Journal of Computer and System Sciences 85 (2017) 132–167

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

The extended equation of Lyndon and Schützenberger ☆

Florin Manea [a,*], Mike Müller [a], Dirk Nowotka [a], Shinnosuke Seki [b]

[a] *Kiel University, Department of Computer Science, D-24098 Kiel, Germany*
[b] *University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, 1828585, Japan*

ARTICLE INFO

Article history:
Received 19 March 2015
Received in revised form 11 November 2016
Accepted 16 November 2016
Available online 28 November 2016

Keywords:
Combinatorics on words
Word equations
Lyndon–Schützenberger equation
Pseudoperiodicity

ABSTRACT

Lyndon and Schützenberger (1962) [3] investigated for which values of ℓ, m, and n, the word-equations $u^\ell = v^m w^n$ have only periodic solutions. Following their result, we determine precisely the values of ℓ, m, and n for which the *generalised Lyndon–Schützenberger word equations* $u_1 \cdots u_\ell = v_1 \cdots v_m w_1 \cdots w_n$, where $u_i \in \{u, \theta(u)\}$ for all $1 \le i \le \ell$, $v_j \in \{v, \theta(v)\}$ for all $1 \le j \le m$, $w_k \in \{w, \theta(w)\}$ for all $1 \le k \le n$, and θ is an antimorphic involution, have only θ-periodic solutions, i.e., $u, v, w \in \{t, \theta(t)\}^*$ for some word t. This answers completely an open problem by Czeizler et al. (2009) [22].

1. Introduction

The study of the classical word equations $u^\ell = v^m w^n$ dates back to the year 1962. Lyndon and Schützenberger [3] showed that for $l, m, n \ge 2$, in all solutions of this equation in a free group, u, v, w are powers of a common element, or in other words, they are elements of the subgroup generated by some element of this free group. Such solutions are also referred to as *periodic* solutions. Their result extends canonically to the case when u, v and w are elements of a free semigroup. In this case however, significantly simpler proofs have been established over the years [4–7].

Lentin [8] studied generalisations of the form $u^\ell = v^m w^n x^p$, while Appel and Djorup [9] looked at equations of the form $u^\ell = v_1^\ell v_2^\ell \cdots v_n^\ell$. Finally, the most general form of these equations, namely $u^\ell = v_1^{k_1} v_2^{k_2} \cdots v_n^{k_n}$ was investigated by Harju and Nowotka [10].

Czeizler et al. [11] introduced a generalisation of Lyndon & Schützenberger's equations of a different kind. They considered equations of the form

$$u_1 u_2 \cdots u_\ell = v_1 v_2 \cdots v_m w_1 w_2 \cdots w_n,$$

where $u_i \in \{u, \theta(u)\}$ for all $1 \le i \le \ell$, $v_j \in \{v, \theta(v)\}$ for all $1 \le j \le m$, and $w_k \in \{w, \theta(w)\}$ for all $1 \le k \le n$, and studied under which conditions $u, v, w \in \{t, \theta(t)\}^+$ for some word t. In other words, they studied the case when u, v, w are generalised powers (more precisely, θ-powers), and thus the solution is what is called θ-*periodic*. Here, θ is a function on the letters of the alphabet, which acts as an antimorphism (i.e., $\theta(uv) = \theta(v)\theta(u)$ for all words u, v) and as an involution (i.e.,

☆ Extended abstracts announcing the results of this paper were presented at the Conference on Foundations of Software Technology and Theoretical Computer Science 2013 [1] and the Symposium on Mathematical Foundations of Computer Science 2014 [2].

* Corresponding author.
E-mail addresses: flm@informatik.uni-kiel.de (F. Manea), mimu@informatik.uni-kiel.de (M. Müller), dn@informatik.uni-kiel.de (D. Nowotka), s.seki@uec.ac.jp (S. Seki).

http://dx.doi.org/10.1016/j.jcss.2016.11.003
0022-0000/

Table 1
Known results about the equations $u_1u_2\cdots u_\ell = v_1v_2\cdots v_m w_1w_2\cdots w_n$.

ℓ	m	n	$u, v, w \in \{t, \theta(t)\}^+$?
≥ 4	≥ 3	≥ 3	Yes [11,23]
3	≥ 5	≥ 5	Open
3	4	≥ 5 and odd	Open
3	4	≥ 4 and even	No [23]
3	3	≥ 3	No [23]
one of $\{\ell, m, n\}$ equals 2			No [11,23]

$\theta(\theta(u)) = u$ for all words u). These so-called *antimorphic involutions* are commonly used to formally model the Watson-Crick complementarity arising in DNA strands. It was this connection that made a systematic study of the combinatorial properties of words that can be expressed as a product of factors and their image under such antimorphic involutions, also called pseudo-repetitions, appealing (see, [11]). Apart from this initial bio-inspired motivation, there is a strong intrinsic mathematical motivation behind the study of such words. Indeed, one of the simplest and most studied operations on words is the reversal, the very basic antimorphic involution. It is thus natural to study equations on words, as well as other combinatorics on words concepts, in which not only powers of variables, but also repeated products of a variable and its mirror image appear. To this end, other topics studied in the context of pseudo-repetitions include: generalisations of the Theorem of Fine and Wilf [12–14], the avoidability of pseudo-repetitions [15–17], as well as algorithmic problems like deciding whether a word is a pseudo-repetition [18,19] or whether a word contains pseudo-repetitions [15,20,21].

The previous results obtained on generalised Lyndon–Schützenberger word equations, which were established by Czeizler et al. [11,22] and Kari, Masson, and Seki [23], are summarised in Table 1. Please note that m and n denote symmetric cases when exchanged. One can observe directly from this table that the more interesting cases in this generalised setting are those in which $\ell, m, n \geq 3$. Moreover, when $\ell = 3$ only several "negative" results have been found so far. By this we mean that there is a series of equations which have non-θ-periodic solutions, but very little is known about those cases of such equations where the θ-periodicity of the solutions is forced, similarly to the classical Lyndon–Schützenberger equations (the only exception was the particular Lemma 23, see Proposition 51 in [23]). Finally, the case $\ell = 3$ seems to be especially intricate and particularly interesting, as it separates the cases when the equation has only θ-periodic solutions ($\ell \geq 4$) from the cases when it may have other solutions as well ($\ell \leq 2$). In this paper, the remaining open cases are solved.

As expected (see the final remarks of [23]), we applied some arguments that have not been used in this context before, but an exhaustive case analysis on the alignments of parts of the equation seems unavoidable and these arguments must be adapted to every case separately.

2. Preliminaries

2.1. Words

Let Σ be a non-empty, finite set, called *alphabet*. We call the elements of Σ *letters*. A *word* over Σ is a (finite or infinite) sequence of letters from Σ. The set of non-empty finite words over Σ, also known as the *free semigroup* generated by Σ, is denoted as Σ^+. Endowing Σ^+ with a unique neutral element, which is called *empty word* and denoted as ε, we obtain the *free monoid* generated by Σ, which is denoted as Σ^* (hence, $\Sigma^* = \Sigma^+ \cup \{\varepsilon\}$). If $S \subseteq \Sigma^*$ is a set of words, then S^+ and S^* denote the free semigroup and free monoid generated by the words in S.

For a finite word w, we denote its *length*, that is the number of letters it consists of, by $|w|$. So, if $w = w_1w_2\cdots w_n$, where $w_i \in \Sigma$ for all $1 \leq i \leq n$, then $|w| = n$. The empty word ε is the unique word of length 0.

If $w = uvz$ for some words u, v and z, then we call u a *prefix*, v a *factor*, and z a *suffix* of w. We denote these relations as follows: $u \leq_p w$, $v \leq_f w$ and $z \leq_s w$. If $u \neq w$ and $u \neq \varepsilon$, then u is called a *proper prefix* of w, and similarly z is a *proper suffix* of w, if $z \neq w$ and $z \neq \varepsilon$. We use the notations $u <_p w$ and $z <_s w$ in this case. A factor that is neither prefix nor suffix of w is called a *proper factor*.

2.2. Periods, repetitions & basic equations on words

One of the most basic properties of a word is expressed by the notion of periodicity. A *period* of a word $w = w_1w_2\cdots w_n$ is a positive integer p, such that $w_i = w_{i+p}$ for all $1 \leq i \leq n - p$.

One of the most well-known, and probably also most frequently used results concerning periods in words, is the Theorem of Fine and Wilf [24]. The theorem, in its slightly generalised version [25], reads as follows, where gcd denotes the *greatest common divisor* of its arguments:

Theorem 1 *(Fine & Wilf, 1965; Shallit, 2008). Let $u, v \in \Sigma^*$ be words. If $\alpha \in u\{u, v\}^*$ and $\beta \in v\{u, v\}^*$ have a common prefix of length at least $|u| + |v| - \gcd(|u|, |v|)$, then $u, v \in \{t\}^+$ for some word t.*