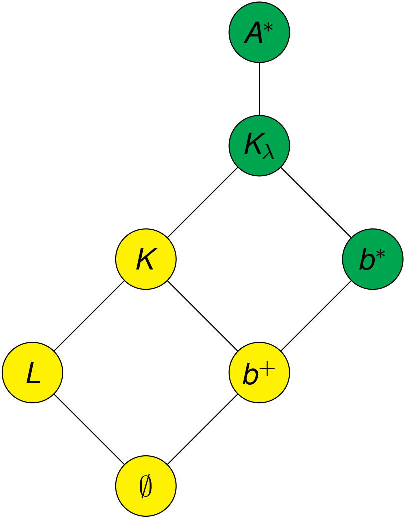
b

$$L = a^+ b^+$$

$$K = a^{-1}L = a^* b^+ = L \cup b^+$$

$$K_\lambda = K \cup b^* = K + \lambda$$

An Example of a Canonical DL-Automaton



$$L = a^+b^+$$
$$K = a^{-1}L = a^*b^+ = L \cup b^+$$
$$K_\lambda = K \cup b^* = K + \lambda$$

Technical Obstacles

- Now actions by letters, and consequently by words, are endomorphisms of the distributive lattice $(W_L, \cap, \cup)$. Unfortunately, if we take two of them, their intersection needn't be an endomorphism.
- Initial state can not be taken an arbitrary state of *DL*-automaton, since they recognize intersections and unions which are not allowed in our varieties now. Therefore we need to keep an information which states comes from the canonical automaton D_L.
- The sum of a potential *DL*-automata is more complicated (e.g. for testing whether the class of DL-automata is closed with respect to this operation). Therefore we prefer to work with a *DL*-automaton generated from a single state.

A DL-Automata – a Formal Definition

Definition (new)

A structure $(i, P, Q, A, \cdot, \wedge, \vee, \perp, \top)$ is a DL-automaton if

- $i \in P \subseteq Q$,
- $(Q, A, \cdot)$ is a DFA,
- $(Q, \wedge, \vee)$ is a distributive lattice with the minimum element $\perp$ and the largest element $\top$,
- actions by letters are endomorphisms of the lattice $(Q, \wedge, \vee)$,
- $\top$ and $\perp$ are absorbing states,
- P is the set of all states reachable from i,
- the lattice Q is generated by the set P.

Languages Recognized by a DL-Automaton

- A *DL*-automaton $(i, P, Q, A, \cdot, \wedge, \vee, \bot, \top)$ recognizes a language L if there are $j \in P, f \in Q$ such that f is a join-irreducible and $L = \{u \in A^* \mid j \cdot u \geq f\}$.
- If the previous automaton is the canonical DL-automaton $(L, D_L, W_L, A, \cdot, \cap, \cup, \emptyset, A^*)$ of a language L, then it recognizes exactly the languages of the form $u^{-1}Lv^{-1}$.
- Morphic images are define in the same way. Definitions of other operators must be modified.
    - Instead of a direct product we must take a substructure generated from the state which contains in all coordinates the initial states.
    - In the definition of the f-subautomaton we must take new initial state in the set P of the original DL-automaton.

An Eilenberg Type Correspondence

Definition (new)

Let $\mathcal{C}$ be a "Straubing" class of morphisms. A weak $\mathcal{C}$-variety of languages $\mathcal{V}$ associates to every finite alphabet A a class $\mathcal{V}(A)$ of regular languages over A in such a way that

- $\mathcal{V}(A)$ is closed under quotients,
- $\mathcal{V}$ is closed under preimages in morphisms from $\mathcal{C}$.

Theorem (new)

There are mutually inverse isomorphisms between the lattice of all $\mathcal{C}$-varieties of DL-automata and the lattice of all weak $\mathcal{C}$-varieties of regular languages.

An Eilenberg Type Correspondence – An Example

Example

Let $\mathcal{V}$ be a class of languages such that
$\mathcal{V}(A) = \{A^*aA^* \mid a \in A\} \cup \{A^*\}$.

- $\mathcal{V}(A)$ is not closed under intersections nor unions, i.e. $\mathcal{V}$ is not a conjunctive (nor disjunctive) variety of languages.
- Let $f : B^* \to A^*$, $a \in A$, $L = A^*aA^*$, then $f^{-1}(L) = B^*DB^*$ where $D = \{d \in B \mid f(d)$ contains $a\}$.
  Therefore we should consider only f's such that
  $$\forall b, c \in B : b \neq c \implies \mathrm{c}(f(b)) \cap \mathrm{c}(f(c)) = \emptyset\,.$$
- $\mathcal{V}$ is a weak $\mathcal{C}$-variety for such morphisms.

An Example

- Since $(A^*aA^*)^c = B^*$ for $B = A \setminus \{a\}$ we can take the dual condition characterizing the conjunctive variety $\mathcal{S}$ given by $\mathcal{S}(A) = \{B^* \mid B \subseteq A\} \cup \{\emptyset\}$, namely

$$\forall q \in Q, a, b \in A \cup \{\lambda\} \; : \; q \cdot ab = q \cdot a \vee q \cdot b \,. \quad (*)$$

- In particular $q \cdot a \geq q$.
- Another property is the following

$$\forall q \in Q, a, b \in A : a \neq b \implies q \cdot a \wedge q \cdot b = q \,. \quad (**)$$

- If $L = A^*a_1A^* \cup \cdots \cup A^*a_nA^*$ for $a_1, \ldots, a_n$ different letters, $n \geq 2$, then $L \cdot a_1 \cap L \cdot a_2 = A^* \neq L$, i.e. the canonical DL-automaton of L does not satisfy the condition $(**)$.
- The equational description of the property $(**)$ is $x \wedge y = 1$ with substitutions from the restricted class $\mathcal{C}$.

A Future Research

- We plan to clarify all details in the Eilenberg type correspondence for DL-automata.
- We are trying to characterize classes of regular languages given by certain models of quantum automata.
- We would like to improve our knowledge of an equational logic of meet automata and of DL-automata.
- A further step: BA-automata
  (automata equipped with a structure of a Boolean algebra).

Thank you.