Journal of Automata, Languages and Combinatorics **26** (2021) 1–2, 67 – 89

WORDS SEPARATION AND POSITIVE IDENTITIES IN SYMMETRIC GROUPS

Olga Karpova[(A)] Arseny M. Shur[(A,B)]

[(A)] *Department of Algebra and Fundamental Informatics, Ural Federal University*
pr. Lenina 51, 620000 Ekaterinburg, Russia
scckleppi@gmail.com (O. Karpova) arseny.shur@urfu.ru (A. M. Shur)

ABSTRACT

We study short positive (i. e., containing no inverses) identities of finite symmetric groups. The interest in such identities is inspired by the problem of separating words with finite automata, in particular, with the automata in which each letter acts on the set of states as a permutation.

We list all positive identities in symmetric groups S_5 and S_6 up to length 40 and prove that the shortest identity in the full transformation semigroup T_5 has length 48. Then we propose the classification of positive identities and describe the identities from some particular classes. We state a conjecture that the shortest positive identity in a sufficiently big symmetric group is "numerical", which implies an $O\big(\frac{\log^2 n}{\log\log n}\big)$ upper bound on the words separation function.

Keywords: symmetric group, identity, finite automaton, words separation

1. Introduction

The problem we study is related to both automata theory and group theory. Let us first introduce the automata-theoretic context. Look at one of the simplest computational tasks: distinguish between two inputs. The simplicity of the task calls for a simple computing device to solve it. Let a deterministic finite automaton (DFA) be such a device; we say that a DFA *separates* two words if it accepts one of them and rejects the other. Two natural problems about separation by DFAs can be formulated. First, *given two input words u and v, what is the minimum size* $\mathsf{Sep}(u,v)$ *of a DFA which separates them?* Second, *given a number n, what is the minimum number* $k = \mathsf{Sep}(n)$ *such that any two words of length at most n can be separated by a DFA with at most k states?* The first problem is known to be NP-hard (NP-complete in the decision version) if the alphabet of the input words is unbounded. This result follows from the complexity studies on a purely algebraic problem of identity checking [1, 7]. As for the case of a constant-size alphabet, we are not aware of any published results.

[(B)]Supported by the Ministry of Science and Higher Education of the Russian Federation (Ural Mathematical Center project No. 075-02-2020-1537/1).