Discrete Mathematics 340 (2017) 443–451

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Coloring factors of substitutive infinite words

André Bernardino [a], Rui Pacheco [a,*], Manuel Silva [b]

[a] Departamento de Matemática, Universidade da Beira Interior, 6200-001 Covilhã, Portugal
[b] Departamento de Matemática, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form 3 September 2016
Accepted 7 September 2016
Available online 16 November 2016

Keywords:
Combinatorics on words
Substitutive words
Recognizability
Ramsey theory

ABSTRACT

In this paper, we consider infinite words that arise as fixed points of primitive substitutions on a finite alphabet and finite colorings of their factors. Any such infinite word exhibits a "hierarchal structure" that will allow us to define, under the additional condition of *strong recognizability*, certain remarkable finite colorings of its factors. In particular we generalize two combinatorial results by Justin and Pirillo concerning arbitrarily large monochromatic k-powers occurring in infinite words; in view of a recent paper by de Luca, Pribavkina and Zamboni, we will give new examples of classes of infinite words $\mathbf{u}$ and finite colorings that do not allow infinite monochromatic factorizations $\mathbf{u} = \mathbf{u}_1\mathbf{u}_2\mathbf{u}_3\ldots$.

1. Introduction

Let $\mathcal{A}$ be a finite alphabet and $\mathcal{A}^*$ be the set of all finite words over $\mathcal{A}$. In this paper we consider infinite words that arise as fixed points of primitive substitutions on $\mathcal{A}$ and finite colorings of $\mathcal{A}^*$. Any such infinite word exhibits a "hierarchal structure", induced by the underlying substitution, that will allow us to define, under the additional condition of *strong recognizability* (in the spirit of the recognizability conditions introduced by B. Host [5] and B. Mossé [9]), certain remarkable finite colorings of its factors.

For a two letter alphabet $\{0, 1\}$, J. Justin and G. Pirillo [7] constructed a finite coloring c of $\{0, 1\}^*$ with respect to which the Thue–Morse word $\mathbf{u}^T$ avoids *uniform monochromatic* 3*-powers*, that is any word of the form $\mathbf{w}_1\mathbf{w}_2\mathbf{w}_3$ occurring in $\mathbf{u}^T$, where the $\mathbf{w}_i$ are of equal length, satisfies $c(\mathbf{w}_i) \neq c(\mathbf{w}_j)$ for some i and j. The Thue–Morse word is a fixed point of the substitution ζ_T defined by $\zeta_T(0) = 01$ and $\zeta_T(1) = 10$. In order to define the coloring c, the authors made implicit use of the recognizability properties of this substitution. In Section 3 we will give a generalization of their result: for a large class of substitutions of constant length on a finite alphabet $\mathcal{A}$, there always exist finite colorings of $\mathcal{A}^*$ with respect to which any fixed word avoids arbitrarily large uniform monochromatic k-powers.

Once this established, it is natural to consider next arbitrarily large monochromatic k-powers $\mathbf{w}_1\mathbf{w}_2\ldots\mathbf{w}_k$ with *uniformly bounded gaps*, that is the lengths of the factors $\mathbf{w}_i$ are bounded by some $p > 0$ and p is independent of k. J. Justin and G. Pirillo [7] showed that there exists a 3-coloring of $\{0, 1\}^*$ with respect to which the fixed point $\mathbf{u}^J$ of the substitution ζ_J defined by $\zeta_J(0) = 00001$ and $\zeta_J(1) = 11110$ avoids arbitrarily large monochromatic k-powers with uniformly bounded gaps. The main ingredients in their argument are the following. Firstly, $\mathbf{u}^J$ avoids *abelian* 5*-powers* [6], that is, there does not exist any finite word of the form $\mathbf{w} = \mathbf{w}_1\ldots\mathbf{w}_5$ occurring in $\mathbf{u}^J$ where any two of the factors $\mathbf{w}_i$ are permutations of each other. Secondly, any very long factor of $\mathbf{u}^J$ contains "about equally many 0's and 1's". In Section 4 we will be able to show that, more generally, if an infinite word $\mathbf{u}$ over a finite alphabet satisfies

(1) $\mathbf{u}$ avoids arbitrarily large abelian k-powers;
(2) each factor of $\mathbf{u}$ occurs in $\mathbf{u}$ with a *uniform frequency*;

* Corresponding author.
E-mail addresses: and_bernardino@hotmail.com (A. Bernardino), rpacheco@ubi.pt (R. Pacheco), mnas@fct.unl.pt (M. Silva).

http://dx.doi.org/10.1016/j.disc.2016.09.013
0012-365X/

then **u** avoids arbitrarily large monochromatic k-powers with uniformly bounded gaps. Again, a large class of substitutions gives rise to infinite words satisfying the above properties. The property of uniform frequencies is closely related to the property of unique ergodicity of the dynamical system associated to the action of the shift map on **u**. It is well known that such dynamical system is uniquely ergodic if **u** is the fixed word of a primitive substitution [11].

In [3,4] the authors discussed the following question: *given an infinite word* **u** *over a finite alphabet, does there exist a finite coloring of its finite factors which avoids monochromatic factorizations of* **u**? They showed that this question, which is ultimately motivated by a result of Schutzenberger [13], has a positive answer for all non-*uniformly recurrent* words and for various classes of uniformly recurrent words. V. Salo and I. Törmä [12] showed that any aperiodic linearly recurrent word **u** admits a coloring of its factors that avoids monochromatic factorizations of **u** into factors of increasing lengths. We will prove, in Section 5, that the question has also a positive answer for a wide class of infinite words arising as fixed points of primitive substitutions and we shall be able to construct examples that do not fit in those classes of infinite words considered in [3,4]. Again, the "hierarchal structure" of these words and the strong recognizability condition will play a fundamental role in the construction of our colorings.

2. Preliminaries

2.1. Monochromatic factorizations

The binary operation obtained by concatenation of two finite words endows $\mathcal{A}^*$ with a monoid structure (the identity is the empty word $\emptyset$). For each $\alpha \in \mathcal{A}$, the α-*length* of a finite word $\mathbf{w} \in \mathcal{A}^*$, which we denote by $|\mathbf{w}|_\alpha$, is the number of occurrences of α in $\mathbf{w}$. The *length* of $\mathbf{w}$ is the sum $|\mathbf{w}| = \sum_{\alpha\in\mathcal{A}} |\mathbf{w}|_\alpha$ of all its α-lengths. If a copy of $\mathbf{w}$ occurs in a word $\mathbf{u}$, we say that $\mathbf{w}$ is a *factor* of $\mathbf{u}$.

Given an infinite word $\mathbf{u} = \alpha_1\alpha_2\alpha_3\ldots$ over $\mathcal{A}$, $\mathcal{L}(\mathbf{u})$ stands for the *language* of $\mathbf{u}$, that is the set of all nonempty finite factors of $\mathbf{u}$. For each i and j satisfying $1 \le i \le j$, we write

$$\mathbf{u}_{[i,j]} = \alpha_i\alpha_{i+1}\ldots\alpha_j \in \mathcal{L}(\mathbf{u}).$$

Let $c : \mathcal{L}(\mathbf{u}) \to \{1, \ldots, r\}$ be a finite coloring of $\mathcal{L}(\mathbf{u})$.

Definition 1 ([7]). A finite word $\mathbf{w} \in \mathcal{L}(\mathbf{u})$ is a *monochromatic* k*-power* if there exists a factorization $\mathbf{w} = \mathbf{w}_1\mathbf{w}_2\ldots\mathbf{w}_k$, with $\mathbf{w}_i \in \mathcal{L}(\mathbf{u})$, such that $c(\mathbf{w}_i) = c(\mathbf{w}_j)$, for all i and j. A monochromatic k-power is *uniform* if $|\mathbf{w}_i| = |\mathbf{w}_j|$ for all i and j. A monochromatic k-power has *gaps bounded* by p if $|\mathbf{w}_i| < p$ for all i.

2.2. Substitutive words

Next we recall some fundamental facts concerning substitutive words. For further details we refer the reader to [11]. A *substitution* ζ of $\mathcal{A}$ is a map from $\mathcal{A}$ to $\mathcal{A}^*$ which associates the letter α to some word $\zeta(\alpha) \in \mathcal{A}^*$. This induces a morphism of the monoid $\mathcal{A}^*$ by putting $\zeta(\emptyset) = \emptyset$ and

$$\zeta(\alpha_1\alpha_2\ldots\alpha_k) = \zeta(\alpha_1)\zeta(\alpha_2)\ldots\zeta(\alpha_k).$$

For each $\alpha \in \mathcal{A}$, let $l_\alpha = |\zeta(\alpha)|$ so that

$$\zeta(\alpha) = \zeta(\alpha)_1\zeta(\alpha)_2\ldots\zeta(\alpha)_{l_\alpha}.$$

The substitution has *constant length* l if $l = l_\alpha$ for all $\alpha \in \mathcal{A}$. The substitution ζ is said to be *primitive* if there exists $n > 0$ such that, for every $\alpha, \beta \in \mathcal{A}$, β occurs in $\zeta^n(\alpha)$ (that is, n can be chosen independent of α and β). Henceforth we will assume that $\alpha = \zeta(\alpha)_1$ for some $\alpha \in \mathcal{A}$.

Consider the space of ζ-*substitutive infinite words* $X_\zeta = \{\mathbf{u} \in \mathcal{A}^{\mathbb{N}} : \ \mathcal{L}(\mathbf{u}) \subseteq \mathcal{L}_\zeta\}$, where

$$\mathcal{L}_\zeta = \bigcup_{n\ge 0, \alpha\in\mathcal{A}} \{\text{factors of } \zeta^n(\alpha)\},$$

and endow $\mathcal{A}^{\mathbb{N}}$ with the metric d defined by: given $\mathbf{u} = \alpha_1\alpha_2\ldots$ and $\mathbf{v} = \beta_1\beta_2\ldots$ in $\mathcal{A}^{\mathbb{N}}$, $d(\mathbf{u}, \mathbf{v}) = 0$ if $\mathbf{u} = \mathbf{v}$ and $d(\mathbf{u}, \mathbf{v}) = 1/N$ if N is the smallest positive integer for which $\alpha_N \ne \beta_N$. Denote by σ the *shift map* in $\mathcal{A}^{\mathbb{N}}$,

$$\sigma(\alpha_1\alpha_2\alpha_3\ldots) = \alpha_2\alpha_3\ldots,$$

which preserves X_ζ. The substitution ζ also induces a map $\zeta : X_\zeta \to X_\zeta$.

Proposition 1 ([11]). *Suppose that* ζ *is a primitive substitution. We have:*

(1) ζ *admits a fixed point:* $\zeta(\mathbf{u}) = \mathbf{u}$, *for some* $\mathbf{u} \in X_\zeta$.
(2) *The fixed word* **u** *is* uniformly recurrent, *that is, for each factor* **w** *of* **u** *there exists* $R > 0$ *such that in any other factor of* **u** *of length* R *there is at least one occurrence of* **w**.