Theoretical Computer Science 550 (2014) 107–112

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Note

On the structure of Thue–Morse subwords, with an application to dynamical systems

Michel Dekking

DIAM, Delft University of Technology, Faculty EEMCS, P.O. Box 5031, 2600 GA Delft, The Netherlands

ARTICLE INFO

Article history:
Received 11 May 2014
Accepted 20 July 2014
Available online 30 July 2014
Communicated by N. Ollinger

Keywords:
Thue–Morse sequence
Thue–Morse factors
Substitution dynamical system
Conjugacy

ABSTRACT

We give an in-depth analysis of the subwords of the Thue–Morse sequence. This allows us to prove that there are infinitely many injective primitive substitutions with Perron–Frobenius eigenvalue 2 that generate a symbolic dynamical system topologically conjugate to the Thue–Morse dynamical system.

1. Introduction

We consider the bi-infinite Thue–Morse sequence $x = \dots 10010110 \cdot 01101001 \dots$ fixed point of the substitution θ given by

$$\theta(0) = 01, \qquad \theta(1) = 10.$$

By taking its orbit closure under the shift map, the sequence x generates a dynamical system called the Thue–Morse dynamical system. In the recent paper [3] it is proved that there are 12 primitive injective substitutions of length 2 that generate a system topologically conjugate to the Thue–Morse system. A natural question is: what is the list of all primitive injective substitutions whose incidence matrix has maximal eigenvalue 2 that generate a system topologically conjugate to the Thue–Morse system?

In general, when α is a substitution on an alphabet A, let $\mathcal{L}_\alpha$ be the language of α, i.e., the collection of all words occurring in some power $\alpha^n(a)$, for some $a \in A$. We write X_α for the set of bi-infinite sequences each of whose finite factors belongs to $\mathcal{L}_\alpha$. Under the left shift it is a minimal symbolic system whenever α is primitive. Two symbolic systems X_α and X_β are called topologically conjugate or simply conjugate if there is a bi-continuous bijective map from one to the other that preserves the shift.

The usual way to generate systems topologically conjugate to a given substitution dynamical system is to consider the N-block substitution associated with the substitution [5,3]. See Section 3 for more details, here we give an example: the 5-block substitution θ_5 associated with the Thue–Morse substitution θ.

There are twelve Thue–Morse subwords of length $N = 5$ (see Example 2.1 in Section 2 for the complete list): $w_1 = 00101, \dots, w_4 = 01011, \dots, w_{12} = 11010$.

E-mail address: F.M.Dekking@TUDelft.nl.

http://dx.doi.org/10.1016/j.tcs.2014.07.024
0304-3975/

The θ_5-image of a w_i is obtained as the prefix of length 5 of $\theta(w_i)$ followed by the prefix of length 5 of $\theta(w_i)$ with the first letter discarded. For example, since $\theta(00101) = 0101100110$, we have $\theta_5(w_1) = w_4 w_{10}$, since $w_{10} = 10110$. In this way one obtains

$$\begin{array}{llll} \theta_5(w_1) = w_4 w_{10}, & \theta_5(w_4) = w_5 w_{11}, & \theta_5(w_7) = w_7 w_1, & \theta_5(w_{10}) = w_8 w_2, \\ \theta_5(w_2) = w_4 w_{10}, & \theta_5(w_5) = w_6 w_{12}, & \theta_5(w_8) = w_7 w_1, & \theta_5(w_{11}) = w_9 w_3, \\ \theta_5(w_3) = w_5 w_{11}, & \theta_5(w_6) = w_6 w_{12}, & \theta_5(w_9) = w_8 w_2, & \theta_5(w_{12}) = w_9 w_3 \end{array}$$

We go from this substitution, which is not injective,[1] to an injective one by redistributing the four letters in the θ_5-images of words of length 2 with odd indices—which always occur in pairs, i.e., the couples $w_5 w_{11}$, $w_7 w_1$, and $w_9 w_3$. Concretely, we define a new substitution ζ_5 by keeping $\zeta_5(w_i) = \theta_5(w_i)$ for all words with an even index, and changing the six others in pairs as, e.g.,

$$\theta_5(w_7)\theta_5(w_1) = w_7 w_1 \quad w_4 w_{10} = w_7 w_1 w_4 \quad w_{10} = \zeta_5(w_7)\zeta_5(w_1).$$

This leads to the substitution given by

$$\begin{array}{llll} \zeta_5(w_1) = w_{10}, & \zeta_5(w_4) = w_5 w_{11}, & \zeta_5(w_7) = w_7 w_1 w_4, & \zeta_5(w_{10}) = w_8 w_2, \\ \zeta_5(w_2) = w_4 w_{10}, & \zeta_5(w_5) = w_6 w_{12} w_9, & \zeta_5(w_8) = w_7 w_1, & \zeta_5(w_{11}) = w_3, \\ \zeta_5(w_3) = w_{11}, & \zeta_5(w_6) = w_6 w_{12}, & \zeta_5(w_9) = w_8 w_2 w_5, & \zeta_5(w_{12}) = w_9 w_3 \end{array}$$

Obviously the substitution ζ_5 is injective, and it is not hard to see that $\zeta_5^n(w_6) = \theta_5^n(w_6)$ for all $n \geq 1$. Thus, *if* ζ_5 would be a primitive substitution, then ζ_5 would generate the same dynamical system as θ_5. However, ζ_5 is *not* primitive, since $\zeta_5^2(w_3) = \zeta_5(w_{11}) = w_3$.

In Section 4 we will repair this defect by defining a substitution η_5 which generates the same dynamical system as θ_5, but *is* primitive. Actually, we give this construction for all η_N, where N is a power of two plus one. For this we need an explicit expression for θ_N, which is given in Section 3, based on the combinatorial analysis in Section 2. Our main result is in Section 6: there exist infinitely many substitutions in the Thue–Morse conjugacy class if we allow also non-constant length substitutions with Perron–Frobenius eigenvalue 2.

2. Combinatorics of Thue–Morse subwords

The subwords of the Thue–Morse sequence have been well studied (see, e.g., [6,1]). We show here that the subwords of length $N = 2^m + 1$ have a particularly elegant structure for $m = 2, 3, \ldots$. Let $\mathcal{A}_m$ be the set of these words. It is well known (and will be reproved here) that the cardinality of $\mathcal{A}_m$ equals $|\mathcal{A}_m| = 3 \cdot 2^m$ (see [6]). We lexicographically order the words in $\mathcal{A}_m$, representing them as

$$w_1^m < w_2^m < \cdots < w_{|\mathcal{A}_m|}^m.$$

Crucial to the following analysis is the partition of $\mathcal{A}_m$ into 4 sets

$$\mathcal{A}_m = \mathcal{Q}_1 \cup \mathcal{Q}_2 \cup \mathcal{Q}_3 \cup \mathcal{Q}_4,$$

where each $\mathcal{Q}_k$ consists of one quarter of consecutive words from $\mathcal{A}_m$. If we want to emphasize the dependence on m we write $\mathcal{Q}_k^m$. Let

$$q_k = \min \mathcal{Q}_k, \quad \text{for } k = 1, 2, 3, 4.$$

Thus

$$q_1^m = w_1^m, \qquad q_2^m = w_{\frac{1}{4}|\mathcal{A}_m|+1}^m, \qquad q_3^m = w_{\frac{1}{2}|\mathcal{A}_m|+1}^m, \qquad q_4^m = w_{\frac{3}{4}|\mathcal{A}_m|+1}^m.$$

Let $f_0^\omega = 0110\ldots$ and $f_1^\omega = 1001\ldots$ be the two infinite fixed points of θ, and let $f_0 = f_0^m$ and $f_1 = f_1^m$ be the length $2^m + 1$ prefixes of f_0^ω and f_1^ω.

Example 2.1. The case $m = 2$. The set $\mathcal{A}_2$ is given by $\{00101, 00110, 01001, 01011, 01100, 01101, 10010, 10011, 10100, 10110, 11001, 11010\}$. Here $q_1 = 00101$, $q_2 = 01011$, $q_3 = 10010$, $q_4 = 10110$, and $f_0 = 01101$, $f_1 = 10010$.

We use frequently mirror invariance of the Thue–Morse words, i.e., if the mirroring operation is defined as the length 1 substitution given by $\widetilde{0} = 1$, $\widetilde{1} = 0$, then u is a Thue–Morse subword if and only if $\widetilde{u}$ is a Thue–Morse subword. This follows directly from $\widetilde{\theta(0)} = \theta(1)$.

The Thue–Morse substitution θ has the following trivial, but important property.

[1] A substitution α is called injective if $a \neq b$ implies $\alpha(a) \neq \alpha(b)$.