Journal of Discrete Algorithms 23 (2013) 119–142

Contents lists available at SciVerse ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

# Computing the partial word avoidability indices of ternary patterns ☆

F. Blanchet-Sadri [a,*], Andrew Lohr [b], Shane Scott [c]

[a] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402-6170, USA
[b] Department of Mathematics, Mathematics Building, University of Maryland, College Park, MD 20742, USA
[c] School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, USA

ARTICLE INFO

*Article history:*
Available online 10 July 2013

*Keywords:*
Combinatorics on words
Partial words
Pattern avoidance
Ternary pattern
Avoidability index

ABSTRACT

We study pattern avoidance in the context of partial words. The problem of classifying the avoidable binary patterns has been solved, so we move on to ternary and more general patterns. Our results, which are based on morphisms (iterated or not), determine all the ternary patterns' avoidability indices or at least give bounds for them.



## 1. Introduction

Pattern avoidance is a topic of interest in Combinatorics on Words. A *pattern* is a sequence over an alphabet of variables, which are denoted by $A$, $B$, $C$, etc. We obtain an occurrence of the pattern if we replace the variables with arbitrary non-empty words in such a way that we replace each occurrence of the same variable with the same word. A pattern $p$ is *avoidable* (resp., *k-avoidable*) if there exists an infinite word (resp., infinite word over a $k$-sized alphabet) that contains no occurrence of $p$; otherwise, $p$ is *unavoidable* (resp., *k-unavoidable*). The *avoidability index* of the pattern is the smallest integer $k$ for which it is $k$-avoidable; if no such $k$ exists, the index is $\infty$.

The problem of deciding whether a given pattern is avoidable has been solved [1,14], but the one of deciding whether it is $k$-avoidable has remained open. An alternative is the problem of classifying all the patterns over a fixed number of variables according to their avoidability indices. This classification has been completed for unary (those over one variable $A$), for binary (those over two variables $A$, $B$), as well as for ternary patterns (those over three variables $A$, $B$, $C$) [7,11,12].

For the lower bounds, we use the so-called backtracking algorithm from [8], while for the upper bounds, we provide *HD0L systems*. For a finite alphabet $\Sigma$, a morphism $f : \Sigma^* \to \Sigma^*$, and $a_0 \in \Sigma$, the tuple $(\Sigma, f, a_0)$ is called a *D0L system (Deterministic 0-sided Lindenmeyer system)* and the *D0L language* generated by the system is the set $\{f^n(a_0) \mid n \in \mathbb{N}\}$. For example, the Thue–Morse morphism $t(a) = ab$ and $t(b) = ba$ gives the D0L system $(\{a, b\}, t, a)$ generating the language

$$\{\varepsilon, a, ab, abba, abbabaab, abbabaabbaababba, \ldots\}.$$

☆ This material is based upon work supported by the National Science Foundation under Grant No. DMS-1060775. Part of this paper was presented at IWOCA'12 (F. Blanchet-Sadri et al., 2012) [4]. We thank the referee as well as Sean Simmons from the Massachusetts Institute of Technology for their very valuable comments and suggestions.

* Corresponding author.
*E-mail addresses:* blanchet@uncg.edu (F. Blanchet-Sadri), alohr1@umd.edu (A. Lohr), scottsha@ksu.edu (S. Scott).

1570-8667/$ – see front matter 
http://dx.doi.org/10.1016/j.jda.2013.06.009

For a D0L system $(\Sigma, f, a_0)$, the *fixed point* is $f^\omega(a_0) = \lim_{n\to\infty} f^n(a_0)$, provided the limit exists. The Thue–Morse word is $t^\omega(a)$. Now, for a morphism $g: \Sigma_1^* \to \Sigma_2^*$ with alphabets $\Sigma_1, \Sigma_2$ and a D0L system $(\Sigma_1, f, a_0)$, the tuple $(\Sigma_1, f, a_0, \Sigma_2, g)$ is called an *HD0L system* whose generated language is the set $\{g \circ f^n(a_0) \mid n \in \mathbb{N}\}$.

In [5], we have completed the classification of the avoidability indices of all the binary patterns in partial words (words with holes) that was started in [6]. The algorithms described in Section 5 of this paper have provided us with the morphisms necessary to complete this classification, which is recalled in the following theorem.

**Theorem 1.** *(See [5].) For partial words, binary patterns fall into three categories*:

1. *The binary patterns $\varepsilon$, $A$, $AA$, $AAB$, $AABA$, $AABAA$, $AB$, $ABA$, and their complements, are unavoidable* (*or have avoidability index* $\infty$).
2. *The binary patterns $AABAB$, $AABB$, $ABAB$, $ABBA$, their reverses, and complements, have avoidability index* 3.
3. *All other binary patterns, and in particular all binary patterns of length six or more, have avoidability index* 2.

In this paper, we investigate the problem of classifying all the avoidable ternary patterns with respect to partial word avoidability. We identify the avoidability indices of almost all of the ternary patterns and show that only four are left in order to complete the classification (for those four we give lower and upper bounds).

The contents of our paper are as follows: In Section 2, we give some background on partial words and patterns (for more information, see [2,11]). In Section 3, we discuss the classification of the ternary patterns. In Section 4, we make some observations for general pattern avoidance. In Section 5, we describe an algorithm to search for an HD0L system avoiding a given pattern. Finally in Section 6, we conclude with some remarks. Note that we have put in Appendix A a ternary lexicon which lists the partial word avoidability indices for the ternary patterns, or at least lists bounds for them.

## 2. Preliminaries

Let $\Sigma$ be a finite alphabet of letters. A *partial word* over $\Sigma$ is a sequence of symbols from $\Sigma_\diamond = \Sigma \cup \{\diamond\}$, where $\Sigma$ is augmented with the "hole" symbol $\diamond$. A *(full) word* is a partial word without holes. The symbol at position $i$ of a partial word $u$ is denoted by $u[i]$, while the *length* of $u$, i.e., the number of symbols in $u$, is denoted by $|u|$. The *empty word* $\varepsilon$ has length zero. The set of all full words (resp., non-empty full words) over $\Sigma$ is denoted by $\Sigma^*$ (resp., $\Sigma^+$), while the set of all partial words (resp., non-empty partial words) over $\Sigma$ is denoted by $\Sigma_\diamond^*$ (resp., $\Sigma_\diamond^+$). The set of all full (resp., partial) words over $\Sigma$ of length $n$ is denoted by $\Sigma^n$ (resp., $\Sigma_\diamond^n$).

A partial word $u$ is a *factor* (resp., *prefix*, *suffix*) of a partial word $v$ if there exist $x$, $y$ such that $v = xuy$ (resp., $v = uy$, $v = xu$). The factor, prefix, or suffix $u$ is *proper* if $u \neq \varepsilon$ and $u \neq v$. We denote by $\mathrm{Pref}(v)$ (resp., $\mathrm{Suf}(v)$) the set of all prefixes (resp., suffixes) of $v$. If $u$ and $v$ are two partial words of equal length, then $u$ is *compatible* with $v$, denoted by $u \uparrow v$, if $u[i] = v[i]$ whenever $u[i], v[i] \in \Sigma$. If $u, v$ are non-empty compatible partial words, then $uv$ is a *square*. A full word compatible with a factor of a partial word $v$ is a *subword* of $v$.

Let $\Delta$, $\Sigma \cap \Delta = \emptyset$, be an alphabet of pattern variables and denote them by $A$, $B$, $C$, etc. A *pattern* is a word over $\Delta$, e.g., $AABAACACCBAACA$ is a ternary pattern. We denote by $\mathrm{alph}(p)$ the set of distinct variables in pattern $p$. For a partial word $w \in \Sigma_\diamond^*$ and pattern $p \in \Delta^*$, we say that $w$ *meets* $p$ or $p$ *occurs in* $w$ if there exists some non-erasing morphism $\varphi: \Delta^* \to \Sigma^*$ such that $\varphi(p)$ is compatible with a factor of $w$; otherwise $w$ *avoids* $p$. These definitions also apply to infinite partial words over $\Sigma$ which are functions from $\mathbb{N}$ to $\Sigma_\diamond$.

A pattern $p$ is *$k$-avoidable* if there is a partial word over a $k$-sized alphabet with infinitely many holes that avoids $p$. We say that $p$ is *avoidable* if it is $k$-avoidable for some $k$. For a given pattern $p$, the *avoidability index* $\mu(p)$ is the minimal $k$ such that $p$ is $k$-avoidable. If $p$ is unavoidable, $\mu(p) = \infty$.

For a given pattern $p$, can we determine $\mu(p)$? A concept useful to answer this question is *division of patterns*. If $p$ occurs in a pattern $q$, then $p$ *divides* $q$. For instance, $ABA\underline{C}BAB\underline{C}$ divides $ABA\underline{BC}BAB\underline{BC}$ (replacing $C$ by $BC$ gives $q$ from $p$). If $p$ divides $q$ and an infinite partial word avoids $p$ then it also avoids $q$, and so $\mu(q) \leqslant \mu(p)$.

## 3. Classification of the ternary patterns

In classifying the avoidability indices of the ternary patterns, it is useful to consider the directed tree of patterns $T$, where the root of $T$ is labeled by $\varepsilon$ and each node has children labeled by every canonical pattern formed by appending $A$, $B$, $C$ to the parent node's pattern, with all edges directed from parent to child. We have a partial order relation defined on the set of canonical ternary patterns by $q > p$ if there is a path in $T$ from the node labeled by pattern $q$ to the node labeled by pattern $p$. Because $q > p$ implies $q \mid p$, we have that $\mu(q) \geqslant \mu(p)$. The classification is complete when every node of $T$ is appended with the avoidability index of the pattern labeling it.

First, we use unavoidability results to rule out known 2-unavoidable patterns, and proceed via a depth-first search to find 2-avoidable patterns which are identified as such using division arguments from the binary patterns and the HD0L finding algorithm described in Section 5. Once a pattern $p$ is known to have avoidability index two, we know its children, grandchildren, etc., also have avoidability index two. We find by exhaustion that every ternary pattern with length twelve