Two-Variable Word Equations
(Extended Abstract)

Lucian Ilie and Wojciech Plandowski
1 Turku Centre for Computer Science TUCS
FIN-20520 Turku, Finland
lucili@cs.utu.fi
2 Institute of Informatics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
wojtekpl@mimuw.edu.pl

Abstract. We consider languages expressed by word equations in two
variables and give a complete characterization for their complexity func-
tions, that is, the functions that give the number of words of a given
length. Specifically, we prove that there are only five types of complex-
ities: constant, linear, exponential, and two in between constant and
linear. For the latter two, we give precise characterizations in terms of
the number of solutions of Diophantine equations of certain types. There
are several consequences of our study. First, we show that the linear up-
per bound on the non-exponential complexities by Karhumäki et al., cf.
[KMP], is optimal. Second, we derive that both of the sets of all finite
Sturmian words and of all finite Standard words are expressible by word
equations. Third, we characterize the languages of non-exponential com-
plexity which are expressible by two-variable word equations as finite
unions of several simple parametric formulae and solutions of a two-
variable word equation with a finite graph. Fourth, we find optimal up-
per bounds on the solutions of (solvable) two-variable word equations,
namely, linear bound for one variable and quadratic for the other. From
this, we obtain an $O(n^2)$ algorithm for testing the solvability of two-
variable word equations.

Keywords: word equation, expressible language, complexity function,
minimal solution, solvability

1 Introduction

Word equations constitute one of the basic parts of combinatorics on words. The
fundamental result in word equations is Makanin’s algorithm, cf. [Ma], which
decides whether or not a word equation has a solution. The algorithm is one of
the most complicated ones existing in the literature. The structure of solutions

\* Research supported by the Academy of Finland, Project 137358. On leave of absence
from Faculty of Mathematics, University of Bucharest, Str. Academiei 14, R-70109,
Bucharest, Romania.
\*\* Supported by KBN grant 8 T11C 039 15.

© Springer-Verlag Berlin Heidelberg 2000
of word equations is not well understood; see [Hm, Raz1, Raz2]. A new light on that topic has been led recently by [KMP] where the languages which are defined by solutions of word equations are studied.

The structure of languages which are defined by equations with one variable is very simple. The infinite languages which are defined by one-variable word equations consist of a finite part and an infinite part which is of the form $A^n A'$ for $A' \text{ a prefix of } A$. The structure of the finite part is not completely known [OGM]. Our analysis deals with languages which are defined by two-variable word equations. We prove that the complexity of those languages, which is measured by the number of words of a given length, belongs to one of five classes: constant, $D_1$-type, $D_2$-type, linear and exponential. The complexities $D_1$-type and $D_2$-type are in between linear and constant and they are related to the number of solutions of certain Diophantine equations. As a side effect of our considerations we prove that the linear upper bound given in [KMP] for languages which do not contain a pattern language is optimal. An interesting related result is that the sets of Sturmian and Standard words are expressible by simple word equations. As another consequence, we characterize the languages of non-exponential complexity which are expressible by two-variable word equations as finite unions of several simple parametric formulae and solutions of simple two-variable word equations.

Based on our analysis, we find optimal upper bounds on the solutions of (solvable) two-variable word equations, namely, linear bound for one variable and quadratic for the other. From this, we obtain an $O(n^6)$ algorithm for testing the solvability of two-variable word equations. We recall that the only polynomial-time algorithm known for this problem is the one given by Charatonik and Pacholski [ChPa]. Its complexity, as computed in [ChPa], is $O(n^{100})$. It should be added that they did not take very much care of the complexity. They mainly intended to prove that the problem can be solved in polynomial time.

Due to space limitations we remove all proofs in particular several lemmas which are used to prove our main theorem Theorem 3.

2 Expressible Languages

In this section we give basic definitions we need later on, as well as recalling some previous results. For an alphabet $\Sigma$, we denote by $\text{card}(\Sigma)$ the number of elements of $\Sigma$; $\Sigma^*$ is the set of words over $\Sigma$ with 1 the empty word. For $w \in \Sigma^*$, $|w|$ is the length of $w$; for $a \in \Sigma$, $|w|_a$ is the number of occurrences of $a$ in $w$. By $\rho(w)$ we denote the primitive root of $w$. If $w = uv$, then we denote $u^{-1}w = v$ and $v u^{-1} = u$. For any notions and results of combinatorics on words, we refer to [Lo] and [ChKa].

Consider two disjoint alphabets, of constants, $\Sigma$, and of variables, $\Xi$. A word equation $e$ is a pair of words $\varphi, \psi \in (\Sigma \cup \Xi)^*$, denoted $e : \varphi = \psi$. The size of $e$, denoted $|e|$, is the sum of the lengths of $\varphi$ and $\psi$. The equation $e$ is said to be reduced if $\varphi$ and $\psi$ start with different letters and end with different letters, as