Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On the aperiodic avoidability of binary patterns with variables and reversals

Robert Mercaş [1]

Department of Computer Science, Loughborough University, UK

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form 12 November 2016
Accepted 19 December 2016
Available online xxxx

Keywords:
Pattern with reversal
Binary avoidability
Aperiodic sequences

ABSTRACT

In this work we present a characterisation of the avoidability of all unary and binary patterns, that do not only contain variables but also reversals of their instances, with respect to aperiodic infinite words. These types of patterns were studied recently in either more general or particular cases.

1. Introduction

The *pattern unavoidability* concept was introduced by Bean, Ehrenfeucht and McNulty in [1] and by Zimin in [32]. A pattern consisting of variables is said to be unavoidable over a k-letter alphabet, if every infinite word over such an alphabet contains an instance of the pattern. That is, there exists a factor of the infinite word which is obtained from the pattern through an assignment of non-empty words to the variables (each occurrence of a variable is substituted with the same word).

The unary patterns, or powers of a single variable α, were investigated by Thue [30,31]: α is unavoidable, $\alpha\alpha$ is 2-unavoidable but 3-avoidable, and α^m with $m \geq 3$ is 2-avoidable. Schmidt proved that there are only finitely many binary patterns, or patterns over $E = \{\alpha, \beta\}$, that are 2-unavoidable [28,29]. Later on, Roth showed that there are no binary patterns of length six or more that are 2-unavoidable [27]. The classification of unavoidable binary patterns was completed by Cassaigne [7] who showed that $\alpha\alpha\beta\beta\alpha$ is 2-avoidable.

In time, the concept of unavoidability was investigated in several other contexts. The ternary patterns were fully characterised in [8,23], the binary patterns in the setting of partial words in [19,3–6], several variations of avoidability of patterns with restrictions on the length of the instances can be found in [26], while the binary patterns avoidable by cube-free words were characterised in [21] together with their growth rates. However, the topic of our work is mostly inspired by [25], where the authors look at the avoidability of words and their reversals, by [14] where the authors show that the pattern $\alpha\alpha\alpha^R$ is avoidable over a binary alphabet, and by the work in [2,10,20], where a more generalised form of avoidability, that of pseudo-repetitions, is investigated.

In this work, we investigate the avoidability of binary patterns, when some of the variables might be reversed. For example, instead of looking only at the pattern $\alpha\alpha$, we shall also investigate the pattern $\alpha\alpha^R$, which does not occur in the

E-mail address: R.G.Mercas@lboro.ac.uk.

[1] Work partially supported by the P.R.I.M.E. programme of DAAD co-funded by BMBF and EU's 7th Framework Programme (grant 605728) and by the Newton International Fellowship with funds from the Royal Society and the British Academy.

http://dx.doi.org/10.1016/j.tcs.2016.12.022
0304-3975/© 2016 Elsevier B.V. All rights reserved.

word 0101, while the former does (take $\alpha = 01$); this is obviously enough for length 2 unary patterns as other variations only consist of complements or mirror images. However, as most of these patterns are avoidable by trivial periodic words (as shown in [9,22]), we extend a bit our interest and focus on the cases when infinite aperiodic words which do not meet these patterns exist.

Our work is structured as follows. In the next section we present basic definitions and notations, as well as some preliminary observations. In Section 3 we give a characterisation for unary patterns with reversals, where the aperiodic constraint is also considered. Section 4 considers the current state of the art regarding avoidability of binary patterns with reversals. Finally, in Section 5 our focus is on the aperiodic avoidability version of the problem in the case of binary patterns with reversals (this section represents the main novelty of this work in comparison to [9,22]).

2. Definitions and preliminaries

Cassaigne's Chapter 3 of [18] provides background on unavoidable patterns, while the handbook itself contains detailed definitions on words.

Let Σ be a non-empty finite set of symbols called an *alphabet*. Each element $0 \in \Sigma$ is called a *letter*. A *word* is a sequence of letters from Σ. The *empty word* is the sequence of length zero, denoted by ε. The set of all finite words (respectively, non-empty finite words) over Σ is denoted by Σ^* (respectively, Σ^+).

A word u is a *factor* of a word v if there exist x, y such that $v = xuy$ (the factor u is *proper* if $u \neq \varepsilon$ and $u \neq v$). We say that u is a *prefix* of v if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$. The *length* of u is denoted by $|u|$ and represents the number of symbols in u. We denote by $u[i..j]$, where $0 \leq i \leq j < |u|$, the factor of u starting at position i in u and ending at position j, inclusive. By $|u|_v$ we denote the number of distinct, possibly overlapping, occurrences of a factor v in u. We denote by $u^R = u[|u|-1]\cdots u[1]u[0]$, the *reversal* or *mirror image* of a word u. A word u is said to be a *palindrome* if $u = u^R$. In this work we only consider palindromes of length greater than 1, as letters are just trivial instances of such.

For a word u, the powers of u are defined recursively by $u^0 = \varepsilon$ and for $n \geq 1$, $u^n = uu^{n-1}$. Furthermore, $\lim_{n\to\infty} u^n$ is denoted by u^ω. For legibility, the 2-powers of words are called *squares*, while 3-powers are called *cubes*. Furthermore, if $u = v^k v'$, where v' is a prefix of v, we say that u is a $\frac{k|v|+|v'|}{|v|}$-power.

A *period* of a word u is an integer p, such that for every defined positions i and $i + p$ of u, we have $u[i] = u[i + p]$. Furthermore, the minimal such p associated to some word is called *the (minimal) period* of the word. This can obviously be extended to infinite words, where the existence condition is dropped. An infinite word for which no such period exists is called *non-periodic*. Observe that in the case of non-periodic infinite words the period will increase the longer a prefix of the word is considered. Finally, if for an infinite word there exists no suffix of it which is periodic, the word is called *aperiodic*. In the case when such a suffix exists, thus the word is of the form uv^ω, the word is called *ultimately periodic*.

Let E be a non-empty finite set of symbols, distinct from Σ, whose elements are denoted by α, β, γ, etc. Symbols in E are called *variables*, and words in E^* are called *patterns*. The *pattern language* over Σ associated with a pattern $p \in E^*$, denoted by $p(\Sigma^+)$, is the subset of Σ^* containing all words of $\varphi(p)$, where φ is any non-erasing morphism that maps each variable in E to an arbitrary non-empty word from Σ^+. A word $w \in \Sigma^*$ *meets* the pattern p (or p *occurs* in w) if for a factorisation $w = xuy$, we have $u \in p(\Sigma^+)$. Otherwise, w *avoids* p.

More precisely, let $p = \alpha_0 \cdots \alpha_m$, where $\alpha_i \in E$ for $i \in \{0, \dots, m\}$. Define an *occurrence* of p in a word w as a factor $u_0 \cdots u_m$ of w, where for $i, j \in \{0, \dots, m\}$, if $\alpha_i = \alpha_j$, then $u_i = u_j$. Stated differently, for all $i \in \{0, \dots, m\}$, $u_i \in \varphi(\alpha_i)$, where φ is any non-erasing morphism from E^* to Σ^* as described earlier. These definitions extend to infinite words w over Σ which are functions from $\mathbb{N}$ to Σ.

Considering the pattern $p = \alpha\beta\beta\alpha$, the language associated with p over the alphabet $\{0, 1\}$ is $p(\{0, 1\}^+) = \{uvvu \mid u, v \in \{0, 1\}^+\}$. The word 001100 meets p (take $\varphi(\alpha) \in \{0, 00\}$ and $\varphi(\beta) = 1$), while the word 01011 avoids p.

Let p and p' be two patterns. If p' meets p, then p *divides* p', which we denote by $p \mid p'$. For example, $\alpha\alpha \nmid \alpha\beta\alpha$, but $\alpha\alpha \mid \alpha\beta\alpha\beta$. When both $p \mid p'$ and $p' \mid p$ hold, the patterns p and p' are *equivalent*, and this happens if and only if they differ by a permutation of E. For instance, $\alpha\alpha$ and $\beta\beta$ are equivalent.

A pattern $p \in E^*$ is *k-avoidable* if in Σ^* there are infinitely many words that avoid p, where Σ is a size k alphabet. On the other hand, if every long enough word in Σ^* meets p, then p is *k-unavoidable* (unavoidable over Σ). Finally, a pattern $p \in E^*$ which is k-avoidable for some k is simply called *avoidable*, and one which is k-unavoidable for every k is called *unavoidable*. The *avoidability index* of p is the smallest k such that p is k-avoidable, or it is ∞ if p is unavoidable.

In the rest of this work, we only consider binary patterns, hence we fix $E = \{\alpha, \beta\}$. Moreover, we define $\overline{\alpha} = \beta$ and $\overline{\beta} = \alpha$, and, similarly, $\overline{0} = 1$ and $\overline{1} = 0$ if Σ is binary, as complementing variables and, respectively, letters. Furthermore, when we talk about reversals we will refer to images of variables, while the term of mirror image will be used to refer to patterns, which might contain variables that are reversed or not, and factors of a word.

Preliminaries. In this paper we are interested in the avoidability of binary patterns in a more general setting. That is, we look at patterns formed not only from variables, but also from their reversals. As it can be seen, the word 0011001 has three occurrences of the pattern $\alpha\alpha$, but also has no fewer than six occurrences of the pattern $\alpha\alpha^R$, when $\alpha \in \{0, 01, 001, 1, 10\}$. Furthermore, it has no occurrence of $\alpha\alpha\alpha$, but has one occurrence of $\alpha\alpha^R\alpha$ for $\alpha = 01$.

Remark 1. Every even length palindrome meets the pattern $\alpha\alpha^R$ and its complement.

Please cite this article in press as: R. Mercaş, On the aperiodic avoidability of binary patterns with variables and reversals, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.12.022