Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 370 (2007) 131–169
Theoretical Computer Science
www.elsevier.com/locate/tcs

An automata-theoretic approach to the word problem for ω-terms over R

Jorge Almeida[a], Marc Zeitoun[b,c,*]

[a] *Centro de Matemática e Dep. Mat. Pura, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal*
[b] *LIAFA, Université Paris 7 and CNRS UMR 7089, France*
[c] *LaBRI, Université Bordeaux 1 and CNRS UMR 5800, 351 cours de la Libération, 33405 Talence Cedex, France*

Received 9 November 2004; received in revised form 18 October 2006; accepted 25 October 2006

Communicated by D. Perrin

Abstract

This paper studies the pseudovariety R of all finite $\mathcal{R}$-trivial semigroups. We give a representation of pseudowords over R by infinite trees, called R-trees. Then we show that a pseudoword is an ω-term if and only if its associated tree is regular (*i.e.* it can be folded into a finite graph), or equivalently, if the ω-term has a finite number of tails. We give a linear algorithm to compute a compact representation of the R-tree for ω-terms, which yields a linear solution of the word problem for ω-terms over R. We finally exhibit a basis for the ω-variety generated by R and we show that there is no finite basis. Several results can be compared to recent work of Bloom and Choffrut on long words.

Keywords: Semigroup; Pseudovariety; Word problem; Automata minimization; Pseudoword; Omega-term; Identity basis

1. Introduction

The main contribution of this paper is the solution of a word problem over R, the pseudovariety of all finite $\mathcal{R}$-trivial semigroups. This pseudovariety corresponds, in Eilenberg's correspondence, to disjoint unions of languages of the form $A_0^*a_1A_1^*a_2\dots a_nA_n^*$, where the a_i's are letters and $a_i \notin A_{i-1}$ for $1 \le i \le n$. Also, finite $\mathcal{R}$-trivial semigroups are the divisors of transition semigroups of the so-called *very weak* automata, that is, automata whose state set is partially ordered and the transition function does not decrease the state. They can even be characterized as the divisors of *extensive* automata, that is, very weak automata where the order on states is total.

Given two terms built from letters of an alphabet A using the concatenation and the ω-power, we show how to decide in linear time whether these terms coincide over all A-generated elements of R, with the usual interpretation of the ω-power in semigroups. We also characterize the set of pseudowords – also known as implicit operations – over R which can be represented by such ω-terms. Since R satisfies the identity in ω-terms $x^{\omega-1} = x^\omega$, all results of this

* Corresponding author at: LaBRI, Université Bordeaux 1 and CNRS UMR 5800, 351 cours de la Libération, 33405 Talence Cedex, France. Tel.: +33 5 40 00 60 94; fax: +33 5 40 00 66 69.

E-mail addresses: jalmeida@fc.up.pt (J. Almeida), mz@labri.fr (M. Zeitoun).

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2006.10.019

paper can be formulated either for ω-terms, or for κ-terms, where $\kappa = \{\_ \cdot \_, \_^{\omega-1}\}$ is the implicit signature consisting of the semigroup multiplication and the unary $(\omega - 1)$-power. We shall state most results using the signature $\{\_ \cdot \_, \_^{\omega}\}$, but this is mainly a matter of style.

The motivation of this work is the κ-tameness property for R. Historically, the notion of tameness was discovered in attempting to find general decidability properties of pseudovarieties which might be preserved under taking semidirect products [5]. It remains open whether it does indeed play such a role, although under certain finiteness hypotheses it has been shown to do so [2].

Proving the κ-tameness of a pseudovariety V consists in solving two subproblems. The first one is the κ-word problem, for which this paper gives an efficient solution. Informally, the second question is whether equation systems[1] with rational constraints having a solution in any semigroup of V also have a *uniform* solution in κ-terms, satisfying the same constraints. This property has proved to be robust and helpful for the solution of the membership problem (see *e.g.* [4], where the κ-tameness of R is used to decide joins involving R). Moreover, if V enjoys it, then V has decidable pointlikes, an important property of pseudovarieties [5,20].

Another motivation for this study comes from the related work of Bloom and Choffrut [11]. Given a finite set A, the collection of all finite or countably infinite A-labeled posets can be endowed with a binary concatenation operation of posets $\_ \cdot \_$, and with a unary ω-power $\_^{\omega}$. Bloom and Choffrut recently proved in [11] that the Birkhoff variety generated by these algebras is not finitely based, and that it is defined by the following set of identities.

$$\begin{cases} (x \cdot y) \cdot z = x \cdot (y \cdot z) \\ (x^r)^{\omega} = x^{\omega}, \quad r \geq 2 \\ (xy)^{\omega} = x(yx)^{\omega}. \end{cases}$$

They also studied ordinal words, that is, labeled ordinals. Among them, they characterized labeled ordinals built from letters of A using the operations $\_ \cdot \_$ and $\_^{\omega}$: these are exactly the ordinals of length less than ω^{ω} and having a finite number of tails (suffixes, in some sense). Finally, they proved that the word problem for two ω-terms u, v can be solved in time $O(|u|^2|v|^2)$, where $|u|$ and $|v|$ denote the lengths of u and v.

Motivated by these results and by the fact that pseudowords over R are labeled ordinals [7], we show that:

- the word problem for ω-terms u, v over R and on an alphabet A can be solved in time $O\big(|A|(|u| + |v|)\big)$, using automata-based techniques. More specifically, we can compute for any ω-term u an automaton $\mathcal{A}(u)$ of size $|A||u|$. Two terms are equal over R if and only if the associated automata recognize the same language. Due to the specific form of these automata, this can again be tested in linear time;
- a pseudoword over R coincides with an ω-term if and only if it has a finite number of distinguished suffixes (resp. factors);
- the variety of ω-semigroups generated by R is not finitely based;
- we exhibit an infinite basis for this variety.

Although these results are very similar to those of [11], the involved word problems are different, and neither set of results seems to directly imply the other one.

The paper is organized as follows. In Section 2, we set up the notation and we recall prerequisites on semigroups and pseudovarieties. In Section 3, we exhibit a sufficient condition for continuity of infinite products in pro-R semigroups and we use it to associate R-trees and R-automata to pseudowords over R. These objects are used in Section 4 to solve the word problem for ω-terms over R and to derive several characterizations of pseudowords having a representation as an ω-term. We then exhibit a canonical form for ω-terms over R, which can be exponentially larger than the original term, in terms of the size of the alphabet, but remains polynomially small, for a fixed alphabet, in terms of the size of the minimal R-automaton of the ω-term. Section 5 presents a linear-time algorithm to compute the canonical R-automaton associated to an ω-term, defined in Section 3, thus proving that the complexity of the word problem for ω-terms over R is linear. We introduce in Section 6 a set of identities in ω-terms. We prove, by a rather involved argument with various levels of nested inductions which uses several key results from previous sections, that this set is a basis for the ω-variety generated by R. We also show that this ω-variety is not finitely based. It should be noted that

[1] Strictly speaking, if the equations are given by arbitrary κ-terms, the property is named complete κ-tameness, whereas κ-tameness stands for a restricted class of equations.