Doctopic: Algorithms, automata, complexity and games ARTICLE IN PRESS TCS:11386

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

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

Avoidability of circular formulas

Guilhem Gamard [a], Pascal Ochem [a,b,*], Gwenaël Richomme [a,c], Patrice Séébold [a,c]

[a] LIRMM, Université de Montpellier and CNRS, France
[b] CNRS, France
[c] Université Paul-Valéry Montpellier 3, France

ARTICLE INFO

Article history:
Received 14 October 2016
Accepted 19 November 2017
Available online xxxx
Communicated by D. Perrin

Keywords:
Words
Pattern avoidance

ABSTRACT

Clark has defined the notion of n-avoidance basis which contains the avoidable formulas with at most n variables that are closest to be unavoidable in some sense. The family C_i of circular formulas is such that $C_1 = AA$, $C_2 = ABA.BAB$, $C_3 = ABCA.BCAB.CABC$ and so on. For every $i \leqslant n$, the n-avoidance basis contains C_i. Clark showed that the avoidability index of every circular formula and of every formula in the 3-avoidance basis (and thus of every avoidable formula containing at most 3 variables) is at most 4. We determine exactly the avoidability index of these formulas.

1. Introduction

A *pattern* p is a non-empty finite word over an alphabet $\Delta = \{A, B, C, \dots\}$ of capital letters called *variables*. An *occurrence* of p in a word w is a non-erasing morphism $h : \Delta^* \to \Sigma^*$ such that $h(p)$ is a factor of w. The *avoidability index* $\lambda(p)$ of a pattern p is the size of the smallest alphabet Σ such that there exists an infinite word over Σ containing no occurrence of p. Bean, Ehrenfeucht, and McNulty [2] and Zimin [13] characterized unavoidable patterns, i.e., such that $\lambda(p) = \infty$. We say that a pattern p is *t-avoidable* if $\lambda(p) \leqslant t$. For more information on pattern avoidability, we refer to Chapter 3 of Lothaire's book [8]. See also this book for basic notions in Combinatorics on Words.

A variable that appears only once in a pattern is said to be *isolated*. Following Cassaigne [3], we associate to a pattern p the *formula* f obtained by replacing every isolated variable in p by a dot. The factors between the dots are called *fragments*.

An *occurrence* of a formula f in a word w is a non-erasing morphism $h : \Delta^* \to \Sigma^*$ such that the h-image of every fragment of f is a factor of w. As for patterns, the avoidability index $\lambda(f)$ of a formula f is the size of the smallest alphabet allowing the existence of an infinite word containing no occurrence of f. Clearly, if a formula f is associated to a pattern p, every word avoiding f also avoids p, so $\lambda(p) \leqslant \lambda(f)$. Recall that an infinite word is *recurrent* if every finite factor appears infinitely many times. If there exists an infinite word over Σ avoiding p, then there exists an infinite recurrent word over Σ avoiding p. This recurrent word also avoids f, so that $\lambda(p) = \lambda(f)$. Without loss of generality, a formula is such that no variable is isolated and no fragment is a factor of another fragment.

Cassaigne [3] began and Ochem [9] finished the determination of the avoidability index of every pattern with at most 3 variables. A *doubled* pattern contains every variable at least twice. Thus, a doubled pattern is a formula with exactly one fragment. Every doubled pattern is 3-avoidable [10]. A formula is said to be *binary* if it has at most 2 variables. The avoidability index of every binary formula has been recently determined [11]. We say that a formula f is *divisible* by a formula

* Corresponding author.
E-mail address: ochem@lirmm.fr (P. Ochem).

https://doi.org/10.1016/j.tcs.2017.11.014
0304-3975/

Please cite this article in press as: G. Gamard et al., Avoidability of circular formulas, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2017.11.014

f' if f does not avoid f', that is, there is a non-erasing morphism h such that the image of every fragment of f' by h is a factor of a fragment of f. If f is divisible by f', then every word avoiding f' also avoids f and thus $\lambda(f) \leqslant \lambda(f')$. Moreover, the reverse f^R of a formula f satisfies $\lambda(f^R) = \lambda(f)$. For example, the fact that $ABA.AABB$ is 2-avoidable implies that $ABAABB$ and $BAB.AABB$ are 2-avoidable. See Cassaigne [3] and Clark [4] for more information on formulas and divisibility.

Clark [4] has introduced the notion of *n-avoidance basis* for formulas, which is the smallest set of formulas with the following property: for every $i \leqslant n$, every avoidable formula with i variables is divisible by at least one formula with at most i variables in the n-avoidance basis.

From the definition, it is not hard to obtain that the 1-avoidance basis is $\{AA\}$ and the 2-avoidance basis is $\{AA, ABA.BAB\}$. Clark obtained that the 3-avoidance basis is composed of the following formulas:

- AA
- $ABA.BAB$
- $ABCA.BCAB.CABC$
- $ABCBA.CBABC$
- $ABCA.CABC.BCB$
- $ABCA.BCAB.CBC$
- $AB.AC.BA.CA.CB$

The following properties of the avoidance basis are derived.

- The n-avoidance basis is a subset of the $(n+1)$-avoidance basis.
- The n-avoidance basis is closed under reverse. (In particular, $ABCA.BCAB.CBC$ is the reverse of $ABCA.CABC.BCB$.)
- Two formulas in the n-avoidance basis with the same number of variables are incomparable by divisibility. (However, AA is divisible $AB.AC.BA.CA.CB$.)
- The n-avoidance basis is computable.

The *circular formula* C_t is the formula over $t \geqslant 1$ variables $A_0, \dots, A_{t-1}$ containing the t fragments of the form $A_i A_{i+1} \dots A_{i+t}$ such that the indices are taken modulo t. Thus, the first three formulas in the 3-avoidance basis, namely $C_1 = AA$, $C_2 = ABA.BAB$, and $C_3 = ABCA.BCAB.CABC$, are also the first three circular formulas. More generally, for every $t \leqslant n$, the n-avoidance basis contains C_t.

It is known that $\lambda(AA) = 3$ [12], $\lambda(ABA.BAB) = 3$ [3], and $\lambda(AB.AC.BA.CA.CB) = 4$ [1]. Actually, $AB.AC.BA.CA.CB$ is avoided by the fixed point $b_4 = 0121032101230321\dots$ of the morphism given below.

$$\begin{aligned} 0 &\mapsto 01 \\ 1 &\mapsto 21 \\ 2 &\mapsto 03 \\ 3 &\mapsto 23 \end{aligned}$$

Clark [4] obtained that b_4 also avoids C_i for every $i \geqslant 1$, so that $\lambda(C_i) \leqslant 4$ for every $i \geqslant 1$. He also showed that the avoidability index of the other formulas in the 3-avoidance basis is at most 4. Our main results finish the determination of the avoidability index of the circular formulas (Theorem 1) and the formulas in the 3-avoidance basis (Theorem 4).

2. Conjugacy classes and circular formulas

In this section, we determine the avoidability index of circular formulas.

Theorem 1. $\lambda(C_3) = 3$. $\forall i \geqslant 4$, $\lambda(C_i) = 2$.

We consider a notion that appears to be useful in the study of circular formulas. A *conjugacy class* is the set of all the conjugates of a given word, including the word itself. The length of a conjugacy class is the common length of the words in the conjugacy class. A word contains a conjugacy class if it contains every word in the conjugacy class as a factor. Consider the uniform morphisms given below.

$$\begin{aligned} g_2(0) &= 000010100111011010 0 & g_3(0) &= 0010 & g_6(0) &= 01230 \\ g_2(1) &= 0011100010100111101 & g_3(1) &= 1122 & g_6(1) &= 24134 \\ g_2(2) &= 0000111000101101000 & g_3(2) &= 0200 & g_6(2) &= 52340 \\ g_2(3) &= 0011110110100111101 & g_3(3) &= 1212 & g_6(3) &= 24513 \end{aligned}$$

Lemma 2.

- *The word $g_2(b_4)$ avoids every conjugacy class of length at least* 5.
- *The word $g_3(b_4)$ avoids every conjugacy class of length at least* 3.
- *The word $g_6(b_4)$ avoids every conjugacy class of length at least* 2.

Please cite this article in press as: G. Gamard et al., Avoidability of circular formulas, Theoret. Comput. Sci. (2018), https://doi.org/10.1016/j.tcs.2017.11.014