Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Symbolic dynamics and automata theory

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

Frits Gerrit Willem Dannenberg
#1285629
Delft, the Netherlands
April 2011

MSc THESIS APPLIED MATHEMATICS

"Symbolic dynamics and automata theory"

Frits Gerrit Willem Dannenberg

Delft University of Technology

Daily supervisor

Dr. R. J. Fokkink

Responsible professor

Prof. dr. F. M. Dekking

Other thesis committee members

Prof. dr. J. W. Klop

Dr. J. Endrullis

April 2011

Delft, the Netherlands

SYMBOLIC DYNAMICS AND AUTOMATA THEORY

FRITS GERRIT WILLEM DANNENBERG

Contents

1. Introduction

This master's thesis investigates the *equivalence of streams* as presented by [EHK11, Ste08]. In the process some morphic properties of the *Toeplitz* words as first described by [JK69] are discovered. The terminology of theoretical computer science and mathematics shows some discrepancies: In theoretical computer science words of infinite length are usually called *streams* and referred to as *symbolic sequences* in mathematics. Similarly, *morphisms* are what mathematicians usually refer to as substitutions. Notation is formally introduced starting in section 2 but will be informally used in the introduction.

The first part of this section is a proper introduction into the subject. The two subsections will be used to describe two original results: The first has to do with prime degrees, the second one has to do with arithmetic subsequences.

Finite state machines were first described by neurophysiologists Warren McCulloch and Walter Pitts in the 1943 paper "A Logical Calculus Immanent in Nervous Activity". Their pioneering work made significant contributions to the neural network theory, theory of automata, theory of computation and cybernetics. Later, two computer scientists, G.H. Mealy and E.F. Moore, generalized the theory to much more powerful machines in separate papers, published in 1955-56. The finite-state machines, *Mealy machine* and the *Moore machine*, are named in recognition of their work.[1]

Date: May 2, 2011.

[1] "The Intellectual Excitement of Computer Science", Eric Roberts, Stanford University

Automata theory emerged to be important in the field of *compiler design*. It is closely tied to the notion of *regular languages*. One focus of this thesis is the ordering obtained by applying tranducers, a special type of finite state machine, to steams. The type of transducer considered in this document has been called *pure sequential transducer* in [Sak09], simply *finite-state transducer* in [AS03] and also *sequential transducer* in other literature. Denote Σ^* the collection of all words over some finite alphabet Σ. In this thesis the term transducer will refer to:

Definition 1.1. *A finite state transducer (which will be abbreviated to FST or transducer) is a sextuple $A = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ where*

- *Q is a finite set of states*
- *Σ is a finite set, the input alphabet*
- *$\delta \subseteq Q \times \Sigma \times \Delta^* \times Q$ is a transition relation*
- *$q_0 \in Q$ is the initial state*
- *Δ is a finite set, the output alphabet*
- *$\lambda : Q \times \Sigma \to \Delta^*$ is the output function*

Suppose $q_1, q_2 \in Q, a \in \Sigma, b \in \Delta^*$ then

$$(q_1, a, b, q_2) \in \delta$$

describes the edge from state q_1 to state q_2 that takes input a and generates output b. Assume the transition relation to be *deterministic*: For all $q_1 \in Q, a \in \Sigma$ there is exactly one $q_2 \in Q, b \in \Delta^*$ such that

$$(q_1, a, b, q_2) \in \delta$$

An elementary example of a transducer that performs the *shift operator* σ on a infinite word (stream) is given in Figure 1. Transitions between states are made by following the arrow, where $a|a'$ takes input a and gives output a'. Let $a, b \in \Sigma^\infty$ be

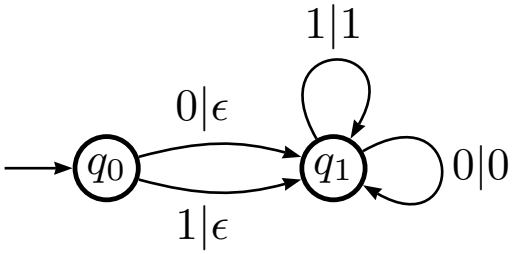

FIGURE 1. Example of a transducer. It follows that $a \triangleright \sigma(a)$ for any stream a.

single-sided streams over some alphabet Σ. Suppose that there is some transducer A such that $A(a) = b$, then this is indicated by the notation:

$$a \triangleright b$$

Suppose there are some transducers A, B such that $A(a) = b$ and $B(b) = a$. Then a, b are *equal under transducers* and share the same *degree*. This is indicated by:

$$a \diamond b$$

Then $\triangleright$ induces a partial ordering on streams, where $\diamond$ indicates equivalence between streams. Note that the alphabets Σ, Δ will be restricted to $\{0, 1\}$, as higher order alphabets do not change the partial ordering as mentioned by [EHK11]. The *degree of stream a* includes the sequences that a can both transduce to *and* which can

be transduced back to a as well. There are countable many different transducers, but there are uncountable many infinite words. So there are uncountably many degrees. Some classes of streams are closed under transducers:

- morphic streams, [AS03]
- eventually recurrent streams, [EHK11]

A special notion is reserved for degrees that include *every* transduction of *every* sequence in the degree, excluding any periodic sequences. A degree is called *prime* if for every sequence in the degree the output under a transducer (the *transduct*) can be transduced back to the original sequence, or is a periodic sequence. Finite sequences are counted as periodic. One of the main results in [EHK11] is the fact that the degree of Π is prime:

$$\Pi = 10100100010000100000\ldots$$

A proof is given in section 9 of this document. Eventually periodic and periodic sequences are transducer equivalent because transducers can add and remove arbitrarily long prefixes. The degree of (eventually) periodic sequences is also prime, as every sequence can be transduced to a periodic sequence. The degree of periodic sequences will be considered the *zero* or *trivial* degree.
Two other known structures exists in the transducer ordering of streams: A collection of streams $\{A_i\}^{i\in\mathbb{N}}$ is called an *infinite ascending chain* if for every i:

$$\begin{aligned} A_{i+1} &\triangleright A_i \\ A_i &\not\triangleright A_{i+1} \end{aligned}$$

Opposed to this is the notion of an *infinite descending chain* for a collection $\{D_i\}^{i\in\mathbb{N}}$ holds:

$$\begin{aligned} D_i &\triangleright D_{i+1} \\ D_{i+1} &\not\triangleright D_i \end{aligned}$$

Examples of such chains are given in [EHK11]. In Section 10 an infinite descending chain based on the classic Fibonacci sequence is constructed, which has not been mentioned before. The degree of the Thue-Morse sequence includes the period-doubling sequence v_2, and is conjectured to be prime by [EHK11]. The Thue-Morse word (TM) is the fixed point of substitution g, starting on zero:

$$\begin{aligned} g(0) &\to 01 \\ g(1) &\to 10 \\ TM &= g^\infty(0) \\ &= 0110100110010110\ldots \end{aligned}$$

The n-th letter of Thue-Morse is equal to the parity of the number of ones in the binary expansion of n, with even giving zero and odd a one. See lemma 3.3 for a proof. The degree of Thue-Morse includes the *period doubling* word v_2:

$$TM \diamond v_2$$

The period doubling word v_2 is the fixed point of the following substitution k:

$$\begin{aligned} k(0) &\to 01 \\ k(1) &\to 00 \\ v_2 &= k^\infty(0) \\ &= \{\max_b 2^b | n\}^{n\in\mathbb{N}} \\ &= \{v_2(n)\}^{n\in\mathbb{N}} \\ &= 0100010101000100\ldots \end{aligned}$$

The n-th letter of the period doubling sequence is given by the parity of the number of leading zeros in the binary expansion of n. A proof is given in lemma 3.4. The notation for v_2 is ambiguous: v_2 refers to the period doubling sequence but also to the function $v_2(n)$ which determines the *2-adic valuation* of n .

1.1. **Prime streams.** It is shown in [Ste08] that the degree of the Thue-Morse sequence includes all of its arithmetic subsequences. That is, for any integers $a, b > 0$ holds:

$$TM \diamond \{TM(a+bn)\}^{n\in\mathbb{N}}$$

For a proof see theorem 7.7. One would expect a similar result for the period doubling sequence, but this turns out to be difficult. For the period doubling sequence it is shown as an original result that for integers $a, b > 1$ with b odd:

$$\begin{aligned} v_2 &\triangleright \{v_2(a+bn)\}^{n\in\mathbb{N}} \\ \{v_2(a+bn)\}^{n\in\mathbb{N}} &\diamond \{v_2(2a+bn)\}^{n\in\mathbb{N}} \end{aligned}$$

The first line however, is true for *any* stream. The second line is proven in section 8. For b even, $\{v_2(a+bn)\}^{n\in\mathbb{N}}$ is either periodic or is transducer equivalent to b odd. At this moment it is not known if

$$\{v_2(a+bn)\}^{n\in\mathbb{N}} \triangleright v_2$$

holds in general. If this were not to be the case, then the degree of Thue-Morse would *not* be prime. If the affirmative would be the case, then this does not assert the prime property of Thue-Morse.

1.2. **Arithmetic subsequences.** Suppose Thue-Morse would be prime. One of the least implications would be that Thue-Morse, the period doubling sequence and all of the arithmetic subsequences would be transducer equivalent. To get a feeling of the objects involved, I tried making a (counter) example to the stated claims. I conjured that perhaps the morphisms associated with these sequences could be indicators for transducer equivalence. As little was known regarding the morphic structure of the mentioned sequences, I decided describing these morphisms would be a first step.
The arithmetic subsequences of the period doubling word v_2 exhibit a property that has not been mentioned in literature before. To understand the property, the following definition is necessary:

Definition 1.2. *The* rotated Toeplitz substitution ρ_{a,k^m} *is a substitution created by taking the iterated substitution* k^m *and rotating it by* a

$$\begin{aligned} k^m(c) &= d_0 d_1 \dots d_n \\ \rho_{1,k^m}(c) &= d_n d_0 \dots d_{n-1} \\ \rho_{a,k^m}(c) &= d_{n-a+1} \dots d_n d_0 \dots d_{n-a} \end{aligned}$$

where $c \in \{0,1\}$.

The rotated Toeplitz substitution is a substitution in its own right, as opposed to being an operator on the Toeplitz substitution. For example:

$$\begin{aligned} k^3(0) &= 01000101 \\ k^3(1) &= 01000100 \\ \rho_{1,k^3}(0) &= 10100010 \\ \rho_{1,k^3}(1) &= 00100010 \\ \rho_{2,k^3}(0) &= 01010001 \\ \rho_{2,k^3}(1) &= 00010001 \end{aligned}$$

Now fix some $m \in \mathbb{N}_{>0}$ and an integer $0 < q < 2^m - 1$. Then by theorem 3.8 $\{v_2(q + (2^m - 1)n)\}^{n\in\mathbb{N}}$ is the fixed point of

$$\rho_{q+1,k^m}$$

starting on $v_2(q)$. In other notation:

$$\begin{aligned} \rho^{\infty}_{q+1,k^m}(v_2(q)) &= \{v_2(q + (2^m - 1)n)\}^{n\in\mathbb{N}} \\ &= \{v_2(\frac{q}{(2^m - 1)} + n)\}^{n\in\mathbb{N}} \end{aligned}$$

For example, $\{v_2(\frac{1}{3} + n)\}^{n\in\mathbb{N}}$ is the fixed point of the substitution ρ_{2,k^2}:

$$\begin{aligned} k^2(0) &= 0100 \\ k^2(1) &= 0101 \\ \rho_{2,2^2}(0) &= 0001 \\ \rho_{2,2^2}(1) &= 0101 \\ \{v_2(1 + 3n)\}^{n\in\mathbb{N}} &= 000100010001010100010001000101010001000100010101 0 \ldots \end{aligned}$$

A 2-adic rational number other from zero may be associated with an arithmetic subsequence of the period doubling word. Zero is not included because $v_2(0)$ is not properly defined. Any non-zero 2-adic rational number

$$\frac{r}{2^l - 1}$$

has an associated Toepliz sequence:

$$\{v_2(\frac{r}{2^l - 1} + n)\}^{n\in\mathbb{N}}$$

The number 1 is associated with the period doubling sequence itself:

$$\begin{aligned} \{v_2(1 + n)\}^{n\in\mathbb{N}} &= \{v_2(n)\}^{n\in\mathbb{N}_{>0}} \\ &= k^{\infty}(0) \end{aligned}$$

For $r \geq 2^l - 1$, the associated sequence is a suffix of

$$\{v_2(\frac{r'}{(2^l - 1)} + n)\}^{n\in\mathbb{N}}$$

where $r' \in \{1, 2, \ldots, 2^l - 1\}$ such that $r' = r \mod 2^l - 1$. Applying a rotated Toeplitz substitution to an arithmetic subsequence of the period doubling word yields another arithmetic subsequence of the period doubling word. Theorem 4.1 shows that:

$$\rho_{q+1,2^m}(\{v_2(\frac{r}{2^l - 1} + n)\}^{n\in\mathbb{N}}) = \{v_2(\frac{2^m r}{2^l - 1} - q + n)\}^{n\in\mathbb{N}}$$

The rotated Toeplitz substitution theorems also apply to a generalization of v_2. Let p be a prime, then:

$$\begin{aligned} k_p(0) &= 0^{p-1}1 \\ k_p(1) &= 0^{p-1}0 \\ v_p &= k_p^{\infty}(0) \\ &= \{\max_b p^b | n \mod 2\}^{n\in\mathbb{N}_{>0}} \end{aligned}$$

For more details see section 3, section 4 and appendix A.

A more general connection between streams and morphisms is provided by a theorem of Cobham. It connects the notion of *k-automatic sequences* with fixed points of morphisms. Thue-Morse, the period doubling word and their arithmetic subsequences are themselves 2-automatic sequences:

Theorem 1.3 (Cobham)**.** *Let $k \geq 2$. Then a sequence is $k-$automatic if and only if it is the image, under a 1-uniform substitution, of a fixed point of a k-uniform substitution.*

For a proof see section 5. The morphisms usually run over higher-order alphabets and a *coding* on the fixed point of said morphism is needed to construct the automatic sequence. The arithmetic subsequences of v_p can be described as *proper* fixed points of rotated morphisms: No coding is needed. This specific property has not been described before. The connection between *operations* on p-adic numbers and applying morphisms on streams as described in theorem 4.1 is novel.
The arithmetic subsequences of *Thue-Morse* can also be described elegantly as *proper* fixed points of morphisms. I have not found the time to include this result in my thesis.
Further research on morphisms could consider general subsequences of streams. Morphic streams different from Thue-Morse and the period doubling sequence should also been considered. One problem would be: characterize streams that are transducer equivalent to all their arithmetic substreams.

2. On words

Let Σ be a finite collection of distinct objects, a so called *set*. Let Σ^m be the set of all finite sequences $a_1 \dots a_m$ with $a_j \in \Sigma$ for $1 \leq i \leq m$. Elements of Σ are called *letters*, and elements of Σ^m are called *words* over Σ of *length* m. Note: m is a natural number; $\Sigma^0 = \epsilon$, where ϵ is the *empty word* having no letters, and Σ^1 can be identified with Σ. The set Σ^m can be identified with the cartesian product $\Sigma \times \Sigma \times \ldots \times \Sigma$, but its elements are written without the usual commas and parentheses. Taken from [Chi09].

Definition 2.1. *Write*

$$\begin{aligned} \Sigma^+ &= \cup_{m \geq 1} \Sigma^m \\ \Sigma^* &= \cup_{m \geq 0} \Sigma^m \\ &= \Sigma^+ \cup \epsilon \end{aligned}$$

If $a = a_1 \dots a_m, b = b_1 \dots b_n \in \Sigma^*$, define $ab \in \Sigma^{m+n}$ to be $a_1 \dots a_m b_1 \dots b_m$. This gives a (closed) binary operation on Σ^* and Σ^+ called *concatenation*. It is associative: $a(bc) = (ab)c$ and $a\epsilon = \epsilon a = a$. Thus Σ^+ is a semigroup called the *free semigroup on* Σ. A semigroup with an identity element (word ϵ) is called a *monoid*. Σ^* is a monoid and is called the *free monoid over* Σ. Recall that a monoid such that every element has a inverse is called a *group*. Denote the length of a word a by $|a|$. Define a^n with $n \in \mathbb{N}$ by: $a^0 = \epsilon, a^{n+1} = a^n a$.
If a is a word over the set Σ (Σ is also called the *alphabet*), a factor of a is a word $c \in \Sigma^*$ such that $a = bcd$ for some $b, d \in \Sigma^*$. If $b = \epsilon$, then c is called a *prefix* of a. If $d = \epsilon$, then c is called a *suffix* of a. If $c \neq a$ then c is called a *pure factor*.
A *language* is any subset $L \subseteq \Sigma^*$. L is a collection of *words over the alphabet* Σ . For example, we may denote $L = \{1^n 0^n | n \in \mathbb{N}\}$. Then L is the language over the alphabet $\{0, 1\}$ containing all words that are constructed by taking ones followed by an equal amount of zeros.
Words may also be of infinite length: Suppose $a \in \Sigma^\infty$. Then a is a (single sided) *infinite word* or *stream* and note $a = a_0 a_1 a_2 \dots$ with $a_i \in \Sigma$ for $\forall i \in \mathbb{N}$.
Denote $s_k(n)$ the expansion of integer n over base k, with the most significant bit to the left. Each $s_k(n)$ starts with an 1 except for $s_k(0) = \epsilon$.
Suppose $a \in \{0, 1\}^*$ then $\overline{a}$ denotes the *conjugate* of a, given by interchanging every 1 with 0 and vice versa.

3. On Toeplitz substitutions

Definition 3.1. *A substitution is a map* $h : \Sigma^+ \to \Delta^*$ *such that*

$$h(ab) = h(a)h(b)$$

for all $a, b \in \Sigma^+$.

In computer science, substitutions are usually referred to as *morphisms*. For example see [AS03]. Suppose that for all $a \in \Sigma$ the length of the substitution is constant: $|h(a)| = k$ for some $k \in \mathbb{N}$. Then the substitution h is of *uniform length* k.

Definition 3.2. *A substitution is* prolongable on symbol $a \in \Sigma$ *if* $h : \Sigma^+ \to \Sigma^*$ *for some alphabet* Σ *and if for some* $a \in \Sigma$ *it holds that* $h(a) = ab$ *for some* $b \in \Sigma^*$ *such that* $h^n(b) \neq \epsilon$ *for all* $n \in \mathbb{N}$.

Denote the limit of a prolongable substitution by

$$\lim_{n\to\infty} h^n(a) \quad = \quad abh(b)h^2(b)\dots.$$

Note that $\lim_{n\to\infty} h^n(a)$ is a *fixed point* of prolongable substitution h, because

$$\begin{aligned} h(\lim_{n\to\infty} h^n(a)) &= h(abh(b)h^2(b)\dots.) \\ &= h(a)h(b)h^2(b)h^3(b)\dots. \\ &= abh(b)h^2(b)h^3(b)\dots. \\ &= \lim_{n\to\infty} h^n(a) \end{aligned}$$

Denote the *Thue-Morse substitution* $g : \{0,1\}^+ \to \{0,1\}^*$:

$$\begin{aligned} g(0) &\to 01 \\ g(1) &\to 10 \end{aligned}$$

Then iterate the substitution on starting word $w = 0$ in the following fashion:

$$\begin{aligned} g(0) &= 01 \\ g^2(0) &= 0110 \\ g^3(0) &= 01101001 \\ g^4(0) &= 0110100110010110 \\ \dots &\quad \dots \end{aligned}$$

The Thue-Morse substitution is prolongable on zero[2] and by taking the limit the infamous *Thue-Morse stream* is obtained:

$$\lim_{n\to\infty} g^n(0) = 0110100110010110100101100110100110010110011010 01\dots$$

The Thue-Morse stream was first implicitly described in a 1851 paper of Prouhet, and was later independently discovered most notably by Thue (1906) and Morse (1921) [AS99]. For historical reasons it is often called the Thue-Morse word. The Thue-Morse word has some amazing properties and shows up in seemingly unrelated subjects. It is an infinite non-periodic word that is *strongly cubefree*: There is no factor in Thue-Morse of the form x^2a where a is the first letter of the non-empty word x. For every factor that *does* occur in the Thue-Morse sequence however, it occurs infinitely often. There are several ways to construct the Thue-Morse word. For the sake of future discussion, we will exhibit one other way to construct the Thue-Morse sequence. For a more complete overview of the Thue-Morse word see [AS99, Ste08, Klo02].

[2]It is also prolongable on 1. The limit is equal to the conjugate of Thue-Morse.

Lemma 3.3. *Denote $s_2(n)$ the base-2 expansion of integer n. Let*

$$TM(n) = \{\textit{number of occurrences of 1 in } s_2(n)\} \mod 2$$

Then $\{TM(n)\}^{n\in\mathbb{N}_0}$ equals the Thue-Morse word.

Proof. Proof by induction. It is clear that

$$g^m(0) = TM(0), TM(1), \ldots, TM(2^m - 1)$$

is true for $m = 2$. Assume that for a given m the equation holds. It is to be shown the equation also holds for $m+1$. From the definition of the function $TM()$ follows that

$$\begin{aligned} TM(2n) &= TM(n) \\ TM(2n+1) &= 1 - TM(n) \end{aligned}$$

Compute $g^{m+1}(0)$ by using the induction hypothesis, expanding the substitution using the above equations:

$$\begin{aligned} g^{m+1}(0) &= g(g^m(0)) \\ &= g(TM(0), TM(1), \ldots, TM(2^m - 1)) \\ &= TM(0), 1 - TM(0), TM(1), 1 - TM(1), \ldots, TM(2^m - 1)), 1 - TM(2^m - 1)) \\ &= TM(0), TM(1), TM(2), TM(3), \ldots, TM(2^{m+1} - 2), TM(2^{m+1} - 1) \end{aligned}$$

□

It follows from a deep result of [Mah29] that $\xi = \sum_{n=0}^{\infty} TM(n)2^{-n}$ is a transcendental number. Later, this fact was proven in a more concise way by [Dek77]. It should be noted that the works of Mahler initially attracted the attention of a small group of specialists. Only relatively recently has his work become more popular. Currently, some of Mahler's results and conjectures are the focus of intense research.[3]

There is another substitution of uniform length 2 that shows similar properties to Thue-Morse, the Toeplitz substitution k:

$$\begin{aligned} k(0) &= 01 \\ k(1) &= 00 \end{aligned}$$

Then call $\lim_{n\to\infty} k^n(0)$ the *Toeplitz word*:

$$\lim_{n\to\infty} k^n(0) = 0100010101000100010001010100010101000101010001000 1\ldots$$

The Toeplitz word is also known as the *period doubling sequence*.

Lemma 3.4. *Define for $n \in \mathbb{N}_{>0}$:*

$$\begin{aligned} ord_2(n) &= \max\{b : 2^b | n\} \\ v_2(n) &= \{ord_2(n)\} \mod 2 \end{aligned}$$

Then $\{v_2(n)\}^{n\in\mathbb{N}_{>0}}$ denotes the period doubling word.

Proof. Proof by formal induction. It is clear that

$$k^m(0) = v_2(1), v_2(1), \ldots, v_2(2^m)$$

[3] *"... Mahler regretted that, apart from his own work, little interest had been shown by 20th century mathematicians in the study of arithmetical properties of decimal expansions."*, A J van der Poorten, Obituary of Kurt Mahler. The Mahler Lectureship for a visiting lectureship by the Australian Mathematical Society has since been awarded to Zagier, Hilton, Conway, Lenstra and Tao amongst others.

holds for $m = 2$ and $m = 1$. Assume the above holds for m and $m-1$. It shall be shown the equation also holds for $m+1$. From the definition of $v_2(n)$ follows that:

$$\begin{aligned} v_2(4n) &= v_2(n) \\ v_2(1+4n) &= v_2(1) = 0 \\ v_2(2+4n) &= v_2(2) = 1 \\ v_2(3+4n) &= v_2(3) = 0 \end{aligned}$$

In particular:

$$\begin{aligned} k^2(0) &= 0100 \\ k^2(1) &= 0101 \end{aligned}$$

Compute $k^{m+1}(0)$ by using the induction hypothesis and expanding the substitution. The final step follows from the relations $v_2(n) = v_2(4n)$ ect. as above.

$$\begin{aligned} k^{m+1}(0) &= k^2(k^{m-1}(0)) \\ &= k^2(v_2(1), \ldots, v_2(2^{m-1})) \\ &= 0, 1, 0, v_2(1), \ldots, 0, 1, 0, v_2(2^{m-1}) \\ &= v_2(1), v_2(2), v_2(3), v_2(4), \ldots, v_2(2^{m+1}-3), v_2(2^{m+1}-2), v_2(2^{m+1}-1), v_2(2^{m+1}) \end{aligned}$$

□

The here mentioned constructions of the Thue-Morse and the period doubling word have long been known in literature. A first generalization of the period doubling word was given by [JK69], who used a construction originally conceived by O. Toeplitz. The period doubling morphism can be generalized in two ways. The first way generalizes v_2 to v_p with p a prime. It is described at the end of this section and also in Appendix A. The second generalization gives rise to the *rotated Toeplitz substitutions* and I think this is a new result.[4] For starters, define the rotated substitution:

Definition 3.5 (rotated substitution)**.** *Let h be a prolongable substitution over alphabet Σ. For all $c \in \Sigma, m \in \mathbb{N}$ there is $n \in \mathbb{N}$ and $d \in \Sigma^*$ so that*

$$\begin{aligned} h^m(c) &= d \\ &= d_0 d_1 \ldots d_n \end{aligned}$$

The rotated substitution ρ_{i,h^m} *is a substitution created by taking substitution h^m and rotating it by i:*

$$\begin{aligned} \rho_{1,h^m}(c) &= d_n d_0 \ldots d_{n-1} \\ \rho_{i,h^m}(c) &= d_{n-i+1} \ldots d_n d_0 \ldots d_{n-i} \end{aligned}$$

For example:

$$\begin{aligned} k^3(0) &= 01000101 \\ k^3(1) &= 01000100 \\ \rho_{1,k^3}(0) &= 10100010 \\ \rho_{1,k^3}(1) &= 00100010 \\ \rho_{2,k^3}(0) &= 01010001 \\ \rho_{2,k^3}(1) &= 00010001 \end{aligned}$$

It should be stressed that ρ_{2,k^3} is a substitution in its own right as opposed to being an operator on another substitution.

[4]The mixed generalization, the *rotated Toeplitz substitution for p prime*, is also covered in Appendix A.

A theorem that shows a relation between the fixed point of a rotated Toeplitz substitution and the v_2 function can now be stated.

Lemma 3.6. *Then for any non-zero integer a such that $-2^m < a < 2^m$ and $n \in \mathbb{N}_{>0}$ it holds that:*

$$\begin{aligned} v_2(n2^m + a) &= v_2(n2^m - a) \\ &= v_2(|a|) \end{aligned}$$

Proof. Immediate. □

Lemma 3.7. *Let k denote the Toeplitz substitution over the $\{0,1\}$ alphabet. Let $0 < a < 2^m - 1$ with $a, m \in \mathbb{N}_{>0}$ fixed. Then*

$$\rho^{\infty}_{a+1,k^m}(v_2(a)) = \{v_2(a + n(2^m - 1))\}^{n\in\mathbb{N}}$$

Proof. First it is shown that the first 2^m symbols of both sides match. Secondly a recurrence relation will be proven to hold for both sides. Together with the first step this completes the proof.
Note that ρ_{a+1,k^m} is a substitution prolongable on $v_2(a)$. The following holds:

$$\begin{aligned} k^m(0) &= v_2(1), v_2(2), \ldots, v_2(2^m) \\ k^m(1) &= v_2(2^m + 1), v_2(2^m + 2), \ldots, v_2(2^{m+1}) \end{aligned}$$

It follows from lemma 3.6:

$$\begin{aligned} k^m(0) &= v_2(1), v_2(2), \ldots, v_2(2^m - 1), v_2(2^m) \\ k^m(1) &= v_2(2^m + 1), v_2(2^m + 2), \ldots, v_2(2^{m+1} - 1), v_2(2^{m+1}) \\ &= v_2(1), v_2(2), \ldots, v_2(2^m - 1), v_2(2^{m+1}) \\ k^m(v_2(a)) &= v_2(1), v_2(2), \ldots, v_2(2^m - 1), v_2(a2^m) \end{aligned}$$

Note that $k^m(v_2(a))$ only depends on a in the 2^m-th (final) entry. It is important for the second step of the proof to see that if m is even it holds $v_2(a2^m) = v_2(a)$ and likewise when m odd $v_2(a2^m) = 1 - v_2(a)$. Rotate $k^m(v_2(a))$ to obtain $\rho_{a+1,k^m}(v_2(a))$ and use lemma 3.6 to rewrite:

$$\begin{aligned} \rho_{a+1,k^m}(v_2(a)) &= v_2(2^m - a), \ldots, v_2(a2^m), v_2(1), \ldots, v_2(2^m - a - 1) \\ &= v_2(a), \ldots, v_2(a + a(2^m - 1)), v_2(a + (a + 1)(2^m - 1)), \ldots, v_2(a + (2^m - 1)(2^m - 1)) \end{aligned}$$

This completes the first step of the proof. Suppose m is even. Let $0 \leq c < 2^m - 1$ with $c \neq a$ and substitute $n = c + r2^m$ for $r \in \mathbb{N}$, then it follows by lemma 3.6 that:

$$\begin{aligned} v_2(a + (2^m - 1)n) &= v_2(a + (2^m - 1)(c + r2^m)) \\ &= v_2(a + c(2^m - 1) + (2^m - 1)r2^m) \\ &= v_2(a + c(2^m - 1)) \end{aligned}$$

Which shows that for $c \neq a$, the $c + 2^m r$-th entry is actually equal to the c-th entry of the sequence. Now suppose m is even and $c = a$:

$$\begin{aligned} v_2(a + (2^m - 1)n) &= v_2(a + (2^m - 1)(c + r2^m)) \\ &= v_2(a + (2^m - 1)(a + r2^m)) \\ &= v_2(2^m(a + (2^m - 1)r)) \\ &= v_2(a + (2^m - 1)r)) \end{aligned}$$

So for $c = a$, the $a + 2^m r$-th entry of $\{v_2(a + n(2^m - 1))\}^{n\in\mathbb{N}}$ is actually equal to the r-th entry of the sequence. Using m even the rotated substitution is given by:

$$\begin{aligned} \rho_{a+1,k^m}(0) &= v_2(2^m - a), \ldots, v_2(2^m - 1), 0, v_2(1), \ldots, v_2(2^m - a - 1) \\ \rho_{a+1,k^m}(1) &= v_2(2^m - a), \ldots, v_2(2^m - 1), 1, v_2(1), \ldots, v_2(2^m - a - 1) \end{aligned}$$

Conclude that indeed the r-th entry of the fixed point of the substitution is equal to its $a + 2^m r$-th entry. Both sequences have the same recurrence rules when m is even. Now consider m odd. As before with $c \neq a$:

$$v_2(a + (2^m - 1)n) = v_2(a + c(2^m - 1))$$

However for $c = a$ it follows:

$$\begin{aligned} v_2(a + (2^m - 1)n) &= v_2(a + (2^m - 1)(c + r2^m)) \\ &= v_2(a + (2^m - 1)(a + r2^m)) \\ &= v_2(2^m(a + (2^m - 1)r)) \\ &= 1 - v_2(a + (2^m - 1)r)) \end{aligned}$$

For m odd the rotated substitution is given by:

$$\begin{aligned} \rho_{a+1,k^m}(0) &= v_2(2^m - a), \ldots, v_2(2^m - 1), 1, v_2(1), \ldots, v_2(2^m - a - 1) \\ \rho_{a+1,k^m}(1) &= v_2(2^m - a), \ldots, v_2(2^m - 1), 0, v_2(1), \ldots, v_2(2^m - a - 1) \end{aligned}$$

The result follows. □

Theorem 3.8. *Let $a, b \in \mathbb{N}_{>0}$ such that $0 < a < b$ and b odd. Then there are some $c, m \in \mathbb{N}_{>0}$ such that*

$$\{v_2(a + bn)\}^{n \in \mathbb{N}} = \lim_{n \to \infty} \rho^n_{ac+1,k^m}(v_2(ac))$$

Proof. Suppose c, m are such that $cb = 2^m - 1$. Then c is odd and by lemma 3.7 write:

$$\begin{aligned} \{v_2(a + bn)\}^{n \in \mathbb{N}} &= \{v_2(ca + cbn)\}^{n \in \mathbb{N}} \\ &= \rho^\infty_{ac+1,k^m}(v_2(ac)) \end{aligned}$$

So now we show that there are some c, m such that $cb = 2^m - 1$. Suppose that $b \nmid 2^m - 1$ for all $m \in \mathbb{N}_{>0}$. Then by the pigeon hole principle there must be some $r, s, v \in \mathbb{N}_{>0}$ such that

$$\begin{aligned} 2^r - 1 &\equiv v \mod b \\ 2^{r+s} - 1 &\equiv v \mod b \end{aligned}$$

It follows that:

$$\begin{aligned} b &\mid 2^{r+s} - 2^r \\ b &\mid 2^r(2^s - 1) \\ b &\mid 2^s - 1 \end{aligned}$$

So then b does divide $2^s - 1$, which is a contradiction. □

There is an obvious generalization to Theorem 3.8. Define for p a prime:

$$\begin{aligned} k_p(0) &= 0^{p-1}1 \\ k_p(1) &= 0^{p-1}0 \\ v_p(n) &= \{\max_b p^b | n\} \mod 2 \end{aligned}$$

Then it follows in a very similar way that

$$\{v_p(n)\}^{n \in \mathbb{N}_{>0}} = k^\infty_p(0)$$

Equally, if $a, b \in \mathbb{N}_{>0}$ are such that $0 < a < b$ and $p \nmid b$, then there are some $c, m \in \mathbb{N}_{>0}$ such that

$$\{v_p(a + bn)\}^{n \in \mathbb{N}} = \rho^\infty_{ac+1,k^m_p}(v_p(ac))$$

The proof can be found the in appendix A, and has the same structure as the non-generalized version. As a corollary to theorem 3.8 and the deep results of [Mah29] and [Dek77] we find

Corollary 3.9. *Let $a, b \in \mathbb{N}_{>0}$ such that $0 < a < b$, $p \nmid b$ and p a prime. Then $\xi = \sum_{n=0}^{\infty} v_p(a+bn)p^{-n}$ is a transcendental number.*

Proof. This is a direct result by applying theorem A.5 and theorem 13.5.4 from p.393 [AS03]. □

4. On p-adic rationals

Consider the non-zero rational numbers of the 2-adic numbering system. Without loss of generality, such a rational number can be written as

$$\frac{q}{2^m - 1}$$

for some $q \in \mathbb{Z}\backslash\{0\}$ and $m \in \mathbb{N}_{>0}$. Every non-zero rational can be associated with a Toeplitz stream in the following fashion:

$$\{v_2(\frac{q}{2^m-1} + n)\}^{n\in\mathbb{N}} := \{v_2(q + n(2^m - 1))\}^{n\in\mathbb{N}}$$

For example, the number 1 is associated with the period doubling sequence:

$$\{v_2(1+n)\}^{n\in\mathbb{N}} = \{v_2(n)\}^{n\in\mathbb{N}_{>0}}$$

The 2-adic rational $\frac{1}{3}$ is associated with the following arithmetic subsequence of the period doubling sequence:

$$\{v_2(\frac{1}{3} + n)\}^{n\in\mathbb{N}} = \{v_2(1+3n)\}^{n\in\mathbb{N}}$$

It follows from section 3 that $\{v_2(\frac{1}{3}+n)\}^{n\in\mathbb{N}}$ is the fixed point of the substitution ρ_{2,k^2}. For $0 < q < (2^m - 1)$ it holds that $\{v_2(\frac{q}{2^m-1}+n)\}^{n\in\mathbb{N}}$ is the fixed point of ρ_{q+1,k^m} as proved in theorem 3.7. What happens if ρ_{q+1,k^m} is applied once to an arbitrarily arithmetic subsequence of v_2? It turns out that:

$$\rho_{q+1,k^m}(\{v_2(\frac{r}{2^l-1}+n)\}^{n\in\mathbb{N}}) = \{v_2(\frac{2^m r}{2^l-1} - q + n)\}^{n\in\mathbb{N}}$$

Note that applying ρ_{q+1,k^m} to the fixed point ρ^{∞}_{q+1,k^m} yields again the fixed point. The following theorem summarizes and extends the discussion up to this point.

Theorem 4.1. *Let ρ_{q+1,k^m} be a rotated Toeplitz substitution as noted in Section 3 with $0 < q < (2^m - 1)$ and $m \in \mathbb{N}_{>0}$. Then for every $r \in \mathbb{Z}\backslash\{0\}$ and $l \in \mathbb{N}_{>0}$ it holds*

$$\rho_{q+1,k^m}(\{v_2(\frac{r}{2^l-1}+n)\}^{n\in\mathbb{N}}) = \{v_2(\frac{2^m r}{2^l-1} - q + n)\}^{n\in\mathbb{N}}$$

Proof. It will be shown that both sides of the equation have the same recurrence rules. From the recurrence rules follows that the first 2^m letters of both sides are the same, completing the proof. But first recall from section 3 the construction of the rotated substitution ρ_{q+1,k^m}:

$$\begin{aligned}
k^m(0) &= v_2(1), v_2(2), v_2(3), \ldots v_2(2^m-1), v_2(2^m)\\
k^m(1) &= v_2(1), v_2(2), v_2(3), \ldots v_2(2^m-1), 1 - v_2(2^m)\\
\rho_{q+1,k^m}(0) &= v_2(2^m-q), v_2(2^m-q+1), \ldots, v_2(2^m), v_2(1), \ldots, v_2(2^m-q-1)\\
\rho_{q+1,k^m}(1) &= v_2(2^m-q), v_2(2^m-q+1), \ldots, 1 - v_2(2^m), v_2(1), \ldots, v_2(2^m-q-1)
\end{aligned}$$

Suppose m is even and consider $n = q + i2^m$ with $i \in \mathbb{N}$. It holds:

$$\begin{aligned} v_2(\frac{2^m r}{2^l - 1} - q + n) &= v_2(\frac{2^m r}{2^l - 1} - q + q + i2^m) \\ &= v_2(2^m(r + i(2^l - 1))) \\ &= v_2(r + i(2^l - 1)) \\ &= v_2(\frac{r}{2^l - 1} + i) \end{aligned}$$

Now consider the left hand side of the statement and denote

$$\rho_{q+1,k^m}(v_2(\frac{r}{2^l - 1} + i)) = \rho_0, \rho_1, \ldots, \rho_{2^m - 1}$$

then $\rho_q = v_2(\frac{r}{2^l - 1} + i)$. Now if m were odd, then $\rho_q = 1 - v_2(\frac{r}{2^l - 1} + i)$ and:

$$v_2(\frac{2^m r}{2^l - 1} - q + n) = 1 - v_2(\frac{r}{2^l - 1} + i)$$

So for $n = q + i2^m$ with $i \in \mathbb{N}$ the equation-to-prove holds. Now consider $n = s + i2^m$ with $s \neq q$ and $0 \leq s < 2^m$. It follows that $|s - q| < 2^m - 1$, which is to be used in the second to final step of the derivation below:

$$\begin{aligned} v_2(\frac{2^m r}{2^l - 1} - q + n) &= v_2(\frac{2^m r}{2^l - 1} - q + s + i2^m) \\ &= v_2(2^m r + (2^l - 1)(s - q + i2^m)) \\ &= v_2(2^m(r + (2^l - 1)i) + (2^l - 1)(s - q)) \\ &= v_2(|s - q|) \end{aligned}$$

It is clear that for $n = s + i2^m$ with $s \neq q$ and $0 \leq s < 2^m$ the expression is dependent on s and independent of i. In fact, the expansion of ρ_{q+1,k^m} given at the start of the proof now shows the two sides are equal. □

5. On automata

Definition 5.1. *An* automaton *(also called a* deterministic finite automaton with output (DFAO) *by [AS03]) is a sextuple* $A = (Q, \Sigma, \delta, q_0, \Delta, \tau)$ *where*

- Q *is a finite set of states*
- Σ *is a finite set, the input alphabet*
- $\delta \subseteq Q \times \Sigma \times Q$ *is a transition function*
- $q_0 \in Q$ *is the initial state*
- Δ *is a finite set, the output alphabet*
- $\tau : Q \to \Delta$ *is the output function*

Assume the transition relation to be deterministic: For all $q \in Q, a \in \Sigma$ *there exists a unique* $q' \in Q$ *such that*

$$\delta(q, a) = q'$$

Also assume the output function to be deterministic: For all $q \in Q$*, there is a unique* $d \in \Delta$ *such that*

$$\tau(q) = d$$

Suppose $q, q' \in Q, a \in \Sigma$. Then

$$\delta(q, a) = q'$$

describes the edge from state q to state q' that takes input a. The automaton starts in the initial state and receives a word as input. The symbols of the word are processed one at a time, such that the state of the machine shifts according to the δ function. For instance, let the word 101 be used as input for the automaton depicted in Figure 2. The sequence of the traversed states starts with initial state

q_0 and subsequent states visited would be q_1, q_1, q_0. Because $\tau(q_0) = 0$, the output would be 0. This is denoted by $A(101) = 0$.

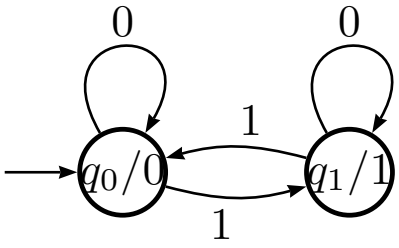

FIGURE 2. An automaton counting modulo 2 the number of ones

Definition 5.2. *Let $s_k(n)$ be the expansion of integer n in base k with the most significant digit first. A sequence $\{a_n\}^{n\in\mathbb{N}}$ is called a* k-automatic sequence *if there is some automaton A such that for all $n \in \mathbb{N}$ we have $a_n = A(s_k(n))$.*

Using the automaton from Figure 2, the Thue-Morse sequence may be constructed. So the Thue-Morse sequence is an 2-automatic sequence. The automaton generating the Toeplitz word is depicted in Figure 3. It is known that if instead $s_k(n)$ would start with the least significant digit first, this would not change the class of automatic sequences [AS03].

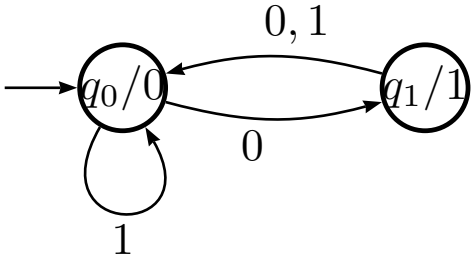

FIGURE 3. An automaton calculating $v_2(n)$ from $s_2(n)$.

Theorem 5.3 (Cobham)**.** *Let $k \geq 2$. Then a sequence u is k-automatic if and only if it is the image, under a 1-uniform substitution, of a fixed point of a k-uniform substitution.*

Proof. Proof using the same construction as [AS03], p 175.
$\Rightarrow$ Suppose $u = \tau(w)$ with $\tau : \Sigma \to \Sigma'$ a coding and $w = w_0 w_1 \ldots$ a fixed point of a k-uniform morphism $h : \Sigma^+ \to \Sigma^*$. Let $s_k(n)$ be as before. The claim is that some automaton $A = (Q, \Sigma, \delta, q_0, \Sigma', \tau)$ exists such that:

$$A(s_k(n)) = \tau(w_n) = u_n$$

Let $q_0 = w_0$ and let $Q = \Sigma$. Define the transition function for every $q \in \Sigma$ and $0 \leq b < k$:

$$\delta(q, b) = \text{the } b\text{-th letter of } h(q)$$

Use induction on the length $|s_k(n)|$. For $|s_k(n)| = 0$ the claim holds because $q_0 = w_0$. Suppose the claim holds for $|s_k(n)| = i$, then it will be shown to hold for $|s_k(n)| = i + 1$. Denote for some n the base-k expansion $s_k(n) = s_0 s_1 \ldots s_{i+1}$. In

particular, $n = kn' + s_{i+1}$ where $s_k(n') = s_0 s_1 \dots s_i$. By the induction step it holds that

$$\delta(q_0, s_0 s_1 \dots s_i) = w_{n'}$$

Because h is k-uniform, it holds that w_n is the s_{i+1}-th letter of $h(w_{n'})$. Then there follows:

$$\begin{aligned} A(s_k(n)) &= \tau(\delta(q_0, s_k(n))) \\ &= \tau(\delta(q_0, s_0 s_1 \dots s_{i+1})) \\ &= \tau(\delta(\delta(q_0, s_0 s_1 \dots s_i), s_{i+1})) \\ &= \tau(\delta(w_{n'}, s_{i+1})) \\ &= \tau(\text{the } s_{i+1}\text{-th symbol of } h(w_{n'})) \\ &= u_n \end{aligned}$$

$\Rightarrow$ Assume that u is a k-automatic sequence, so that $u_n = \tau(w_n) = \tau(\delta(q_0, s_k(n)))$ for some automaton $A = (Q, \Sigma_k, \delta, q_0, \Sigma', \tau)$. Assume $\delta(q_0, 0) = q_0$. Define the morphism $h : Q^+ \to Q^*$ in the following fashion:

$$h(q) = \delta(q, 0)\delta(q, 1) \dots \delta(q, k-1)$$

Then morphisms h is prolongable on q_0 with some fixed point w'. Then it is clear that $w'_n = \delta(q_0, s_k(n))$ holds for $|s_k(n)| \leq 1$. Assume $w'_n = \delta(q_0, s_k(n))$ holds for $|s_k(n)| \leq i$. Then showing it also holds for $|s_k(n)| \leq i+1$ completes the proof. Denote for some n the base-k expansion $s_k(n) = s_0 s_1 \dots s_{i+1}$. In particular there holds $n = kn' + s_{i+1}$ where $s_k(n') = s_0 s_1 \dots s_i$. So it follows:

$$\begin{aligned} \delta(q_0, s_k(n)) &= \delta(q_0, s_0 s_1 \dots s_{i+1}) \\ &= \delta(w'_{n'}, s_{i+1}(n)) \\ &= \text{the } s_{i+1}\text{-th letter of } h(w'_{n'}) \\ &= w_{kn'+s_{i+1}} \\ &= w_n \end{aligned}$$

□

Theorem 5.4. *Let h be a k-uniform substitution over alphabet $\Sigma_r = \{0, 1, \dots, r-1\}$ prolongable on 0. Denote*

$$\begin{aligned} h^\infty(0) &= w \\ &= w_0 w_1 w_2 \dots \end{aligned}$$

Then for any $a, b \in \mathbb{N}$ $\{w_{a+bn}\}^{n \in \mathbb{N}}$ is a k-automatic sequence.

Proof. This is a direct result from the definition of automatic sequences and lemma 6.6. An alternative using Cobhams theorem is:
Let $\Sigma_r^b = \{0, 1, \dots, r-1\}^b$ be an *alphabet* consisting of all words of length b over Σ_r. Construct the substitution $h^* : \{\Sigma_r^b\}^+ \to \{\Sigma_r^b\}^+$ of uniform length k in the following fashion: Let $d \in \Sigma_r^b$, then

$$h^*(d) = h(d)$$

Then h^* is prolongable on $w_0 w_1 \dots w_{b-1} \in \Sigma_r^b$ with fixed point w. Let the 1-uniform substitution (*coding*) $c : \Sigma_r^b \to \{0, 1, \dots, r\}$ be:

$$c(d) = \{ \text{ the } a\text{-th letter of } d\}$$

Then applying the coding to w indeed yields $\{w_{a+bn}\}^{n \in \mathbb{N}}$. □

An example of the above theorem is given: The 2-uniform prolongable substitution that has as fixed point $\{TM(1+3n)\}^{n\in\mathbb{N}}$ is given by:

$$\begin{aligned} g^*(A) &= AB \\ g^*(B) &= AC \\ g^*(C) &= BD \\ g^*(D) &= EC \\ g^*(E) &= FD \\ g^*(F) &= FE \\ c(A) &= 1, \quad c(B)=1, \quad c(C)=0 \\ c(D) &= 1, \quad c(E)=0, \quad c(F)=0 \\ g^\infty(A) &= ABACABBDABACACEC\ldots \\ TM(1+3n) &= 1110111111101000\ldots \end{aligned}$$

The above substitution can be derived by expanding the Thue-Morse word in blocks of three digits: The 2-uniform substitution for Thue-Morse is:

$$\begin{aligned} g(0) &= 01 \\ g(1) &= 10 \end{aligned}$$

From this the substitution $g^*()$ over the alphabet of words of length 3 and coding $c()$ follows:

$$\begin{aligned} g^*(\boxed{011}) &= \boxed{011}\boxed{010} \\ g^*(\boxed{010}) &= \boxed{011}\boxed{001} \\ g^*(\boxed{001}) &= \boxed{010}\boxed{110} \\ g^*(\boxed{110}) &= \boxed{101}\boxed{001} \\ g^*(\boxed{101}) &= \boxed{100}\boxed{110} \\ g^*(\boxed{100}) &= \boxed{100}\boxed{101} \\ c(\boxed{011}) &= 1, \quad c(\boxed{010})=1, \quad c(\boxed{001})=0 \\ c(\boxed{110}) &= 1, \quad c(\boxed{101})=0, \quad c(\boxed{100})=0 \end{aligned}$$

Substitution g^* is prolongable on $\boxed{011}$ and has fixed point TM:

TM :	011	010	011	001	011	010	010	110	011	010	011	001	011	001	101	001	…
$TM(1+3n)$:	1	1	1	0	1	1	1	1	1	1	1	0	1	0	0	1	…

6. On transducers

Definition 6.1. *A finite state transducer (which will be abbreviated to FST or transducer) is a sextuple $T=(Q,\Sigma,\delta,q_0,\Delta,\lambda)$ where*

- *Q is a finite set of states*
- *Σ is a finite set, the input alphabet*
- *$\delta \subseteq Q\times\Sigma\times\Delta^*\times Q$ is a transition relation*
- *$q_0\in Q$ is the initial state*
- *Δ is a finite set, the output alphabet*
- *$\lambda : Q\times\Sigma\to\Delta^*$ is the output function*

Suppose $q_1,q_2\in Q, a\in\Sigma, b\in\Delta^*$ then

$$(q_1,a,b,q_2)\in\delta$$

describes the edge from state q_1 to state q_2 that takes input a and generates output b. Assume the transition relation to be *deterministic*: For all $q_1 \in Q, a \in \Sigma$ there is exactly one $q_2 \in Q, b \in \Delta^*$ such that

$$(q_1, a, b, q_2) \in \delta$$

Now define the *extended transition relation* $\delta^* \subseteq Q \times \Delta^* \times \Gamma^* \times Q$ to be the closure of δ:

- $\delta \subseteq \delta^*$
- $(q, \epsilon, \epsilon, q) \in \delta^*$ for all $q \in Q$
- If $(q_1, a, b, q_2) \in \delta^*$ and $(q_2, c, d, q_3) \in \delta$ then $(q_1, ac, bd, q_3) \in \delta^*$.

Then it is clear that δ^* contains all the possible paths on the graph of A. Also define the *transition function* $\delta : Q \times \Sigma \to Q$ and the *output function* $\lambda : Q \times \Sigma \to \Delta^*$:

$$\delta(q_1, a) = q_2 \text{ and } \lambda(q_1, a) = b \text{ if and only if } (q_1, a, b, q_2) \in \delta^*$$

It should be clear from the notation that $\delta(.)$ denotes the function and δ, δ^* the sets. For transducer T note that $T(a) = b$ if $\lambda(q_0, a) = b$.
Let L be a language. Then

$$\begin{aligned} T(L) &= \{T(a) | a \in L\} \\ T^{-1}(L) &= \{a \in \Sigma^* | T(a) \in L\} \end{aligned}$$

Lemma 6.2. *There are countably many finite state transducers.*

Proof. The transducer is entirely determined by the transition relation δ. In other notation, denote

$$F : Q \times \Sigma \times Q \to \Delta^*$$

such that $F(q_1, a, q_2) = b$ if and only if:

$$(q_1, a, b, q_2) \in \delta$$

Then the number of such functions F is countable. □

Lemma 6.3. *Let $a, b \in \Sigma^+$ and T a finite state transducer. Then $T(a)$ is a prefix of $T(ab)$.*

Proof. By the deterministic property it follows that the path defined by input ab on T is unique: There is no other path on T that takes input ab. Then it should be clear that $T(a)$ is a prefix of $T(ab)$. □

Definition 6.4. *Denote $a, b \in \Sigma^\infty$ and T a finite state transducer. Denote $T(a) = b$ if and only if for every prefix b' of b there is a prefix a' of a such that b' is a prefix of $T(a')$.*

Lemma 6.5. *Suppose T is a transducer with output language Σ' and R a transducer with input language Σ', and let $w \subseteq \Sigma^*$ be a word. Then there exists a transducer T' such that $R(T(w)) = T'(w)$.*

Proof. Let

$$\begin{aligned} T &= (Q, q_0, \Sigma, \Sigma', \delta, \lambda) \\ R &= (Q', q_0', \Sigma', \Sigma'', \delta', \lambda') \\ T' &= (Q'', q_0'', \Sigma, \Sigma'', \delta'', \lambda'') \end{aligned}$$

with $Q'' = Q \times Q'$ and $q_0'' = [q_o, q_0']$. Define for all $a \in \Sigma, [q, q'] \in Q''$:

$$([q, q'], a, \lambda'(q, \lambda(q', a)), [\delta(q, a), \delta(q', \lambda(a))]) \in \delta''$$

This shows T' is a properly defined deterministic FST. Take $a \in \Sigma$ and note that:

$$R(T(a)) = R(\lambda(q_0, a)) = \lambda'(q_0', \lambda(q_0, a)) = T'(a)$$

By induction on the length of w it can be shown that for all $w \in \Sigma^*$ it holds $R(T(w)) = T'(w)$. □

The construction used in the proof of lemma 6.5 is called *wreath product* in [AS03].

Lemma 6.6. *Let T be a FST with input and output language $\Sigma_k = \{0, 1, \ldots, k-1\}$. Let A be an automaton with input language Σ_k. Then there is an automaton A' such that for all $w \in \Sigma_k^*$ it holds that $A(T(w)) = A'(w)$.*

Proof. The idea here is to create a automaton A' that has state space $Q \times Q'$ and emulates the path of both T and A at the same time. Let

$$\begin{aligned} T &= (Q, \Sigma_k, \delta, q_0, \Sigma_k, \lambda) \\ A &= (Q', \Sigma_k, \delta', q_0', \Delta, \tau) \\ A' &= (Q \times Q', \Sigma_k, \delta'', [q_0, q_0'], \delta'', \tau') \end{aligned}$$

be an automaton such that for all $a \in \Sigma_k$ and $[q, q'] \in Q \times Q'$ we have

$$\begin{aligned} ([q, q'], a, [\delta(a), \delta'(T(a)) &\in \delta'' \\ \tau'([q, q']) &= \tau(\delta'(q')) \end{aligned}$$

On input w the automaton A' emulates the FST T and automaton A simultaneously:

$$\begin{aligned} A'(w) &= \tau'([q_0, q_0'], T(w)) \\ &= \tau(\delta'(q_0', T(w))) \\ &= A(T(w)) \end{aligned}$$

□

7. Equivalence of sequences under transducers

Definition 7.1. *Let a, b be streams. Then*

$$a \triangleright b$$

if there exists a transducer such that $A(a) = b$. Clearly, this defines a partial ordering. Denote

$$a \diamond b$$

if and only if $a \triangleright b$ and $b \triangleright a$.

Lemma 7.2 (Transitivity)**.** *Let $a, b, c \in \Sigma^\infty$ and suppose $a \triangleright b$, $b \triangleright c$. Then $a \triangleright c$. Alternatively, suppose $a \diamond b$, $b \diamond c$. Then $a \diamond c$.*

Proof. Follows from lemma 6.5 □

Lemma 7.3. *$: \diamond$ defines an* equivalence relation*.*

Proof. By using the identity transducer for any stream a there holds $a \diamond a$, so reflexivity is satisfied. Symmetry is also satisfied: for any streams a, b by definition $a \diamond b$ implies $b \triangleright a$ and $a \triangleright b$. By the previous lemma there is also transitivity. □

Theorem 7.4. *There is an uncountable number of degrees of streams.*

Proof. There are uncountable many infinite sequences over $\{0, 1\}$. There are countably many transducers. Therefore, the number of degrees is uncountable. □

Lemma 7.5. *Let $a \in \Sigma^\infty$ and let σ be the* shift operator*: If $a = a_0 a_1 a_2 \ldots$ then $\sigma a = a_1 a_2 a_3 \ldots$. For all $n \in \mathbb{N}$ there holds:*

$$\sigma^n a \diamond a$$

Proof. $\triangleright$: For some $w \in \Sigma^n$ we have that $a = w\sigma^n a$. Build a transducer such the output for the first letter is w followed by the letter itself. For every subsequent letter read as input, the transducer outputs the same letter. $\triangleleft$: We may build an transducer that outputs ϵ for the first n symbols read. For every subsequent letter read as input, the transducer outputs the same letter. □

Lemma 7.6. *Let $a, b \in \{0, 1, \ldots, k\}^\infty$ be streams, with b purely periodic. Let $c = \{a_n + b_n \mod k\}^{n \in \mathbb{N}}$. Then*

$$a \diamond c$$

Proof. Let the period of b be $p \in \mathbb{N}_{>0}$. Then a transducer with exactly p states can be constructed such that the value of $a_n + b_n \mod k$ is computed correctly. In particular for $0 \leq i < p$ and $j \in \{0, 1, \ldots k-1\}$:

$$\begin{aligned} T &= (Q, \{0, 1, \ldots k-1\}, \delta, q_0, \{0, 1, \ldots k-1\}, \lambda) \\ Q &= \{q_0, q_1, \ldots, q_{p-1}\} \\ (q_i, j, b_i + j \mod k, q_{i+1}) &\in \delta \\ (q_{p-1}, j, b_{p-1} + j \mod k, q_0) &\in \delta \end{aligned}$$

This shows that $T(a) = c$. In a similar fashion a transducer R can be constructed such that $R(c) = a$. □

Some examples of transducers now follow. Figure 4 and 5 show that $TM \diamond v_2$.

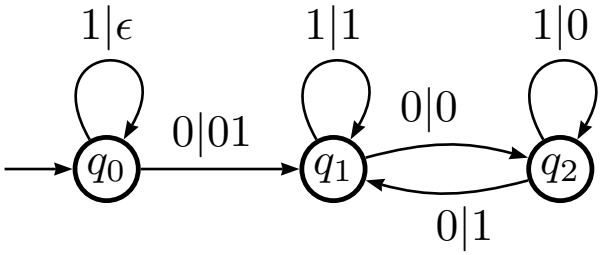

FIGURE 4. $v_2 \triangleright TM$

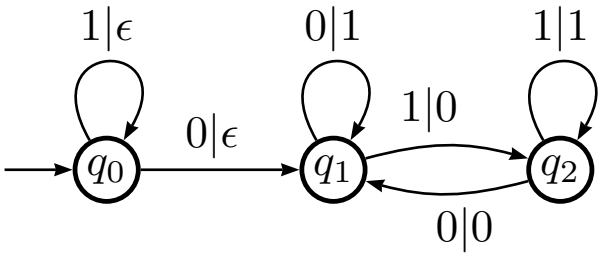

FIGURE 5. $TM \triangleright v_2$

Denote $\{f_i | f_i \in \mathbb{N}_{>0}\}^{i \in \mathbb{N}_0}$ a positive integer-valued sequence. From Figure 6 and 7 it is clear that:

$$1^{f_0}0^{f_1}1^{f_2}0^{f_3}1^{f_4} \ldots \diamond 10^{f_0}10^{f_1}10^{f_2}10^{f_3}10^{f_4} \ldots$$

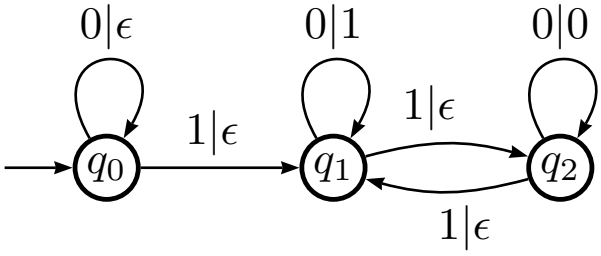

FIGURE 6. $10^{f_0}10^{f_1}10^{f_2}10^{f_3} \triangleright 1^{f_0}0^{f_1}1^{f_2}0^{f_3}$

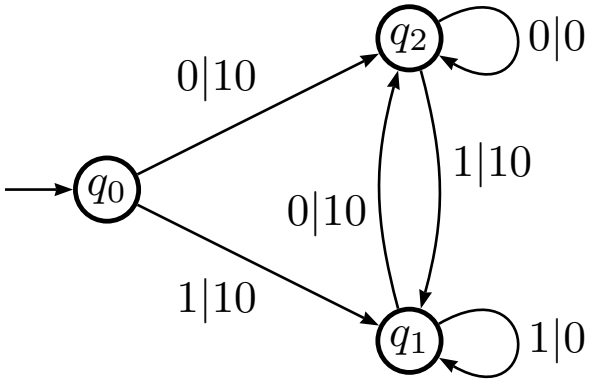

FIGURE 7. $10^{f_0}10^{f_1}10^{f_2}10^{f_3} \triangleleft 1^{f_0}0^{f_1}1^{f_2}0^{f_3}$

This illustrates that the sequence $f_0 f_1 f_2 \ldots$ is more significant than the words $w_0^{f_0} w_1^{f_1} w_2^{f_2} \ldots$. A transducer can also do simple arithmetic. For example: denote by $f_i \mod 3$ the remainder of f_i after devision by 3. Figure 8 shows how an transducer can do this calculation. It is a result first published in [Ste08] that

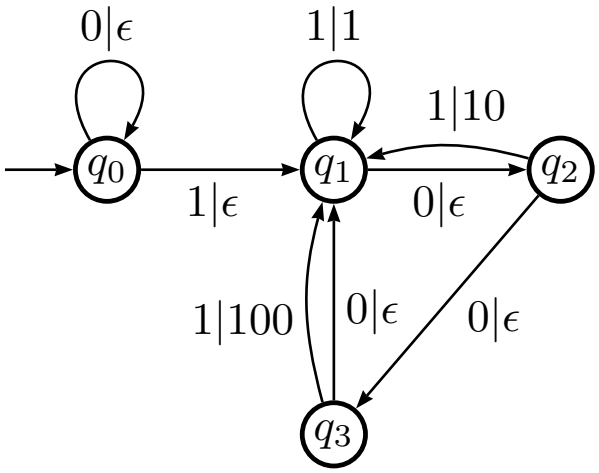

FIGURE 8. $10^{f_0}10^{f_1}10^{f_2}10^{f_3} \triangleright 10^{f_0 \mod 3}10^{f_1 \mod 3}10^{f_2 \mod 3}10^{f_3 \mod 3}$

every subsequence of TM is equivalent to the TM. Recall one of the definitions of the Thue-Morse sequence:

$$TM = \{ \text{ number of occurrences of 1 in } s_2(n)\}^{n\in\mathbb{N}}$$

Theorem 7.7. *Suppose $a, b \in \mathbb{N}_{>0}$ and $0 \leq a < b$. Then:*

$$\{TM(a+bn)\}^{n\in\mathbb{N}} \diamond TM$$

Proof. The $\triangleleft$ implication can be shown by a transducer that outputs every b th input, and omits everything else. The $\triangleright$ implication can be shown using another transducer. First note that the Thue-Morse sequence is the fixed point of substitution g:

$$\begin{aligned} g(0) &= 01 \\ g(1) &= 10 \end{aligned}$$

It is also the fixed point of g^m for any $m \in \mathbb{N}_{>0}$. Note that $g^m(0)$ is the conjugate of $g^m(1)$:

$$g^m(0) = \overline{g^m(1)}$$

The Thue-Morse sequence may be written in the following fashion:

$$TM = g^m(TM(0))g^m(TM(1))g^m(TM(2))\dots$$

In other words, the Thue-Morse word is a concatenation of the words $g^m(0)$ and $g^m(1)$. Suppose there is a mystery word x that can either be $g^m(0)$ or $g^m(1)$. Suppose the qth letter of mystery word x is known. Then the mystery word x is known: Because $g^m(0) = \overline{g^m(1)}$, there is only one candidate that will have a matching letter in the q-th position. Choose m such that there is a letter known for every block of length 2^m. Because $a+bn$ is periodic modulo 2^m, a finite transducer can be build that reconstructs the Thue-Morse sequence from one of its arithmetic subsequences:
Fix $m \in \mathbb{N}_{>0}$ to be the smallest m such that $b \leq 2^m$. Define

$$c_n = a + bn \mod 2^m$$

so that $0 \leq c_n < 2^m$ for all $n \in \mathbb{N}$. Then $\{c_n\}^{n\in\mathbb{N}}$ is periodic with period p, say. Let the transducer $T = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ have p states: $q_0, q_1, \dots, q_{p-1}$. The initial state is q_0. The transducer reads one letter of input at a time, and then moves to the next state regardless of output. For all j such that $c_j \geq c_{j+1}$ define the transition

$$\begin{aligned} (q_j, a, x, q_{j+1}) &\in \delta \\ x &= \begin{cases} g^m(0) & \text{if the } c_j\text{-th letter of } g^m(0) \text{ is equal to a} \\ g^m(1) & \text{if the } c_j\text{-th letter of } g^m(0) \text{ is } \textit{not} \text{ equal to a} \end{cases} \end{aligned}$$

Now suppose that $c_j < c_{j+1}$ for some j. Then this would mean that the same block would be outputted twice. So for this case define the transition:

$$(q_j, a, \epsilon, q_{j+1}) \in \delta$$

Regardless of the input, with period p the transducer returns to state q_0. $\square$

8. On Toeplitz equivalence classes

The purpose of this section is to investigate the equivalence under transducers of the arithmetic subsequences of the Toeplitz word. In other words, the equivalence of sequences given by $\{v_2(a+bn)\}^{n\in\mathbb{N}}$ is explored. As a final result the results are generalized to $\{v_p(a+bn)\}^{n\in\mathbb{N}}$ with p a prime.

Lemma 8.1. *Let $b \in \mathbb{N}_{>0}$. Then*

$$\{v_2(bn)\}^{n\in\mathbb{N}_{>0}} \diamond v_2$$

Proof. By the base-2 expansion of bn it follows that:

$$\{v_2(bn)\}^{n\in\mathbb{N}_{>0}} = \{v_2(b) + v_2(n) \mod 2\}^{n\in\mathbb{N}_{>0}}$$

The result follows by lemma 7.6. □

Lemma 8.2. *Let $a, b \in \mathbb{N}_{>0}$ such that $0 < a < b$ with a odd and b even. There holds:*

$$\begin{aligned}\{v_2(a+bn)\}^{n\in\mathbb{N}} &= 0^\infty \\ &= 00000000000..\end{aligned}$$

Proof. For the given a, b it follows that $a + bn$ is odd for all $n \in \mathbb{N}$. □

Suppose $a, b \in \mathbb{N}_{>0}$ such that $0 < a < b$ with a even and b even. Then there is some $r \in \mathbb{N}_{>0}$ such that

$$\begin{aligned}a &= 2^r a^* \\ b &= 2^r b^*\end{aligned}$$

with $a^*, b^* \in \mathbb{N}$ appropriately. There holds:

$$\{v_2(a+bn)\}^{n\in\mathbb{N}} = \{v_2(2^r) + v_2(a^* + b^* n)\}^{n\in\mathbb{N}}$$

So there follows $\{v_2(a+bn)\}^{n\in\mathbb{N}} \diamond \{v_2(a^* + b^* n)\}^{n\in\mathbb{N}}$.

Theorem 8.3. *Let $a, b \in \mathbb{N}_{>0}$ with $a > 0$ and b odd. It holds:*

$$\{v_2(a+bn)\}^{n\in\mathbb{N}} \diamond \{v_2(2a+bn)\}^{n\in\mathbb{N}}$$

Proof. $\triangleleft$: The first step below follows from creating a transducer that skips every second letter. The second step follows from the definition of v_2:

$$\begin{aligned}\{v_2(2a+bn)\}^{n\in\mathbb{N}} &\triangleright \{v_2(2a+2bn)\}^{n\in\mathbb{N}} \\ &\triangleright \{v_2(a+bn)\}^{n\in\mathbb{N}}\end{aligned}$$

$\triangleright$: Because b is odd, there is some $c, k \in \mathbb{N}$ such that $cb = 2^k - 1$ (see proof of theorem 3.8). It holds:

$$\begin{aligned}\{v_2(a+bn)\}^{n\in\mathbb{N}} &\triangleright \{v_2(a+cba+bn)\}^{n\in\mathbb{N}} \\ &\triangleright \{v_2(2^k a+bn)\}^{n\in\mathbb{N}} \\ &\triangleright \{v_2(2a+bn)\}^{n\in\mathbb{N}}\end{aligned}$$

□

Theorem 8.4. *Let $a, b \in \mathbb{N}_{>0}$ with $0 < a < b$. Let p be a prime and $p \nmid b$, then:*

$$\{v_p(a+bn)\}^{n\in\mathbb{N}} \diamond \{v_p(pa+bn)\}^{n\in\mathbb{N}}$$

Proof. The proof of theorem 8.3 may be generalized to p prime in an obvious way. □

9. A prime sequence

The goal of this section is to show that

$$\Pi = 1010^2 10^3 10^4 10^5 10^6 \dots$$

is a prime sequence, which is one of the main results by [EHK11]. By theorem 9.7 it holds that for any transducer T

$$T(\Pi) = w \prod_{i=0}^{\infty} \prod_{j=0}^{n} c_j d_j^i$$

where w, c_j, d_j are transducer-specific words and n a transducer-specific integer. To show that Π is a prime sequence, it has to be shown that either

$$w \prod_{i=0}^{\infty} \prod_{j=0}^{n} c_j d_j^i \triangleright \Pi$$

or the left-hand side of the above equation is periodic. Lemma 9.9 and lemma 9.10 show that in fact for any transducer T with non-eventually periodic output

$$T(\Pi) = w \prod_{i=0}^{\infty} \prod_{j=0}^{n} c_j d_j^i$$

theorem 9.8 can be applied. So Π is indeed prime.

Definition 9.1. *A sequence $\{f_i\}^{i\in\mathbb{N}}$ with $f_i \in \mathbb{N}$ is called* periodic *if there exists a $p \in \mathbb{N}$ such that $\forall i, n \in \mathbb{N}$ it holds $f_{i+np} = f_i$. A sequence is* eventually periodic *if some suffix of the sequence is periodic. A sequence f is said to be* periodic modulo every positive integer *if $\{f_i \mod n\}^{i\in\mathbb{N}}$ is periodic for every $n \in \mathbb{N}_{>0}$.*

Lemma 9.2. *For any $a, b \in \mathbb{N}$ the sequences $f = \{a + bi\}^{i\in\mathbb{N}}$ and $g = \{ab^i\}^{i\in\mathbb{N}}$ are eventually periodic modulo every integer.*

Proof. Periodicity modulo m will be demonstrated. Denote:

$$\begin{aligned} f_0 &= a \\ G(x) &= x + b \mod m \\ f_{i+1} &= G(f_i) \end{aligned}$$

Then f_i is just an iteration of map G on a finite set. There exists an $I \in \mathbb{N}_{>0}$ and a $p \in \mathbb{N}_{>0}$ such that $f_I \equiv f_{I+p} \mod m$. So f is eventually periodic. The same argument can also be given for g. □

It can be shown that f is purely periodic modulo every integer: Fix $I > 0$, then there follows:

$$\begin{aligned} f_I &\equiv f_{I+p} \mod m \\ f_I - bI &\equiv f_{I+p} - bI \mod m \\ f_0 &\equiv f_p \mod m \end{aligned}$$

This property is in contrast with g, which is *not* necessarily purely periodic. A similar argument shows that the Fibonacci sequence is also periodic modulo every positive integer, due to [Wal60]. The problem of classifying all sequences that are periodic modulo every integer has been studied in [DP08, PZ82].

Definition 9.3 (zero-loop)**.** *For any state $q_i \in Q$ of a finite state transducer define its* zero-loop*. The* zero-loop *is the path traversed when starting in state q_i and feeding only zeros as input. Because there are only finitely many states, such a path will return to a previously visited state eventually. In other words, the sequence $\{\delta(q_i, 0^j)\}^{j\in\mathbb{N}}$ is eventually periodic with period $n \in \mathbb{N}$. Denote the first state to*

be visited twice as state q. The sequence $\{q, \delta(q,0), \ldots, \delta(q,0^{n-1})\}$ is the zero-loop *belonging to q_i.*

Lemma 9.4 (Pumping lemma for transducers)**.** *Denote $|Q|$ the number of states in given transducer $T = (Q, \{0,1\}, \delta, q_0, \Delta, \lambda)$. Denote Z the least common multiple of the periods of all zero-loops in T. Let $\{f_i\}^{i\in\mathbb{N}}$ be a sequence that is periodic modulo Z with period p such that for all $i, j \in \mathbb{N}$ $f_{i+jp} \geq f_i$. Let there be a $N \in \mathbb{N}$ such that for all $m > N$ it holds $f_m > |Q|$. Then for every m and $q \in Q$ fixed there exists another fixed state $q' \in Q$ and words $c, d \in \Delta^*$ so that for every $i \in \mathbb{N}$ holds:*

$$(q, 10^{f_{m+ip}}, cd^{a_i}, q') \in \delta^*$$

with

$$a_i = \frac{f_{m+ip} - f_m}{Z}$$

Proof. Consider the statement-to-prove for $i = 0$ at first. State q accepts input 10^{f_m}, so there must be $q' \in Q, c \in \Delta^*$ such that

$$(q, 10^{f_m}, c, q') \in \delta^*.$$

By the pigeonhole principle it holds that there must occur a zero-loop for the input 10^{f_m}, because $f_m > |Q|$. The state q' as defined above must be part of this zero-loop. Then there must be some $d \in \Delta^*$ such that

$$(q', 0^Z, d, q') \in \delta^*$$

Then it is clear that

$$(q, 10^{f_{m+ip}}, cd^{a_i}, q') \in \delta^*$$

□

Lemma 9.5. *Denote $|Q|$ the number of states in given transducer $T = (Q, \{0,1\}, \delta, q_0, \Delta, \lambda)$. Let Z be the least common multiple of the periods of all zero-loops in T. Let $\{f_i\}^{i\in\mathbb{N}}$ be a sequence that is periodic modulo Z with period p such that for all $i, j \in \mathbb{N}$ $f_{i+jp} \geq f_i$. Let there be an $N \in \mathbb{N}$ such that for all $m > N$ it holds $f_m > |Q|$. Then for every m and $q \in Q$ fixed there is another fixed state $q' \in Q$ and words $c_j, d_j \in \Delta^*$ so that for all $i \in \mathbb{N}$*

$$(q, 10^{f_{m+ip+1}} 10^{f_{m+ip+2}} \cdots 10^{f_{m+(i+1)p}}, \prod_{j=1}^{p} c_j d_j^{a_{i,j}}, q') \in \delta^*$$

where

$$a_{i,j} = \frac{f_{m+j+ip} - f_{m+j}}{Z}$$

Proof. Begin by applying lemma 9.4 for state q and input $10^{f_{m+ip+1}}$. We now have

$$(q, 10^{f_{m+1+ip}}, c_1 d_1^{a_{i,1}}, q_1) \in \delta^*$$

Equally for state q_1 there will be some state q_2 so that there holds:

$$(q_1, 10^{f_{m+2+ip}}, c_2 d_2^{a_{i,2}}, q_2) \in \delta^*$$

Combining these findings it follows that:

$$(q, 10^{f_{m+1+ip}} 10^{f_{m+2+ip}}, c_1 d_1^{a_{i,1}} c_2 d_2^{a_{i,2}}, q_2) \in \delta^*$$

Or one more term:

$$(q, 10^{f_{m+1+ip}} 10^{f_{m+2+ip}} 10^{f_{m+3+ip}}, c_1 d_1^{a_{i,1}} c_2 d_2^{a_{i,2}} c_3 d_3^{a_{i,3}}, q_3) \in \delta^*$$

Continue all up to $1, 2, 3, \ldots, p$ such that for a certain q' there holds:

$$(q, 10^{f_{m+ip+1}} 10^{f_{m+ip+2}} \cdots 10^{f_{m+(i+1)p}}, \prod_{j=1}^{p} c_j d_j^{a_{i,j}}, q') \in \delta^*$$

This proves the lemma. □

Lemma 9.6. *Denote $|Q|$ the number of states in a given transducer $T = (Q, \{0,1\}, \delta, q_0, \Delta, \lambda)$. Let Z be the least common multiple of the periods of all zero-loops in T. Let $\{f_i\}^{i \in \mathbb{N}}$ be a sequence that is eventually periodic modulo Z with period p such that for all $i, j \in \mathbb{N}$ $f_{i+jp} \geq f_i$. Let there be an $N \in \mathbb{N}$ such that for all $m > N$ it holds $f_m > |Q|$ and $\{f_n\}^{n \geq m}$ periodic. Then there must be some m and $q \in Q$ with words $w, c_j, d_j \in \Delta^*$ and $k \in \mathbb{N}_{>0}$ such that there holds for all $i \in \mathbb{N}$*

$$(q_0, 10^{f_0} 10^{f_1} 10^{f_2} \cdots 10^{f_m}, w, q) \in \delta^*$$

$$(q, 10^{f_{m+1+ikp}} 10^{f_{m+2+ikp}} \cdots 10^{f_{m+(i+1)kp}}, \prod_{j=1}^{kp} c_j d_j^{a_{i,j}}, q) \in \delta^*$$

with

$$a_{i,j} = \frac{f_{m+j+ip} - f_{m+j}}{Z}$$

Proof. Fix some m as described above. There is some $q^* \in Q$ such that:

$$(q_0, 10^{f_0} 10^{f_1} 10^{f_2} \cdots 10^{f_m}, w, q^*) \in \delta^*$$

Define a function $t : Q \to Q$ such that $t(q) = q'$ if and only if:

$$(q, 10^{f_{m+1+ip}} 10^{f_{m+2+ip}} \cdots 10^{f_{m+(i+1)p}}, \prod_{j=1}^{p} c_j d_j^{a_{i,j}}, q') \in \delta^*$$

Then by Lemma 9.5, $t(q)$ is defined for every $q \in Q$. By the same lemma t is invariant in i. Consider the sequence:

$$q^*, t(q^*), t^2(q^*), t^3(q^*), t^4(q^*), \ldots$$

Then let state q denote the first state that reoccurs in the above sequence. This means that there is some $k \in \mathbb{N}$ such that:

$$t^k(q) = q$$

In other words, q is the fixed point of the function t^k. There holds:

$$(q, 10^{f_{m+1+ikp}} 10^{f_{m+2+ikp}} \cdots 10^{f_{m+(i+1)kp}}, \prod_{j=1}^{kp} c_j d_j^{a_{i,j}}, q) \in \delta^*$$

Denote r to be the smallest integer such that $t^r(q^*) = q$. There holds:

$$(q_0, 10^{f_0} 10^{f_1} 10^{f_2} \cdots 10^{f_{m+rp}}, w, q) \in \delta^*$$

Renaming the constants completes the proof. □

Theorem 9.7. *Let $\{f_i\}^{i \in \mathbb{N}}$ be a sequence that is eventually periodic modulo Z with period p such that for all $i, j \in \mathbb{N}$ $f_{i+jp} \geq f_i$. Denote $T = (Q, \Sigma, \delta, q_0, \Delta, \lambda)$ a finite state transducer. Then there are some $k \in \mathbb{N}$ and $w, c_j, d_j \in \Sigma^*$ such that*

$$T(\prod_{i=0}^{\infty} 10^{f_i}) = w \prod_{i=0}^{\infty} \prod_{j=1}^{kp} c_j d_j^{a_{i,j}}$$

with

$$a_{i,j} = \frac{f_{m+j+ip} - f_{m+j}}{Z}$$

where Z is the least common multiple of all zero-loops in T.

Proof. By lemma 9.6 the theorem holds. □

Theorem 9.8. *Let $w\prod_{i=0}^{\infty}\prod_{j=0}^{n}c_jd_j^i$ be a non-eventually periodic word for $n\geq 1$ and some $w,c_j,d_j\in\{0,1\}^*$. Assume there is for all $0\leq j\leq n$ the following holds:*

- $d_j\neq\epsilon$
- $d_j^\infty\neq c_{j+1}d_{j+1}^\infty$

Equally assume $d_n^\infty\neq c_0d_0^\infty$. Then there holds:

$$w\prod_{i=0}^{\infty}\prod_{j=0}^{n}c_jd_j^i\triangleright\Pi$$

Proof. A transducer R can be constructed that transforms the left-hand side into the right-hand side. The starting word w may be transducer to a suitable prefix of Π. The transducer may detect every transition from $\ldots d_jd_j$ to $c_{j+1}d_{j+1}d_{j+1}\ldots$ as the two words are not equal at some position. Upon detecting a switch in the input a "1" may be given as output and for each subsequent d a "0". □

The subsequent lemmas will show that Theorem 9.8 can be applied to any non-periodic transduct of Π.

Lemma 9.9. *Let $w\prod_{i=0}^{\infty}\prod_{j=0}^{n}c_jd_j^i$ be a non-eventually periodic word for $n\geq 0$ and some $w,c_j,d_j\in\{0,1\}^*$. Then it may be assumed that $d_j\neq\epsilon$ for $0\leq j\leq n$.*

Proof. Suppose there is some k such that $d_k=\epsilon$. Suppose $n=0$, then the word would be periodic, so assume $n\geq 1$. Then replace c_{k+1} with c_kc_{k+1} and the word may be written as

$$w\prod_{i=0}^{\infty}\prod_{j=0,j\neq k}^{n}c_jd_j^i=w\prod_{i=0}^{\infty}\prod_{j=0}^{n-1}c'_j{d'_j}^i$$

for appropriate $c'_j,d'_j\in\{0,1\}^*$. □

The subsequent lemma has been called the "removing confusion lemma" in [EHK11].

Lemma 9.10. *Let $w\prod_{i=0}^{\infty}\prod_{j=0}^{n}c_jd_j^i$ be a non-eventually periodic word for $n\geq 0$ and some $w,c_j\in\{0,1\}^*$ and $d_j\in\{0,1\}^+$. Then it may be assumed that for all $0\leq j<n$ it holds that:*

$$d_j^\infty\neq c_{j+1}d_{j+1}^\infty$$

Equally it holds that $d_n^\infty\neq c_0d_0^\infty$.

Proof. Suppose that for a certain j it holds:

$$d_j^\infty=c_{j+1}d_{j+1}^\infty$$

Suppose $n=0$, then the above would read:

$$d_0^\infty=c_0d_0^\infty$$

This would suggest:

$$c_0d_0^\infty=d_0c_0d_0^\infty$$

So surely $c_0d_0=d_0c_0$. Then

$$w\prod_{i=0}^{\infty}c_0d_0^i$$

would be periodic, so assume $n \geq 1$. Let r be the largest $r \in \mathbb{N}_{>0}$ such that for some $x \in \Sigma^+$:

$$d_j = x^r$$

There is some $v \in \mathbb{N}$ such that

$$c_{j+1} = x^v c^*_{j+1}$$

where c^*_{j+1} is the appropriate suffix of c_{j+1} such that x is no prefix of c^*_{j+1}. Conclude that c^*_{j+1} is a prefix of x: $x = c^*_{j+1}b$ holds for some $b \in \Sigma^*$. But then it is clear that b is a prefix of d_{j+1}. Summarizing this yields:

$$\begin{aligned} c^*_{j+1}b &= x \\ d_{j+1} &= bd^*_{j+1} \end{aligned}$$

where d^*_{j+1} is the appropriate suffix of d_{j+1}. Then it holds that:

$$\begin{aligned} d_j^\infty &= c_{j+1}d_{j+1}^\infty \\ x^\infty &= c_{j+1}(bd^*_{j+1})^\infty \\ &= x^v c^*_{j+1} b (d^*_{j+1}b)^\infty \\ &= x^{v+1}(d^*_{j+1}b)^\infty \\ &= (d^*_{j+1}b)^\infty \end{aligned}$$

Recall that $c^*_{j+1}b = x$ and conclude that $d_{j+1} = bx^s c^*_{j+1}$ for some $s \in \mathbb{N}$. Combining the above yields:

$$\begin{aligned} c_j d_j^k c_{j+1} d_{j+1}^k c_{j+2} &= c_j x^{kr} x^v c^*_{j+1} b x^{k(s+1)-1} c^*_{j+1} c_{j+2} \\ &= c_j x^{kr+v} x^{k(s+1)} c^*_{j+1} c_{j+2} \\ &= c_j x^{k(s+r+1)+v} c^*_{j+1} c_{j+2} \end{aligned}$$

Renaming appropriately shows that the word can be written as

$$T(\Pi) = w \prod_{i=0}^{\infty} \prod_{j=0}^{n-1} c_j d_j^i$$

with $n \geq 1$. □

10. On the Fibonacci sequence

Denote $a \not\triangleright b$ to be the negation of $a \triangleright b$. The following construction has been introduced by [EHK11] and has appropriately been called an *infinite descending chain*:

$$\begin{aligned} & \dots \\ \not\triangleright \quad & 10^{2^0}10^{2^4}10^{2^8}10^{2^{12}}10^{2^{16}}10^{2^{20}}10^{2^{24}}\dots \\ \not\triangleright \quad & 10^{2^0}10^{2^2}10^{2^4}10^{2^6}10^{2^8}10^{2^{10}}10^{2^{12}}\dots \\ \not\triangleright \quad & 10^{2^0}10^{2^1}10^{2^2}10^{2^3}10^{2^4}10^{2^5}10^{2^6}\dots \end{aligned}$$

More formally define:

$$I_i = \prod_{k=0}^{\infty} 10^{(2^{k2^i})}$$

It holds that for $i \geq 0$

$$\begin{aligned} I_i &\triangleright I_{i+1} \\ I_{i+1} &\not\triangleright I_i \end{aligned}$$

This section will now go on to introduce a second infinite descending chain. The construction involves using the Fibonacci sequence:

$$\begin{aligned} F_{n+2} &= F_n + F_{n+1} \\ F_0 &= 0 \\ F_1 &= 1 \end{aligned}$$

Lemma 10.1. *Fix $i \in \mathbb{N}_{>0}$, then there holds:*

$$\prod_{k=0}^{\infty} 10^{F_{ik}} \triangleright \prod_{k=0}^{\infty} 10^{F_{2ik}}$$

Proof. Figure 9 shows the appropriate transducer. It removes every second instance of 10^n and leaves the others untouched. □

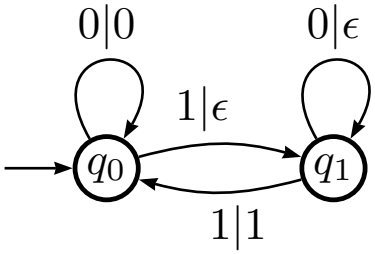

FIGURE 9. $\prod_{k=0}^{\infty} 10^{F_{ik}} \triangleright \prod_{k=0}^{\infty} 10^{F_{2ik}}$

Theorem 10.2. *Fix $i \in \mathbb{N}_{>0}$, then there holds:*

$$\prod_{k=0}^{\infty} 10^{F_{2ik}} \not\triangleright \prod_{k=0}^{\infty} 10^{F_{ik}}$$

Proof. The Fibonacci sequence $\{F_k\}^{k\in\mathbb{N}}$ is period modulo any positive integer, as shown in [Wal60]. Then the sequences $\{F_{ik}\}^{k\in\mathbb{N}}$ and $\{F_{2ik}\}^{k\in\mathbb{N}}$ will also be periodic modulo every integer. From Theorem 9.7 follows that for any transducer $T = (Q, \{0,1\}, \delta, q_0, \Delta, \lambda)$ there are some $c_j, d_j \in \Delta^*$ and $n, m \in \mathbb{N}$ such that

$$T(\prod_{k=0}^{\infty} 10^{F_{2ik}}) = w \prod_{i=0}^{\infty} \prod_{j=0}^{n} c_j d_j^{a_{i,j}}$$

with

$$a_{i,j} = \frac{F_{2k(m+j+ip)} - F_{2k(m+j)}}{Z}$$

where Z is the least common multiple of all zero-loops in T and p the period of $\{F_{2ik} \bmod Z\}^{k\in\mathbb{N}}$. In accordance with the proof in Theorem 9.8, assume that $d_j \neq \epsilon$ for $0 \leq j \leq n$. Suppose that:

$$w \prod_{i=0}^{\infty} \prod_{j=0}^{n} c_j d_j^{a_{i,j}} = \prod_{k=0}^{\infty} 10^{F_{ik}}$$

Then for the right hand side every factor $10^n 1$ occurs at most once for every $n \in \mathbb{N}$. This implies that every d_j contains only zeros. It also implies that every c_j contains at most one one. The claim now is that there can be no d_j that satisfy the above

equation. If this were true, then the theorem would hold. The proof of the claim follows from the fact that for $0 \leq j \leq n$:

$$\begin{aligned}\lim_{k\to\infty} \frac{a_{i,j}}{F_{ik}} &= \lim_{k\to\infty} \frac{F_{2k(m+j+ip)} - F_{2k(m+j)}}{ZF_{ik}} \\ &= \lim_{k\to\infty} \frac{F_{2k(m+j+ip)}}{F_{ik}} \\ &= \infty\end{aligned}$$

This shows that the growth of F_{2ki} in terms of F_{ik} is more than linear. The word length of $c_j d_j^{F_{2ik}}$ is a linear function for argument F_{2ik}, so it cannot adjust for the more-than-linear growth of F_{2ik} over F_{ik}.[5] □

So then $\prod_{k=0}^{\infty} 10^{F_{ik}}$ is an infinite descending chain:

$$\prod_{k=0}^{\infty} 10^{F_{ik}} \quad \triangleright \quad \prod_{k=0}^{\infty} 10^{F_{2ik}}$$
$$\prod_{k=0}^{\infty} 10^{F_{2ik}} \quad \not\triangleright \quad \prod_{k=0}^{\infty} 10^{F_{ik}}$$

11. Acknowledgments

First and foremost I would like to thank my thesis supervisor Robbert Fokkink for all the effort and pastoral support. He has been the best supervisor I could ever imagine. Further thanks goes to Michel Dekking, Joerg Endrullis, Wan Fokkink, Dimitri Hendriks and Jan Willem Klop for additional supervision and support.

References

[AS99] Jean-Paul Allouche and Jeffrey Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and their applications, Proceedings of SETA98*, pages 1–16. Springer, 1999.

[AS03] Jean-Paul Allouche and Jeffery Shallit. *Automatic sequences: Theory, Applications, Generalizations*. Cambridge university press, 2003.

[Chi09] I. Chiswell. *A course in Formal Languages, Automata and Groups*. Universitext, 2009.

[Dek77] Michel Dekking. Transcendance du nombre de Thue-Morse. *C. R. Acad. Sci. Paris*, 285:157–160, 1977.

[DP08] Arturas Dubickas and Tomas Plankis. Periodicity of some recurrence sequences modulo m. *INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY*, 8(A42), 2008.

[EHK11] Joerg Endrullis, Dimitri Hendriks, and Jan Willem Klop. Degrees of streams. To appear in *Integers*, 2011.

[JK69] Konrad Jacobs and Michael Keane. 0-1-sequences of toeplitz type. *Probability Theory and Related Fields*, 13:123–131, 1969. 10.1007/BF00537017.

[Klo02] Jan Willem Klop. A wonderful stream for Jaco, 2002.

[Mah29] Kurt Mahler. Arithmetische eigenschaften der lösungen einer klasse von funktionalgleichungen. *Mathematische Annalen*, 101:342–366, 1929. 10.1007/BF01454845.

[PZ82] A. Perelli and U. Zannier. On periodic mod p sequences. *Journal of Number Theory*, 15(1):77 – 82, 1982.

[Sak09] Jean-Paul Sakarovitch. *Automata theory*. Cambridge university press, 2009.

[Ste08] Sebastian Stern. The Thue-Morse sequence. Vrije Universiteit Amsterdam, master thesis, 2008.

[Wal60] D. D. Wall. Fibonacci series modulo m. *The American Mathematical Monthly*, 67(6):pp. 525–532, 1960.

[5] Special thanks to Joerg Endrullis for providing useful comments regarding the proof.

Appendix A. Period doubling sequence

Let the generalization of Toeplitz substitution k and v_2 be with p a prime:

Definition A.1.

$$\begin{aligned} k_p(0) &= 0^{p-1}1 \\ k_p(1) &= 0^{p-1}0 \\ v_p(n) &= \{\max_b p^b | n\} \mod 2 \end{aligned}$$

Lemma A.2. *For $n \in \mathbb{N}_{>0}$ and p prime holds that:*

$$\{v_p(n)\}^{n\in\mathbb{N}_{>0}} = k_p^\infty(0)$$

Proof. We shall prove this claim by showing the recurrence relations are the same for both the substitution and the function. From the definition of $v_p(n)$ follows that for all $0 < c < p^2$

$$\begin{aligned} v_p(p^2 n) &= v_p(n) \\ v_p(c + p^2 n) &= v_p(c) \end{aligned}$$

It is clear that

$$k_p^m(0) = v_p(1), v_p(2), \ldots, v_p(p^m)$$

holds for $m = 2$ and $m = 1$. We note that

$$\begin{aligned} k_p^2(0) &= v_p(1), v_p(2), \ldots, v_p(p^2-1), 0 \\ k_p^2(1) &= v_p(1), v_p(2), \ldots, v_p(p^2-1), 1 \end{aligned}$$

and it follows that both sequences have the same recurrence relations. □

Lemma A.3. *Let p be a prime. Then for $-p^m < a < p^m$ with $a' \neq 0$ and all $n \in \mathbb{N}_{>0}$ holds that:*

$$\begin{aligned} v_p(n2^m + a) &= v_p(n2^m - a) \\ &= v_p(|a|) \end{aligned}$$

Proof. Follows from the base-p expansion of $a + n2^m$. □

Lemma A.4. *Let k_p denote the Toeplitz substitution with prime p over the $\{0,1\}$ alphabet. Let $0 < a < p^m - 1$ with $a, m \in \mathbb{N}_{>0}$ fixed. Then it holds:*

$$\rho^\infty_{a+1,k_p^m}(v_p(a)) = \{v_p(a + n(2^m - 1))\}^{n\in\mathbb{N}}$$

Proof. First we will show that the first p^m symbols of both sides are matching. After that, we will show a recurrence relation that holds for both sides. Together with the first step this will complete the proof.
Note that ρ_{a+1,k_p^m} is a substitution prolongable on $v_p(a)$. The following holds:

$$\begin{aligned} k^m(0) &= v_p(1), v_p(2), \ldots, v_p(2^m) \\ k^m(1) &= v_p(2^m+1), v_p(2^m+2), \ldots, v_p(2^{m+1}) \end{aligned}$$

It follows from lemma A.3:

$$\begin{aligned} k^m(0) &= v_p(1), v_p(2), \ldots, v_p(p^m-1), v_p(p^m) \\ k^m(1) &= v_p(2^m+1), v_p(2^m+2), \ldots, v_p(p^{m+1}-1), v_p(p^{m+1}) \\ &= v_p(1), v_p(2), \ldots, v_p(p^m-1), v_p(p^{m+1}) \\ k^m(v_p(a)) &= v_p(1), v_p(2), \ldots, v_p(p^m-1), v_p(ap^m) \end{aligned}$$

We note that $k_p^m(v_p(a))$ only depends on a in the p^m-th (final) entry. It is important for the second step of the proof to see that if m is even it holds $v_p(ap^m) = v_p(a)$

and likewise when m odd $v_p(ap^m) = 1 - v_p(a)$. We may rotate $k_p^m(v_p(a))$ to obtain $\rho_{a+1,k_p^m}(v_p(a))$ and use lemma 3.6 to rewrite:

$$\begin{aligned}\rho_{a+1,k_p^m}(v_P(a)) &= v_p(p^m - a), \ldots, v_p(ap^m), v_p(1), \ldots, v_p(p^m - a - 1) \\ &= v_p(a), \ldots, v_p(a + a(p^m - 1)), v_p(a + (a+1)(p^m - 1)), \ldots, v_p(a + (p^m - 1)(p^m - 1))\end{aligned}$$

This completes the first step of the proof. Suppose m is even. Let $0 \leq c < p^m - 1$ with $c \neq a$ and substitute $n = c + rp^m$ for $r \in \mathbb{N}$, then it follows by lemma A.3 that:

$$\begin{aligned}v_p(a + (p^m - 1)n) &= v_p(a + (p^m - 1)(c + rp^m)) \\ &= v_p(a + c(p^m - 1) + (p^m - 1)rp^m) \\ &= v_p(a + c(p^m - 1))\end{aligned}$$

Which shows that for $c \neq a$, the $c + p^m r$-th entry is actually equal to the c-th entry of the sequence. Now suppose m is even and $c = a$. Then we find:

$$\begin{aligned}v_p(a + (p^m - 1)n) &= v_p(a + (p^m - 1)(c + rp^m)) \\ &= v_p(a + (p^m - 1)(a + rp^m)) \\ &= v_p(p^m(a + (p^m - 1)r)) \\ &= v_p(a + (p^m - 1)r))\end{aligned}$$

So for $c = a$, we find that the $a + p^m r$-th entry of $\{v_p(a + n(p^m - 1))\}^{n \in \mathbb{N}}$ is actually equal to the r-th entry of the sequence. Using m even we find

$$\begin{aligned}\rho_{a+1,k_p^m}(0) &= v_p(p^m - a), \ldots, v_p(p^m - 1), 0, v_p(1), \ldots, v_p(p^m - a - 1) \\ \rho_{a+1,k_p^m}(1) &= v_p(p^m - a), \ldots, v_p(p^m - 1), 1, v_p(1), \ldots, v_p(p^m - a - 1)\end{aligned}$$

and we conclude that indeed the r-th entry of the fixed point of the substitution is equal to its $a + p^m r$-th entry. Both sequences have the same recurrence rules for m is even. Now consider m odd. We still find that for $c \neq a$:

$$v_p(a + (p^m - 1)n) = v_p(a + c(p^m - 1))$$

However for $c = a$ it follows:

$$\begin{aligned}v_p(a + (p^m - 1)n) &= v_p(a + (p^m - 1)(c + rp^m)) \\ &= v_p(a + (p^m - 1)(a + rp^m)) \\ &= v_p(p^m(a + (p^m - 1)r)) \\ &= 1 - v_p(a + (p^m - 1)r))\end{aligned}$$

Equally we find

$$\begin{aligned}\rho_{a+1,k_p^m}(0) &= v_p(p^m - a), \ldots, v_p(p^m - 1), 1, v_p(1), \ldots, v_p(p^m - a - 1) \\ \rho_{a+1,k_p^m}(1) &= v_p(p^m - a), \ldots, v_p(p^m - 1), 0, v_p(1), \ldots, v_p(p^m - a - 1)\end{aligned}$$

and the result follows. □

Theorem A.5. *Let $a, b \in \mathbb{N}_{>0}$ be such that $0 < a < b$, $p \nmid b$ and p a prime, then there are some $c, m \in \mathbb{N}_{>0}$ such that*

$$\{v_p(a + bn)\}^{n \in \mathbb{N}} = \rho_{ac+1,k_p^m}^{\infty}(v_p(ac))$$

Proof. Suppose c, m are such that $cb = p^m - 1$. Then we know $p \nmid c$ and invoke lemma A.4 to write:

$$\begin{aligned}\{v_p(a + bn)\}^{n \in \mathbb{N}} &= \{v_p(ca + cbn)\}^{n \in \mathbb{N}} \\ &= (\rho_{ac+1,k_p^m})^{\infty}(v_p(ac))\end{aligned}$$

So now we need to show that there are some c, m such that $cb = p^m - 1$. Suppose that $b \nmid p^m - 1$ for all $m \in \mathbb{N}_{>0}$. Then by pigeon hole principle there must be some $r, s, v \in \mathbb{N}_{>0}$ such that

$$\begin{aligned} p^r - 1 &\equiv v \mod b \\ p^{r+s} - 1 &\equiv v \mod b \end{aligned}$$

It follows that:

$$\begin{aligned} b &\mid p^{r+s} - p^r \\ b &\mid p^r(p^s - 1) \\ b &\mid p^s - 1 \end{aligned}$$

So then b does divide $p^s - 1$, which is an contradiction. □

Theorem A.6. *Let ρ_{q+1,k_p^m} be a rotated Toeplitz substitution as noted in section 3 with $0 < q < (p^m - 1)$, $m \in \mathbb{N}_{>0}$ and p a prime. Then for every $r \in \mathbb{Z}\backslash\{0\}$ and $l \in \mathbb{N}_{>0}$ it holds*

$$\rho_{q+1,k_p^m}(\{v_p(\frac{r}{p^l - 1} + n)\}^{n\in\mathbb{N}}) = \{v_p(\frac{p^m r}{p^l - 1} - q + n)\}^{n\in\mathbb{N}}$$

Proof. It will be shown that both sides of the equation have the same recurrence rules. From the recurrence rules follows that the first 2^m letters of both sides are the same, completing the proof. But first recall from section 3 the construction of the rotated substitution ρ_{q+1,k_p^m}:

$$\begin{aligned} k_p^m(0) &= v_p(1), v_p(2), v_p(3), \ldots v_p(p^m - 1), v_p(p^m) \\ k_p^m(1) &= v_p(1), v_p(2), v_p(3), \ldots v_p(p^m - 1), 1 - v_p(p^m) \\ \rho_{q+1,k_p^m}(0) &= v_p(p^m - q), v_p(p^m - q + 1), \ldots, v_p(p^m), v_p(1), \ldots, v_p(p^m - q - 1) \\ \rho_{q+1,k_p^m}(1) &= v_p(p^m - q), v_p(p^m - q + 1), \ldots, 1 - v_p(p^m), v_p(1), \ldots, v_p(p^m - q - 1) \end{aligned}$$

Suppose m is even and consider $n = q + ip^m$ with $i \in \mathbb{N}$. It holds:

$$\begin{aligned} v_p(\frac{p^m r}{p^l - 1} - q + n) &= v_p(\frac{p^m r}{p^l - 1} - q + q + ip^m) \\ &= v_p(p^m(r + i(p^l - 1))) \\ &= v_p(r + i(p^l - 1)) \\ &= v_p(\frac{r}{p^l - 1} + i) \end{aligned}$$

Now consider the left hand side of the statement and denote

$$\rho_{q+1,k_p^m}(v_p(\frac{r}{p^l - 1} + i)) = \rho_0, \rho_1, \ldots, \rho_{p^m - 1}$$

then $\rho_q = v_p(\frac{r}{p^l - 1} + i)$. Now if m were odd, then $\rho_q = 1 - v_p(\frac{r}{p^l - 1} + i)$ and:

$$v_p(\frac{p^m r}{p^l - 1} - q + n) = 1 - v_p(\frac{r}{p^l - 1} + i)$$

So for $n = q + ip^m$ with $i \in \mathbb{N}$ the equation-to-prove holds. Now consider $n = s + ip^m$ with $s \neq q$ and $0 \leq s < p^m$. It follows that $|s - q| < p^m - 1$, which is to be used in the second to final step of the derivation below:

$$\begin{aligned} v_p(\frac{p^m r}{p^l - 1} - q + n) &= v_p(\frac{p^m r}{p^l - 1} - q + s + ip^m) \\ &= v_p(p^m r + (p^l - 1)(s - q + ip^m)) \\ &= v_p(p^m(r + (p^l - 1)i) + (p^l - 1)(s - q)) \\ &= v_p(|s - q|) \end{aligned}$$

It is clear that for $n = s + ip^m$ with $s \neq q$ and $0 \leq s < p^m$ the expression is dependent on s and independent of i. In fact, the expansion of ρ_{q+1,k_p^m} given at the start of the proof now shows the two sides are equal. □