Topological conjugacy of constant length substitution dynamical systems

Coven, Ethan M.; Dekking, Michel; Keane, Mike

DOI
10.1016/j.indag.2016.11.006

Publication date
2017

Document Version
Final published version

Published in
Indagationes Mathematicae

Available online at www.sciencedirect.com

ScienceDirect

Indagationes Mathematicae 28 (2017) 91–107

indagationes mathematicae

www.elsevier.com/locate/indag

Topological conjugacy of constant length substitution dynamical systems

Ethan M. Coven[b], F. Michel Dekking[a,*], Michael S. Keane[a,c]

[a] DIAM, Delft University of Technology, Faculty EEMCS, P.O. Box 5031, 2600 GA Delft, The Netherlands
[b] Department of Mathematics and Computer Science, Wesleyan University, 265 Church Street, Middletown, CT 06459-0128, USA
[c] New York University Shanghai, China

Abstract

Primitive constant length substitutions generate minimal symbolic dynamical systems. In this article we present an algorithm which can produce the list of injective substitutions of the same length that generate topologically conjugate systems. We show that each conjugacy class contains infinitely many substitutions which are not injective. As examples, the Toeplitz conjugacy class contains three injective substitutions (two on two symbols and one on three symbols), and the length two Thue–Morse conjugacy class contains twelve substitutions, among which are two on six symbols. Together, they constitute a list of all primitive substitutions of length two with infinite minimal systems which are factors of the Thue–Morse system.

Keywords: Substitution dynamical system; Conjugacy; Sliding block code; Thue–Morse substitution; Toeplitz substitution

1. Prologue

In the article [6] published in 1971, the minimal dynamical systems arising from primitive substitutions on a binary alphabet having the same constant length were classified, yielding for

* Corresponding author.
E-mail address: F.M.Dekking@TUDelft.nl (F. Michel Dekking).

http://dx.doi.org/10.1016/j.indag.2016.11.006

a given such substitution a list of all substitutions of the same length generating topologically conjugate systems.

Topological conjugacy is the natural isomorphism notion for topological dynamical systems. Two systems (X, S) and (Y, T) are called *topologically conjugate* iff there exists homeomorphism $\psi : X \to Y$ such that $\psi \circ S = T \circ \psi$. The map ψ is called a *semi-conjugacy* if it is equivariant and merely continuous and onto. In this case (Y, T) is called a *factor* of (X, S).

In this paper we extend the classification of substitution dynamical systems to arbitrary finite alphabets. More recently, the articles [7,4] exhibit characterizations of such systems; these only implicitly yield corresponding topological conjugacies, and do not result in lists of conjugate systems. Also, in [17,19] a related goal has been partially accomplished – a classification of measure-theoretic conjugacy – for a restricted class of constant length substitutions.

If two constant length substitution systems are conjugate, then by Cobham's theorem, the lengths of the substitutions have a common non-zero integer power (see [5] for a short proof of the more simple version that we need here). Therefore, by taking suitable powers we can, and do, restrict our attention to substitutions of the same length L.

In this contribution we address the following two problems, in which L denotes a fixed integer larger than one, and σ is the left shift transformation.

Problem 1.1. Let α and β be two substitutions of the same length L, both primitive. Decide whether the dynamical systems (X_α, σ) and (X_β, σ) are topologically conjugate.

Problem 1.2. Let α be a primitive substitution of length L. Give a list of all the injective substitutions β of length L such that the dynamical systems (X_α, σ) and (X_β, σ) are topologically conjugate.

Finite systems are elementary, and we restrict attention everywhere to the non-periodic case of primitive substitutions with corresponding infinite minimal sets.

We show that to any primitive substitution of constant length whose minimal set is infinite, there are always infinitely many primitive substitutions of the same constant length having topologically conjugate minimal systems, but only finitely many of these are injective. Thus, the list produced by our algorithm for attacking Problem 1.2 will, starting from any given primitive substitution of constant length, consists of all injective substitutions of that length with dynamical systems topologically conjugate to the initial system. Clearly, since the list in Problem 1.2 is finite, Problem 1.1 has then also been solved, since there is a simple algorithm to associate to a substitution an injective substitution generating a conjugate system (cf. Section 6). This contrasts with the situation for the natural generalization of our problem to the collection of *all* substitutions. In [10] it is shown that there may be infinitely many primitive injective (non-constant length) substitutions that generate systems conjugate to a system generated by a substitution with the same Perron–Frobenius eigenvalue for its incidence matrix.

Recently a completely different solution has been obtained for Problem 1.1. in the paper [8]. Actually, because of Theorem 5.1, a solution of Problem 1.1 also yields a solution of Problem 1.2. However, it seems unfeasible – using the algorithm of [8] – to obtain the Thue–Morse list by hand, as we do in Section 11.

2. Substitutions and standard forms

We begin by recalling the basic definitions and known results without proof for primitive substitutions and their corresponding minimal systems, referring the reader to the standard Ref. [22].

Let A be a finite set (an *alphabet*) with $c \geq 2$ elements which are *symbols*, or *letters*. Elements of $A^* = \cup_{n=0}^{\infty} A^n$ are called *words*. A *substitution* is a mapping

$$\alpha : A \to A^*.$$

The substitution α is of constant length L if $\alpha(a) \in A^L$ for each $a \in A$. It is natural to view A^* as a monoid under juxtaposition, thus extending α to mappings from A^* to A^*, $A^{\mathbf{N}}$ to $A^{\mathbf{N}}$, and $A^{\mathbb{Z}}$ to $A^{\mathbb{Z}}$—no confusion results if we also denote them by α, and they can be iterated, defining α^n for each $n \in \mathbf{N}$.

Definition 2.1. The substitution α is *primitive* if for some $n > 0$ and for every $a \in A$ the word $\alpha^n(a)$ contains each of the letters of A. The *language* of α is the subset $\mathcal{L}_\alpha$ of A^* consisting of those words appearing as consecutive letters, *subwords*, or *factors*, of images under powers of α. We denote by $\mathcal{L}_\alpha^N$ the set of words of length N in $\mathcal{L}_\alpha$.

We write X_α for the compact subset of $A^{\mathbb{Z}}$ of bilaterally infinite sequences each of whose finite factors belongs to the language of α. Under the left shift σ on $A^{\mathbb{Z}}$, it is a minimal symbolic system whenever α is primitive. If in addition, X_α is infinite, then α is *recognizable* [20]. This means that there exists an integer K such that any word w of length larger than K from the language of α can be written in a unique way as $u\,\alpha(a_1)\ldots\alpha(a_\ell)\,v$, where the lengths of u and v are less than L, and $a_1 \ldots a_\ell \in \mathcal{L}_\alpha^\ell$.

For substitutions, it is clear that the names we give to the individual symbols of their alphabets are not essential—different namings will produce conjugate systems. This leads us to restricting an alphabet of c symbols to the alphabet $A = \{1, \ldots, c\}$. Even then, there is a permutational ambiguity, since permuting A will yield up to $c!$ different substitutions, which we view as essentially the same. We find it useful in the following to single out one of these permutations as the one yielding the *standard form* of a substitution, as follows. If α is a constant length L substitution on the alphabet of size c, then we define its *characteristic word* to be the word $\alpha(1)\cdots\alpha(c)$ of length Lc. For constant length[1] substitutions, permutations yielding different substitutions then possess different characteristic words, and we call the substitution with the lexicographically smallest characteristic word the *standard form* of the substitution α. For well-known substitutions like the Thue Morse substitution and the Toeplitz substitution we will stick to the usual alphabet $\{0, 1\}$.

3. Letter-to-letter maps

Let A and B be finite alphabets. A map

$$\pi : A \to B$$

is called a *letter-to-letter map*; by juxtaposition it clearly extends to maps from (finite or infinite) sequences on A to sequences of the same lengths on B. We also denote this extension by the same symbol π. It will appear that the following easily proved lemma is the key to understanding the properties of conjugacies.

Lemma 3.1. *If $\alpha : A \to A^*$ and $\beta : B \to B^*$ are substitutions, and if π satisfies the intertwining equation $\pi\,\alpha = \beta\,\pi$, then for each positive integer n*

$$\pi\,\alpha^n = \beta^n\,\pi.$$

[1] See [11] for a standard form for arbitrary substitutions.

Under the hypotheses of the lemma, the word $\alpha^n(a)$ is mapped by π to the word $\beta^n(b)$, with $b = \pi(a)$, for any positive n. In particular, the language of α is mapped to the language of β, and we have:

Corollary 3.1. *$\pi(X_\alpha) \subseteq X_\beta$, with equality whenever π is surjective. In particular, if the map $\pi : A \to B$ is surjective, then primitivity of α implies primitivity of β and minimality of (X_α, σ) implies minimality of (X_β, σ).*

When $\pi\,\alpha = \beta\,\pi$ and $\pi : A \to B$ is surjective, we call β an *amalgamation* of α.

4. N-Block presentations and N-Block substitutions

Let A be a finite alphabet, and let $N \geq 2$ denote a positive integer. We consider the elements $a_0 a_1 \dots a_{N-1}$ of A^N as symbols $[a_0 a_1 \dots a_{N-1}]$ in an alphabet denoted $A^{[N]}$ by defining the N-block map

$$\Psi(a_0 a_1 \dots a_{N-1}) = [a_0 a_1 \dots a_{N-1}].$$

If X is a closed σ-invariant subset of $A^{\mathbb{Z}}$, then $X^{[N]} := \psi(X)$ is called the *N-block presentation* of X, where ψ is the conjugacy from (X, σ) to $(X^{[N]}, \sigma)$ associated with the sliding block code Ψ (see e.g. [25]). Sequences y from $X^{[N]}$ are obtained by

$$y_k = \psi(x)_k = \Psi(x_{k-P} x_{k-P+1} \dots x_{k-P+N-1}) \quad \text{for all integers } k.$$

Here P is called the *memory* of the sliding block code.

The inverse of ψ when $P = 0$ is associated with the letter-to-letter map π_0 given by

$$\pi_0([a_0 a_1 \dots a_{N-1}]) = a_0.$$

We now concentrate our attention on $X = X_\alpha$, where α is a primitive substitution on A with constant length L. A pleasant property is that the N-block presentation of (X_α, σ) is again a substitution dynamical system. This means that we are looking for a primitive substitution β on the alphabet $B = A_\alpha^{[N]} := \{[a_0 \dots a_{N-1}] : a_0 \dots a_{N-1} \in \mathcal{L}_\alpha\}$, such that $X_\beta = X_\alpha^{[N]}$. Such a β exists, and has been introduced for completely different reasons in Queffélec's book [22] on page 95. However, we want a whole family of substitutions generating the N-block presentation of (X_α, σ). We denote the members of this family by $\widehat{\alpha}_{N,M}$. The $\widehat{\alpha}_{N,0}$ are considered in [22], and the $\widehat{\alpha}_{2,M}$ play a key role in [17].

If $[a_0 \dots a_{N-1}]$ is an element of B, we can apply α, obtaining a word

$$v = v_0 v_1 \dots v_{LN-1} := \alpha(a_0 \dots a_{N-1}).$$

Now choose any integer M with $0 \leq M \leq (L-1)(N-1)$, so that the factor w of length $L+N$ of $v = \alpha(a_0 \dots a_{N-1})$ starting with the symbol v_M is well-defined. Then we define

$$\widehat{\alpha}_{N,M}([a_0 \dots a_{N-1}]) = [v_M \dots v_{M+N-1}][v_{M+1} \dots v_{M+N}] \dots [v_{M+L-1} \dots v_{M+L+N-2}].$$

We call the parameter M the *lag* of $\widehat{\alpha}_{N,M}$.

Example 4.1. Let $A = \{1, 2, 3\}$, and let α be given by

$$\alpha(1) = 1233, \qquad \alpha(2) = 2313, \qquad \alpha(3) = 3123.$$

Then the words of length $N = 2$ in the language of α are 12, 13, 23, 31, 32 and 33. We construct the 2-block substitution $\beta = \widehat{\alpha}_{2,1}$ on the alphabet $A_\alpha^{[2]}$ with lag $M = 1$. Since $\alpha(12) = 12332313$, we have $\beta([12]) = [23][33][32][23]$. Coding the $[aa']$ in lexicographical order to a standard alphabet gives $B = \{1, 2, \ldots, 6\}$. On[2] B we have $\beta(1) = 3653, \beta(2) = 3664, \ \beta(3) = 4264, \ \beta(4) = 1341, \ \beta(5) = 1353, \ \beta(6) = 1364$.

The following lemma links the iterates of $\widehat{\alpha}_{N,M}$ to the iterates of α. For notational reasons we define the *hat operator* $\mathcal{H}$ by $\mathcal{H}_{N,M}(\alpha) = \widehat{\alpha}_{N,M}$.

Lemma 4.1. *For all $n \geq 1$, $[\mathcal{H}_{N,M}(\alpha)]^n = \mathcal{H}_{N,M(L^n-1)/(L-1)}(\alpha^n)$.*

Proof. It is easily seen that for two lags M and M' we obtain for the composition $\mathcal{H}_{N,M}(\alpha) \circ \mathcal{H}_{N,M'}(\alpha) = \mathcal{H}_{N,M'L+M}(\alpha^2)$. Iterating, one finds that the cumulative lag in $\alpha^n(a_0 \ldots a_{N-1})$ is $L^n M + L^{n-1} M + \cdots + LM + M = M(L^n - 1)/(L - 1)$. $\square$

A weaker form of the following proposition for the case $N = 2$ can be found in the paper [17].

Proposition 4.1. *Let α be a primitive substitution of length L on an alphabet A. For a positive integer N, and any M with $0 \leq M \leq (L-1)(N-1)$ let $\beta = \widehat{\alpha}_{N,M}$ on the alphabet $A_\alpha^{[N]}$. Then $X_\beta = X_\alpha^{[N]}$, and so the system (X_β, σ) is conjugate to (X_α, σ).*

Proof. From Lemma 4.1 it can be deduced that $\beta = \widehat{\alpha}_{N,M}$ on $B = A_\alpha^{[N]}$ is a primitive substitution (choose n so large that all words from $\mathcal{L}_\alpha^N$ occur in any word of length L^n from $\mathcal{L}_\alpha$). Lemma 4.1 can also be used to see that for all n there are words of length L^n from $\mathcal{L}_\beta$ which are words of the N-block presentation. Hence $X_\beta \subseteq X^{[N]}$, and by minimality, the sets are equal. $\square$

If X_α is infinite, then clearly the alphabets $A_\alpha^{[N]}$ grow larger and larger with N. So by Proposition 4.1 one obtains

Proposition 4.2. *For any primitive constant length substitution with infinite associated symbolic system there exist infinitely many primitive substitutions of the same length with symbolic systems topologically conjugate to the given system.*

5. For constant length substitution minimal sets 3-block codes suffice

In general a semi-conjugacy from a system (X, σ) to (Y, σ) can always be obtained as a sliding block code from X to Y (see [15]).

Here we give a new proof of a known result (see [4], Theorem 3).

Theorem 5.1. *Let α and β each be primitive substitutions of constant length $L > 1$, whose minimal systems (X_α, σ) and (X_β, σ) are infinite. If there exists a semi-conjugacy from (X_α, σ) to (X_β, σ), and β is injective then there is such a semi-conjugacy which is given by a 3-block code.*

Proof. Denote by ϕ the hypothesized semi-conjugacy. We may assume without loss of generality that the associated sliding block code Φ is an L^n-block code with memory 0 for some integer n. Recall that $\mathcal{L}_\alpha^3$ denotes the set of words of length three in $\mathcal{L}_\alpha$, and let B be the alphabet of β. First note that injectivity of β implies injectivity of β^n for all n.

[2] In the sequel we will often identify the alphabet $A_\alpha^{[N]}$ with its standard form.

We construct a three-block code Ψ from $\mathcal{L}_\alpha^3$ to B.

Let $w = w_1 \dots w_K$ be a word in the language of α that contains all ijk from $\mathcal{L}_\alpha^3$, and such that the length of w is large enough so that $\Phi(\alpha^n(w))$ is β^n-recognizable. Then there is a unique decomposition

$$\Phi(\alpha^n(w)) = u\, \beta^n(b_1)\beta^n(b_2) \dots \beta^n(b_{K-2})\, v$$

with the lengths of u and v smaller than L^n and $b_1 b_2 \dots b_{K-2}$ in $\mathcal{L}_\beta$. Here the sum of the lengths $|u| + |v| = L^n + 1$, except when $|u| = 0$ or $|u| = 1$, in which case $|u| + |v| = 1$, and there are actually $K - 1$ β^n-blocks in the decomposition above.

It follows that for each block $\alpha^n(ijk)$ there is a unique β^n-block, say $\beta^n(p)$, coded by $\alpha^n(ijk)$, in $\Phi(\alpha^n(w))$. By injectivity of β^n, $\Psi(ijk) := p$ is then well-defined.

We give a more detailed description of this with aid of the notation $x[s,t] = x_s x_{s+1} \dots x_t$ for a word $x = x_1 \dots x_m$, and $1 \le s < t \le m$.

If $\ell = |u|$ is the length of the prefix u, then for $2 \le \ell \le L^n - 1$ we have

$$\beta^n(\Psi(w_k w_{k+1} w_{k+2})) = \Phi\big(\alpha^n(w_k)[\ell+1, L^n]\, \alpha^n(w_{k+1})\, \alpha^n(w_{k+2})[1, \ell-1]\big).$$

In case $\ell = 0$ or $\ell = 1$ the right hand side equals $\Phi\big(\alpha^n(w_k)\, \alpha^n(w_{k+1})[1, L^n - 1]\big)$, respectively $\Phi\big(\alpha^n(w_k)[2, L^n]\, \alpha^n(w_{k+1})\big)$.

Now note that $\Psi(w_1 w_2 w_3)\, \Psi(w_2 w_3 w_4) \dots \Psi(w_{K-2} w_{K-1} w_K) = b_1 b_2 \dots b_{K-2}$ is in the language of β. This is true in a similar way for words $w \in \mathcal{L}_\alpha$ with a length larger than K. So $\psi(X_\alpha) \subseteq X_\beta$, and by minimality $\psi(X_\alpha) = X_\beta$. $\square$

Corollary 5.1. *If the semi-conjugacy of Theorem 5.1 is a conjugacy, then the 3-block code which results from the proof is also a conjugacy.*

Proof. If x and x' are different points in X_α, it is obvious that their images under ψ are also different, so that a conjugacy results. $\square$

Remark 5.1. In [17] it is shown for a special class of substitutions that the measure-theoretic semi-conjugacies are given by 2-block codes. The example of the Thue–Morse substitution (see Section 10) shows that 3-block codes are sometimes necessary.

Remark 5.2. Let τ be the Toeplitz substitution given by $\tau(0) = 01$, $\tau(1) = 00$. It is easily checked that for any $n \ge 1$ $\tau^n(0)$ and $\tau^n(1)$ differ only in their final letters. This implies that for this substitution the 3-block codes can be replaced by 2-block codes, since $\Psi(ijk)$ can be replaced by $\Psi(ij)$ as $\tau^n(0)[1, \ell-1] = \tau^n(1)[1, \ell-1]$ for all possible $\ell = 2, \dots, 2^n - 1$.

6. Injective substitutions

A key ingredient in our classification result is that we may suppose that the substitutions are injective. This is based on the following result.

Theorem 6.1 ([1]). *Any system generated by a primitive, non-periodic substitution which is not injective is conjugate to a system generated by a primitive substitution that* is *injective.*

The proof given in [1] is constructive, and yields what we call the *injectivization* of a substitution. It is an amalgamation of the original substitution. The construction amounts to identifying (iteratively) those letters which have equal images. For example, the substitution β given by

$$\beta(1) = 46, \qquad \beta(2) = 45, \qquad \beta(3) = 26, \qquad \beta(4) = 25,$$
$$\beta(5) = 13, \qquad \beta(6) = 13$$

amalgamates in a first step to

$$\beta'(1) = 45, \qquad \beta'(2) = 45, \qquad \beta'(3) = 25, \qquad \beta'(4) = 25, \qquad \beta'(5) = 13,$$

and then in a second step to the injective substitution

$$\beta''(1) = 35, \qquad \beta''(3) = 15, \qquad \beta''(5) = 13.$$

7. Substitutions and graph homomorphisms

Let x be an infinite two-sided sequence over an alphabet A. Here we study the general question whether x can be generated by a substitution of length L.

We consider graphs $\mathcal{G} = (V, E)$, $\mathcal{G}' = (V', E')$, and graph homomorphisms $\varphi : \mathcal{G} \to \mathcal{G}'$, i.e., maps $\varphi : V \to V'$ having the property that $(u, v) \in E$ implies that $(\varphi(u), \varphi(v)) \in E'$.

Let $W_2 = W_2(x) = \{ab : ab = x_k x_{k+1}$ for some $k \in \mathbb{Z}\}$, be the set of 2-blocks occurring in x, and for $0 \le M \le L-1$ let $W_{L,M} = W_{L,M}(x) = \{a_1 \dots a_L : a_1 \dots a_L = x_{kL+M} \dots x_{kL+M+L-1}$ for some $k \in \mathbb{Z}\}$ be the set of L-blocks occurring in x at positions M mod L.

With x we associate a family of graphs—cf. [18], Section 1.3.4. The simplest is $\mathcal{G}_1^x = (V_1, E_1)$, the factor graph of order 1 of x, given by

$$V_1 = A, \qquad E_1 = \{(a, b) : ab \in W_2\}.$$

The graphs $\mathcal{G}_{L,M}^x = (V_{L,M}, E_{L,M})$ for $M = 0, \dots, L-1$ are defined by

$$V_{L,M} = W_{L,M},$$
$$E_{L,M} = \{(a_1 \dots a_L, b_1 \dots b_L) : a_1 \dots a_L b_1 \dots b_L \in W_{2L,M} \cup W_{2L,M+L}\}.$$

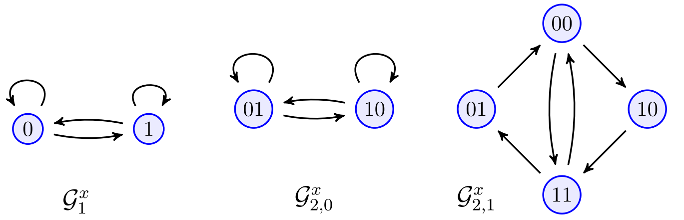
We follow the convention of calling a surjective homomorphism an *epimorphism*. This requires that both the map on vertices and the map on edges are surjective.

Lemma 7.1. *Let x be sequence over A, and let φ be a primitive substitution of constant length L over A. If $x \in X_\varphi$, then there exists an integer $M \in \{0, 1, \dots, L-1\}$ such that φ induces a graph epimorphism from $\mathcal{G}_1^x$ onto $\mathcal{G}_{L,M}^x$.*

Proof. When x is in X_φ, x can be written as a concatenation of φ-blocks. Define M as the first cutting position at or after 0, i.e., $M = \min\{m \ge 0 \mid \sigma^m(x) \in \varphi(X_\varphi)\}$. Let y be such that $x = \sigma^M \varphi(y)$. By minimality of X_φ, all letters of A occur in y, and the substitution defines a surjective map from $V_1 = A$ to $V_{L,M} = W_{L,M}(x)$. By minimality of X_φ, one has $W_2(y) = W_2(x)$, and if ab occurs in y, then $\varphi(a)\varphi(b)$ is in $W_{2L,M}(x)$. Thus φ can be seen as a graph homomorphism, and is also surjective on the edges, since any $w \in W_{2L,M}(x)$ must come from (at least) one word ab in $W_2(y)$ as $w = \varphi(ab)$. $\square$

Note that to avoid cumbersome notation we do not distinguish between φ as a map on words and φ as a graph epimorphism.

As a simple example we consider the Thue–Morse sequence $x = 0110100110\dots$. It is easy to write down the graphs of the letters and the 2-blocks:



Note that $\mathcal{G}_{2,1}^x$ has too many vertices to admit a graph epimorphism φ. With $\mathcal{G}_{2,0}^x$ we find *two* surjective graph homomorphisms: $\varphi(0) = 01$, $\varphi(1) = 10$, corresponding to the usual substitution, but also $\varphi^\flat$ given by $\varphi^\flat(0) = 10$, $\varphi^\flat(1) = 01$. Note that the standard forms $1 \to 12,\ 2 \to 21$ and $1 \to 21,\ 2 \to 12$ are different.

Remark 7.1. In the sequel we consider only sequences x from a substitution minimal set X_α. Then the graphs $\mathcal{G}_{L,M}^x$ will be basically the same for all $x \in X_\alpha$, except that they will be cyclically permuted if one passes from x to $\sigma(x)$. Our canonical choice is $x \in \alpha(X_\alpha)$.

8. The list problem

In this section we first describe an algorithm to find for a given primitive substitution α all primitive injective substitutions β of the same length whose associated systems are factors of (X_α, σ).

Procedure 8.1. *By Theorem 5.1 we may suppose that the factor map is a* 3*-block map. Start with the* 3*-block presentation* $X_\alpha^{[3]}$ *of* α *from Section 4. All factors of* (X_α, σ) *can be obtained by going through all (including the identity) letter-to-letter maps* π *from* $X_\alpha^{[3]}$ *to another shift space. To see whether such a factor* $X := \pi(X_\alpha^{[3]})$ *is generated by a primitive substitution of length L, take any sequence u from* $X_\alpha^{[3]}$*, and define* $x := \pi(u)$*. Determine the graph* $\mathcal{G}_1^x$ *and the graphs* $\mathcal{G}_{L,M}^x$ *for all* $M = 0, \dots, L-1$*. Then determine all epimorphisms* φ *from* $\mathcal{G}_1^x$ *to* $\mathcal{G}_{L,M}^x$*. By Lemma 7.1 this gives a list of all possible candidates* φ *that might generate X. Discard the* φ *which are not primitive. Then check whether all subwords that appear in sequences of X also occur in sequences of* X_φ*. If not, discard* φ*. Else,* $X = X_\varphi$*, and* (X_φ, σ) *is a factor of* (X_α, σ)*.*

The last step in this procedure is algorithmic because of minimality and Theorem 34 in [2] (based on earlier work in [14]). A computer program for this can be found at [21]. In some cases the procedure can be executed by hand. We shall do this in Section 9 for the Toeplitz substitution, and in Section 10 for the Thue–Morse substitution.

It is useful in practice that the last step in the procedure may be supplemented (and in many cases replaced) by checking whether there exists an integer p with $1 \le p \le \mathrm{Card}(A_\alpha^{[3]})$ and an integer M with $0 \le M \le 2(L-1)$, such that φ^p is an amalgamation of $(\widehat{\alpha}_{3,M})^p$, i.e., such that $\pi \circ (\widehat{\alpha}_{3,M})^p = \varphi^p \circ \pi$ holds for some letter-to-letter map π.

For an algorithm for the list problem for conjugacy we still need another ingredient. A dynamical system is called *coalescent* if every endomorphism is an automorphism, i.e., every topological semi-conjugacy from the system onto itself is a topological conjugacy. It was shown for a two symbol alphabet in [3] and for a general alphabet in [12,13] that primitive, not necessarily constant length, substitutions generate coalescent dynamical systems.

Procedure 8.2. *Use Procedure 8.1 to determine all primitive injective substitutions β with the same length that generate factors of (X_α, σ). Make the list for β, and check whether α is on it. If it is, then (X_α, σ) is conjugate to (X_β, σ), by coalescence; if not, then (X_α, σ) is not conjugate to (X_β, σ).* $\square$

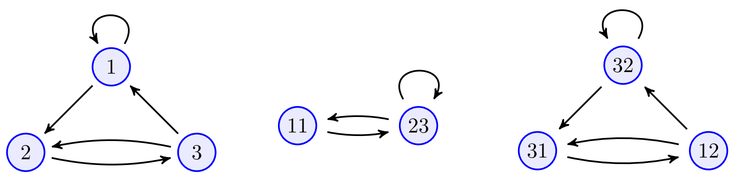
9. The conjugacy class of the Toeplitz substitution

We use Procedure 8.1 to determine the injective substitutions of length two that generate factors of the Toeplitz system (X_τ, σ) where τ is the substitution

$$\tau(0) = 01, \qquad \tau(1) = 00.$$

According to Remark 5.2 we can restrict ourselves to 2-block codes for τ. The set of words of length two in $\mathcal{L}_\tau$ is equal to $\mathcal{L}_\tau^2 = \{00, 01, 10\}$, so we code the 2-blocks lexicographically by $A_\tau^{[2]} = \{1, 2, 3\}$.

We first consider the case where the letter-to-letter map π is the identity. The graphs $\mathcal{G}_1 = \mathcal{G}_1^x$, $\mathcal{G}_{2,0} = \mathcal{G}_{2,0}^x$ and $\mathcal{G}_{2,1} = \mathcal{G}_{2,1}^x$ of a sequence $x \in \psi(\tau(X_\tau))$ in the 2-block presentation $X_\tau^{[2]}$ are given by



$\mathcal{G}_1$ $\mathcal{G}_{2,0}$ $\mathcal{G}_{2,1}$

There are two surjective graph homomorphisms $\varphi : \mathcal{G}_1 \to \mathcal{G}_{2,0}$ which give a primitive substitution:

$$\varphi(1) = 23,\ \varphi(2) = 23,\ \varphi(3) = 11, \quad \text{and} \quad \varphi(1) = 23,\ \varphi(2) = 11,\ \varphi(3) = 23.$$

The first φ generates the 2-block presentation, since it may be checked that φ is equal to $\widehat{\tau}_{2,0}$. After injectivization it gives the substitution α given by $\alpha(1) = 13,\ \alpha(3) = 11$, whose standard form is the Toeplitz substitution. The second one is not equal to a $\widehat{\tau}_{2,M}$, and so we will postpone the answer to the question whether it generates a factor. It injectivizes to the substitution α given by $\alpha(1) = 21,\ \alpha(2) = 11$, which we call the *rotated Toeplitz substitution*.

There is exactly one surjective graph homomorphism $\varphi : \mathcal{G}_1 \to \mathcal{G}_{2,1}$, which gives the primitive substitution:

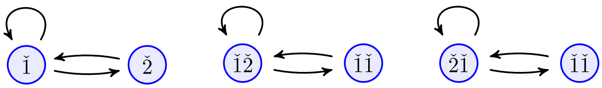
$$\varphi(1) = 32, \qquad \varphi(2) = 31, \qquad \varphi(3) = 12,$$

which has the standard form given by $\alpha(1) = 23,\ \alpha(2) = 13,\ \alpha(3) = 12$. We call this substitution 3-*symbol Toeplitz*. It may be checked that this φ is equal to $\widehat{\tau}_{2,1}$, and so the system generated by this φ is conjugate to the Toeplitz system by Proposition 4.1.

To finish, we still have to examine the possibilities of letter-to-letter maps $\pi : \{1, 2, 3\} \to \{\check{1}, \check{2}\}$, where $\{\check{1}, \check{2}\}$ is a two letter alphabet. There are three of these maps π_k given by

$$\pi_1 : 1 \to \check{1},\ 2 \to \check{1},\ 3 \to \check{2}, \qquad \pi_2 : 1 \to \check{1},\ 2 \to \check{2},\ 3 \to \check{1},$$
$$\pi_3 : 1 \to \check{2},\ 2 \to \check{1},\ 3 \to \check{1}.$$

Let t_k for $k = 1, 2, 3$ be a sequence from $\pi_k(\widehat{\tau}_{2,1}(X_{\widehat{\tau}_{2,1}}))$. The graphs $\mathcal{G}_1^1 = \mathcal{G}_1^{t_1}$, $\mathcal{G}_{2,0}^1 = \mathcal{G}_{2,0}^{t_1}$ and $\mathcal{G}_{2,1}^1 = \mathcal{G}_{2,1}^{t_1}$ are given by



$\mathcal{G}_1^1$ $\mathcal{G}_{2,0}^1$ $\mathcal{G}_{2,1}^1$

There are obvious graph epimorphisms from $\mathcal{G}_1^1$ to $\mathcal{G}_{2,0}^1$ and to $\mathcal{G}_{2,1}^1$. The first one again yields the Toeplitz substitution, the second one yields the substitution

$$\check{\varphi}(\check{1}) = \check{2}\check{1}, \qquad \check{\varphi}(\check{2}) = \check{1}\check{1},$$

whose standard form is rotated Toeplitz. Since here we have the intertwining relation

$$\pi_1 \circ \widehat{\tau}_{2,1} = \check{\varphi} \circ \pi_1,$$

$\check{\varphi}$ is an amalgamation of $\widehat{\tau}_{2,1}$, so $(X_{\check{\varphi}}, \sigma)$ *is* a factor of the Toeplitz substitution system. It actually is conjugate to the Toeplitz system, since Toeplitz will be in the list of factors of the rotated Toeplitz substitution.

One can check that the letter-to-letter map π_2 gives similar results, and that the graph $\mathcal{G}_1^3$ has two loops, which prevents graph homomorphisms in the case of π_3.

Conclusion: the conjugacy class of the injective substitutions of the Toeplitz system consists of three substitutions:

Toeplitz, rotated Toeplitz, and 3-symbol Toeplitz.

We will examine the properties of the minimal set $Y := \pi_3(X_\tau^{[2]}) \subseteq \{\check{1}, \check{2}\}^{\mathbb{Z}}$ in more detail. We showed that Y is not generated by a substitution of length 2. We will prove more: Y is not generated by *any* substitution of constant length. The only other example we know of this kind is the Rudin–Shapiro minimal set, cf. [24], page 1613.

First we prove the rather surprising fact that the sequences in Y are essentially obtained by doubling the letters in the sequences of the Toeplitz minimal set. Define the doubling morphism $\delta : \{0, 1\}^* \to \{\check{1}, \check{2}\}^*$ by

$$\delta(0) = \check{1}\check{1}, \qquad \delta(1) = \check{2}\check{2}.$$

Lemma 9.1. *Let τ be the Toeplitz substitution on $\{0, 1\}$, let $\beta := \widehat{\tau}_{2,1}$, and let $\pi = \pi_3$ be the projection $1 \to \check{2}$, $2 \to \check{1}$, $3 \to \check{1}$. Then for all $n \geq 1$*

$$\pi(\beta^{2n}(1)) = \check{2}\,\delta(\tau^{2n-1}(0))\,\check{2}^{-1},$$
$$\pi(\beta^{2n}(2)) = \check{2}\,\delta(\tau^{2n-1}(1))\,\check{1}^{-1},$$
$$\pi(\beta^{2n}(3)) = \check{1}\,\delta(\tau^{2n-1}(0)\,)\check{2}^{-1}.$$

Proof. By induction. For $n = 1$ we have $\pi(\beta^2(1)) = \pi(1231) = \check{2}\check{1}\check{1}\check{2}$. On the other hand, $\check{2}\delta(\tau(0))\check{2}^{-1} = \check{2}\delta(01)\check{2}^{-1} = \check{2}\check{1}\check{1}\check{2}\check{2}\check{2}^{-1} = \check{2}\check{1}\check{1}\check{2}$. Now the induction step:

$$\begin{aligned}\pi(\beta^{2(n+1)}(1)) &= \pi(\beta^{2n}(1231)) = \pi(\beta^{2n}(1))\pi(\beta^{2n}(2))\pi(\beta^{2n}(3))\pi(\beta^{2n}(1))\\ &= \check{2}\delta(\tau^{2n-1}(0))\check{2}^{-1}\check{2}\delta(\tau^{2n-1}(1))\check{1}^{-1}\check{1}\delta(\tau^{2n-1}(0))\check{2}^{-1}\check{2}\delta(\tau^{2n-1}(0))\check{2}^{-1}\\ &= \check{2}\,\delta(\tau^{2n-1}(0100))\check{2}^{-1} = \check{2}\,\delta(\tau^{2n+1}(0))\check{2}^{-1}.\end{aligned}$$

For the letters 2 and 3 a similar computation yields the corresponding formula. $\square$

It follows from Lemma 9.1 that Y is the closed orbit of the sequence $y = \delta(t)$, where t is the Toeplitz sequence.

We need another combinatorial lemma.

Lemma 9.2. *Let t be the Toeplitz sequence, and let M be a fixed integer with $0 \le M < 2^n$ for some $n \ge 1$. Then there is at most one word w of length 2^n such that the square ww occurs at some position M mod 2^n in t. The same property holds for the sequence $y = \delta(t)$.*

Proof. For even n (for odd n exchange the suffixes 0 and 1) the words

$$\tau^n(0) =: a_1 a_2 \dots a_{2^n-1} 0, \quad \text{and} \quad \tau^n(1) =: a_1 a_2 \dots a_{2^n-1} 1$$

only differ in the last letters. Therefore the only two words of length 2^n occurring in t at position M mod 2^n are

$$v_M := a_{M+1} \dots a_{2^n-1} 0 a_1 \dots a_M, \quad \text{and} \quad w_M := a_{M+1} \dots a_{2^n-1} 1 a_1 \dots a_M.$$

Since 11 does not occur in t, $\tau^n(11)$ does not occur in t, and this implies that $v_M v_M$ is the only square occurring at positions M mod 2^n.

Now note that this implies that the same property holds for $\delta(t)$ for all words occurring at the even positions $2M$ mod 2^n. But then it also holds for positions $2M + 1$ mod 2^n, since if a square occurred at such an odd position, then we could shift 1 to the left, obtaining a square at an even position (the words in $\delta(t)$ in even positions have prefix $\check{1}\check{1}$ or $\check{2}\check{2}$). □

We are now ready to prove the announced result.

Proposition 9.1. *Let $Y := \pi_3(X_\tau^{[2]}) \subseteq \{\check{1}, \check{2}\}^{\mathbb{Z}}$ be the projection of the 2-block presentation of the Toeplitz minimal set considered before. Then (Y, σ) is not a substitution dynamical system.*

Proof. First note that if Y would be generated by a substitution γ, then, by Cobham's Theorem, the length of γ would be a power of 2. Recall $y = \delta(t)$. We use Lemma 7.1. The graph $\mathcal{G}_1^y$ is the complete graph on the nodes $\check{1}$ and $\check{2}$. For each n and for all $M = 0, \dots, 2^n - 1$ the graphs $\mathcal{G}_{2^n,M}^y$ have *only* one loop, because y has only one square at position M mod 2^n, by Lemma 9.2. But then an epimorphism from $\mathcal{G}_1^y$ to $\mathcal{G}_{2^n,M}^y$ is impossible. □

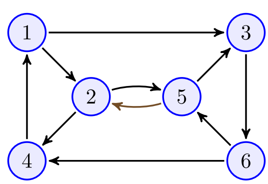
10. The length 2 substitution factors of the Thue–Morse system

Let θ and $\theta^\flat$ be the Thue–Morse substitutions of length 2 on $A = \{0, 1\}$ given by

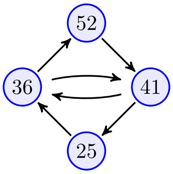
$$\theta(0) = 01, \qquad \theta(1) = 10, \qquad \theta^\flat(0) = 10, \qquad \theta^\flat(1) = 01.$$

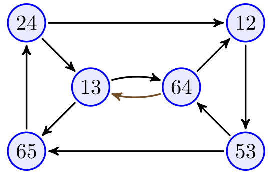
The set of words of length 3 in the language of θ is $\mathcal{L}_\theta^3 = \{001, 010, 011, 100, 101, 110\}$. The usual lexicographic coding – which happens to be the binary coding – gives the 3-block alphabet $A_\theta^{[3]} := \{1, 2, 3, 4, 5, 6\}$. The graph $\mathcal{G}_1 = \mathcal{G}_1^x$ of a sequence $x \in \psi(\theta(X_\theta))$ from the 3-block presentation $X_\theta^{[3]}$ is given by

$\mathcal{G}_1$



The graphs $\mathcal{G}_{2,0} = \mathcal{G}^x_{2,0}$ and $\mathcal{G}_{2,1} = \mathcal{G}^x_{2,1}$ describing the 2-blocks in a sequence x from the 3-block presentation $X_\theta^{[3]}$ are given by

$\mathcal{G}_{2,0}$ 

$\mathcal{G}_{2,1}$ 

To find all graph epimorphisms from $\mathcal{G}_1$ to $\mathcal{G}_{2,0}$ and $\mathcal{G}_{2,1}$, we exploit the following simple lemma.

Lemma 10.1. *Let $\varphi : \mathcal{G} \to \mathcal{G}'$ be a graph homomorphism. Suppose $\mathcal{G}'$ has no loops. Then 2-cycles and 3-cycles in $\mathcal{G}$ are mapped to 2-cycles, respectively 3-cycles in $\mathcal{G}'$.*

It will appear that all these graph epimorphisms are either a 3-block substitution of θ or of $\theta^\flat$. This will be indicated below.

We start with finding all $\varphi : \mathcal{G}_1 \to \mathcal{G}_{2,0}$. By Lemma 10.1, $\{\varphi(2), \varphi(5)\}$ equals $\{36, 41\}$. If $\varphi(2) = 36$, then $\varphi(4) = 52$ and $\varphi(1) = 41$, *or* $\varphi(4) = 41$ and $\varphi(1) = 25$. In the first case necessarily $(5, 3, 6) \to (41, 25, 36)$ by Lemma 10.1, and we obtain

$$\widehat{\theta^\flat}_{2,2} :\ 1\ \to 41,\ 2 \to 36,\ 3 \to 25,\ 4 \to 52,\ 5 \to 41,\ 6 \to 36.$$

In the second case $(5, 3, 6) \to (41, 36, 52)$, and we obtain

$$\widehat{\theta}_{2,0} :\ 1\ \to 25,\ 2 \to 36,\ 3 \to 36,\ 4 \to 41,\ 5 \to 41,\ 6 \to 52.$$

If $\varphi(2) = 41$, then in the same way we obtain third and fourth epimorphisms

$$\widehat{\theta}_{2,2} :\ 1 \to 36,\ 2 \to 41, 3 \to 52,\ 4 \to 25,\ 5 \to 36,\ 6 \to 41,$$
$$\widehat{\theta^\flat}_{2,0} :\ 1 \to 52,\ 2 \to 41, 3 \to 41,\ 4 \to 36,\ 5 \to 36,\ 6 \to 25.$$

Next we consider all $\varphi : \mathcal{G}_1 \to \mathcal{G}_{2,1}$. Now $\{\varphi(2), \varphi(5)\}$ equals $\{13, 64\}$.

If $\varphi(2) = 13$, then $\varphi(4) = 65$ and $\varphi(1) = 24$, and also $\varphi(5) = 64$, $\varphi(3) = 12$ and $\varphi(6) = 53$, since $(2, 4, 1)$ and $(5, 3, 6)$ form 3-cycles. In this way we obtain

$$\widehat{\theta^\flat}_{2,1} :\ 1 \to 24,\ 2 \to 13, 3 \to 12,\ 4 \to 65,\ 5 \to 64,\ 6 \to 53.$$

If $\varphi(2) = 64$, then in the same way we obtain an epimorphism

$$\widehat{\theta}_{2,1} :\ 1\ \to 53,\ 2 \to 64,\ 3 \to 65,\ 4 \to 12,\ 5 \to 13,\ 6 \to 24.$$

We now do the letter-to-letter maps. This is much more involved than in the case of the Toeplitz substitution.

Note that the letter-to-letter maps from $A_\theta^{[3]} = \{1, 2, 3, 4, 5, 6\}$ to another alphabet are in one to one correspondence with the set of all partitions of $\{1, 2, 3, 4, 5, 6\}$. Hence there are $B_6 = 203$

of such maps, where B_6 is the sixth Bernoulli number. Since M can take the values 0 and 1, this means that there are 406 cases of candidate epimorphisms to consider.

To reduce this number, we note that there is the mirror symmetry $0 \to 1, 1 \to 0$, which at the level of 3-blocks corresponds to the permutation $\mathcal{P} = (16)(25)(34)$. Obviously a partition and its permuted version will generate (if any) a substitution with the same standard form.

To further speed up the process we can apply the following three simple tools.

(T1) If $\mathcal{G}_{L,M}$ has more nodes than $\mathcal{G}_1$, then an epimorphism is not possible.

If there *is* an epimorphism from $\mathcal{G}_1$ to $\mathcal{G}_{L,M}$, then:

(T2) If the graph $\mathcal{G}_1$ contains a loop then $\mathcal{G}_{L,M}$ contains a loop.

(T3) If $\mathcal{G}_1$ and $\mathcal{G}_{L,M}$ have the same number of nodes, then they also must have the same number of edges.

With aid of the tools one finds 15 candidate substitutions to generate factors of the Thue–Morse system generated by injective substitutions of length 2:

Nr.	Partition	M	Substitution
θ_1	$\{1,2,3\}\{4,5,6\}$	0	$1 \to 14, 4 \to 41$
θ_2	$\{1,2,3\}\{4,5,6\}$	0	$1 \to 41, 4 \to 14$
θ_3	$\{1,2,5,6\}\{3,4\}$	0	$1 \to 31, 3 \to 11$
θ_4	$\{1,6\}\{2,3,4,5\}$	0	$1 \to 22, 2 \to 21$
θ_5	$\{1,4,5\}\{2,3\}\{6\}$	1	$1 \to 12, 2 \to 61, 6 \to 21$
θ_6	$\{1,4,5\}\{2,6\}\{3\}$	1	$1 \to 21, 2 \to 13, 3 \to 12$
θ_7	$\{1,6\}\{2,5\}\{3,4\}$	1	$1 \to 23, 2 \to 13, 3 \to 12$
θ_8	$\{1\}\{2,3\}\{4,5\}\{6\}$	0	$1 \to 24, 2 \to 26, 4 \to 41, 6 \to 42$
θ_9	$\{1\}\{2,3\}\{4,5\}\{6\}$	0	$1 \to 42, 2 \to 41, 4 \to 26, 6 \to 24$
θ_{10}	$\{1,5\}\{2,6\}\{3\}\{4\}$	0	$1 \to 41, 2 \to 32, 3 \to 21, 4 \to 12$
θ_{11}	$\{1,5\}\{2,6\}\{3\}\{4\}$	0	$1 \to 32, 2 \to 41, 3 \to 12, 4 \to 21$
θ_{12}	$\{1,5\}\{2\}\{3\}\{4\}\{6\}$	1	$1 \to 13, 2 \to 64, 3 \to 61, 4 \to 12, 6 \to 24$
θ_{13}	$\{1\}\{2,3\}\{4\}\{5\}\{6\}$	1	$1 \to 24, 2 \to 12, 4 \to 65, 5 \to 64, 6 \to 52$
θ_{14}	$\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}$	1	$1 \to 24, 2 \to 13, 3 \to 12, 4 \to 65, 5 \to 64, 6 \to 53$
θ_{15}	$\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}$	1	$1 \to 53, 2 \to 64, 3 \to 65, 4 \to 12, 5 \to 13, 6 \to 24$

Thue–Morse Factor List—direct projections

All 15 do generate a factor by the following arguments. The systems generated by $\theta_1, \theta_2, \theta_3$ and θ_4 are well known factors of the Thue–Morse system, and θ_8, $\theta_9, \theta_{10}, \theta_{11}, \theta_{14}$, and θ_{15} actually give conjugate systems, because they are injectivizations of 3-block substitutions of θ or of $\theta^\flat$. All others turn out to be amalgamations of either θ_{14} or θ_{15}. For example $\theta_5 \circ \pi = \pi \circ \theta_{15}$, where the partition representation of π is $\{1,4,5\}\{2,3\}\{6\}$. In the same way $\theta_6, \theta_7, \theta_{12}$ and θ_{13} are amalgamations of respectively $\theta_{14}, \theta_{14}, \theta_{15}$ and θ_{14} by projections whose partition representation can be found in the table.

At an early stage of our research we found more than 15 substitutions in the factor list, failing to see that some were essentially the same. For example, θ_8 can also be obtained as the substitution generated by the partition $\{1,5\}\{2,6\}\{3\}\{4\}$, but now for $M = 1$. It is therefore important to transform all θ_k to their standard forms $\theta_k^\diamond$. The standard forms of the substitutions in the Thue–Morse factor list are given in the following table.

Nr.	Standard form	Nr.	Standard form
$\theta_1^\diamond$	$1 \to 12, 2 \to 21$	$\theta_9^\diamond$	$1 \to 23, 2 \to 14, 3 \to 21, 4 \to 12$
$\theta_2^\diamond$	$1 \to 21, 2 \to 12$	$\theta_{10}^\diamond$	$1 \to 21, 2 \to 13, 3 \to 43, 4 \to 31$
$\theta_3^\diamond$	$1 \to 21, 2 \to 11$	$\theta_{11}^\diamond$	$1 \to 23, 2 \to 13, 3 \to 41, 4 \to 31$
$\theta_4^\diamond$	$1 \to 12, 2 \to 11$	$\theta_{12}^\diamond$	$1 \to 12, 2 \to 31, 3 \to 45, 4 \to 35, 5 \to 14$
$\theta_5^\diamond$	$1 \to 12, 2 \to 31, 3 \to 21$	$\theta_{13}^\diamond$	$1 \to 21, 2 \to 13, 3 \to 45, 4 \to 51, 5 \to 43$
$\theta_6^\diamond$	$1 \to 21, 2 \to 13, 3 \to 12$	$\theta_{14}^\diamond$	$1 \to 23, 2 \to 14, 3 \to 21, 4 \to 56, 5 \to 63, 6 \to 54$
$\theta_7^\diamond$	$1 \to 23, 2 \to 13, 3 \to 12$	$\theta_{15}^\diamond$	$1 \to 23, 2 \to 13, 3 \to 41, 4 \to 56, 5 \to 46, 6 \to 25$
$\theta_8^\diamond$	$1 \to 12, 2 \to 31, 3 \to 34, 4 \to 13$		

11. The Thue–Morse conjugacy list

Three substitutions (θ_3, θ_4 and θ_7) in the Thue–Morse factor list generate systems that are certainly not conjugate to the Thue–Morse system, as they are in the Toeplitz conjugacy class. Obviously θ_1 and θ_2 are in the conjugacy list, and we already know that the substitutions θ_8, $\theta_9, \theta_{10}, \theta_{11}, \theta_{14}$, and θ_{15} generate systems conjugate to the Thue–Morse system. To see whether the 4 remaining substitutions yield systems conjugate to the Thue–Morse system, according to Procedure 8.2 we would have to construct the factor list of each of these. This is quite involved, for example the 3-block presentations of the two factors on 5 symbols have 11 symbols.

However, there is a quicker way to determine whether these factors are conjugate to the Thue–Morse system, by finding explicit semi-conjugacies from these factors to the Thue–Morse system. Then by coalescence the systems are conjugate.

It is quickly verified that indeed each of $\theta_5, \theta_6, \theta_{12}$ and θ_{13} amalgamates to Morse or Morse flat. For example for θ_{12} one takes $1, 4 \to 0,\ 2, 3, 6 \to 1$.

Conclusion: there are 12 primitive injective substitutions of length 2 that generate a system conjugate to the Thue–Morse dynamical system.

12. Proper factors

We have seen that for the Thue–Morse system all factors are actually conjugate to the system, if there are no spectral obstructions (discrete spectrum systems cannot be conjugate to systems with partially continuous spectrum). In this section we present in a simple way a system with partially continuous spectrum which has another system with partially continuous spectrum as a proper factor.

Let α be the Mephisto Waltz substitution given by $\alpha(1) = 112,\ \alpha(2) = 221$. Let β be the substitution on four symbols given by

$$\beta(1) = 123, \qquad \beta(2) = 124, \qquad \beta(3) = 341, \qquad \beta(4) = 431.$$

Both substitutions have column number 2, and hence their systems have partially continuous spectrum (cf. [9]).

Proposition 12.1. *The system (X_α, σ) is a proper factor of (X_β, σ).*

Proof. Note that α is an amalgamation of β under the projection map

$$\pi(1) = \pi(2) = 1, \qquad \pi(3) = \pi(4) = 2.$$

Therefore (X_α,σ) is a factor of (X_β,σ). However, (X_β,σ) is not a factor of (X_α,σ). To see this, note that 13 and 14 are in $\mathcal{L}_\beta$, and that 1 is suffix of $\beta^2(1)$. It follows that the two sequences $z := (\beta^2)^\infty(1)\cdot(\beta^2)^\infty(3)$ and $z' := (\beta^2)^\infty(1)\cdot(\beta^2)^\infty(4)$ are in X_β. Next, note that $z \neq z'$ and that $\pi(z) = \pi(z')$, since for all n

$$\pi\beta^n(3) = \alpha^n\pi(3) = \alpha^n(2) = \alpha^n\pi(4) = \pi\beta^n(4).$$

Now suppose $\psi : X_\alpha \to X_\beta$ is a semi-conjugacy. Then, by coalescence, $\pi\circ\psi$ is a conjugacy. But this contradicts our finding that π is 2 to 1 somewhere. $\square$

We remark that it is quite a delicate matter whether a factor is proper or not. For example, let α be the Mephisto Waltz, and let δ be the substitution defined by

$$\delta(1) = 123, \qquad \delta(2) = 124, \qquad \delta(3) = 431, \qquad \delta(4) = 432.$$

Then (X_δ,σ) is conjugate to (X_α,σ), since it may be easily checked that δ is the injectivization of the 3-block substitution $\widehat{\alpha}_{3,0}$.

However, suppose we would follow the approach above, noting that α is an amalgamation of δ with the same π map as above. Now δ^2 has fixed prefix letters 1 and 4 and fixed suffix letters 1,2,3 and 4. This implies that the eight sequences $z_{b,a} := (\delta^2)^\infty(b)\cdot(\delta^2)^\infty(a)$ are well-defined for $a = 1, 4$ and $b = 1, 2, 3, 4$. But, similarly as above, we have $\pi(z_{b,1}) = \pi(z_{b',1})$ and $z_{b,1} \neq z_{b',1}$ for $b = 1, b' = 2$ and for $b = 3, b' = 4$, yielding several points where π is 2 to 1. However, this does not contradict conjugacy of the two systems, since neither $z_{2,1}$ nor $z_{4,1}$ are elements of X_δ, simply because the words 21 and 41 are not in the language of δ.

13. Epilogue

Related work can be found in the thesis of Joseph Herning [16] which mainly concentrates on bijective substitutions, which generate a relatively small subclass of systems with partially continuous spectrum. A *bijective substitution* α on an alphabet A is defined by $\{\alpha(a)_i : a \in A\} = A$ for all $1 \le i \le L$. One of the major results in [16] is that there exist substitution dynamical systems that do not have discrete spectrum factors generated by substitutions. As an example Herning gives the substitution α on three symbols, which also occurs in [22], defined by

$$\alpha(1) = 121, \qquad \alpha(2) = 233, \qquad \alpha(3) = 312.$$

We have reproved his result by computing the factor list of α. It consists of nine injective substitutions, on alphabets of size three to eight, all (indeed!) generating systems with partially continuous spectrum. Without doing any computations, it follows from Theorem 8 in [19] that these factors are in fact all conjugate to the system generated by α, since the substitution α has no non-trivial amalgamations.

An interesting extension of our result would be to consider also non-constant length substitutions. For example, let θ be the ternary Thue–Morse substitution, defined by

$$\theta(1) = 123, \qquad \theta(2) = 13, \qquad \theta(3) = 2.$$

An application of Theorem 1 in Section V of [9] shows that (X_θ,σ) is conjugate to a substitution of constant length 2 on 6 symbols. Its injectivization is a substitution on 5 symbols, and taking the standard form of this substitution we find that it is on the Thue–Morse list.

The paper [23] considers conjugacies between systems generated by two primitive substitutions whose matrices have the same Perron–Frobenius eigenvalue: it is shown there that modulo powers of the shift there are only finitely many conjugacies between such systems. Nevertheless, it has been shown in [10] that there are infinitely many systems on the Thue–Morse list, all generated by primitive injective substitutions with Perron–Frobenius eigenvalue 2.

Primitive substitutions generate dynamical systems with a unique shift invariant measure. One can consider Problem 1.2 for measure-theoretic conjugacy. When a substitution of length L generates a system with discrete spectrum, then obviously there are infinitely many primitive injective substitutions in the measure-theoretic conjugacy class (in fact *all* pure (see [9]) substitutions of length L). When there is partially continuous spectrum, we believe that the equivalence class will be finite, and the same as for topological conjugacy. This has been proved for a subclass of such constant length substitutions in [17].

Acknowledgments

We have profited from electronic discussions on coalescence with Fabien Durand and Reem Yassawi, and from useful remarks by Michelle Lemasurier. Finally, we thank a referee for his comments which have improved our paper.

Some of this work was done while the first author visited Delft and some while the second author was van Vleck Visiting Professor of Mathematics at Wesleyan. We thank these institutions for their support.

References

[1] François Blanchard, Fabien Durand, Alejandro Maass, Constant-length substitutions and countable scrambled sets, Nonlinearity 17 (3) (2004) 817–833.

[2] Émilie Charlier, Narad Rampersad, Jeffrey Shallit, Enumeration and decidable properties of automatic sequences, Internat. J. Found. Comput. Sci. 23 (5) (2012) 1035–1066.

[3] Ethan M. Coven, Endomorphisms of substitution minimal sets, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 20 (1971–1972) 129–133.

[4] Ethan M. Coven, Andrew Dykstra, Michael Keane, Michelle LeMasurier, Topological conjugacy to given constant length substitution minimal systems, Indag. Math. (NS) 25 (4) (2014) 646–651.

[5] Ethan M. Coven, Andrew Dykstra, Michelle Lemasurier, A short proof of a theorem of Cobham on substitutions, Rocky Mountain J. Math. 44 (1) (2014) 19–22.

[6] Ethan M. Coven, Michael S. Keane, The structure of substitution minimal sets, Trans. Amer. Math. Soc. 162 (1971) 89–102.

[7] Ethan M. Coven, Michael Keane, Michelle Lemasurier, A characterization of the Morse minimal set up to topological conjugacy, Ergodic Theory Dynam. Systems 28 (5) (2008) 1443–1451.

[8] Ethan M. Coven, Anthony Quas, Reem Yassawi, Computing automorphism groups of shifts using atypical equivalence classes, Diskret. Anal. (2016) 28 pages. Paper No. 611.

[9] F.M. Dekking, The spectrum of dynamical systems arising from substitutions of constant length, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 41 (3) (1977–1978) 221–239.

[10] Michel Dekking, On the structure of Thue-Morse subwords, with an application to dynamical systems, Theoret. Comput. Sci. 550 (2014) 107–112.

[11] F.Michel Dekking, Morphisms, symbolic sequences, and their standard forms, J. Integer Seq. 19 (1) (2016) 8. Article 16.1.1.

[12] Fabien Durand, Linearly recurrent subshifts have a finite number of non-periodic subshift factors, Ergodic Theory Dynam. Systems 20 (4) (2000) 1061–1078.

[13] Fabien Durand, Corrigendum and addendum to: "Linearly recurrent subshifts have a finite number of non-periodic subshift factors"[Ergodic Theory Dynam. Systems, 20 (2000), no. 4, 1061–1078; MR1779393 (2001m:37022)], Ergodic Theory Dynam. Systems 23 (2) (2003) 663–669.

[14] Isabelle Fagnot, Sur les facteurs des mots automatiques, Theoret. Comput. Sci. 172 (1–2) (1997) 67–89.

[15] G.A. Hedlund, Endomorphisms and automorphisms of the shift dynamical system, Math. Syst. Theory 3 (1969) 320–375.

[16] Joseph Herning, Spectrum and factors of substitution dynamical systems (Ph.D. thesis), George Washington University, 2013.

[17] B. Host, F. Parreau, Homomorphismes entre systèmes dynamiques définis par substitutions, Ergodic Theory Dynam. Systems 9 (3) (1989) 469–477.

[18] M. Lothaire, Algebraic Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002, A collective work by Jean Berstel et al., with a preface by Berstel and Perrin.

[19] Mieczysław K. Mentzen, Invariant sub-σ-algebras for substitutions of constant length, Studia Math. 92 (3) (1989) 257–273.

[20] Brigitte Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques, Bull. Soc. Math. France 124 (2) (1996) 329–346.

[21] H. Mousavi, J. Shallit, Walnut. Technical report, 2013, https://cs.uwaterloo.ca/~shallit/papers.html.

[22] Martine Queffélec, Substitution Dynamical Systems—Spectral Analysis, second ed., in: Lecture Notes in Mathematics, vol. 1294, Springer-Verlag, Berlin, 2010.

[23] Ville Salo, Ilkka Törmä, Block maps between primitive uniform and Pisot substitutions, Ergodic Theory Dynam. Systems 35 (7) (2015) 2292–2310.

[24] Luke Schaeffer, Jeffrey Shallit, The critical exponent is computable for automatic sequences, Internat. J. Found. Comput. Sci. 23 (8) (2012) 1611–1626.

[25] Susan G. Williams, Introduction to symbolic dynamics, in: Symbolic Dynamics and its Applications, in: Proc. Sympos. Appl. Math., vol. 60, Amer. Math. Soc., Providence, RI, 2004, pp. 1–11.