Quest for Short Identities in Transformation Semigroups and Symmetric Groups

Olga Karpova
Ural Federal University

Joint work with:
Andrei Bulatov (Simon Fraser University),
Arseny Shur (Ural Federal University),
and Konstantin Startsev (Ural Federal University)

Basic definitions

- $\mathcal{A} = (\Sigma, Q, \delta, s, T)$ - deterministic finite automaton (DFA), where
  - Σ - finite input alphabet,
  - Q - finite set of states,
  - $\delta : (Q \times \Sigma) \to Q$ - full transition function,
  - $s \in Q$ - initial state,
  - $T \subseteq Q$ - set of terminal states.
- $q.w \in Q$ is the state obtained by reading the word $w \in \Sigma^*$ from the state q
- $\mathcal{A}$ accepts w if $s.w \in T$
- $\mathcal{A}$ separates u from v if $\mathcal{A}$ accepts exactly one of u and v
- A simplified view: $\mathcal{A}$ separates u from v if $s.u \neq s.v$, terminal states do not matter. $\mathcal{A} = (\Sigma, Q, \delta, s)$ is a quadruple

Separating words problem

- A general question: given two words u and v, how big is the smallest automaton separating u from v?
- Let $Sep(u, v)$ be the minimum number of states in a DFA separating u from v.
- $n = \max(|u|, |v|)$
- $Sep(n) = max_{u \neq v; |u|,|v| \leq n} Sep(u, v)$
- Problem: find the asymptotics of the function $Sep(n)$.
  - stated by Goralcik and Koubek in 1986.

Known results

Separation function:

- Lower bound: $Sep(n) = \Omega(\log n)$ (Goralcik and Koubek)
  - More precisely, the lower bound is $\log n + o(\log n)$ (natural logarithm)
- Upper bound: $Sep(n) = O(n^{2/5} \log^{3/5}(n))$ (Robson, 1989)
- No advances on the problem since then.

Known results

Separation function:

- Lower bound: $Sep(n) = \Omega(\log n)$ (Goralcik and Koubek)
  - More precisely, the lower bound is $\log n + o(\log n)$ (natural logarithm)
- Upper bound: $Sep(n) = O(n^{2/5} \log^{3/5}(n))$ (Robson, 1989)
- No advances on the problem since then.

Related results:

- A pair of random words of any length needs, on expectation, less than 3 states to separate;
- The words that are hard to separate (if exist) should have
  - a long common prefix;
  - a long common suffix;
  - the same number of occurrences of each short factor;
  - a big enough Hamming distance;
  - ...

Identities in full transformation semigroup

- Full transformation semigroup T_k is the semigroup of all selfmaps of the set $\{1, 2, ..., k\}$ under the composition of maps.
- Example. T_2 consists of four maps: $f_1(x) = 1$, $f_2(x) = 2$, $f_3(x) = x$, $f_4(x) = 3 - x$, and has the Cayley table

$\circ$	f_1	f_2	f_3	f_4
f_1	f_1	f_2	f_1	f_2
f_2	f_1	f_2	f_2	f_1
f_3	f_1	f_2	f_3	f_4
f_4	f_1	f_2	f_4	f_3

- Identity in T_k: $u \equiv_k v$ if for every subtitution $\sigma : \Sigma \to T$ one has $\sigma(u) = \sigma(v)$
- Given an automaton $\mathcal{A}$, the set of all maps $w : q \to q.w$ is a subsemigroup of $T_{|Q|}$ called the transition semigroup of $\mathcal{A}$.
- All these semigroups are in fact monoids.

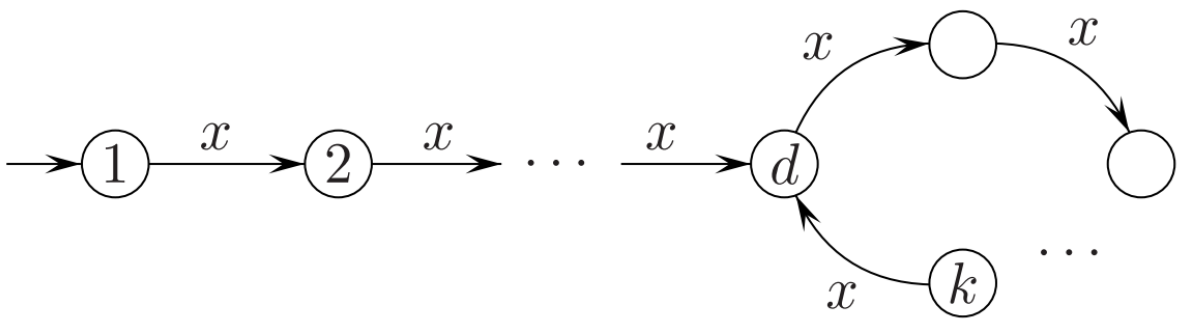
Identities in T_k and words separation

Observation. For any words $u, v \in \Sigma^*$ $u \equiv_k v$ iff $Sep(u, v) > k$.

- Decision problem "given u, v, and k, is $u \equiv_k v$?" is coNP-complete for $k \geq 3$. (Almeida, Volkov, Goldberg 2007).
  - Then the decision problem "given u, v, and k, is $Sep(u, v) \leq k$?" is NP-complete.
- Finding the asymptotics of $Sep(n)$ is equivalent to finding the asymptotics of the length of the shortest identity in T_k.

Unary identity

- The shortest known identity in T_k is $x^{k-1} \equiv_k x^{k-1+lcm\{1,\ldots,k\}}$, where *lcm* is the least common multiple.
- This means that for any $\mathcal{A} = (\Sigma, Q, \delta, s)$ such that $|Q| \leq k$, any $q \in Q$ and any $x \in \Sigma$ the states $q.x^{k-1}$ and $q.x^{k-1+lcm\{1,\ldots,k\}}$ coincide:



- Since $lcm\{1, .., k\} = e^{k+o(k)}$ from the Prime Number Theorem, one gets $Sep(n) \geq \log n + o(\log n)$.
- Does T_k have shorter (non-unary) identities?

Separating words with permutational DFA

- DFA $\mathcal{A}$ is permutational if every letter acts on the set of states as permutation: $|Q.a| = |Q|$.
- $Sepp(n)$ is the analog of function $Sep(n)$ for permutational automata.
- $Sepp(u, v) > k$ iff (u, v) is an identity in the symmetric group S_k. Such an identity $u \cong_k v$ is called positive (contains no inverses).

Separating words with permutational DFA

- DFA $\mathcal{A}$ is permutational if every letter acts on the set of states as permutation: $|Q.a| = |Q|$.
- $Sepp(n)$ is the analog of function $Sep(n)$ for permutational automata.
- $Sepp(u, v) > k$ iff (u, v) is an identity in the symmetric group S_k. Such an identity $u \cong_k v$ is called positive (contains no inverses).

Known results for permutational automata:

- Upper bound: $Sepp(n) = O(n^{1/2})$ (Robson, 1996).
- Lower bound: $Sepp(n) \geq \log(n) + o(\log(n))$
  - from the identity $x^{lcm\{1,\ldots,k\}} = 1$
- Are there short identities in S_k?
  - Yes, of length $O(e^{\log^4 n \log\log n})$ (Kozma, Thom, 2016)
  - No positive short identities is known

Our goals

- To find short identities in T_k and S_k for small k (as far as the optimized exhaustive search is feasible).
- Using these identities, find some general series of identities for arbitrary k and use them to update lower bounds on Sep and $Sepp$

Restrictions for examined words

- The shortest non-unary identity in T_k (S_k) is binary; so $\Sigma = \{x, y\}$.
- If $u \equiv_k v$, then $u = pu_1r$, $v = pv_1r$, where $|p| \geq k-2$, $|r| \geq k-1$, and $u_1 \cong_k v_1$.
  Idea:
  Identity in S_k + common prefixes and suffixes = identity in T_k
- If an identity is shorter than the unary identity, then $|u| = |v|$ and $|u|_x = |v|_x$.

Shortest Identities for $k = 3, 4$

In S_3:

- $x^2y^2 = y^2x^2$ (length 4).

In T_3:

- $x^2 = x^8$ (length 8).
  - ⋆ the shortest binary identity has length 10

In S_4:

- $x^6y^2xy^2 = y^2xy^2x^6$ (length 11).

In T_4:

- $x^3 = x^{15}$ (length 15)
  - ⋆ the shortest binary identity has length 18

Shortest Identities for k=5,6

In S_5:

- $(xy)(xyyx)^3(yxxy)^2(yx)(yxxy)^2 = (yxxy)^2(xy)(yxxy)^2(xyyx)^3(yx)$ (length 32)
- $(xy)^4(yx)^5(xy)^6(yx) = (yx)(xy)^6(yx)^5(xy)^4$ (length 32)
- No irreducible identities of length 33.

In T_5:

- $(xy)^{15}(yx)^5(xy)^4 = (xy)^3(yx)^5(xy)^{16}$ (length 48)
  - suspected (but not proved) to be the shortest identity

In S_6

- $(xy)^4(yx)^5(xy)^6(yx) = (yx)(xy)^6(yx)^5(xy)^4$ (length 32)

In T_6

- No luck yet.

Identities in T_k

Theorem. Semigroup T_k satisfies the following identity of length $2lcm\{1,..,k-1\}+6(k-1)$:

$$(xy)^{k-2+lcm\{1,..,k-1\}}(yx)^k(xy)^{k-1} \equiv_k (xy)^{k-2}(yx)^k(xy)^{k-1+lcm\{1,..,k-1\}}$$

Corollary. If $k \geq 5$ is either a prime or an odd prime power, the semigroup T_k satisfies an identity which is shorter than the unary identity.

- The result improves a little the lower bound for $Sep(n)$, but the improvement is swallowed by the o-term.

Identities in S_k

Proposition. $(xy)^a(yx)^b \cong_k (yx)^b(xy)^a$, if the order of any element of S_k divides at least one of a, b.

Theorem. $Sepp(n) \geq \frac{3}{2}\log(n) + O(\frac{\log n}{\log\log n})$

- This is the first (though small) asymptotic improvement of the lower bound on words seperation.

Proposition. $(xy)^a(yx)^b(xy)^c(yx)^d \cong_k (yx)^d(xy)^c(yx)^b(xy)^a$, if any order h of an element of S_k satisfies at least one of the following conditions or their counterparts obtained by swapping b with c and a with d:

- $h|a$ and $h|c$
- $h|(a+c)$ and $h|b$
- $h|a$ and $b \equiv d \pmod{h}$

Short identities of special forms in S_k

	Identities of type "abcd"					Identities of type "ab"			
k	a	b	c	d	Len	a	b	Len	$lcm(k)$
5, 6	1	6	5	4	32	12	5	34	60
7	2	14	12	10	76	60	7	134	420
8	23	60	7	24	228	60	56	232	840
9	18	60	42	24	288	180	56	472	2520
10	18	60	42	24	288	120	126	492	2520
11	48	180	132	84	888	840	198	2076	27720
12	24	222	420	198	1728	840	198	2076	27720
13						2520	286	5612	360360
14						2520	858	6756	360360
15						2520	1716	8472	360360
16						5040	8580	27240	720720
17						27720	10608	76656	12252240
18						55440	13260	137400	12252240
19						55440	251940	614760	232792560
20						360360	15504	751728	232792560
21						360360	77520	875760	232792560
22						360360	77520	875760	232792560
23						720720	445740	2332920	5354228880

Thank you for attention!