# Arithmetic Self-Similarity of Infinite Sequences *

Dimitri Hendriks Frits G.W. Dannenberg Jörg Endrullis

Mark Dow Jan Willem Klop

December 21, 2013

**Abstract**

We define the arithmetic self-similarity (AS) of a one-sided infinite sequence $\sigma$ to be the set of arithmetic subsequences of $\sigma$ which are a vertical shift of $\sigma$. We study the AS of several families of sequences, viz. completely additive sequences, Toeplitz words and Keane's generalized Morse sequences. We give a complete characterization of the AS of completely additive sequences, and classify the set of single-gap Toeplitz patterns that yield completely additive Toeplitz words. We show that every arithmetic subsequence of a Toeplitz word generated by a one-gap pattern is again a Toeplitz word. Finally, we establish that generalized Morse sequences are specific sum-of-digits sequences, and show that their first difference is a Toeplitz word.

## 1 Introduction

Some infinite sequences are similar to a part of themselves. Zooming in on a part of the structure reveals the whole structure again. Of special interest are what we may call *scale-invariant* sequences. A sequence $\boldsymbol{w} = (w(n))_{n\geq 1}$ (over some additive group $\Sigma$) is scale-invariant if for all 'dilations' $k \geq 1$ the compressed sequence $\boldsymbol{w}/k = (w(kn))_{n\geq 1}$ is a 'vertical shift' of $\sigma$, that is, $w(kn) - w(n)$ is constant. An example of a scale-invariant sequence is the period doubling sequence $\boldsymbol{p}$ which can be defined as $\boldsymbol{p} = (v_2(n) \bmod 2)_{n\geq 1}$ with $v_2(n)$ the 2-adic valuation of $n$. We have that $\boldsymbol{p}/k = \boldsymbol{p} + \boldsymbol{p}(k) \bmod 2$ for every $k \geq 1$:

$$\begin{aligned}
\boldsymbol{p} &= 0100010101000100010001010100010101000101010\ldots \\
\boldsymbol{p}/2 &= 1\ 0\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 1\ \ldots \\
\boldsymbol{p}/3 &= 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ \ldots \\
\boldsymbol{p}/5 &= 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ \ldots
\end{aligned}$$

*This research has been partially funded by the Netherlands Organisation for Scientific Research (NWO) under grant numbers 612.000.934 and 639.021.020.

Scale-invariant sequences $\sigma$ with first term $\sigma_1 = 0$ are known as *completely additive* sequences, that is, sequences $\sigma$ such that $\sigma(nm) = \sigma(n) + \sigma(m)$ for all positive integers $n, m$.

In general, given an infinite sequence $\sigma$ one may wonder what are the arithmetic subsequences of $\sigma$ similar to $\sigma$ itself. This leads to the notion of what we call 'arithmetic self-similarity': For an equivalence relation $\sim$ on $\Sigma^\omega$, the *arithmetic self-similarity* of a sequence $\sigma \in \Sigma^\omega$, which we denote by $\mathcal{AS}^\sim(\sigma)$, is the set of pairs $\langle a, b\rangle$ such that the subsequence of $\sigma$ indexed by the arithmetic progression $a + bn$ is equivalent to $\sigma$ :

$$\mathcal{AS}^\sim(\sigma) = \{ \langle a, b\rangle \mid \sigma_{a,b} \sim \sigma \} \qquad \text{where } \sigma_{a,b} = \sigma(a)\,\sigma(a+b)\,\sigma(a+2b)\,\ldots$$

For instance, in [12] it is shown that the arithmetic self-similarity of the Thue–Morse sequence $\boldsymbol{m} = 01101001\cdots$ [3] with respect to 'transducer-equivalence' $\diamond$ is the full space $\mathcal{AS}^\diamond(\boldsymbol{m}) = \mathbb{N} \times \mathbb{N}_{>0}$, i.e., for every arithmetic subsequence $\boldsymbol{m}_{a,b}$ of $\boldsymbol{m}$ there is a finite state transducer which reconstructs $\boldsymbol{m}$ from $\boldsymbol{m}_{a,b}$.

In this paper we focus on cyclic groups $\langle \Sigma, +, 0\rangle$ and on a particular equivalence $\sim$ on sequences over $\Sigma$, namely $\sigma \sim \tau$ if and only if $\sigma = \tau + c$ for some $c \in \Sigma$. Our main results are as follows:

- We give a complete characterization of the arithmetic self-similarity of completely additive sequences $\sigma \in \Sigma^{\mathbb{N}_{>0}}$; we show that $\langle a, b\rangle \in \mathcal{AS}(\sigma)$ if and only if $a = b$.
- We define Toeplitz patterns to be finite words over $\Sigma \cup \mathfrak{S}_\Sigma$, where $\mathfrak{S}_\Sigma$ is the set of permutations over $\Sigma$ which play the role of 'gaps'. The composition operation defined on Toeplitz patterns (Definition 4.4) forms a monoid (Proposition 4.6).
- We give a complete characterization of one-gap Toeplitz patterns that yield completely additive sequences (Theorem 5.13). These patterns are constructed using discrete logarithms. The completely additive sequences they generate are determined by infinite sets of primes.
- We show that every arithmetic subsequence of a Toeplitz word generated by a one-gap pattern is again a Toeplitz word (Theorem 6.9).
- We prove that the first difference of a generalized Morse sequence [16] is a Toeplitz word (Theorem 7.8). This gives rise to an embedding from Keane's monoid of block products (extended to additive groups) to that of Toeplitz pattern composition (Theorem 7.7).
- We show how generalized Morse sequences [16] are specific sum-of-digits sequences (Theorem 7.5).
- For the Thue–Morse sequence $\boldsymbol{m} = 01101001\ldots$ we show that $\boldsymbol{m}(a+bn) = \boldsymbol{m}(n) + \boldsymbol{m}(a) \bmod 2$ ($n \in \mathbb{N}$) if and only if $0 \le a < b = 2^m$ for some $m \in \mathbb{N}$ (Theorem 7.10).

## 2 Basic Definitions

We write $\mathbb{N} = \{0, 1, 2, \ldots\}$ for the set of natural numbers, and $\mathbb{N}_{>0}$ for the set of positive integers, $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$. For $k \geq 0$ we define $\Sigma_k = \{0, \ldots, k-1\}$. The sets of finite words and finite non-empty words over an alphabet $\Sigma$ are denoted by $\Sigma^*$ and $\Sigma^+$, respectively. We identify a (one-sided) infinite word with a map from $\mathbb{N}$ to $\Sigma$, or, when this is more convenient, from $\mathbb{N}_{>0}$ to $\Sigma$. We write $\Sigma^{\mathbb{N}}$, or $\Sigma^{\mathbb{N}_{>0}}$, for the set of infinite words over $\Sigma$:

$$\Sigma^{\mathbb{N}} = \{\, \sigma \mid \sigma : \mathbb{N} \to \Sigma \,\} \qquad\qquad \Sigma^{\mathbb{N}_{>0}} = \{\, \sigma \mid \sigma : \mathbb{N}_{>0} \to \Sigma \,\}$$

We let $\Sigma^\infty$ denote the set of all finite and infinite words over $\Sigma$. We write $\varepsilon$ for the empty word, and $xy$ for concatenation of words $x \in \Sigma^*$, $y \in \Sigma^\infty$. For $u \in \Sigma^*$ we let $u^0 = \varepsilon$ and $u^{k+1} = uu^k$, and, for $u \in \Sigma^+$ we write $u^\omega$ for the infinite sequence $uuu \ldots$. For a word $w \in \Sigma^\infty$ we use $w(n)$ to denote the value at index $n$ (if defined). The *reversal* $w^R$ of $w \in \Sigma^*$ is defined by $\varepsilon^R = \varepsilon$ and $(au)^R = u^R a$, for all $a \in \Sigma$ and $u \in \Sigma^*$.

We shall primarily deal with infinite words over finite cyclic groups. A cyclic group is a group generated by a single element (or its inverse). Every finite cyclic group of order $k$ is isomorphic to the additive group $\langle \Sigma_k, +, 0\rangle$ with $+$ denoting addition modulo $k$.

Let $\Sigma = \langle \Sigma, +_\Sigma, 0_\Sigma \rangle$ be a finite cyclic group. Let $w \in \Sigma^\infty$, and $a \in \Sigma$. Then $w +_\Sigma a$ is defined by $(w +_\Sigma a)(n) = w(n) +_\Sigma a$, whenever $w(n)$ is defined. Furthermore, we write $\sigma +_\Sigma \tau$ for the sequence obtained from $\sigma, \tau \in \Sigma^{\mathbb{N}}$ by pointwise addition: $(\sigma +_\Sigma \tau)(n) = \sigma(n) +_\Sigma \tau(n)$ for all $n \in \mathbb{N}$; so $\sigma +_\Sigma a = \sigma +_\Sigma a^\omega$. When the group $\Sigma$ is clear from the context, we will write just $+$ and $0$ for $+_\Sigma$ and $0_\Sigma$.

For $k \in \Sigma_n$ (or $k \in \mathbb{N}$) and $c \in \Sigma_m$ (with $n, m \in \mathbb{N}_{>0}$), we write $k \odot c$ for the letter $c + c + \cdots + c$ ($k$ times) in the alphabet $\Sigma_m$. For words $w = a_0 \ldots a_r \in \Sigma_n^*$ we use $w \odot c$ to denote the word $(a_0 \odot c) \ldots (a_r \odot c) \in \Sigma_m^*$. Likewise for infinite sequences $\sigma = a_0 a_1 \ldots \in \Sigma_n^{\mathbb{N}}$ we write $\sigma \odot c$ for $(a_0 \odot c)(a_1 \odot c) \ldots \in \Sigma_m^{\mathbb{N}}$. Finally, by the partial application $\odot c$ we denote the map from $\Sigma_n$ to $\Sigma_m$ defined by $k \mapsto k \odot c$ for all $k \in \Sigma_n$. For $n, m \in \mathbb{N}_{>0}$ we define

$$\xi(n, m) = \frac{m}{\gcd(n, m)} = \frac{\operatorname{lcm}(n, m)}{n}$$

**Lemma 2.1.** *Let $c \in \Sigma_m$ s.t. $c$ is divisible by $\xi(n, m)$, i.e., $c = \xi(n, m) \odot c'$ for some $c' \in \Sigma_m$. Then the map $\odot c$ is a group homomorphism from $\langle \Sigma_n, +_{\Sigma_n}, 0_{\Sigma_n}\rangle$ to $\langle \Sigma_m, +_{\Sigma_m}, 0_{\Sigma_m}\rangle$.*

*Proof.* Obviously $0_{\Sigma_n} \odot c = 0_{\Sigma_m}$. Note that $n \odot c = n \odot (\xi(n, m) \odot c') = (n \cdot \xi(n, m)) \odot c' = 0_{\Sigma_m}$ since $n \cdot \xi(n, m)$ is a multiple of $m$. Let $k, \ell \in \Sigma_n$ arbitrary. If $k +_{\mathbb{N}} \ell < n$ (here $+_{\mathbb{N}}$ indicates that we use addition on the natural numbers), then trivially $(k +_{\Sigma_n} \ell) \odot c = k \odot c +_{\Sigma_m} \ell \odot c$. Otherwise $k +_{\mathbb{N}} \ell = n +_{\mathbb{N}} u$ for some $u \in \Sigma_n$ and $k \odot c +_{\Sigma_m} \ell \odot c = n \odot c +_{\Sigma_m} u \odot c = u \odot c = (k +_{\Sigma_n} \ell) \odot c$. $\square$

Let $\sigma \in \Sigma^{\mathbb{N}}$ be an infinite sequence, and let $a \geq 0$ and $b \geq 1$ be integers. The *arithmetic subsequence* $\sigma_{a,b}$ of $\sigma$ is defined by[1]

$$\sigma_{a,b}(n) = \sigma(a + bn) \qquad\qquad (n \in \mathbb{N})$$

By composition of arithmetic progressions we have $(\sigma_{a,b})_{c,d} = \sigma_{a+bc,bd}$ for all $a, c \geq 0$ and $b, d \geq 1$.

A sequence $\sigma \in \Sigma^{\mathbb{N}}$ is *ultimately periodic* if there are $n_0 \in \mathbb{N}$ and $t \in \mathbb{N}_{>0}$ such that $\sigma(n+t) = \sigma(n)$ for all $n \geq n_0$. Here $t$ is called the *period* and $n_0$ the *preperiod*. If $n_0 = 0$, $\sigma$ is called *(purely) periodic*.

**Lemma 2.2.** *Every arithmetic subsequence of an ultimately periodic sequence is ultimately periodic.*

*Proof.* Let $\sigma \in \Sigma^{\mathbb{N}}$, $n_0 \in \mathbb{N}$, $t \in \mathbb{N}_{>0}$ with $(\forall n \geq n_0)(\sigma(n+t) = \sigma(n))$. Further let $a \in \mathbb{N}$, $b \in \mathbb{N}_{>0}$. We show that $\sigma_{a,b}$ is ultimately periodic. Let $n_0' \in \mathbb{N}$ be such that $a + bn_0' \geq n_0$, and let $t' \in \mathbb{N}_{>0}$ with $bt' \equiv 0 \pmod{t}$ (such $t'$ with $t' \leq t$ always exists). Then we have $\sigma_{a,b}(n+t') = \sigma(a+bn+bt') = \sigma(a+bn) = \sigma_{a,b}(n)$ for all integers $n \geq n_0'$. ☐

We use $\mathbf{P} = \{2, 3, 5, \ldots\}$ for the set of prime numbers. We write $[a]_k$ for the congruence (or residue) class of $a$ modulo $k$: $[a]_k = \{n \in \mathbb{N} \mid n \equiv a \pmod{k}\}$. The set of residue classes modulo $k$ is denoted by $\mathbb{Z}/k\mathbb{Z}$. Moreover, for the union of $[a_1]_k, \ldots, [a_q]_k \in \mathbb{Z}/k\mathbb{Z}$ we write

$$[a_1, \ldots, a_q]_k = [a_1]_k \cup \cdots \cup [a_q]_k$$

We let $\mathbb{Z}_k = (\mathbb{Z}/k\mathbb{Z})^{\times} = \{\, [h]_k \mid \gcd(h, k) = 1\}$, the multiplicative group of integers modulo $k$. A *primitive root modulo* $k$ is a generator of the group $\mathbb{Z}_k$. That is, a primitive root is an element $g$ of $\mathbb{Z}_k$ such that for all $a \in \mathbb{Z}_k$ there is some integer $e$ with $a \equiv g^e \pmod{k}$. For integers $e_1, e_2$ we have that $g^{e_1} \equiv g^{e_2} \pmod{k}$ if and only if $e_1 \equiv e_2 \pmod{k}$, and so we have a bijection $e \mapsto g^e$ on $\mathbb{Z}_k$, called *discrete exponentation (with base* $g$*)*. Its inverse, $a \mapsto \log_g(a) = e$ with $e$ such that $a \equiv g^e \pmod{k}$, is called *discrete logarithm (to the base* $g$*)*.

Groups $\mathbb{Z}_p = \{\, [1]_p\,, [2]_p\,, \ldots, [p-1]_p\} \simeq \{1, 2, \ldots, p-1\}$ with $p$ an odd prime number contain all nonzero residue classes modulo $p$, always have a primitive root and are cyclic. For example the primitive roots modulo 7 are 3 and 5. The powers of 3 modulo 7 are $3, 2, 6, 4, 5, 1, 3 \ldots$, and of 5 they are $5, 4, 6, 2, 3, 1, 5, \ldots$, listing every number modulo 7 (except 0). On the other hand, 2 is not a primitive root modulo 7: the powers of 2 modulo 7 are $2, 4, 1, 2, \ldots$, missing several values from $\{1, \ldots, 6\}$. Note that a primitive root is not necessarily a prime number, e.g., 6 is a primitive root modulo 11.

**Lemma 2.3.**

*(i) (Base change) Let* $g$*,* $h$ *be primitive roots mod* $n$*,* $a \in \mathbb{Z}_n$*. Then* $\log_g(a) = \log_g(h) \cdot \log_h(a)$.

[1] For infinite sequences $\sigma \in \Sigma^{\mathbb{N}_{>0}}$ (indexed over the positive integers) we require both $a$ and $b$ to be positive integers, and then $\sigma_{a,b}(n) = \sigma(a + b(n-1))$ for all $n \in \mathbb{N}_{>0}$.

*(ii)* *(Fermat's little theorem) Let $p$ be a prime number and $a \in \mathbb{Z}_p$. Then $a^{p-1} \equiv 1 \pmod{p}$.*

*(iii)* *Let $p > 2$ be prime and $g$ be a primitive root modulo $p$. Then $g^{(p-1)/2} \equiv -1 \pmod{p}$.*

*Proof.* Folklore. We only prove item (iii): by $g$ being a primitive root we know $g^k \equiv -1 \pmod{p}$ for some unique $1 \le k \le p-1$, and so $g^{2k} \equiv 1 \pmod{p}$ which in turn implies $2k \equiv p - 1 \pmod{p}$ by item (ii). □

## Arithmetic Self-Similarity

Throughout the paper we fix an arbitrary finite cyclic group $\langle \Sigma, +, 0 \rangle$.

We define the 'arithmetic self-similarity' of an infinite sequence $\sigma$, as the set of pairs $\langle a, b \rangle$ such that the arithmetic subsequence $\sigma_{a,b}$ is similar to $\sigma$, where we interpret 'similar to' as 'is a vertical shift of'.

**Definition 2.4.** Let the relation $\sim \subseteq \Sigma^{\mathbb{N}} \times \Sigma^{\mathbb{N}}$ be defined by

$$\sigma \sim \tau \quad \text{if and only if} \quad \sigma = \tau + c \text{ for some } c \in \Sigma \qquad (\sigma, \tau \in \Sigma^{\mathbb{N}})$$

Then, we define the *arithmetic self-similarity* of a sequence $\sigma \in \Sigma^{\mathbb{N}}$ by

$$\mathcal{AS}(\sigma) = \{ \langle a, b \rangle \mid \sigma_{a,b} \sim \sigma \}$$

Clearly the relation $\sim$ is an equivalence relation on $\Sigma^{\mathbb{N}}$ as the elements of the group $\Sigma$ are invertible. Equivalent sequences have the same arithmetic self-similarity.

**Lemma 2.5.** *For all $\sigma, \tau \in \Sigma^{\mathbb{N}}$, if $\sigma \sim \tau$ then $\mathcal{AS}(\sigma) = \mathcal{AS}(\tau)$.* □

Arithmetic self-similarity is closed under composition of aritmetic progressions.

**Lemma 2.6.** *Let $\sigma \in \Sigma^{\mathbb{N}}$ be an infinite sequence and assume $\langle a, b \rangle, \langle c, d \rangle \in \mathcal{AS}(\sigma)$. Then also $\langle a + bc, bd \rangle \in \mathcal{AS}(\sigma)$.*[2] □

If for some $a \ge 1$ the shift $\sigma_{a,1}$ of a sequence $\sigma \in \Sigma^{\mathbb{N}}$ is similar to $\sigma$, then $\sigma$ is periodic.

**Lemma 2.7.** *Let $\sigma \in \Sigma^{\mathbb{N}}$ and $a \ge 1$. If $\langle a, 1 \rangle \in \mathcal{AS}(\sigma)$, then $\sigma$ is periodic.*[3] □

[2] For sequences $\sigma \in \Sigma^{\mathbb{N}_{>0}}$, indexed by the positive integers, the statement has to be reformulated: for all $a, b, c, d \in \mathbb{N}_{>0}$, if $\langle a, b \rangle, \langle c, d \rangle \in \mathcal{AS}(\sigma)$, then $\langle a + b(c-1), bd \rangle \in \mathcal{AS}(\sigma)$.

[3] For sequences $\sigma \in \Sigma^{\mathbb{N}_{>0}}$ this reads: if $a \ge 2$ and $\langle a, 1 \rangle \in \mathcal{AS}(\sigma)$, then $\sigma$ is periodic.

## 3 Completely Additive Sequences

In this section we investigate the arithmetic self-similarity of completely additive sequences.

**Definition 3.1.** A sequence $\sigma \in \Sigma^{\mathbb{N}_{>0}}$ is *completely additive (with respect to* $\Sigma$*)* if it is a homomorphism from the multiplicative monoid $\langle \mathbb{N}_{>0}, \cdot, 1 \rangle$ of positive integers to the additive group $\langle \Sigma, +, 0 \rangle$, that is

$$\sigma(1) = 0 \qquad\qquad \sigma(nm) = \sigma(n) + \sigma(m) \quad (n, m \in \mathbb{N}_{>0})$$

The constant zero sequence $\boldsymbol{z} \in \Sigma^{\mathbb{N}_{>0}}$, defined by $\boldsymbol{z}(n) = 0_\Sigma$ $(n \in \mathbb{N}_{>0})$, is called *trivially additive*, or just *trivial*.

**Lemma 3.2.** *Every sequence that is both completely additive and ultimately periodic is trivial.*

*Proof.* Let $\sigma \in \Sigma^{\mathbb{N}_{>0}}$ be a completely additive sequence. First we assume $\sigma$ to be purely periodic with period $t \in \mathbb{N}_{>0}$. Then for all $n \in \mathbb{N}_{>0}$ we have $\sigma(nt) = \sigma(n) + \sigma(t)$ by complete additivity and $\sigma(nt) = \sigma(t)$ by $t$-periodicity, whence $\sigma(n) = 0$.

Now let $\sigma$ be ultimately periodic with period $t \in \mathbb{N}_{>0}$ and preperiod $n_0 \in \mathbb{N}_{>0}$, i.e., $\sigma(n + t) = \sigma(n)$ for all integers $n \geq n_0$. It is clear that the arithmetic subsequence $\sigma/n_0 = \sigma(n_0)\sigma(2n_0)\ldots$ is periodic (see Lemma 2.2). By complete additivity we know $\sigma(nn_0) = \sigma(n) + \sigma(n_0)$, and so $\sigma$ is periodic as well. Then, by the observation above, we conclude that $\sigma$ is the constant zero sequence. □

Completely additive sequences are uniquely determined by their values at prime number positions, i.e., by a map $\mu : \mathbf{P} \to \Sigma$, as follows. First we recall the definition of the $p$-adic valuation of an integer $n \geq 1$, that is, the multiplicity of prime $p$ in the prime factorization of $n$:

**Definition 3.3.** Let $p$ be a prime number. The *$p$-adic valuation* of $n \in \mathbb{N}_{>0}$ is defined by

$$v_p(n) = \max \{ a \in \mathbb{N} \mid p^a \text{ divides } n \}$$

The infinite sequence $\boldsymbol{v}_p = v_p(1)\, v_p(2)\, \ldots$ is completely additive (with respect to the cyclic group $\mathbb{Z}$).

**Definition 3.4.** Let $\mu : \mathbf{P} \to \Sigma$ and define the sequence $\boldsymbol{v}_\mu \in \Sigma^{\mathbb{N}_{>0}}$ for all $n \in \mathbb{N}_{>0}$ by

$$\boldsymbol{v}_\mu(n) = \sum_{p \in \mathbf{P}} v_p(n) \odot \mu(p)$$

(Here summation is the generalized form of addition of the group $\Sigma$.) If the set $\{p \in \mathbf{P} \mid \mu(p) \neq 0\}$ is finite, then $\boldsymbol{v}_\mu$ is called *finitely prime generated*. Otherwise $\boldsymbol{v}_\mu$ is called an *infinitely* prime generated sequence.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{v}_2$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | ... |
| $\boldsymbol{v}_3$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | ... |
| $\boldsymbol{v}_{2,3}$ | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | ... |

Table 1: An instance of Lemma 3.7: $\boldsymbol{v}_{2,3} = \boldsymbol{v}_2 + \boldsymbol{v}_3$.

Every completely additive sequence $\sigma$ has a generator $\mu : \mathbf{P} \to \Sigma$, viz. $\mu = \sigma|_{\mathbf{P}}$, the domain restriction of the function $\sigma : \mathbb{N}_{>0} \to \Sigma$ to the set of primes. Hence, for every finite cyclic group $\Sigma$ with $|\Sigma| \geq 2$, the set of completely additive sequences over $\Sigma$ has the cardinality of the continuum.

**Example 3.5.** Define $\mu : \mathbf{P} \to \Sigma_4$ by $\mu(p) = 0, 1, 3, 2, 1$ if $p \equiv 1, 2, 3, 4, 0 \pmod 5$, respectively. Then $\boldsymbol{v}_\mu = 01321\,01322\,01320\,01323\,01322\ldots$ is generated by the infinite set $\{p \mid \mu(p) \neq 0\} = \mathbf{P} \cap [2, 3, 4]_5 \cup \{5\}$.

If $\Sigma = \Sigma_2 = \{0, 1\}$ then $\mu : \mathbf{P} \to \Sigma_2$ is the characteristic function of a set $X \subseteq \mathbf{P}$, and we simply write $\boldsymbol{v}_X(n)$ $(= \sum_{p \in X} v_p(n) \mod 2)$. We sometimes write $\boldsymbol{v}_X$ where $X \subseteq \mathbb{N}$ to denote $\boldsymbol{v}_{\mathbf{P} \cap X}$. For instance $\boldsymbol{v}_{[a]_b}$ denotes the completely additive sequence generated by the set of prime numbers congruent to $a$ modulo $b$, which is infinite if $\gcd(a, b) = 1$ by [10].

**Example 3.6.** The sequence $\boldsymbol{t} = \zeta^\omega(0)$ with $\zeta : \Sigma_2^* \to \Sigma_2^*$ the morphism given by $\zeta(0) = 010$ and $\zeta(1) = 011$, is the sequence of turns of the *Terdragon curve* [9] ([17, `A080846`]). From Theorem 5.13 (and Proposition 4.13) it follows that $\boldsymbol{t}$ is a completely additive sequence, generated by the infinite set of primes congruent to 2 modulo 3, i.e., $\boldsymbol{t} = \boldsymbol{v}_{[2]_3}$.

**Lemma 3.7.** *Let* $\mu_1, \mu_2 : \mathbf{P} \to \Sigma_2$. *Then* $\boldsymbol{v}_{\mu_1 + \mu_2} = \boldsymbol{v}_{\mu_1} + \boldsymbol{v}_{\mu_2}$. ☐

For sets $A, B \subseteq \mathbf{P}$ Lemma 3.7 says $\boldsymbol{v}_{A \Delta B} = \boldsymbol{v}_A + \boldsymbol{v}_B$, where $\Delta$ is symmetric difference: $A \,\Delta\, B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$, and so we have $\boldsymbol{v}_{A \cup B} = \boldsymbol{v}_A + \boldsymbol{v}_B$ if $A \cap B = \varnothing$. See for example Table 1, where the 2 and 3-adic valuation sequences in base 2 are added modulo 2.

If $w \in \Sigma^{\mathbb{N}_{>0}}$ is additive, it is clear that $\langle b, b\rangle \in \mathcal{AS}(w)$ for all $b \in \mathbb{N}_{>0}$. The question arises whether the arithmetic self-similarity of $w$ possibly contains other progressions. This is not the case, a result due to Kevin Hare and Michael Coons [13].

**Theorem 3.8.** *Let* $w \in \Sigma^{\mathbb{N}_{>0}}$ *be a non-trivial completely additive sequence. Then:*

$$\mathcal{AS}(w) = \{\, \langle b, b\rangle \mid b \in \mathbb{N}_{>0} \,\}$$

*Proof (K. Hare, M. Coons [13]).*

($\supseteq$) Let $w$ be a completely additive sequence over $\Sigma$, and let $b \in \mathbb{N}_{>0}$. For all $n \in \mathbb{N}_{>0}$ we have $w_{b,b}(n) = w(bn) = w(n) + w(b)$. Hence $w_{b,b} \sim w$ and so $\langle b, b\rangle \in \mathcal{AS}(w)$.

($\subseteq$) Let $w$ be a completely additive sequence over $\Sigma$ and assume $\langle a, b\rangle \in \mathcal{AS}(w)$ with $a \neq b$. We will show that $w$ is trivially additive, i.e., $w(n) = 0$ for all $n \in \mathbb{N}_{>0}$.

**Claim 1.** Without loss of generality we may assume $|a - b| = \pm 1$.

*Proof of Claim 1.* By the assumption $\langle a, b\rangle \in \mathcal{AS}(w)$ we have

$$w_{a,b}(n) = w(a + b(n-1)) = w(bn + a - b) = w(n) + c \qquad (n \in \mathbb{N}_{>0})$$

for some $c \in \Sigma$. So in particular it holds for $n = |a - b| \cdot N$, where $N$ is some arbitrary positive integer. This implies that

$$w(b \cdot |a-b| \cdot N + a - b) = w(|a-b|) + w(N) + c$$

but

$$w(b \cdot |a-b| \cdot N + a - b) = w(|a-b|) + w(bN \pm 1))$$

Subtracting $w(|a - b|)$ from both right-hand sides, we have

$$w(bN + \operatorname{sign}(a-b)) = w(N) + c \qquad (a > b)$$

which concludes the proof of Claim 1. ☐

We employ the techniques used in the proof of Theorem 4 in [5] for showing that in both cases $w$ is trivially additive.

**Case 1.** If there is some $c \in \Sigma$ such that $w(bn + 1) = w(n) + c$ for all $n \in \mathbb{N}_{>0}$, then $w$ is trivial.

*Proof of Case 1.* Let $c \in \Sigma$ be such that $w(bn + 1) = w(n) + c$ for all $n$. We will show by induction that

$$w(bn + i) = w(n) + c \qquad (n \in \mathbb{N}_{>0}) \tag{‡}$$

for all $i = 1, 2, \ldots$. Hence, taking $n = 1$, the sequence $w$ is constant from position $b + 1$ onward. With Lemma 3.2 it then follows that $w$ is trivially additive.

For $i = 1$ equation (‡) holds by assumption. Now assume (‡) holds for $i = 1, 2, \ldots, j$. Then we find:

$$w(bn + j) + w(bn + 1) = 2 \cdot w(n) + 2c$$

and

$$\begin{aligned} w(bn + j) + w(bn + 1) &= w((bn + j)(bn + 1)) \\ &= w(b^2n^2 + bn(j + 1) + j) \\ &= w(b(bn^2 + n(j + 1)) + j) \\ &= w(bn^2 + n(j + 1)) + c \\ &= w((bn + j + 1)n) + c \\ &= w(bn + j + 1) + w(n) + c \end{aligned}$$

Hence $w(bn + j + 1) = w(n) + c$ and the result follows by induction. This concludes the proof of Case 1. □

**Case 2.** If there is some $c \in \Sigma$ such that $w(bn + 1) = w(n) + c$ for all $n \in \mathbb{N}_{>0}$, then $w$ is trivial.

*Proof.* Assume, for some $c \in \Sigma$ that $w(bn - 1)) = w(n) + c$ for all integers $n \in \mathbb{N}_{>0}$. Then

$$
\begin{aligned}
w(b(bn^2) - 1)) &= w((bn - 1)(bn + 1)) \\
&= w(bn - 1) + w(bn + 1) \\
&= w(n) + c + w(bn + 1)
\end{aligned}
$$

and

$$
\begin{aligned}
w(b(bn^2) - 1)) &= w(bn^2) + c \\
&= w(b) + 2 \cdot w(n) + c
\end{aligned}
$$

Hence $w(bn+1) = w(n)+w(b)$, for all $n \in \mathbb{N}_{>0}$. By Case 1 it then follows that $w$ is trivially additive. □

To establish the direction $\subseteq$, we have shown that if $w$ is completely additive and $\langle a, b\rangle \in \mathcal{AS}(w)$ for some $a \neq b$, then $w$ is trivially additive. □

**Remark 3.9.** We were pleasantly surprised to receive an e-mail from Kevin Hare with the beautiful proof of Theorem 3.8 [13]. At the time we had some partial results: we had a proof of the statement for finitely prime generated sequences $\sigma$, and also for some specific infinitely prime generated sequences, namely those produced by one-gap Toeplitz patterns (see Sections 4 and 5). For these partial results we refer to the first version of the current arXiv report, available at the following url: `http://arxiv.org/pdf/1201.3786v1`. Finally, we had also established the statement for completely additive sequences over the infinite cyclic group $\mathbb{Z}$.

# 4 Toeplitz Words

Toeplitz words were introduced in [15], see also [1, 6]. A Toeplitz word over an alphabet $\Sigma$ is an infinite sequence iteratively constructed as follows: Given is a starting sequence $\sigma_0$ over the alphabet $\Sigma \cup \{?\}$, where we may think of the symbol $? \notin \Sigma$ as 'undefined', and of $\sigma_0$ as a 'partially defined' sequence. The occurrences of ? in a sequence $\sigma$ are called the 'gaps' of $\sigma$. For $i > 0$ the sequence $\sigma_i$ is obtained from $\sigma_{i-1}$ by 'filling its sequence of gaps' (i.e., consecutively replacing the occurrences of ? in $\sigma_{i-1}$) by the sequence $\sigma_{i-1}$ itself, as made precise in Definition 4.1 below. In the limit we then obtain a totally defined sequence (i.e., without gaps) if and only if the first symbol of the start sequence $\sigma_0$ is defined (i.e., is in $\Sigma$).

As in [1] we allow the application of any bijective map $f : \Sigma \to \Sigma$ to the symbols that replace the gaps. Thus we let the above $\sigma_i$ be sequences over $\Sigma \cup \mathfrak{S}_\Sigma$ where $\mathfrak{S}_\Sigma$ denotes the symmetric group of bijections (or permutations) on $\Sigma$, and let the elements from $\mathfrak{S}_\Sigma$ play the role of gaps. Filling a gap $f$ with a letter $a \in \Sigma$ then results in $f(a)$, and filling a gap $f$ by a gap $g$ results in the function composition $f \circ g$. Viewed in this way the symbol ? stands for $\mathrm{id}_\Sigma$, the identity element of $\mathfrak{S}_\Sigma$, and we will use it in that way. From the finiteness of the alphabet $\Sigma$ and the $f$'s being one-to-one it directly follows that extending the set of Toeplitz patterns does *not* increase the expressive power of the system [1]: every Toeplitz word generated by a pattern with gaps from $\mathfrak{S}_\Sigma$ can already be defined using a (longer) pattern with gaps ? only. In other words, it is a conservative extension.

**Definition 4.1.** For infinite words $\sigma, \tau \in (\Sigma \cup \mathfrak{S}_\Sigma)^{\mathbb{N}_{>0}}$ we define $\sigma[\tau]$ recursively by

$$(a\,x)[y] = a\,x[y] \qquad (f\,x)[b\,y] = f(b)\,x[y] \qquad (f\,x)[g\,y] = (f \circ g)\,x[y]$$

where $a, b \in \Sigma$, $f, g \in \mathfrak{S}_\Sigma$, and $x, y \in (\Sigma \cup \mathfrak{S}_\Sigma)^{\mathbb{N}_{>0}}$. Further let $P \in \Sigma(\Sigma \cup \mathfrak{S}_\Sigma)^*$ (first symbol not a gap) and, for $k = 0, 1, 2, \ldots$, define $T_k(P) \in (\Sigma \cup \mathfrak{S}_\Sigma)^{\mathbb{N}_{>0}}$ by

$$T_0(P) = ?^\omega = ???\ldots \qquad\qquad T_{k+1}(P) = P^\omega[T_k(P)]$$

Then $T(P) \in \Sigma^{\mathbb{N}_{>0}}$, the *Toeplitz word generated by* $P$, is defined as the limit of these words, as follows:

$$T(P) = \lim_{k\to\infty} T_k(P)$$

We let $|P|_? = |\{h \mid P(h) \in \mathfrak{S}_\Sigma\}|$ denote *the number of gaps in* $P$, and, following [6], we call $T(P)$ a Toeplitz word *of type* $\langle r, q\rangle$ when $r = |P|$ and $q = |P|_?$.

**Example 4.2.** Let $P = 0?1? \in (\Sigma_2 \cup \mathfrak{S}_{\Sigma_2})^*$. The sequence of '$P$-generations' $T_0(P), T_1(P), \ldots$ starts as follows:

$$?^\omega,\ (0?1?)^\omega,\ (001?011?)^\omega,\ (0010011?0011011?)^\omega,\ \ldots$$

The limit of this sequence of sequences is the well-known regular paperfolding sequence: $T(P) = \boldsymbol{f}$, see [1]; $\boldsymbol{f}$ is additive and generated by the (infinite) set of prime numbers congruent to 3 modulo 4.

**Proposition 4.3.** *The set* $(\Sigma \cup \mathfrak{S}_\Sigma)^{\mathbb{N}_{>0}}$ *with the operation* $\langle \sigma, \tau\rangle \mapsto \sigma[\tau]$ *and identity element* $?^\omega$ *forms a monoid.*

The Toeplitz word $T(P)$ is the unique solution of $x$ in the equation $x = P^\omega[x]$. The construction can thus be viewed as 'self-reading' in the sense that the sequence under construction itself is read to fill the gaps. For example, the period doubling sequence $\boldsymbol{p}$ is the Toeplitz word over $\Sigma_2$ generated by the pattern 010?, as follows:

$$(010?)^\omega = 010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}010\mathbf{?}\ldots$$
$$\boldsymbol{p} = T(010?) = 010\mathbf{0}010\mathbf{1}010\mathbf{0}010\mathbf{0}010\mathbf{0}010\mathbf{1}010\mathbf{0}010\mathbf{1}010\mathbf{0}010\mathbf{1}010\mathbf{0}010\mathbf{0}\ldots$$

For $d \in \Sigma$ let us denote by $?^{+d}$ the rotation $a \mapsto a +_\Sigma d$ (so $? = ?^{+0}$). Then the pattern 010? can be simplified to $0?^{+1}$, because $0?^{+1} =_T 010?$, where the relation $=_T$ is defined by:

$$P =_T Q \quad \text{if and only if} \quad T(P) = T(Q)$$

We call the elements of the set $(\Sigma \cup \mathfrak{S}_\Sigma)^+$ $\Sigma$-*patterns*. We continue with a definition of a composition operation $\circ$ on $\Sigma$-patterns $P, Q$ such that $(P \circ Q)^\omega = P^\omega[Q^\omega]$. This then explains the equivalence $0?^{+1} =_T 010?$ since $(0?^{+1}) \circ (0?^{+1}) = 010?$ and because the equivalence classes induced by $=_T$ are closed under composition. The idea of composing patterns $P, Q$ is to first take copies $P^n$ and $Q^m$ such that $|P^n|_? = |Q^m|$ and then fill the sequence of gaps through $P^n$ by $Q^m$. Recall that we defined $\xi(n,m) = \frac{m}{\gcd(n,m)}$ (so $n \cdot \xi(n,m) = m \cdot \xi(m,n)$).

**Definition 4.4.** Let $P, Q$ be $\Sigma$-patterns, and define their *Toeplitz composition* as follows:

$$P \circ Q = \begin{cases} P & \text{if } |P|_? = 0 \\ P^{d_1}[Q^{d_2}] & \text{if } |P|_? > 0,\ d_1 = \xi(|P|_?, |Q|), \text{ and } d_2 = \xi(|Q|, |P|_?). \end{cases}$$

where $u[v]$ is defined for all words $u, v \in (\Sigma \cup \mathfrak{S}_\Sigma)^*$ such that $|u|_? = |v|$ by

$$x[\varepsilon] = x \quad (a\,x)[y] = a\,x[y] \quad (f\,x)[b\,y] = f(b)\,x[y] \quad (f\,x)[g\,y] = (f \circ g)\,x[y]$$

where $a, b \in \Sigma$, $f, g \in \mathfrak{S}_\Sigma$, and $x, y \in (\Sigma \cup \mathfrak{S}_\Sigma)^*$.[4]

For a $\Sigma$-pattern $P$ and integer $k \geq 0$ we define $P^{(k)}$ by $P^{(0)} = ?$ and $P^{(n+1)} = P \circ P^{(n)}$.

**Example 4.5.** To compute $(0??1??) \circ (234567)$ we take $d_1 = 3$ and $d_2 = 2$:

$$(0??1??) \circ (234567) = (0??1??)^3[(234567)^2] = 023145\,067123\,045167$$

**Proposition 4.6.** *The set of $\Sigma$-patterns forms a monoid with pattern composition as its operation and ? as its identity element.*

**Lemma 4.7.** *The map $P \mapsto P^\omega$ is a monoid homomorphism from $\langle(\Sigma \cup \mathfrak{S}_\Sigma)^*, \circ, ?\rangle$ to $\langle(\Sigma \cup \mathfrak{S}_\Sigma)^{\mathbb{N}_{>0}}, \langle\sigma,\tau\rangle \mapsto \sigma[\tau], ?^\omega\rangle$.*

This immediately implies $(P^{(k)})^\omega = T_k(P)$, and hence $T(P) = \lim_{k\to\infty} P^{(k)}$.

The length and number of gaps of a composed pattern $P \circ Q$ are computed as follows:

**Lemma 4.8.** $|P \circ Q| = \xi(|P|_?, |Q|) \cdot |P|$ *and* $|P \circ Q|_? = \xi(|Q|, |P|_?) \cdot |Q|_?$

The congruence classes induced by $=_T$ are closed under concatenation and composition.

**Lemma 4.9.** *Let $P$ be a $\Sigma$-pattern and $k \geq 1$. Then $P^k =_T P$ and $P^{(k)} =_T P$.*

[4] Note that the recursive calls of $u[v]$ preserve the requirement $|u|_? = |v|$.

**Example 4.10.** The classical Hanoi sequence [17, A101607] is the sequence $\boldsymbol{h}$ of moves such that the prefix of length $2^N - 1$ of $\boldsymbol{h}$ transfers $N$ disks from peg $A$ to peg $B$ if $N$ is odd, and to peg $C$ if $N$ is even. We represent moving the topmost disk from peg $X$ to $Y$ by the pair $\langle X, Y\rangle$, and map moves to $\Sigma_6$ using

$$\begin{array}{lll} \langle A,B\rangle \mapsto 0 & \langle B,C\rangle \mapsto 2 & \langle C,A\rangle \mapsto 4 \\ \langle B,A\rangle \mapsto 1 & \langle C,B\rangle \mapsto 5 & \langle A,C\rangle \mapsto 3 \end{array}$$

In [2, 1] it is shown that the sequence $\boldsymbol{h}$ is the Toeplitz word generated by the pattern 032?450?214? :

$$\begin{aligned} \boldsymbol{h} &= T(032?450?214?) \\ &= 03204503214203204504214503204503214203214504214 2\ldots \end{aligned}$$

Now let $f = ?^{+3}$, the involution on $\Sigma_6$ corresponding to swapping $B$s and $C$s in moves $\langle X, Y\rangle$. Then the above pattern can be simplified to $0f2f4f$, that is,

$$\boldsymbol{h} = T(0f2f4f)$$

as composing the pattern $0f2f4f$ with itself yields the pattern 032?450?214? from above:

$$(0f2f4f) \circ (0f2f4f) = 032f^2450f^2214f^2 = 032?450?214?$$

The sequence of directions obtained by taking $\boldsymbol{h}$ modulo 2 (we took even (odd) numbers to represent (counter)clockwise moves) is the period doubling sequence $\boldsymbol{p}$, see [1]:

$$(\boldsymbol{h} \bmod 2) = (T(0f2f4f) \bmod 2) = T(0?^{+1}0?^{+1}0?^{+1}) = T(0?^{+1}) = \boldsymbol{p}$$

The recurrence equations for a Toeplitz word are easy to establish.

**Lemma 4.11.** *Let $P = a_1 \ldots a_r$ be a $\Sigma$-pattern with $a_1 \in \Sigma$, $r \geq 2$. Let $h_1 < h_2 < \ldots < h_q$ be the sequence of indices $h$ such that $a_h \in \mathfrak{S}_\Sigma$ (so $P$ has $q < r$ gaps). Then for all $n \in \mathbb{N}$:*

$$\begin{array}{rll} T(P)(rn + i) &= a_i & \textit{if } 1 \leq i \leq r \textit{ and } a_i \in \Sigma\,, \textit{ and} \\ T(P)(rn + h_j) &= f(T(P)(qn + j)) & \textit{for all } 1 \leq j \leq q \textit{ with } a_{h_j} = f\,. \end{array}$$

**Example 4.12.** Consider the $\Sigma$-pattern $P = afbgh$ for some $a, b \in \Sigma$ and $f, g, h \in \mathfrak{S}_\Sigma$. With Lemma 4.11 we obtain the following recurrence equations for $\sigma = T(P)$, for all $n \in \mathbb{N}$:

$$\begin{array}{lll} \sigma(5n+1) = a & \sigma(5n+2) = f(\sigma(3n+1)) & \sigma(5n+3) = b \\ \sigma(5n+4) = g(\sigma(3n+2)) & \sigma(5n+5) = h(\sigma(3n+3)) & \end{array}$$

Toeplitz words of type $\langle r, 1\rangle$ can be obtained by iterating an $r$-uniform morphism [6], whence they are $r$-automatic [4].

**Proposition 4.13.**

*(i)* *Let* $P \in \Sigma(\Sigma \cup \mathfrak{S}_\Sigma)^*$ *and define* $h : \Sigma^* \to \Sigma^*$ *by* $h(a) = P \circ a$ $(a \in \Sigma)$*. Then* $h^\omega(a) = T(P)$*.*

*(ii)* *Let* $h : \Sigma^* \to \Sigma^*$ *be the morphism defined by* $h(a) = buf(a)v$ $(a \in \Sigma)$ *for some fixed* $b \in \Sigma$*,* $u, v \in \Sigma^*$*, and* $f \in \mathfrak{S}_\Sigma$*. Define the* $\Sigma$*-pattern* $P$ *by* $P = bufv$*. Then* $T(P) = h^\omega(b)$*.*

# 5 Additive Toeplitz Words Generated by Single-Gap Patterns

We characterize additive Toeplitz words of type $\langle r, 1\rangle$. More precisely, we characterize the set $X$ of one-gap Toeplitz patterns $P$ such that $T(P)$ is additive if and only if $P$ is in $X$. As it turns out, the discrete logarithm to the base $g$ modulo a prime number $p$ (see Section 2) plays a key role in the construction of these patterns.

We adopt the following convention: Whenever a $\Sigma$-pattern $P$ generates a non-surjective Toeplitz word (i.e., $T(P)(\mathbb{N}_{>0}) \subsetneq \Sigma$), we identify gaps $f$ occurring in $P$ with all bijections that coincide with $f$ on the letters occurring in $T(P)$.

**Lemma 5.1.** *Let* $P = a_1 a_2 \dots a_\ell$ *such that* $T(P)$ *is non-trivially additive. Then* $a_1 = 0$ *and* $a_\ell = ?^{+d}$ *for some* $d \in \Sigma$*.*

*Proof.* Let $\sigma = T(P)$. Additive sequences have intial value 0, so $\sigma(1) = a_1 = 0$. Moreover if $a_\ell \in \Sigma_k$ then it follows from the definition of Toeplitz words that $\sigma/\ell$ is a constant sequence: $\sigma(n\ell) = a_\ell$ for all $n \in \mathbb{N}_{>0}$. On the other hand, we also have $\sigma/\ell \sim \sigma$ by complete additivity of $\sigma$, that is, $\sigma(n\ell) = \sigma(n) + \sigma(\ell)$ for all $n \in \mathbb{N}_{>0}$. This combination of facts only occurs if $\sigma$ is constant zero, contradicting the assumption that $\sigma$ is non-periodic. Finally, by additivity of $\sigma$ it follows that $\sigma/\ell = \sigma + \sigma(\ell)$ and hence the bijection at position $\ell$ has to be a rotation for all elements in the image $\sigma(\mathbb{N}_{>0})$. ☐

We let $\mathfrak{P}$ denote the set of one-gap $\Sigma$-patterns generating additive sequences:

$$\mathfrak{P} = \{\, P \in (\Sigma \cup \mathfrak{S}_\Sigma)^* \mid T(P) \text{ is additive and } |P|_? = 1 \,\}$$

**Lemma 5.2.** *Let* $P \in \mathfrak{P}$ *with* $|P| = nm$ *for some* $n, m \geq 2$*. Then the arithmetic subsequence* $T(P)_{i,n}$ *is constant for every* $i$ *with* $1 \leq i < n$*.*

*Proof.* Let $\sigma = T(P)$. By Lemma 5.1 we know that the (only) gap of $P$ is at the end. Hence $a_i \in \Sigma$ and from Lemma 4.11 it follows that $\sigma_{im,nm} = (\sigma_{m,m})_{i,n}$ is constant. Moreover $(\sigma_{m,m})_{i,n} = (\sigma + \sigma(m))_{i,n} = \sigma_{i,n} + \sigma(m)$ by additivity of $\sigma$. Hence also $\sigma_{i,n}$ is constant. ☐

**Lemma 5.3.** *Let* $P \in \mathfrak{P}$ *with* $|P|$ *not a prime power. Then* $T(P)$ *is the constant zero sequence.*

*Proof.* Let $P$ be non-trivial and $|P| = nm$ for some $n, m > 1$ with $\gcd(n, m) = 1$ and $\sigma = T(P)$. From Lemma 5.2 it follows that all subsequences $\sigma_{i,n}$ and $\sigma_{j,m}$ are constant for $1 \leq i < n$ and $1 \leq j < m$. Since $n$ and $m$ are relatively prime we have that for every $j < m$ there is a $k \in \mathbb{N}_{>0}$ such that $kn + 1 \equiv j \pmod{m}$. Hence the subsequence $\sigma_{1,n}$ contains elements of every subsequence $\sigma_{j,m}$ $(1 \leq j < m)$, and thus $\sigma_{1,m} = \sigma_{2,m} = \ldots = \sigma_{m-1,m}$. Similarly we obtain $\sigma_{1,n} = \sigma_{2,n} = \ldots = \sigma_{n-1,n}$. Consequently we have $\sigma(k) = 0$ if $k \not\equiv 0 \pmod{m}$ or $k \not\equiv 0 \pmod{n}$. Hence $\sigma(k) = 0$ if $k \not\equiv 0 \pmod{\operatorname{lcm}(n, m)}$. From Lemma 5.1 we know that the only gap is at position $nm$, and since $\operatorname{lcm}(n, m) = nm = |P|$ we get $P(k) = 0$ for all $1 \leq k < nm$. Finally the gap at $nm$ has to map 0 to 0 because $\sigma(nm) = \sigma(n) + \sigma(m) = 0$. □

**Lemma 5.4.** *Let $P = a_1 a_2 \ldots a_r \in \mathfrak{P}$ with $r = p^k$ for some prime $p$ and $k \geq 2$, and $T(P) \neq 0^\omega$. Then $P = Q^{(k)}$ with $Q = a_1 a_2 \ldots a_{p-1} ?^{+a_p} \in \mathfrak{P}$. (Hence $P =_T Q$ by Lemma 4.9.)*

*Proof.* With Lemma 5.1 we have that $a_1, \ldots, a_{p^k-1} \in \Sigma$ and $a_{p^k} \in \mathfrak{S}_\Sigma$ $(*)$. From Lemma 5.2 we know that all subsequences $\sigma_{i,p}$ with $1 \leq i < p$ are constant. Hence, using $(*)$, $P$ and $Q^{(k)}$ coincide on all positions $j \not\equiv 0 \pmod{p}$. Also, by $T(P)$ being additive and the definition of pattern composition (Definition 4.4), we have $P(mp) = a_m + a_p = Q^{(k)}(mp)$ for every $1 \leq m < p$. Finally, $P(p^k) = a_p + a_{p^{k-1}} = \ldots = k \cdot a_p = Q^{(k)}(p^k)$. □

So far we have shown that all one-gap patterns which generate non-trivial additive Toeplitz words have the gap at the end (Lemma 5.1), of prime power length (Lemma 5.3), and can always be decomposed to a pattern of prime length (Lemma 5.4). Next we give the exact shape of these atomic patterns, which are defined using discrete logarithms.

**Definition 5.5.** Let $g$ be a primitive root of some prime $p > 2$. We define words $\lambda_{p,g} \in \Sigma_{p-1}^*$ and $\lambda_{2,1} \in \Sigma_1^*$ by

$$\lambda_{p,g} = 0 \, \log_g(2) \, \log_g(3) \, \ldots \, \log_g(p-1) \qquad\qquad \lambda_{2,1} = 0$$

We show that every atomic pattern in $\mathfrak{P}$ is of the form $(\lambda_{p,g} \odot c)?^{+d}$ for some $c, d \in \Sigma$.

**Theorem 5.6.** *Let $P \in \mathfrak{P}$ with $|P| = p$ for some prime $p$, $T(P)$ be non-trivial, and $g$ a primitive root modulo $p$. Then $P = (\lambda_{p,g} \odot P(g))\, ?^{+d}$ for some $d \in \Sigma$.*

*Proof.* By Lemma 5.1 we know that the only gap is at the end and that it is a rotation. Let $1 \leq i < p$. By Lemma 4.11 and additivity of $T(P)$ we obtain $T(P)(g^{\lambda_{p,g}(i)}) = \lambda_{p,g}(i) \cdot P(g)$. □

**Example 5.7.** Let $\Sigma = \Sigma_4$, and $P = 002022?^{+3}$. This gives rise to the following completely additive sequence[5]:

$$T(P) = 0020223\,0020223\,0020221\,0020223\,0020221\,0020221\,0020222\,\ldots$$

[5] Additivity of $T(P)$ can be checked with Lemma 5.10.

Theorem 5.6 gives

$$P = (\lambda_{7,3} \odot P(3))\,?^{+3} = (021453 \odot 2)\,?^{+3}$$
$$P = (\lambda_{7,5} \odot P(5))\,?^{+3} = (045213 \odot 2)\,?^{+3}$$

The question remains whether every pattern of the form $P = (\lambda_{p,g} \odot c)?^{+d}$ gives rise to an additive sequence $T(P)$. The following example shows this is not the case. Theorem 5.12 formulates the exact requirement for $c$ so that $T(P)$ is additive.

**Example 5.8.** Let $\Sigma = \Sigma_4$, and $Q = (\lambda_{7,3} \odot 3)\,?^{+3} = (021453 \odot 3)\,?^{+3} = 023031\,?^{+3}$. Then the sequence $T(Q) = 0230313\,0230311\ldots$ is not additive: $T(Q)(8) \neq T(Q)(4) + T(Q)(2)$.

We note that an additive sequence over $\Sigma_n$ (with $n \in \mathbb{N}_{>0}$) is surjective if and only if it contains the element $1_{\Sigma_n}$.

**Lemma 5.9.** *Let $n, m \in \mathbb{N}_{>0}$, $\sigma \in \Sigma_n^\omega$ an additive sequence, and $c \in \Sigma_m$. If $c$ is divisible by $\xi(n,m)$, then $\sigma \odot c \in \Sigma_m^\omega$ is an additive sequence. If, moreover, the sequence $\sigma$ is surjective, then the converse direction holds as well.*

*Proof.* For all $i, j \in \mathbb{N}$ we have $(\sigma \odot c)(ij) = \sigma(ij) \odot c = (\sigma(i) + \sigma(j)) \odot c = \sigma(i) \odot c + \sigma(j)) \odot c = (\sigma \odot c)(i) + (\sigma \odot c)(j)$ since $\sigma$ is additive and by Lemma 2.1.

For the converse direction, assume that the sequence $\sigma$ is surjective and $\sigma \odot c$ additive. There exist $i, j \in \mathbb{N}$ such that $\sigma(i) = 1$ and $\sigma(j) = n - 1$. Then $\sigma(ij) = \sigma(i) + \sigma(j) = 0$. As a consequence $\sigma(ij) \odot c = \sigma(i) \odot c + \sigma(j) \odot c = 0$. Then $(i +_{\mathbb{N}} j) \odot c = n \odot c = 0$. Hence $n \cdot_{\mathbb{N}} c \equiv 0 \pmod{m}$, and thus $c$ must be a multiple of $\xi(n,m) = \frac{m}{\gcd(n,m)}$. ☐

**Lemma 5.10.** *Let $w \in \Sigma^*$ such that $p = |w| + 1$ is prime, and $d \in \Sigma$. Then $w\,?^{+d} \in \mathfrak{P}$ if and only if $w(k) = w(i) + w(j)$ for all $i, j, k$ with $0 < i, j, k < p$ such that $k \equiv i \cdot j \pmod{p}$.*

*Proof.* For the direction '⇒', let $P = w\,?^{+d} \in \mathfrak{P}$. Let $0 < i, j, k < p$ such that $k \equiv i \cdot j \pmod{p}$. Then $w(k) = T(P)(k) = T(P)(ij)$ by Lemma 4.11 since $k \equiv i \cdot j \pmod{p}$. Moreover, we have $T(P)(ij) = T(P)(i) + T(P)(j) = w(i) + w(j)$. Hence $w(k) = w(i) + w(j)$.

For the direction '⇐', let $P = w\,?^{+d}$, and assume $w(k) = w(i) + w(j)$ for all $0 < i, j, k < p$ such that $k \equiv i \cdot j \pmod{p}$. We show additivity of $\sigma = T(P)$. Let $n, m \in \mathbb{N}_{>0}$, we distinguish two cases:

(i) Case: $p \nmid nm$. There exist $n', m', k' \in \Sigma_{p-1}$ such that $n' \equiv n \pmod{p}$, $m' \equiv m \pmod{p}$ and $k' \equiv nm \pmod{p}$. We have:

$$\begin{aligned}
\sigma(nm) &= \sigma(k') && \text{by Lemma 4.11 and } k' \equiv nm \pmod{p}\\
&= w(k')\\
&= w(n') + w(m')\\
&= \sigma(n') + \sigma(m')\\
&= \sigma(n) + \sigma(m) && \text{by Lemma 4.11}
\end{aligned}$$

(ii) Case: $p \mid n$ or $p \mid m$. We use induction on the exponent of the prime factor $p$ in $nm$. For symmetry assume $p \mid n$ (the other case follows analogously). We have:

$$
\begin{aligned}
\sigma(nm) &= \sigma(\frac{n}{p}m) + d && \text{by Lemma 4.11} \\
&= \sigma(\frac{n}{p}) + \sigma(m) + d && \text{by induction hypothesis} \\
&= \sigma(n) + \sigma(m) && \text{by Lemma 4.11}
\end{aligned}
$$

This concludes the proof. □

**Lemma 5.11.** *The Toeplitz word $T(\lambda_{p,g}\, ?^{+d})$ is additive for every prime $p$, primitive root $g$ modulo $p$ and $d \in \Sigma_{p-1}$.*

*Proof.* Let $p$ be a prime, $g$ a primitive root modulo $p$, $d \in \Sigma$, and $P = \lambda_{p,g}\, ?^{+d}$. For $P \in \mathfrak{P}$ it suffices to check $\lambda_{p,g}(k) = \lambda_{p,g}(i) + \lambda_{p,g}(j)$ for all $0 < i, j, k < p$ such that $k \equiv i \cdot j \pmod{p}$ by Lemma 5.10. This property follows from $\lambda_{p,g}$ being the discrete logarithm modulo $p$. □

**Theorem 5.12.** *Let $p$ be prime, $g$ a primitive root modulo $p$, and $c, d \in \Sigma$. Then we have that the pattern $(\lambda_{p,g} \odot c)\, ?^{+d} \in \mathfrak{P}$ if and only if $\xi(p-1, |\Sigma|)$ divides $c$.*

*Proof.* First, we observe that $(*)$ $T(\lambda_{p,g}\, ?^{+d}) \odot c = T(\lambda_{p,g} \odot c\, ?^{+d\odot c})$, and by Lemma 5.10 we have $(\lambda_{p,g} \odot c)\, ?^{+d\odot c} \in \mathfrak{P}$ if and only if $(\lambda_{p,g} \odot c)\, ?^{+d} \in \mathfrak{P}$. For the direction '$\Leftarrow$', let $c \in \Sigma$ such that $\xi(p-1, |\Sigma|)$ divides $c$. Then $T(\lambda_{p,g}\, ?^{+d}) \odot c$ is additive by Lemmas 5.11 and 5.9, which implies the claim by $(*)$. For the direction '$\Rightarrow$', let $(\lambda_{p,g} \odot c)\, ?^{+d} \in \mathfrak{P}$. By $(*)$ we obtain that $T(\lambda_{p,g}\, ?^{+d}) \odot c$ is additive, and by Lemma 5.10 together with surjectivity of $T(\lambda_{p,g}\, ?^{+d})$ it follows that $\xi(p-1, |\Sigma|)$ divides $c \in \Sigma$. □

Theorem 5.13 summarizes the results of this section.

**Theorem 5.13.** *For every $p \in \mathbf{P}$, let $g(p)$ denote a primitive root modulo $p$. Then we have*

$$
\begin{aligned}
\mathfrak{P} &\stackrel{def}{=} \{\, P \in (\Sigma \cup \mathfrak{S}_\Sigma)^* \mid T(P) \text{ is additive and } |P|_? = 1 \,\} \\
&= \left\{\, ((\lambda_{p,g(p)} \odot c)\, ?^{+d})^{(k)} \;\middle|\; p \in \mathbf{P},\ k \geq 1,\ c, d \in \Sigma,\ \xi(p-1, |\Sigma|) \mid c \,\right\} \\
&\quad \cup \{\, 0^+ \, ? \, 0^* \,\}
\end{aligned}
$$

For additive $\langle r, 1\rangle$-type Toeplitz words we can now give the exact generator $\mu : \mathbf{P} \to \Sigma$.

**Corollary 5.14.** *Let $T(P)$ be additive with $|P|_? = 1$, so that $P = (\lambda_{p,g} \odot c)\, ?^{+d}$ as in Theorem 5.12. Then $T(P) = \boldsymbol{v}_\mu$ holds, where $\mu : \mathbf{P} \to \Sigma$ is defined for all $q \in \mathbf{P}$ by*

$$
\mu(q) = \begin{cases} \lambda_{p,g}(i) \odot c & \text{if } q \equiv i \pmod{p}, \text{ for some } i \text{ with } 1 \leq i < p \\ d & \text{if } q = p \end{cases}
$$

*The sequence $T(P)$ is infinitely prime generated if and only if $c \neq 0$ and $p > 2$, by Dirichlet's Theorem on arithmetic progressions [10].*

**Example 5.15.** The pattern $P = 001011?^{+1}$ derived from $\lambda_{7,3}$ yields an additive binary sequence whose underlying (infinite) prime set is $X = \mathbf{P} \cap [3,5,6]_7 \cup \{7\}$, that is, $T(P) = \boldsymbol{v}_X$.

**Example 5.16.** Toeplitz patterns for single prime generated sequences have a simple shape: Let $p \in \mathbf{P}$, $d \in \Sigma$ and define $\mu(p) = d$ and $\mu(q) = 0$ for $q \neq p$. Then $\boldsymbol{v}_\mu = T(0^{p-1}?^{+d})$ by Corollary 5.14 (take $c = 0_\Sigma$). E.g., for the period doubling sequence $\boldsymbol{p} = \boldsymbol{v}_2$ we have $\boldsymbol{p} = T(0?^{+1})$.

## 6 Toeplitz Permutations

There is a strong connection between a Toeplitz pattern $P$ and the arithmetic progressions through $T(P)$. For $|P|_? = 1$ we show that every subsequence $T(P)_{a,b}$ is a Toeplitz word $T(P_{a,b})$ where $P_{a,b}$ is derived from $P$. Thus the classification of the arithmetic self-similarity of $\langle r, 1\rangle$ Toeplitz words is reduced to a problem of analyzing patterns. And, using the results of Section 5, we conclude that $\mathcal{AS}(\sigma) = \{ \langle b, b\rangle \mid b \in \mathbb{N}_{>0} \}$ for all non-periodic additive Toeplitz words $\sigma \in \Sigma^{\mathbb{N}_{>0}}$ of type $\langle r, 1\rangle$.

**Lemma 6.1.** *Fix $b, r \in \mathbb{N}_{>0}$ s.t. $\gcd(b, r) = 1$ and $r \geq 2$. Then $\exists\, c, m \in \mathbb{N}_{>0}$ s.t. $cb = r^m - 1$.*

*Proof.* Suppose that $b \nmid r^m - 1$ for all $m \in \mathbb{N}_{>0}$. Then by the pigeon hole principle there must be some $p, s, v \in \mathbb{N}_{>0}$ such that

$$r^p - 1 \equiv v \pmod{b}$$
$$r^{p+s} - 1 \equiv v \pmod{b}$$

It follows that:

$$b \mid r^{p+s} - r^p$$
$$b \mid r^p(r^s - 1)$$
$$b \mid r^s - 1$$

So then $b$ divides $r^s - 1$, contradicting the assumption. □

In the sequel we index Toeplitz patterns $P$ (and sequences $T(P)$) by positive integers, i.e., $P = P(1)P(2)\ldots P(r)$ with $r = |P|$. We write $P(k)$ to denote $P^\omega(k) = P(k \mod r)$ where we take $1, 2, \ldots, r$ as the representatives of the congruence classes modulo $r$.

**Definition 6.2.** Let $P$ be a one-gap $\Sigma$-pattern of length $r \geq 2$ with its single gap $f \in \mathfrak{S}_\Sigma$ at the end. Fix $a, b \in \mathbb{N}_{>0}$ such that $a \leq b$ and $\gcd(b, r) = 1$. Let $c, m \in \mathbb{N}_{>0}$ with $m$ minimal such that $cb = r^m - 1$ (Lemma 6.1). The *arithmetic permutation $P_{a,b}$ of $P$* is defined by

$$P_{a,b} = P^{(m)}(a)P^{(m)}(a+b)\ldots P^{(m)}(a + b(r^m - 1))$$

Note that the pattern $P_{a,b}$ is a *permutation* of $P^{(m)}$ (so $|P_{a,b}|_? = 1$) because $b$ is a generator of the additive group $\{1, 2, \ldots, r^m\}$. The following lemma generalizes [8, Theorem 3.8].

Recall that we index Toeplitz patterns $P$ (and sequences $T(P)$) by positive integers, i.e., $P = P(1)P(2)\ldots P(r)$ with $r = |P|$. We write $P(k)$ to denote $P^\omega(k) = P(k \mod r)$ where we take $1, 2, \ldots, r$ as the representatives of the congruence classes modulo $r$.

**Lemma 6.3.** *Let $P \in (\Sigma \cup \mathfrak{S}_\Sigma)^+$ be a pattern of length $r \geq 2$ with its single gap $f$ at the end. Let $a, b \in \mathbb{N}_{>0}$ such that $a \leq b$ and $\gcd(b, r) = 1$. Then $T(P)_{a,b} = T(P_{a,b})$.*

*Proof.* Let $c, m \in \mathbb{N}_{>0}$ with $m$ minimal such that $bc = r^m - 1$ so that $P_{a,b}$ is the permutation

$$P_{a,b} = P^{(m)}(a)P^{(m)}(a+b)\ldots P^{(m)}(a+(r^m-1)b)$$

of $P^{(m)} = P^{(m)}(1)P^{(m)}(2)\ldots P^{(m)}(r^m)$ as in Definition 6.2. Recall that $T(P) = T(P^{(m)})$ (Lemma 4.9), and that the single gap $f^m$ of $P^{(m)}$ is at the end, $P^{(m)}(r^m) = f^m$. Let $j$ be the index $1 \leq j \leq r^m$ such that $P_{a,b}(j) = f^m$. From $P_{a,b}(j) = P^{(m)}(a+(j-1)b)$ we infer $a + (j-1)b \equiv r^m \pmod{r^m}$. Note that $j$ is unique with this property among $1, 2, \ldots, r^m$. We also have that $a + ((ac+1)-1))b = a + acb = a + a(r^m - 1) = ar^m$ with $1 \leq ac + 1 \leq r^m$, and so $a + (j-1)b = ar^m$ and $j = ac + 1$. We prove

$$T(P)_{a,b}(n) = T(P_{a,b})(n)$$

by induction on $n \in \mathbb{N}$. (We tacitly make use of Lemma 4.11.)

Let $n = sr^m + i$ for some $s \in \mathbb{N}$ and $1 \leq i \leq r^m$. Consider the base case $n \leq r^m$, that is, $s = 0$. For $i \neq j$, the claim follows from the construction of $P_{a,b}$. For $i = j$ we find

$$\begin{aligned}
T(P)_{a,b}(j) &= T(P^{(m)})_{a,b}(j) \\
&= T(P^{(m)})(a + (j-1)b) \\
&= T(P^{(m)})(ar^m) \\
&= f^m(T(P^{(m)})(a)) \\
&= T(P_{a,b})(j)
\end{aligned}$$

Now let $s \geq 1$. For $i \neq j$ we obtain

$$\begin{aligned}
T(P)_{a,b}(n) &= T(P)(a + b(n-1)) \\
&= T(P)(a + b(n - 1 - r^m)) \\
&= T(P)_{a,b}(n - r^m)) \\
&\overset{\text{IH}}{=} T(P_{a,b})(n - r^m) \\
&= T(P_{a,b})(n)
\end{aligned}$$

If $i = j = ac + 1$ we find

$$\begin{aligned}
T(P)_{a,b}(n) &= T(P)(a + b(n-1)) \\
&= T(P)(a + b(sr^m + i - 1)) \\
&= T(P)(a + b(sr^m + ac)) \\
&= T(P)(a + acb + sbr^m) \\
&= T(P)(a + a(r^m - 1) + sbr^m) \\
&= T(P)((a + sb)r^m) \\
&= f^m(T(P)(a + sb)) \\
&= f^m(T(P)_{a,b}(s + 1)) \\
&\overset{\text{IH}}{=} f^m(T(P_{a,b})(s + 1)) \\
&= T(P_{a,b})(sr^m + j) \\
&= T(P_{a,b})(n) \qquad \square
\end{aligned}$$

**Example 6.4.** Let $P = 0123f$ with $f(n) = n + 2 \mod 5$ and $\sigma = T(P)$. Then $\sigma_{2,2} = T(P_{2,2})$ with $P_{2,2} = 1302f$, and $\sigma_{2,12} = T(P_{2,12})$ with $P_{2,12} = 1302f^2 13020130221302413021$; for the latter note that $cb = 2 \cdot 12 = 5^2 - 1 = r^m - 1$ and $P^{(2)} = 012320123401234012300123f^2$.

**Lemma 6.5.** *Let $Q$ be a one-gap Toeplitz pattern of length $r$ of the form $Q = ufv$ for some $f \in \mathfrak{S}_\Sigma$, $u \in \Sigma^+$ and $v \in \Sigma^*$. Then $Q = P_{a,b}$ for $P = u^R v^R f$, $a = |u| \leq b = |Q| - 1$.*

*Proof.* Immediate from the definitions. $\square$

**Definition 6.6.** Let $P = ufv$ with $u \in \Sigma^+$, $f \in \mathfrak{S}_\Sigma$, and $v \in \Sigma^*$. Let $j = |u| + 1$ the index of the single gap $f$. Let $a, b \in \mathbb{N}_{>0}$ s.t. $a < j$ and $cb = r = |P|$ for some $c \in \mathbb{N}_{>0}$, and define

$$B = P(a)P(a + b) \dots P(a + (c-1)b)$$

$$P_{a,b} = \begin{cases} (B \circ u)\ (?^{j'-1} f\ ?^{c-j'})\ (B \circ v) & \text{if } (\exists j')(j \equiv_r a + (j'-1)b) \\ B & \text{otherwise (so } f \text{ is not in } B) \end{cases}$$

In the first case ($(\exists j')(j \equiv_r a + (j'-1)b)$) we have $B(j') = f$. We also note that $b|B| = r$ and $b|P_{a,b}| = r^2$.

**Lemma 6.1.** Let $f, g \in \mathfrak{S}_\Sigma$ such that $f \circ g = g \circ f$, and $P = ufv$ a pattern of length $r$ for some $u \in \Sigma^+$ and $v \in \Sigma^*$. Define $P^g = (g \circ u)\ f\ (g \circ v) = g(u)\ f\ g(v)$. Then $T(P^g) = g(T(P))$.

*Proof.* We show $T(P^g)(n) = g(T(P))(n)$ by induction on $n \in \mathbb{N}_{>0}$. Let $j = |u| + 1$ (so $P(j) = P^g(j) = f$). Let $n = rn' + i$ for some $n' \in \mathbb{N}$ and $1 \leq i \leq r$. Using the recurrence relations for Toeplitz words (Lemma 4.11) we obtain:

$$\begin{aligned}
T(P^g)(n) &= P^g(i) = g(P(i)) = g(T(P)(i)) && \text{if } i \neq j \\
T(P^g)(n) &= f(T(P^g)(n'+1)) \overset{\text{IH}}{=} f(g(T(P)(n'+1))) \\
&= g(f(T(P)(n'+1))) = g(T(P)(n)) && \text{if } i = j \qquad \square
\end{aligned}$$

**Lemma 6.7.** *Let $Q$ be a one-gap pattern, and $a, b \in \mathbb{N}_{>0}$ s.t. $b$ divides $|Q|$. Then $T(Q)_{a,b} = T((Q^{(m)})_{a,b})$ where $m$ is minimal such that the (single) gap in $Q^{(m)}$ is at index $j$ with $j > a$.*

*Proof.* Let $P = Q^{(m)} = ufv$ of length $r = |Q|^m$ for some $u \in \Sigma^+$, $f \in \mathfrak{S}_\Sigma$, and $v \in \Sigma^*$, and let $j|u| + 1$ be the index of the gap in $P$. Note that $T(P) = T(Q)$ by Lemma 4.9 and $P_{a,b} = Q_{a,b}$ by definition. Let $n \in \mathbb{N}_{>0}$. We prove $T(P)_{a,b}(n) = T(P_{a,b})(n)$.

Let $c \in \mathbb{N}_{>0}$ be such that $bc = r$ and define $B = P(a)P(a+b)\dots P(a+(c-1)b)$. We distinguish the following cases:

(i) If $j \not\equiv a + (j'-1)b \pmod r$ for all $j' \in \mathbb{N}_{>0}$, then $P_{a,b} = B$, and so $T(P_{a,b}) = B^\omega$ because $B$ is free of gaps. By Lemma 4.11 we obtain

$$\begin{aligned} T(P)_{a,b}(n) &= T(P)(a + (n-1)b) \\ &= P(a + (n-1)b) \\ &= P_{a,b}(n) = B(n) \\ &= T(P_{a,b})(n) \end{aligned}$$

(ii) Assume there is $j'$ with $1 \le j' \le c$ such that $j \equiv a + (j'-1)b \pmod r$, so that $B(j') = f$. Let $r' = |P_{a,b}| = rc = \frac{r^2}{b}$. We make a further case distinction.

(a) Assume $n \not\equiv j' \pmod c$. Let $n = r'n' + i$ for some $n' \in \mathbb{N}$ and $1 \le i \le r'$, so that $i \not\equiv j' \pmod c$. We prove $T(P_{a,b})(r'x+i') = B(i')$ for all $i' \equiv i \pmod c$ by induction on $x \in \mathbb{N}$. Using Lemma 4.11 we obtain (note that $|B \circ x|_? = 0$ for all $x \in \Sigma^*$)

$$\begin{aligned} &T(P_{a,b})(r'x + i') = (B \circ u)(i') = B(i') && \text{if } 1 \le i' \le c|u| \\ &T(P_{a,b})(r'x + i') = (B \circ v)(i'') = B(i') && \text{if } i' = c|u| + c + i'' \\ & && \text{for some } 1 \le i'' \le c|v| \end{aligned}$$

For $i' = c|u| + i''$ for some $1 \le i'' \le c$ we find with Lemma 4.11

$$\begin{aligned} T(P_{a,b})(r'x + i') &= T(P_{a,b})(cx + i'') \\ &= T(P_{a,b})(r'x' + i''') \quad \text{for some } x' \in \mathbb{N}, \text{ and } i''' \equiv_c i \qquad (*) \\ &\overset{\text{IH}}{=} B(i''') \\ &= B(i') \end{aligned}$$

The identity ($*$) holds because $i'' \equiv i' \equiv i \pmod c$ and $r' \equiv 0 \pmod c$.

For $T(P)_{a,b}(n)$ we find, again with Lemma 4.11

$$\begin{aligned} T(P)_{a,b}(r'n' + i) &= T(P)(a + (r'n' + i - 1)b) \\ &= T(P)(r^2 n' + a + (i-1)b) \\ &= P(a + (i-1)b) \qquad\qquad = B(i) \end{aligned}$$

For the third step, note that we have $a + (i-1)b \not\equiv_r j$ since $i \not\equiv_c j'$. Hence we have shown $T(P)_{a,b}(n) = T(P_{a,b})(n)$.

(b) For $n \equiv j' \pmod{c}$, let $n = tc + j'$ for some $t \in \mathbb{N}$. We have that

$$\begin{aligned} T(P)_{a,b}(n) &= T(P)(a + (n-1)b) \\ &= T(P)(a + (tc + j' - 1)b) \\ &= T(P)(rt + a + (j' - 1)b) \\ &= T(P)(rt + j) \\ &= f(T(P)(t+1)) \end{aligned}$$

We also have

$$P_{a,b}(n) = G(t+1) \qquad \text{where } G = (f \circ u)\, f\, (f \circ v) = f(u)\, f\, f(v)$$

because $(B \circ x)(kc + j') = (f \circ x)(k+1)$ for all $x \in \Sigma^*$, $k \in \mathbb{N}$, and $(?^{j'-1} f\, ?^{c-j'})(j') = f$.

We show $T(P_{a,b})(xc + j') = T(G)(x+1)$ by induction on $x$. Next we distinguish two cases for $x$. If $x \not\equiv |u| \pmod{r}$, then

$$\begin{aligned} T(P_{a,b})(xc + j') &= P_{a,b}(xc + j') && \text{since } P_{a,b}(xc + j') \in \Sigma \\ &= G(t+1) && \text{since } G(x+1) \in \Sigma \\ &= T(G)(x+1) \end{aligned}$$

If $x \equiv |u| \pmod{r}$, then $x = x'r + |u|$ for some $x' \in \mathbb{N}$, and

$$\begin{aligned} T(P_{a,b})(xc + j') &= T(P_{a,b})((x'r + |u|)c + j') \\ &= T(P_{a,b})(x'r' + |u|c + j') \\ &= f(T(P_{a,b})(cx' + j')) \\ &\overset{\text{IH}}{=} f(T(G)(x'+1)) \\ &= T(G)(x+1) \end{aligned}$$

We have shown $T(P)_{a,b}(n) = f(T(P)(t+1))$, and $T(P_{a,b})(n) = T(G)(t+1)$. It thus remains to be shown that $f(T(P)(t+1)) = T(G)(t+1)$, which follows from Lemma 6.1.

□

**Example 6.8.** Let $Q = 0g12$ with $g(n) = n + 1 \mod 3$. We construct $Q_{2,2}$ as follows: First take $Q^{(2)} = 01120g^2 1202120012$ so that the gap index $j = 6 > a = 2$. Then we find

$$Q_{2,2} = (B \circ 01120)(??g^2?????)(B \circ 1202120012) \qquad \text{where } B = 12g^2 22202$$

**Theorem 6.9.** *Every arithmetic subsequence of an $\langle r, 1\rangle$-type Toeplitz word is a Toeplitz word.*

*Proof.* Let $P$ be a pattern with $r = |P|$ and $|P|_? = 1$, and let $a, b \in \mathbb{N}_{>0}$. Without loss of generality:

(i) We assume the gap of $P$ to be at the end. For otherwise, using Lemma 6.5 we have a pattern $Q$ of length $r$ with its gap at the end, and $1 \leq c \leq r-1$ such that $Q_{c,r-1} = P$. By Lemma 6.3 we have that $T(Q)_{c,r-1} = T(P)$, and hence $T(P)_{a,b} = T(Q)_{c+(r-1)(a-1),b(r-1)}$. Then we proceed with constructing a pattern generating $T(Q)_{c+(r-1)(a-1),b(r-1)}$.

(ii) We assume $b < r$, since $T(P) = T(P^{(k)})$ for all $k \geq 1$ by Lemma 4.9.

Let $b = b_1 b_2$ such that $b_1$ is maximal with $\gcd(b_1, r) = 1$. Note that $(*)$ $b_1$ contains all primes in the factorization of $b$ that do not occur in the prime factorization of $r$. Moreover let $a_2 \in \mathbb{N}_{>0}$ such that $1 \leq a_1 \leq b_1$ where $a_1 = a - (a_2 - 1)b_1$. Then

$$T(P)_{a,b} = (T(P)_{a_1,b_1})_{a_2,b_2}$$

by composition of arithmetic progressions. The sequence $T(P)_{a_1,b_1}$ is generated by the Toeplitz pattern $P_{a_1,b_1}$ of length $r^m$ with one gap by Lemma 6.3 (for some $m \geq 1$). Thus it suffices to show that $T(P_{a_1,b_1})_{a_2,b_2}$ is generated by a Toeplitz pattern. By $(*)$ all primes in the factorization of $b_2$ occur in $r$, so also in $r^m$. Hence there exists $n \geq 1$ such that $b_2$ divides $(r^m)^n = r^{mn}$. Since $T(P)_{a_1,b_1} = T(P^{(n)}_{a_1,b_1})$, $|P^{(n)}_{a_1,b_1}|_? = 1$, and $|P^{(n)}_{a_1,b_1}| = r^{nm}$ the claim follows by Lemma 6.7. □

# 7 Keane Words

We generalize (binary) *generalized Morse sequences* as introduced by Keane in [16], to sequences over the cyclic additive group $\Sigma$, but restrict to 'uniform' infinite block products $u \times u \times u \times \cdots$, or 'Keane words' as we call them.

**Definition 7.1.** The *Keane product* is the binary operation $\times$ on $\Sigma^*$ defined as follows:

$$u \times \varepsilon = \varepsilon \qquad\qquad u \times av = (u + a)(u \times v)$$

for all $u, v \in \Sigma^*$ and $a \in \Sigma$. We define $u^{(n)}$ by $u^{(0)} = 0$ and $u^{(n+1)} = u \times u^{(n)}$.

Let $u \in \Sigma^*$ with $|u| \geq 2$ and $u(0) = 0$. The *Keane word generated by* $u$ is defined by

$$K(u) = \lim_{n\to\infty} u^{(n)}$$

The product $u \times v$ is formed by concatenation of $|v|$ copies of $u$ (so $|u \times v| = |u| \cdot |v|$), taking the $i$th copy as $u + v(i)$ $(0 \leq i \leq |v| - 1)$. As 0 is the identity with respect to the $\times$-operation, $u \times v$ is a proper extension of $u$ whenever $v(0) = 0$ and $|v| \geq 2$. Hence $K(u)$ is well-defined and is the unique infinite fixed point of $x \mapsto u \times x$. Also note that $K(u)$ is the iterative limit of the $|u|$-uniform morphism $h$ defined by $h(a) = u + a$, for all $a \in \Sigma$; thus Keane words are automatic sequences [4].

**Proposition 7.2.** *$\langle \Sigma^*, \times, 0\rangle$ is a monoid.*

Note that $\times$ is not commutative, e.g., $00\times 01 = 0011$ whereas $01\times 00 = 0101$.

**Lemma 7.3.** *Let $u \in \Sigma^*$ with $u(0) = 0$ and $k = |u| \geq 2$, and let $0 \leq i < k$. Then for all $n \geq 0$ we have $K(u)(nk + i) = K(u)(n) + u(i)$, and so $K(u)_{i,k} \sim K(u)$. Hence (by Lemma 2.6) $\langle i, k^n\rangle \in \mathcal{AS}(K(u))$ for all $n \geq 0$ and $0 \leq i < k^n$.*

*Proof.* Because of associativity of $\times$ we have that $u \times u^{(n)} = u^{(n)} \times u$, which means that we can view $u^{(n+1)}$ as consisting of $|u|^n$ variants of $u$, as well as $|u|$ variants of $u^{(n)}$, respectively:

$$u^{(n+1)} = u \times u^{(n)} = (u + u^{(n)}(0)) \cdots (u + u^{(n)}(|u|^n - 1)) \tag{1}$$

$$u^{(n+1)} = u^{(n)} \times u = (u^{(n)} + u(0)) \cdots (u^{(n)} + u(|u| - 1)) \tag{2}$$

Let $0 \leq i < k$ and $0 \leq j < k^n$. Since we have $u^{(n)} \sqsubseteq K(u)$ for all $n \geq 0$, by the use of (1) and (2) we conclude

$$u^{(n+1)}((j|u| + i) = (u + u^{(n)}(j))(i) = u^{(n)}(j) + u(i)$$

$$u^{(n+1)}(i|u|^n + j) = (u^{(n)} + u(i))(j) = u^{(n)}(j) + u(i)$$

□

**Proposition 7.4.** *The only completely additive Keane word is the constant zero sequence.*

*Proof.* Let $u = u_0u_1 \dots u_{k-1} \in \Sigma^*$ with $u_0 = 0$ and $k \geq 2$, and assume that $\sigma = K(u) \in \Sigma^{\mathbb{N}}$ is additive (note that, as $\sigma$ is indexed $0, 1, \dots$, additivity of $\sigma$ means $\sigma(nm-1) = \sigma(n-1)+\sigma(m-1)$ for all $n, m > 0$). Then $\sigma_{k,k+1} = \sigma+\sigma(k)$, that is,

$$\sigma_{k,k+1}(n) = \sigma(k + (k + 1)n) = \sigma(k(n + 1) + n) = \sigma(n) + \sigma(k) \qquad (n \in \mathbb{N}) \tag{3}$$

Using the recurrence relations for Keane words (Lemma 7.3) we obtain

$$\sigma(k(i + 1) + i) = \sigma(i + 1) + u_i \qquad (0 \leq i < k) \tag{4}$$

Combining (3) and (4) gives $\sigma(i+1) = \sigma(k)$ for all $0 \leq i < k$. Hence $u_1 = u_2 = \dots = u_{k-1}$.

To see that $u_i = 0$ for all $0 \leq i < k$, we show $u_1 = u_1 + u_1$ in each of the following three cases: For $k \geq 4$ this is immediate: $u_3 = u_1 + u_1$ follows from additivity of $\sigma$. For $k = 3$ we get $\sigma(3) = u_0 + u_1 = u_1$ from equation (4) and $\sigma(3) = u_1+u_1$ by additivity. Finally, for $k = 2$, we get $\sigma(8) = u_1$ and $\sigma(2) = u_1$, since by Lemma 7.3 we have $K(v)(\ell^n) = v_1$ for all blocks $v = v_0v_1 \dots v_\ell$ and $n \in \mathbb{N}$. Moreover, by additivity, $\sigma(8) = \sigma(2) + \sigma(2) = u_1 + u_1$. □

In what follows we let $(n)_k$ denote the base $k$-expansion of $n$, and $|w|_a$ the number of occurrences of a letter $a \in \Sigma$ in a word $w \in \Sigma^*$.

**Theorem 7.5.** *Let $u = u_0u_1 \dots u_{k-1} \in \Sigma^*$ be a $k$-block with $u_0 = 0$ and $k \geq 2$. Then*

$$K(u)(n) = \sum_{i \in \Sigma_k} |(n)_k|_i \odot u_i \qquad (n \geq 0)$$

*Proof.* We prove $(\forall n \in \mathbb{N})(|(n)_k| = r \implies K(u)(n) = \sum_{i\in\Sigma_k} |(n)_k|_i \odot u_i)$ by induction on $r \in \mathbb{N}_{>0}$. If $r = 1$, then $n \in \Sigma_k$ and $K(u)(n) = u_n$. For $r > 1$, let $(n)_k = n_{r-1}n_{r-2} \dots n_1n_0$ and define $n' = \frac{n-n_0}{k}$. Then $(n')_k = n_{r-1}n_{r-2} \dots n_1$ and $|(n')_k| = r - 1$, and we find

$$\begin{aligned} K(u)(n) &= K(u)(n'k + n_0) \\ &= K(u)(n') + u_{n_0} && \text{(Lemma 7.3)} \\ &\stackrel{\text{IH}}{=} \left( \sum_{i\in\Sigma_k} |(n')_k|_i \odot u_i \right) + u_{n_0} \\ &= \sum_{i\in\Sigma_k} |(n)_k|_i \odot u_i \end{aligned}$$

□

**Example 7.6.** For the Morse sequence $\boldsymbol{m} = K(01) = 01\times01\times\cdots$, Theorem 7.5 gives another well-known definition of $\boldsymbol{m}$, due to J.H. Conway [7]: $\boldsymbol{m}(n)$ is the parity of the number of 1s in the binary expansion of $n$. Likewise, for the generalized Morse sequence $\boldsymbol{w} = K(001)$ discussed in [16, 14], and called the 'Mephisto Waltz' in [14, p. 105], we find that $\boldsymbol{w}(n)$ is the parity of the number of 2s in the ternary expansion of $n$.

Let $x = x_0x_1 \dots x_{k-1}$ with $k \geq 1$ be a word over $\Sigma$. The *first difference* $\Delta(x)$ of $x$ is defined by $\Delta(x) = \varepsilon$ if $|x| = 1$ and $\Delta(x) = (x_1 - x_0)(x_2 - x_1) \dots (x_{k-1} - x_{k-2})$, otherwise.

We give an embedding of the Keane monoid (Prop. 7.2) into the monoid of Toeplitz pattern composition (Prop. 4.6).

Let $\mathcal{B} = \{\, u \in \Sigma^+ \mid u_0 = 0 \,\}$, and define the map $\Delta_T : \mathcal{B} \to (\Sigma \cup \mathfrak{S}_\Sigma)^*$ by

$$\Delta_T(u) = \Delta(u)?^{+d} \qquad \text{for all } u \in \mathcal{B} \text{ and } d = u_0 - u_{|u|-1} = -u_{|u|-1}$$

**Theorem 7.7.** $\Delta_T : \langle \Sigma^+, \times, 0\rangle \hookrightarrow \langle (\Sigma \cup \mathfrak{S}_\Sigma)^+, \circ, ?\rangle$ *is an injective homomorphism:*

$$\Delta_T(0) = ? \qquad\qquad \Delta_T(u \times v) = \Delta_T(u) \circ \Delta_T(v)$$

*Proof.* We first show that $\Delta_T$ preserves structure. The preservation of identity is immediate. For $c \in \Sigma$ let $\Delta_{T,c} : \Sigma^+ \to (\Sigma \cup \mathfrak{S}_\Sigma)^+$ be defined by $\Delta_{T,c}(x) = \Delta(x)\,?^{+c}$ for all $x \in \Sigma^+$ (so that $\Delta_T(x) = \Delta_{T,c}(x)$ iff $c = -x_{|x|-1}$, for all $x \in \mathcal{B}$). We prove the following equation, for all $u, v \in \Sigma^+$ and $d \in \Sigma$:

$$\Delta_{T,d'}(u \times v) = \Delta_T(u) \circ \Delta_{T,d}(v) \qquad \text{where } d' = d_u + d \text{ and } d_u = u_0 - u_{|u|-1},$$

by induction on the length of $v$. The claim then follows by taking $d = d_v = v_0 - v_{|v|-1} = -v_{|v|-1}$ since $\Delta_T(u \times v) = \Delta_{T,d_u+d_v}(u \times v)$.

If $|v| = 1$, then $v = a$ for some $a \in \Sigma$, and

$$\begin{aligned}
\Delta_{T,d'}(u \times a) &= \Delta_{T,d'}(u + a) \\
&= \Delta(u + a)?^{+d'} \\
&= \Delta(u)?^{+d'} \\
&= (\Delta(u)?^{+d_u}) \circ ?^{+d} \\
&= \Delta_T(u) \circ \Delta_{T,d}(a)
\end{aligned}$$

If $|v| > 1$, then $v = aa'v'$ for some $a, a' \in \Sigma$ and $v' \in \Sigma^*$, and we find

$$\begin{aligned}
\Delta_{T,d'}(u \times aa'v') &= \Delta_{T,d'}((u + a)(u \times a'v')) \\
&= \Delta((u + a)(u \times a'v'))?^{+d'} \\
&= \Delta(u + a)(u_0 + a' - (u_{|u|-1} + a))\Delta(u \times a'v')?^{+d'} && (5) \\
&= \Delta(u)(d_u + a' - a)\Delta_{T,d'}(u \times a'v') \\
&= \Delta(u)(d_u + a' - a)(\Delta_T(u) \circ \Delta_{T,d}(a'v')) && \text{(IH)} \\
&= (\Delta(u)?^{+d_u}) \circ ((a' - a)\Delta_{T,d}(a'v')) \\
&= \Delta_T(u) \circ \Delta_{T,d}(aa'v')
\end{aligned}$$

where (5) follows from $\Delta(xy) = \Delta(x)(y_0 - x_{|x|-1})\Delta(y)$ for all $x, y \in \Sigma^+$, and $(u \times a'v')(0) = u_0 + a'$.

Finally, we show that $\Delta_T$ is injective by defining a retraction $\int_T : \Delta_T(\mathcal{B}) \to \mathcal{B}$. Let $P$ be a $\Sigma$-pattern $P \in \Delta_T(\mathcal{B})$. Then $P$ has the form $P = a_1 a_2 \dots a_n$ where $a_i \in \Sigma$ for $1 \leq i < n$ and $a_n = ?^{+d}$ with $d = -\sum_{1 \leq i < n} a_i$. Define $\int_T(P) = u_0 u_1 \dots u_{n-1}$ by $u_0 = 0$, $u_i = u_{i-1} + a_i$ $(1 \leq i \leq n-1)$.

We show that $\int_T$ is a left-inverse of $\Delta_T$. Let $u = u_0 u_1 \dots u_{n-1} \in \mathcal{B}$. Then $\Delta_T(u) = \Delta(u)?^{+d}$ with $d = -u_{n-1}$. Let $\Delta(u) = a_1 a_2 \dots a_{n-1}$. Then $a_i = u_i - u_{i-1}$ $(1 \leq i < n)$ and so $\int_T(\Delta_T(u)) = u$. □

It follows that the first difference sequence of a Keane word is a Toeplitz word. Here the difference operator $\Delta$ is extended to $\Sigma^\infty \to \Sigma^\infty$ in the obvious way.

**Theorem 7.8.** *$\Delta(K(u)) = T(\Delta_T(u))$, for all blocks $u \in \mathcal{B}$.*

*Proof.* Let $u = u_0 u_1 \dots u_{k-1}$ be a $k$-block over $\Sigma$ with $u_0 = 0$ and $k \geq 2$. Let $n \in \mathbb{N}_{>0}$ be a positive integer and $(n)_k = n_{r-1} n_{r-2} \dots n_0$ the base $k$-expansion of $n$.

The first nonzero digit of $(n)_k$ (reading from right to left) is $n_a$ with $a = v_k(n)$ and so

$$\begin{array}{rcccccccc}
(n)_k & = & n_{r-1} & n_{r-2} & \dots & n_a & 0 & 0 & \dots & 0 \\
(n-1)_k & = & n_{r-1} & n_{r-2} & \dots & (n_a - 1) & (k-1) & (k-1) & \dots & (k-1)
\end{array}$$

Let $\kappa = K(u)$. From Theorem 7.5 we then obtain

$$\begin{aligned}\Delta(\kappa)(n) &= \kappa(n) - \kappa(n-1) \\ &= (u_{n_a} + a \cdot u_0) - (u_{n_a - 1} + a \cdot u_{k-1}) \\ &= u_{n_a} - u_{n_a-1} - a \cdot u_{k-1}\end{aligned}$$

whence

$$\Delta(\kappa)(n) = \begin{cases} u_1 - u_0 & \text{if } n \equiv 1 \pmod{k} \\ u_2 - u_1 & \text{if } n \equiv 2 \pmod{k} \\ \vdots & \vdots \\ u_{k-1} - u_{k-2} & \text{if } n \equiv k-1 \pmod{k} \\ \Delta(\kappa)(n/k) - u_{k-1} & \text{if } n \equiv 0 \pmod{k} \end{cases}$$

From Lemma 4.11 we infer that this is exactly the recurrence equation of the Toeplitz word generated by the pattern $\Delta_T(u) = \Delta(u)?^{-u_{k-1}}$. □

**Example 7.9.** The first difference of the Morse sequence $\boldsymbol{m}$ is the conjugate of the period doubling sequence $\boldsymbol{p}$: $\Delta(\boldsymbol{m}) = \Delta(K(01)) = T(\Delta(01)?^{+1}) = T(1?^{+1}) = \boldsymbol{p} + 1$.

The first difference of the Mephisto Waltz $\boldsymbol{w}$ [16]: $\Delta(\boldsymbol{w}) = \Delta(K(001)) = T(\Delta(001)?^{+1}) = T(01?^{+1})$ turns out to be the sequence of turns (folds) of the alternate Terdragon curve [9]! Moreover we also know $\Delta(\boldsymbol{w}) = \boldsymbol{s}/2$, where $\boldsymbol{s} = T(00?^{+1}11?^{+1})$ is the sequence of turns of the Sierpiński curve [11, 12]! Note that $T(01?^{+1})$ is the additive sequence $\boldsymbol{v}_{[2]_3 \cup \{3\}}$ derived from $\lambda_{3,2}$, see the construction in Section 5.

We conclude with a complete characterization of the arithmetic self-similarity of the Thue–Morse sequence $\boldsymbol{m}$, and leave the characterization of the entire class of Keane words as future work. We know that at least $\langle i, |u|^n \rangle \in \mathcal{AS}(K(u))$ for all $n \geq 0$ and $0 \leq i < |u|^n$ by Lemma 7.3.

## The Arithmetic Self-Similarity of the Thue–Morse Sequence

**Theorem 7.10.** $\mathcal{AS}(\boldsymbol{m}) = \{ \langle a, b \rangle \mid 0 \leq a < b = 2^m \textit{ for some } m \geq 0 \}$.

We have the following recurrence equations for $\boldsymbol{m} = K(01) = 01101001\ldots$:

$$\boldsymbol{m}(2n) = \boldsymbol{m}(n) \qquad\qquad \boldsymbol{m}(2n+1) = \overline{\boldsymbol{m}(n)} \tag{6}$$

*Carry-free addition* $\alpha \uplus \beta$ of binary numbers $\alpha = \alpha_p \ldots \alpha_0$ and $\beta = \beta_q \ldots \beta_0$ is defined by:

$$\alpha \uplus \beta = \begin{cases} \alpha + \beta & \text{if there is no } i \leq \min(p,q) \text{ with } \alpha_i = \beta_i = 1 \\ \textit{undefined} & \text{otherwise} \end{cases}$$

Carry-free addition preserves the total number of ones: $|\alpha \uplus \beta|_1 = |\alpha|_1 + |\beta|_1$. Hence we have that $\boldsymbol{m}(x+y) = \boldsymbol{m}(x) + \boldsymbol{m}(y)$ whenever $(x)_2 \uplus (y)_2$ is defined.

**Lemma 7.11.** *Let $x > 0$ with $x \neq 2^k$ for any $k \geq 0$. Then $\boldsymbol{m}(xy) = 1$ for some $y > 0$ with $\boldsymbol{m}(y) = 0$.*

*Proof.* Without loss of generality, we let $x$ be an odd number. Otherwise, if $x = 2x'$, then $\boldsymbol{m}(2x'y) = \boldsymbol{m}(x'y)$. Put differently, trailing 0's of $(x)_2$ can be removed, as they do not change the parity of the number of 1's in $(xy)_2$.

Let $(x)_2 = x_n \dots x_1 x_0$ be the binary representation of $x$ (with $x_n = x_0 = 1$). Note that $n > 0$ for otherwise $x = 2^0$. We distinguish two cases:

(i) If $x_1 = 0$, then we take $(y)_2 = y_n \dots y_0$ with $y_n = y_0 = 1$ and $y_i = 0$ for $0 < i < n$. So $y = 2^n + 2^0$ and $xy = x2^n + x$ ; in binary representation:

$$
\begin{array}{cccccccc}
 & & & x_n & x_{n-1} & \dots & x_1 & x_0 \\
 & & & 1 & 0 & \dots & 0 & 1 \\
\hline
 & & & x_n & x_{n-1} & \dots & x_1 & x_0 \\
x_n & \dots & x_1 & x_0 & 0 & \dots & 0 & 0 \\
\hline
x_n & \dots & 1 & 0 & x_{n-1} & \dots & x_1 & x_0
\end{array}
$$

To see that $\boldsymbol{m}(xy) = 1$, i.e., that $(xy)_2$ has an odd number of 1's, observe that $(xy)_2$ (the bottom line) contains the subword $x_{n-1} \dots x_2$ twice. These cancel each other out, and, as $x_1 = 0$, the only remaining 1's are $x_n$ and $x_0$ on the outside and the created 1 at position $n + 1$. That makes three.

(ii) In case $x_1 = 1$, we take $y = 3y'$ with $y'$ the result of case (i) for $3x$. First of all note that $(3x)_2$ ends in 01 and so case (i) applies. Secondly, we get that $\boldsymbol{m}(3xy') = 1$, and hence $\boldsymbol{m}(xy) = 1$. Finally, to see that $\boldsymbol{m}(y) = 0$, note that $y' = 2^m + 1$ for some $m > 0$, and so $\boldsymbol{m}(y) = \boldsymbol{m}(3y') = \boldsymbol{m}(2^{m+1} + 2^m + 2 + 1) = 1 - \boldsymbol{m}(2^m + 2^{m-1} + 1)$ by the recurrence equations (6) for $\boldsymbol{m}$. To see that $\boldsymbol{m}(2^m + 2^{m-1} + 1) = 1$, one distinguishes cases $m = 1$ (then $\boldsymbol{m}(4) = 1$) and $m > 1$ (then we again have three 1's in the binary expansion). ☐

*Proof of Theorem 7.10.* The direction '$\supseteq$' follows from Lemma 7.3.

For the direction '$\subseteq$', assume $\langle a, b \rangle \in \mathcal{AS}(\boldsymbol{m})$. We distinguish two cases: (i) $b$ is a power of 2 and $a \geq b$, and (ii) $b$ is not a power of 2, and show that they lead to a contradiction.

(i) Let $m \geq 0$ and $a \geq b = 2^m$. Consider $k \geq 1$ such that $bk \leq a < b(k+1)$ and let $a' = a - bk$. Then $\boldsymbol{m}_{a,b}(n) = \boldsymbol{m}(a + bn) = \boldsymbol{m}(a' + b(k+n)) = \boldsymbol{m}_{a',b}(k+n)$, in short: $\boldsymbol{m}_{a,b} = (\boldsymbol{m}_{a',b})_{k,1}$ $(*)$. As we have that $a' < b$, it follows from ($\supseteq$) above that $\langle a', b \rangle \in \mathcal{AS}(\boldsymbol{m})$, and so $\boldsymbol{m}_{a',b} \sim \boldsymbol{m}$ $(**)$. Combining $(*)$ and $(**)$ with the assumption $\boldsymbol{m}_{a,b} \sim \boldsymbol{m}$, we obtain $\boldsymbol{m}_{k,1} \sim \boldsymbol{m}$.

Suppose $\boldsymbol{m}_{s,1} = \boldsymbol{m}$ for some $s \geq 1$ (the proof of $\boldsymbol{m}_{s,1} \neq \overline{\boldsymbol{m}}$ is analogous).[6] This means $\boldsymbol{m}(s + n) = \boldsymbol{m}(n)$ for all $n$. So $\boldsymbol{m}(s) = \boldsymbol{m}(0) = 0$. Now let $p$ be maximal such that $2^p \leq s$ and $(s)_2 = s_p s_{p-1} \dots s_0$ (so $s_p = 1$). Then

[6]We may also conclude this case by using the fact that $\boldsymbol{m}$ is non-periodic: No positive iteration $k \geq 1$ of the shift operator of $\boldsymbol{m}$ equals $\boldsymbol{m}$ or its conjugate, because a sequence $\sigma \in A^{\mathbb{N}}$ such that $\sigma_{k,1} = \sigma$ with $k \geq 1$ is $k$-periodic, see Lemma 2.7.

$(s+2^p)_2 = 10s_{p-1}\dots s_0$, and hence $\boldsymbol{m}(s+2^p) = \boldsymbol{m}(s) = 0$. But the assumption $\boldsymbol{m}_{s,1} = \boldsymbol{m}$ tells us $\boldsymbol{m}(s+2^p) = \boldsymbol{m}(2^p) = 1$.

(ii) Assume $b > 0$ is not a power of 2, and let $a \geq 0$ be arbitrary. By Lemma 7.11 there exists $y > 0$ such that $\boldsymbol{m}(y) = 0$ and $\boldsymbol{m}(by) = 1$. Let $m$ be minimal such that $a < 2^m$. Then also $\boldsymbol{m}(by2^m) = 1$, and, since $(a + by2^m)_2 = (a)_2 \uplus (by2^m)_2$, we get $\boldsymbol{m}(a + by2^m) = \boldsymbol{m}(a) + 1 \pmod 2$.

From the assumption $\langle a, b\rangle \in \mathcal{AS}(\boldsymbol{m})$ it follows that either $(\forall n)(\boldsymbol{m}(a + bn) = \boldsymbol{m}(n))$, or $(\forall n)(\boldsymbol{m}(a + bn) = \overline{\boldsymbol{m}}(n))$. In the first case we have $\boldsymbol{m}(a) = \boldsymbol{m}(0) = 0$ and $\boldsymbol{m}(a + by2^m) = \boldsymbol{m}(y2^m) = 0$. In the second case, we have $\boldsymbol{m}(a) = \overline{\boldsymbol{m}}(0) = 1$ and $\boldsymbol{m}(a + by2^m) = \overline{\boldsymbol{m}}(y2^m) = 1$. Both cases contradict $\boldsymbol{m}(a + by2^m) = \boldsymbol{m}(a) + 1 \pmod 2$. □

# References


[1] J.-P. Allouche and R. Bacher. Toeplitz Sequences, Paperfolding, Towers of Hanoi, and Progression-Free Sequences of Integers. *L'Enseignement Mathématique*, 38:315–327, 1992.

[2] J.-P. Allouche and F. Dress. Tours de Hanoï et Automates. *Informatique Théorique et Applications*, 24:1–16, 1990.

[3] J.-P. Allouche and J. Shallit. The Ubiquitous Prouhet–Thue–Morse Sequence. In *Proceedings of International Conference on Sequences and Their Applications (SETA 1998)*, pages 1–16. Springer, 1999.

[4] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations.* Cambridge University Press, New York, 2003.

[5] J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat, and A. Sárközy. On Finite Pseudorandom Binary Sequences IV: The Liouville Function, II. *Acta Arithmetica*, 95:349–359, 2000.

[6] J. Cassaigne and J. Karhumäki. Toeplitz Words, Generalized Periodicity and Periodically Iterated Morphisms. *European Journal of Combinatorics*, 18(5):497–510, 1997.

[7] J.H. Conway. *On Numbers and Games.* London Mathematical Society Monographs (No. 6). Academic Press, 6 edition, 1976.

[8] F.G.W. Dannenberg. Symbolic Dynamics and Automata Theory. Master's thesis, Delft University of Technology, 2011.

[9] C. Davis and D.E. Knuth. Number Representations and Dragon Curves. *Journal of Recreational Mathematics*, 3:66–81, 133–149, 1970.

[10] P.G.L. Dirichlet. Beweis des Satzes, daß jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält. *Abhandlungen Akademische Wissenschaft Berlin*, 48:45–81, 1837.

[11] J. Endrullis, D. Hendriks, and J.W. Klop. Let's Make a Difference!

[12] J. Endrullis, D. Hendriks, and J.W. Klop. Degrees of Streams. *Integers, Electronic Journal of Combinatorial Number Theory*, 2011. Accepted for publication.

[13] K. Hare and M. Coons. *Personal communication*, 2011.

[14] K. Jabobs. *Invitation to Mathematics.* Princeton University Press, 1992.

[15] K. Jacobs and M. Keane. 0-1-Sequences of Toeplitz Type. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 13(2):123–131, 1969.

[16] M. Keane. Generalized Morse Sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 10(4):335–353, 1968.

[17] N.J.A. Sloane. Online Encyclopedia of Integer Sequences. http://oeis.org/.