Journal of Computer and System Sciences 81 (2015) 1298–1310

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Cubic patterns with permutations ☆

Florin Manea *, Mike Müller *, Dirk Nowotka *

Christian-Albrechts-Universität zu Kiel, Institut für Informatik, D-24098 Kiel, Germany

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 11 April 2014
Accepted 7 November 2014
Available online 21 April 2015

Keywords:
Combinatorics on words
Avoidable patterns
Pseudorepetitions
Permutations

ABSTRACT

We consider a generalisation of the classical problem of pattern avoidance in infinite words with functional dependencies between pattern variables. More precisely, we consider patterns involving permutations. The foremost remarkable fact regarding this new setting is that the notion of avoidability index (the smallest alphabet size for which a pattern is avoidable) is meaningless, since a pattern with permutations that is avoidable in one alphabet can be unavoidable in a larger alphabet. We characterise the (un-)avoidability of all patterns of the form $\pi^i(x)\pi^j(x)\pi^k(x)$, called cubic patterns with permutations here, for all alphabet sizes in both the morphic and antimorphic case.

1. Introduction

The avoidability of patterns in infinite words is an old area of interest with a first systematic study going back to Thue [2,3]. Basically, this domain is concerned with the existence of infinite words whose factors do not have a given particular form, or, in other words, are not instances of some particular patterns. This field includes studies by many authors over the last one hundred years; some of these, as well as some of their applications, are surveyed in [4, Ch. 3] and [5], for examples.

The very first results, obtained by Thue at the beginning of the last century, concerned the existence of infinite words avoiding very specific patterns, namely repetitions. A repetition is a word of the form x^k for some non-empty word x, and their study lies at the very centre of combinatorics on words. This notion has been generalised recently to the so called pseudorepetitions [6]; these are words from sets of the form $x\{x, f(x)\}^+$, where f is some function. The idea behind this new notion is that of working with strings that are not repetitions but, however, have an intrinsic repetitive structure. For example, the word *acgttgca* is not a repetition, but it is a pseudorepetition for the non-empty word $x = acgt$ and the morphic involution f with $f(a) = t$, $f(t) = a$, $f(c) = g$, and $f(g) = c$ or the antimorphic identity function (also known as reversal or mirror image); more precisely, it is a pseudosquare $xf(x)$. Various aspects of the combinatorial properties of these pseudorepetitions have been investigated [6–9]; papers like [6,7] also discuss the biological motivation of introducing pseudorepetitions, as well as their possible applications in bio-inspired computer science or bioinformatics.

In this article, we are concerned with studying avoidability questions considering patterns with functional dependencies between variables. In particular, we define and thoroughly investigate the case when these functions are permutations. More precisely, we do allow function variables in the pattern that are substituted by either morphic or antimorphic extensions of permutations on the alphabet. Consider the following pattern, for example:

☆ An extended abstract of this paper was presented at the conference Developments in Language Theory 2012 [1].
* Corresponding authors. Fax: +49 4318 805745.
E-mail addresses: flm@informatik.uni-kiel.de (F. Manea), mimu@informatik.uni-kiel.de (M. Müller), dn@informatik.uni-kiel.de (D. Nowotka).

http://dx.doi.org/10.1016/j.jcss.2015.04.001
0022-0000/

$$x\pi(x)x$$

where an instance of the pattern is a word uvu that consists of three parts of equal length, i.e., $|u| = |v|$, and v is the image of (the reversal of) u under any permutation on the alphabet. For example, $aab|bba|aab$ ($aab|abb|aab$) is an instance of $x\pi(x)x$ for the morphic (respectively, antimorphic) extension of the permutation $a \mapsto b$ and $b \mapsto a$.

Recently, there has been some initial work on avoidance of patterns with involutions which is a special case of the permutation setting considered in this paper (as involutions are permutations of order at most two); see [10,11]. The original interest of investigating patterns with involutions was motivated by possible applications in biology where the Watson–Crick complement corresponds to an antimorphic involution over four letters. A very restricted class of patterns involving only a specific permutation of the alphabet has also been studied previously [12]. However, our considerations here are much more general.

Since these are the very first considerations on this kind of pattern avoidance at all, we restrict ourselves to cubic patterns, following somehow the initial approach of Thue. The cube xxx is the most basic and well-investigated pattern that lends itself to nontrivial considerations on patterns with functional dependencies (a square is clearly not interesting in that context). So, we have one variable, occurring three times, and only one function variable, as the number of function symbols has to be strictly less than the length of the pattern minus one, otherwise the pattern is trivially unavoidable. Therefore, we investigate patterns of the form:

$$\pi^i(x)\pi^j(x)\pi^k(x)$$

where $i, j, k \geq 0$.

It is worth noting that the notion of avoidability index plays no role in the setting of patterns involving permutations. Contrary to the classical setting, where once a pattern is avoidable for some alphabet size it remains avoidable in larger alphabets, a pattern with permutations may be avoidable in some alphabet and become unavoidable in a larger alphabet. Moreover, the set of numbers defining the sizes of the alphabets over which a pattern with permutations is avoidable is a contiguous interval of natural numbers. This is a new and somewhat unexpected phenomenon in the field of pattern avoidance. It does not occur, for example, in the involution setting but requires permutations of higher order.

The paper is organised as follows. In Section 2 we fix our terminology, recall some basic results on avoidable (classical) patterns that we use, and give an algebraic definition of the avoidability setting we are concerned with; while this definition is not necessary to understand and follow our paper, this is aimed to establish a precise and sound formal framework for our investigation. In Section 3 we study avoidability questions when the function variables in the patterns are replaced by morphic extensions of a permutation. The corresponding questions for the case of antimorphic extensions are investigated in Section 4.

In some of our proofs we refer to computer programs which were used to search for occurrences of a pattern in a finite set of words. These programs are available at http://zs.uni-kiel.de/en/team/mueller/cpwp.zip.

2. Preliminaries

2.1. Definitions and notation

We define $\Sigma_m = \{0, \ldots, m-1\}$ to be an alphabet with m letters. For words u and w, we say that u is a *prefix* (resp. *suffix*) of w, if there exists a word v such that $w = uv$ (resp. $w = vu$). We denote that by $u \leq_p w$ (resp. $u \leq_s w$). Furthermore, u is a *factor* of w, if $w = xuv$ for some words x and v. As usual, the length of a word w is denoted by $|w|$.

For a word w and an integer i with $1 \leq i \leq |w|$ we denote the i-th letter of w by $w[i]$. We also denote the factor that starts with the i-th letter and ends with the j-th letter in w by $w[i..j]$. For any word w, we define w^R, the reversal of w, to be the word $w[|w|]w[|w|-1]\cdots w[1]$.

We say that a function $f : \Sigma_m^* \to \Sigma_\ell^*$ is a *morphism* if $f(uv) = f(u)f(v)$, for all words u and v over Σ_m. Further, f is an *antimorphism* if $f(uv) = f(v)f(u)$, for all words u and v over Σ_m. Clearly, when we want to define a morphism or an antimorphism f it is enough to give the definitions of $f(a)$, for all $a \in \Sigma_m$. A morphism or antimorphism $f : \Sigma_m^* \to \Sigma_m^*$ is an *involution* when $f^2(a) = a$ for all $a \in \Sigma_m$. A morphism (resp. antimorphism) $f : \Sigma_m^* \to \Sigma_m^*$ is called a morphic (resp. antimorphic) permutation if the restriction of f to Σ_m, denoted $f|_{\Sigma_m}$, is a permutation on the alphabet Σ_m.

If $f : \Sigma_m \to \Sigma_m$ is a permutation, we say that the order of f, denoted $\mathbf{ord}(f)$, is the minimal positive integer such that $f^{\mathbf{ord}(f)}$ is the identity. If $a \in \Sigma_m$ is a letter, the order of a with respect to f, denoted $\mathbf{ord}_f(a)$, is the minimal positive integer such that $f^{\mathbf{ord}_f(a)}(a) = a$.

A word w is called *k-power-free* for some integer k if it has no factor of the form u^k for some word u. For the cases $k = 2$ and $k = 3$, we use the terms *square-free* and *cube-free*.

A morphism f is k-power-free, if it preserves k-power-freeness, that is, $f(w)$ is k-power-free if w is k-power-free. We call a morphism *uniform*, if the images of all letters have the same length.

2.2. Some known results on k-power-free words and morphisms

We now recall a series of results on k-power-free words and morphisms from combinatorics on words: The infinite Thue–Morse word $\boldsymbol{t}$ is defined as