



THE ABELIAN COMPLEXITY OF INFINITE WORDS AND THE FROBENIUS PROBLEM

Ian Kaye¹

Narad Rampersad²

*Department of Mathematics and Statistics, University of Winnipeg, Winnipeg,
Manitoba, Canada*

n.rampersad@uwinnipeg.ca

Received: 7/23/19, Revised: 8/25/20, Accepted: 4/13/21, Published: 4/20/21

Abstract

We study the following problem, first introduced by Dekking. Consider an infinite word \mathbf{x} over an alphabet $\{0, 1, \dots, k-1\}$ and a semigroup homomorphism $S : \{0, 1, \dots, k-1\}^* \rightarrow \mathbb{N}$. Let $\mathcal{L}_{\mathbf{x}}$ denote the set of factors of \mathbf{x} . What conditions on S and the abelian complexity of \mathbf{x} guarantee that $S(\mathcal{L}_{\mathbf{x}})$ contains all but finitely many elements of \mathbb{N} ? We examine this question for some specific infinite words \mathbf{x} having different abelian complexity functions.

1. Introduction

It is well-known that if a and b are two co-prime positive integers then all sufficiently large positive integers n can be written as a linear combination $n = xa + yb$, where x and y are non-negative integers. Frobenius posed the problem of determining the largest positive integer that cannot be so represented; Sylvester [12] was the first to give a solution to Frobenius' problem: he showed that the largest non-representable number is

$$ab - a - b. \tag{1}$$

Ramírez Alfonsín [10] has written a monograph devoted entirely to this problem.

Dekking [6] studied the following variation of this problem. Let $S : \{0, 1\}^* \rightarrow \mathbb{N}$ be a *semigroup homomorphism*, i.e., there are non-negative integers a and b such that S is defined by $S(0) = a$, $S(1) = b$, and $S(uv) = S(u) + S(v)$ for any words u and v over the binary alphabet $\{0, 1\}$. Given an infinite word \mathbf{w} over the alphabet $\{0, 1\}$, let $\mathcal{L}_{\mathbf{w}}$ denote the set of all factors of \mathbf{w} and let $\mathcal{L}_{n, \mathbf{w}}$ denote the set of all

¹The author was supported by an NSERC USRA.

²The author was supported by NSERC Discovery Grants 418646-2012 and RGPIN-2019-04111.

length- n factors of \mathbf{w} . Define

$$S(\mathcal{L}_{\mathbf{w}}) = \{S(u) : u \in \mathcal{L}_{\mathbf{w}}\}.$$

What conditions on \mathbf{w} and S ensure that $S(\mathcal{L}_{\mathbf{w}})$ is *co-finite* (contains all but finitely many elements of \mathbb{N})?

Certainly a and b must be co-prime (and so we will assume this to be the case for the remainder of the paper). The set $S(\mathcal{L}_{\mathbf{w}})$ is closely related to the *abelian complexity* [14] of \mathbf{w} (as well as the *additive complexity* [2] of \mathbf{w}). For any word u over an alphabet A , we write $|u|_a$ to denote the number of occurrences of a letter $a \in A$ in the word u and $|u|$ to denote the length of n . If $A = \{a_1, \dots, a_k\}$, the *Parikh vector* of u is the vector $\psi(u)$ whose i -th entry equals $|u|_{a_i}$. Let $A = \{0, 1\}$. For any n , we have $n \in S(\mathcal{L}_{\mathbf{w}})$ exactly when there is a factor u of \mathbf{w} such that $n = xa + yb$ and $\psi(u) = (x, y)$. The *abelian complexity function* of \mathbf{w} is the function $\rho_{\mathbf{w}}(n)$ that maps n to the cardinality of the set $\psi(\mathcal{L}_{n, \mathbf{w}})$. If $\psi(\mathcal{L}_{n, \mathbf{w}}) = \{(0, n), (1, n - 1), \dots, (n - 1, 1), (n, 0)\}$ for all n (i.e. $\rho_{\mathbf{w}}(n) = n + 1$), then \mathbf{w} has *maximal abelian complexity* and it is clear that in this case $S(\mathcal{L}_{\mathbf{w}})$ is co-finite. Indeed, in this case the problem is the classical one stated by Frobenius. On the other hand, for words with lower abelian complexity functions, this may not be the case.

Dekking studied the case where \mathbf{w} is a *Sturmian word*. Sturmian words are the aperiodic words with the smallest possible abelian complexity; i.e., if \mathbf{w} is an aperiodic binary word then \mathbf{w} is Sturmian if and only if $\rho_{\mathbf{w}}(n) = 2$ for all $n \geq 1$ [5]. Dekking gave an explicit formula for $S(\mathcal{L}_{\mathbf{w}})$ for any Sturmian word \mathbf{w} ; this formula implies that for any given \mathbf{w} there are only finitely many maps S such that $S(\mathcal{L}_{\mathbf{w}})$ is co-finite. For the Fibonacci word, Dekking characterized exactly the set of such maps S . Given the close relationship between Sturmian words and Beatty sequences, we also mention the work of Steuding and Stumpf [11] concerning the Frobenius problem and Beatty sequences.

The general question we are interested in then is, “What conditions on the abelian complexity of \mathbf{w} are sufficient to ensure that $S(\mathcal{L}_{\mathbf{w}})$ is co-finite for all maps S ?” (Remember, we are assuming that $S(0)$ and $S(1)$ are relatively prime.) If $S(\mathcal{L}_{\mathbf{w}})$ is co-finite for all maps S , we say that \mathbf{w} has the *Frobenius property*. As previously noted, if \mathbf{w} has maximal abelian complexity, then \mathbf{w} has the Frobenius property, and if \mathbf{w} is Sturmian, then \mathbf{w} does not have the Frobenius property. In this paper we analyze some example of words \mathbf{w} whose abelian complexity is intermediate between these two extremes.

Finally, we note that the Frobenius problem can be extended from two given positive integers a and b to any number of given positive integers. Similarly, we can extend the notions defined above to words over larger alphabets. Recall that Dekking studied $S(\mathcal{L}_{\mathbf{w}})$ for Sturmian words \mathbf{w} , which are infinite binary words with constant abelian complexity $\rho_{\mathbf{w}}(n) = 2$. We examine $S(\mathcal{L}_{\mathbf{t}})$ for a certain infinite

ternary word \mathbf{t} with constant abelian complexity $\rho_{\mathbf{t}}(n) = 3$.

To summarize, in the next sections we study:

- the *paperfolding word* \mathbf{pf} , which has abelian complexity $\rho_{\mathbf{pf}}(n) = O(\log n)$; this word does not have the Frobenius property.
- a pure morphic binary word Φ with abelian complexity $\rho_{\Phi}(n) = \Theta(n^{\log_5 2})$; this word has the Frobenius property.
- a balanced ternary word \mathbf{t} with abelian complexity $\rho_{\mathbf{t}}(n) = 3$ for all $n \geq 1$; this word does not have the Frobenius property.

2. The Paperfolding Word

In this section we examine whether the (*ordinary*) *paperfolding word* has the Frobenius property. This is a word whose abelian complexity function is unbounded, unlike that of the Sturmian words. For a nice introduction to the paperfolding words and their properties, see the series of papers by Dekking, Mendès France, and Poorten [7]. There are a number of equivalent definitions of the paperfolding word \mathbf{pf} . If $w = w_1w_2 \dots w_k$ is a word over $\{0, 1\}$ then the *complement* of w is the word $\bar{w} = (1 - w_1)(1 - w_2) \dots (1 - w_k)$ and the *reversal* of w is the word $w^R = w_kw_{k-1} \dots w_1$. The word \mathbf{pf} may be constructed as the limit of the following process: Let $f^{(1)} = 0$. Having constructed $f^{(n)}$, we define $f^{(n+1)} := f^{(n)} 0 \overline{f^{(n)}^R}$. Then $\mathbf{pf} = \lim_{n \rightarrow \infty} f^{(n)}$.

The next construction of the paperfolding word is known as the *Toeplitz construction*. We begin with a sequence of empty spaces and fill every *second* space with the alternating sequence $(01)^\omega$. After infinitely many repetitions of this process, we obtain the ordinary paperfolding word \mathbf{pf} . Beginning with $_ _ _ _ _ _ \dots$, the first few steps in this process are

$$\begin{array}{cccccccccccc}
 \bar{_} & \bar{_} & \bar{1} & \bar{_} & \bar{0} & \bar{_} & \bar{1} & \bar{_} & \bar{0} & \bar{_} & \bar{1} & \bar{_} & \dots \\
 0 & 0 & 1 & \bar{_} & 0 & 1 & 1 & \bar{_} & 0 & 0 & 1 & \bar{_} & \dots \\
 0 & 0 & 1 & 0 & 0 & 1 & 1 & \bar{_} & 0 & 0 & 1 & 1 & \dots \\
 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & \dots
 \end{array}$$

This construction implies the following recursive definition of $\mathbf{pf} = (f_n)_{n \geq 1}$:

$$(f_{2n-1})_{n \geq 1} = (01)^\omega \text{ and } (f_{2n})_{n \geq 1} = \mathbf{pf}. \tag{2}$$

We may also define the n -th term f_n of \mathbf{pf} from the binary representation of n . Let $n = m \cdot 2^j$ be given, where m is odd. Then define

$$f_n = \begin{cases} 0 & \text{if } m \equiv 1 \pmod{4} \\ 1 & \text{if } m \equiv 3 \pmod{4}. \end{cases}$$

Madill and Rampersad [9] studied the abelian complexity of \mathbf{pf} . They proved that $\rho_{\mathbf{pf}}(n) = O(\log n)$; however, it is also the case that $\rho_{\mathbf{pf}}$ takes the value 3 infinitely often. In particular, we have

$$\rho_{\mathbf{pf}}(2^n) = 3 \text{ for } n \geq 1. \tag{3}$$

This can be proved by induction on n , using [9, Claim 5] (which states that $\rho_{\mathbf{pf}}(4m) = \rho_{\mathbf{pf}}(2m)$). As we will see, these low values of the abelian complexity function prevent \mathbf{pf} from having the Frobenius property.

We define $\Delta : \mathcal{L}_{\mathbf{pf}} \rightarrow \mathbb{Z}$ by $\Delta(w) = |w|_0 - |w|_1$ and $M : \mathbb{N} \rightarrow \mathbb{Z}$ by $M(n) = \max\{\Delta(\mathcal{L}_{n,\mathbf{pf}})\}$.

Example 1. For $n = 2$ we have

- $\mathcal{L}_{n,\mathbf{pf}} = \{00, 01, 10, 11\}$,
- $\psi(\mathcal{L}_{n,\mathbf{pf}}) = \{(2, 0), (1, 1), (0, 2)\}$,
- $\Delta(\mathcal{L}_{n,\mathbf{pf}}) = \{2, 0, -2\}$, and
- $M(n) = 2$.

Note that for any $w \in \mathcal{L}_{n,\mathbf{pf}}$ we have $\bar{w}^R \in \mathcal{L}_{n,\mathbf{pf}}$, so $-M(n) \leq \Delta(w) \leq M(n)$. We need the following two facts [9, Claims 3 and 4 (and their proofs)]:

$$\rho_{\mathbf{pf}}(n) = M(n) + 1 \tag{4}$$

$$M(n + 1) = M(n) \pm 1. \tag{5}$$

Lemma 1. For $n \geq 2$, the Parikh vectors $(2^{n-1} \pm 2, 2^{n-1} \mp 2)$ do not occur in $\psi(\mathcal{L}_{2^n,\mathbf{pf}})$.

Proof. Since $(1, 3), (2, 2), (3, 1)$ are all elements of $\psi(\mathcal{L}_{4,\mathbf{pf}})$, we can apply the recursive definition (2) inductively to show that $(2^{n-1} \pm 1, 2^{n-1} \mp 1)$ and $(2^{n-1}, 2^{n-1})$ are elements of $\psi(\mathcal{L}_{2^n,\mathbf{pf}})$. From (3), we see that these three vectors are the only vectors in $\psi(\mathcal{L}_{2^n,\mathbf{pf}})$, which establishes the claim. \square

Theorem 1. If $S(0) = a$ and $S(1) = b$ and $4 \leq a < b$ then $\mathbb{N} \setminus S(\mathcal{L}_{\mathbf{pf}})$ is an infinite set. In particular, the word \mathbf{pf} does not have the Frobenius property.

Proof. Suppose that $2 \leq a < b$ and consider a positive integer m with representation $m = a \cdot (2^{n-1} - 2) + b \cdot (2^{n-1} + 2)$ for some (large) n . By Lemma 1, \mathbf{pf} does not contain any factor with Parikh vector $(2^{n-1} - 2, 2^{n-1} + 2)$, so so we must look for another representation $m = a \cdot (2^{n-1} - 2 + tb) + b \cdot (2^{n-1} + 2 - ta)$ for some non-zero integer t . This representation will correspond to a factor w of length $|w| = 2^n + t(b - a)$ with

Parikh vector $(u, v) = (2^{n-1} - 2 + tb, 2^{n-1} + 2 - ta)$. Then $\Delta(w) = u - v = t(b+a) - 4$. Now by (4), we have

$$|t(a + b) - 4| + 1 = |\Delta(w)| + 1 \leq M(2^n + t(b - a)) + 1 = \rho(2^n + t(b - a)) \quad (6)$$

Furthermore, by (4)–(5), we have $\rho(|w| + 1) \leq \rho(|w|) + 1$, which implies

$$\rho(2^{n-1} + t(b - a)) \leq \rho(2^{n-1}) + |t|(b - a) = 3 + |t|(b - a). \quad (7)$$

The inequalities (6) and (7) give

$$|t(a + b) - 4| + 1 \leq 3 + |t|(b - a). \quad (8)$$

If $t < 0$ we get a contradiction immediately, since $|t(a+b)-4| = |t|(a+b)+4$ and (8) becomes $a|t| \leq -1$, which is impossible. If $t > 0$ we have $|t(a+b)-4| = |t|(a+b)-4$ (since $a + b \geq 4$), and (8) becomes $a|t| \leq 3$. Since $t \geq 1$ we find that $a \leq 3$. We conclude that if $a \geq 4$, there are infinitely many $m \notin S(\mathcal{L}_{\mathbf{pf}})$. \square

3. A Binary Word With Abelian Complexity $\Theta(n^{\log_5 2})$

In the last section we saw that the ordinary paperfolding word \mathbf{pf} does not have the Frobenius property, and that in this case this is due to the fact that $\liminf_{n \rightarrow \infty} \rho_{\mathbf{pf}}(n)$ is bounded. This suggests that to produce an (interesting) example of an infinite word with the Frobenius property, we should consider a word Φ with less than maximal abelian complexity but for which

$$\liminf_{n \rightarrow \infty} \rho_{\Phi}(n) = \infty.$$

Let $\phi := \{0, 1\}^* \rightarrow \{0, 1\}^*$ be the morphism that sends $0 \rightarrow 00101$ and $1 \rightarrow 11011$. Let Φ be the fixed point of ϕ that starts with 0: that is, let $\Phi = \phi^\omega(0) = \lim_{n \rightarrow \infty} \phi^n(0)$.

For a general morphism $h : \{0, 1, \dots, k - 1\}^* \rightarrow \{0, 1, \dots, k - 1\}^*$ we define the *incidence matrix* of h as the matrix M_h whose i^{th} column is the Parikh vector of $h(i)$. Blanchet-Sadri et al. [4] conducted an extensive study of the asymptotic abelian complexities of binary words generated by iterating morphisms. We will make use of several ideas from their paper in this section. Following the notation of [4], we will use $z(u)$ to denote the number of zeroes that appear in the factor u . Let $z_0 = z(\phi(0))$ and $z_1 = z(\phi(1))$. We will also use $z_M(n)$ (resp. $z_m(n)$) to denote the maximum (resp. minimum) number of zeroes among factors of length n in Φ . The *difference* and *delta* functions are defined in [4] for a general ℓ -uniform morphism; for our morphism ϕ we have $d = |z_0 - z_1| = 2$ and $\Delta = z_M(\ell) - \max\{z_0, z_1\} = 3 - 3 = 0$.

Example 2. For ϕ as defined above, we have $\Phi = 0010100101110110010111011 \dots$, $z_0 = 3$, $z_1 = 1$, $d = 2$, $\Delta = 0$, $z_m(2) = 0$, $z_M(2) = 2$, and

$$M_\phi = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}.$$

The following result allows us to determine the asymptotic abelian complexity of Φ .

Theorem 2. ([4, Theorem 7]). *Let φ be an ℓ -uniform binary morphism, and define $f(n) = 1$ if $\log_2 n \in \mathbb{Z}$, and $f(n) = \log n$ otherwise. Then the following hold:*

$$\rho^{ab}(n) = \begin{cases} \Theta(1), & d = 0; \\ \Theta(1), & \varphi = ((01)^{\lfloor \frac{\ell}{2} \rfloor} 0, (10)^{\lfloor \frac{\ell}{2} \rfloor} 1) \text{ if } \ell \text{ is odd}; \\ \Theta(1), & \varphi = (01^{\ell-1}, 1^\ell); \\ \tilde{\Theta}(f(n)), & d = 1, \Delta = 0, \text{ and not earlier cases}; \\ O(\log n), & d = 1, \Delta > 0; \\ \Theta(n^{\log_\ell d}), & d > 1. \end{cases}$$

We thus find that $\rho_\Phi(n) = \Theta(n^{\log_5 2})$, which is certainly not maximal. The following is the main result of this section.

Theorem 3. *The word Φ has the Frobenius property.*

We need a preliminary result (Proposition 1 below). In the proof of this result, and again later in this section, we will need to determine, by computer search, the Parikh vectors of all factors of Φ of length r for r up to some specified bound. In order to perform this computation we make use of the following fact:

If $r \leq 5^t$ for some $t \in \mathbb{N}$, then each factor of Φ of length r appears in some $\phi^t(x)$, where $|x| = 2$.

We also note that when performing such a computation there is no need to save all Parikh vectors for factors of length r . This is due to the following result:

Lemma 2. ([14, Lemma 2.1]). *If an infinite word ω has two factors u and v of the same length n for which the i -th entry of the Parikh vector are p and $p + c$ respectively, for some p and c , then for all $\ell = 0, \dots, c$, there exist factors u_ℓ of ω whose i th entry is $p + \ell$.*

Lemma 2 thus implies that the Parikh vectors of factors of length r in Φ are completely determined by the pair $(z_m(r), z_M(r))$.

We need the following result.

Lemma 3. ([4, Lemma 13]). *For a uniform binary morphism with a fixed point at 0 such that $z_0 \geq z_1$,*

$$z_M(\ell n) = dz_M(n) + z_1 n + \Delta \text{ and } z_m(\ell n) = dz_m(n) + z_1 n - \Delta.$$

Similarly, for a uniform binary morphism with a fixed point at 0 such that $z_0 \leq z_1$,

$$z_M(\ell n) = -dz_m(n) + z_1 n + \Delta \text{ and } z_m(\ell n) = -dz_M(n) + z_1 n - \Delta,$$

(and the formulas are equal if $d = 0$).

Proposition 1. For each integer $C \geq 4$, we have

1. $z_M(n) \geq \frac{n}{3} + C$ whenever $n \geq 29 \cdot 5^{C-4}$ and
2. $z_m(n) \leq \frac{n}{3} - C$ whenever $n \geq 132 \cdot 5^{C-4}$.

Proof. To prove part 1 we define $N_C = 29 \cdot 5^{C-4}$. The proof of part 2 is similar, but with the definition $N_C = 132 \cdot 5^{C-4}$. For clarity, we parametrize the property

$$P(j, C) : \left[5^j \cdot N_C \leq n \leq 5^{j+1} \cdot N_C \Rightarrow z_M(n) \geq \frac{n}{3} + C \right].$$

Clearly, if $P(j, C)$ holds for a given C and for all $j \in \mathbb{N}$ then our proposition holds for that C . Thus, we proceed by double-induction on j and C .

We first verify by computer that $29 \leq n \leq 145 \Rightarrow z_M(n) \geq \frac{n}{3} + 4$ and thus $P(0, 4)$ is satisfied. Suppose that $P(j, 4)$ holds for some $j \in \mathbb{N}$ and let $5^{j+1} \cdot N_4 \leq n \leq 5^{j+2} \cdot N_4$. We may write $n = 5k + r$ for some integers k, r with $0 \leq r \leq 4$. Then $5^j \cdot N_4 \leq k + \frac{r}{5} \leq 5^{j+1} \cdot N_4$ and we have two cases: either $k < 5^{j+1} \cdot N_4$ or $k = 5^{j+1} \cdot N_4$.

If $k < 5^{j+1} \cdot N_4$ then $5^j \cdot N_4 \leq k + 1 \leq 5^{j+1} \cdot N_4$ and by $P(j, 4)$ we have $z_M(k + 1) \geq \frac{k+1}{3} + 4$. One of the inequalities (for an ℓ -uniform morphism) in the proof of [4, Lemma 18] is

$$z_M(\ell k + r) \geq dz_M(k + 1) + z_1(k + 1) + \Delta - z_M(\ell - r),$$

which, after substituting the appropriate values for the constants for ϕ , becomes

$$z_M(5k + r) \geq 2z_M(k + 1) + k + 1 - z_M(5 - r) \geq 2z_M(k + 1) + k - 2, \tag{9}$$

since $z_M(1) \leq \dots \leq z_M(5) = 3$. Thus, we have

$$\begin{aligned} z_M(n) &= z_M(5k + r) \\ &\geq 2z_M(k + 1) + k - 2 && \text{(by (9))} \\ &\geq 2 \left(\frac{k + 1}{3} + 4 \right) + k - 2 && \text{(by } P(j, 4)) \\ &= \frac{1}{3}(5k + 4) + \frac{16}{3} \\ &\geq \frac{1}{3}(5k + r) + 4 \\ &= \frac{n}{3} + 4, && \text{(since } 0 \leq r \leq 4) \end{aligned}$$

as required.

If $k = 5^{j+1} \cdot N_4$ then by Lemma 3 we get

$$\begin{aligned} z_M(n) &= z_M(5^{j+2}N_4) = d \cdot z_M(5^{j+1}N_4) + 5^{j+1}N_4 + \Delta \\ &= 2z_M(5^{j+1}N_4) + 5^{j+1}N_4 \\ &\geq 2\left(\frac{5^{j+1}N_4}{3} + 4\right) + 5^{j+1}N_4 \\ &= \frac{5}{3}(5^{j+1}N_4) + 8 \\ &= \frac{n}{3} + 8 \\ &\geq \frac{n}{3} + 4, \end{aligned}$$

as required, and so in either case, $P(j+1, 4)$ holds and by induction we have $P(j, 4)$ for all $j \in \mathbb{N}$.

Suppose that there exist $C \geq 4$ and N_C with $(\forall j \in \mathbb{N})[P(j, C)]$. Now if $n \geq 5N_C$ we may write $n = 5k + r$ where $k \geq N_C$ and $0 \leq r \leq 4$. Then we have

$$\begin{aligned} z_M(n) &= z_M(5k + r) \\ &\geq 2z_M(k + 1) + k - 2 && \text{(by (9))} \\ &\geq 2\left(\frac{k + 1}{3} + C\right) + k - 2 \\ &= \frac{1}{3}(5k + 4) + 2C - \frac{8}{3} \\ &\geq \frac{1}{3}(5k + r) + C + 1 \\ &= \frac{n}{3} + (C + 1), \end{aligned}$$

so $N_{C+1} = 5N_C$ and the result holds by induction. □

Corollary 1. *For each $C \geq 4$ and $N_C = 132 \cdot 5^{C-4}$, we have*

$$\left\{ \left(\left\lfloor \frac{n}{3} \right\rfloor + D, n - \left\lfloor \frac{n}{3} \right\rfloor - D \right) : -C \leq D \leq C \right\} \subseteq \psi(\mathcal{L}_{n, \Phi})$$

for all $n \geq N_C$.

We will use Corollary 1 to show that, given a and b , every sufficiently large integer has a representation $ax + by$ where $(x, y) \in \psi(\mathcal{L}_\Phi)$. Theorem 3 therefore follows from the next lemma.

Lemma 4. *Let $C = \lceil \max \{1 + \frac{a+2b}{3}, b, \frac{b-a}{3}, 4\} \rceil$. Then every integer*

$$M \geq M_{a,b} := \max \left\{ (a + 2b) \cdot \max\{a, b\}, \frac{a + 2b}{3}(132 \cdot 5^{C-4} + |a - b|) \right\}$$

has a representation $M = a(x - tb) + b(y + ta)$ where $(x - tb, y + ta) \in \psi(\mathcal{L}_\Phi)$ for some $t \in \mathbb{Z}$.

Proof. Suppose that $(a, b) = 1$ is given and let $M = ax + by$ for some non-negative integers x, y (note that M is larger than the quantity from (1), so such a representation exists). For each $t \in \mathbb{Z}$ we have $M = a(x - tb) + b(y + ta)$. Our aim is to show that there is a choice of t for which $(x - tb, y + ta) \in \psi(\mathcal{L}_\Phi)$. Note that, from Corollary 1, if we look at large enough factors of Φ we eventually obtain a factor that is roughly one third 0's. Thus, if we define $n(t) = x + y + t(a - b)$, then we seek a t_0 such that $x - t_0b = \frac{1}{3}n(t_0)$ and thus let $t_0 = \frac{2x-y}{2b+a}$. However, t_0 is not necessarily an integer, so we will use either the floor or ceiling $\lfloor t_0 \rfloor$ and show the existence of a subword with length $n(\lfloor t_0 \rfloor)$ and $x - \lfloor t_0 \rfloor b$ zeroes.

We first claim that $x - \lfloor t_0 \rfloor b$ and $y + \lfloor t_0 \rfloor a$ are nonnegative (and thus it is possible to speak of a factor with length $n(\lfloor t_0 \rfloor)$ and $x - \lfloor t_0 \rfloor b$ zeroes). We have

$$x - t_0b = \frac{1}{3}n(t_0) = \frac{1}{2}(y + t_0a)$$

and so $x - t_0a$, $n(t_0)$, and $y + t_0a$ each have the same sign. As well,

$$x - t_0b = \frac{ax + by}{2b + a} = \frac{M}{2b + a} \geq 0$$

so the three integers are nonnegative. Now note that replacing t_0 with $\lfloor t_0 \rfloor$ only changes each expression by a small amount:

$$|x - t_0b - (x - \lfloor t_0 \rfloor b)| < b \text{ and } |y + t_0a - (y + \lfloor t_0 \rfloor a)| < a.$$

Thus if $M > (2b + a) \cdot \max\{a, b\}$ then we have

$$x - \lfloor t_0 \rfloor b > (x - t_0b) - b > \frac{M}{2b + a} - b > \left(\frac{2b + a}{2b + a}\right) \max\{a, b\} - b \geq 0$$

and

$$\begin{aligned} y + \lfloor t_0 \rfloor a &> y + t_0a - a = 2(x - t_0b) - a = 2\left(\frac{M}{2b + a}\right) - a \\ &> 2\left(\frac{2b + a}{2b + a}\right) \max\{a, b\} - a \\ &= 2 \cdot \max\{a, b\} - a \\ &> 0, \end{aligned}$$

and thus both $x - \lfloor t_0 \rfloor b$ and $y + \lfloor t_0 \rfloor a$ are nonnegative as required.

We now show that the corresponding factor exists within Φ . We have two cases:

Case 1: $a > b$. Then we have

$$\begin{aligned} \left\lfloor \frac{1}{3}n(\lfloor t_0 \rfloor) \right\rfloor &\leq \frac{1}{3}(x + y + \lfloor t_0 \rfloor(a - b)) \leq \frac{1}{3}(x + y + t_0(a - b)) \\ &= x - t_0b \text{ by our choice of } t_0 \\ &\leq x - \lfloor t_0 \rfloor b \end{aligned}$$

and

$$\begin{aligned} \left\lfloor \frac{1}{3}n(\lfloor t_0 \rfloor) \right\rfloor &\geq \frac{1}{3}(x + y + \lfloor t_0 \rfloor(a - b)) - 1 \\ &\geq \frac{1}{3}(x + y + (t_0 - 1)(a - b)) - 1 \\ &= x - t_0b - \frac{1}{3}(a - b) - 1 \text{ by our choice of } t_0 \\ &\geq x - (\lfloor t_0 \rfloor + 1)b - \frac{1}{3}(a - b) - 1 \\ &= x - \lfloor t_0 \rfloor b - \left(\frac{a + 2b}{3} + 1 \right) \end{aligned}$$

so

$$\left\lfloor \frac{1}{3}n(\lfloor t_0 \rfloor) \right\rfloor \leq x - \lfloor t_0 \rfloor b \leq \left\lfloor \frac{1}{3}n(\lfloor t_0 \rfloor) \right\rfloor + \left(\frac{a + 2b}{3} + 1 \right). \tag{10}$$

Case 2: $a < b$. Then we have

$$\begin{aligned} \left\lfloor \frac{1}{3}n(\lceil t_0 \rceil) \right\rfloor &\leq \frac{1}{3}(x + y + \lceil t_0 \rceil(a - b)) \leq \frac{1}{3}(x + y + t_0(a - b)) \\ &= x - t_0b \text{ by our choice of } t_0 \\ &\leq x - (\lceil t_0 \rceil - 1)b \\ &= x - \lceil t_0 \rceil b - b \end{aligned}$$

and

$$\begin{aligned} \left\lfloor \frac{1}{3}n(\lceil t_0 \rceil) \right\rfloor &\geq \frac{1}{3}(x + y + \lceil t_0 \rceil(a - b)) - 1 \geq \frac{1}{3}(x + y + (t_0 + 1)(a - b)) - 1 \\ &= x - t_0b + \frac{1}{3}(a - b) \text{ by our choice of } t_0 \\ &\geq x - \lceil t_0 \rceil b + \frac{1}{3}(a - b) \end{aligned}$$

so

$$\left\lfloor \frac{1}{3}n(\lceil t_0 \rceil) \right\rfloor - b \leq x - \lceil t_0 \rceil b \leq \left\lfloor \frac{1}{3}n(\lceil t_0 \rceil) \right\rfloor + \frac{1}{3}(b - a). \tag{11}$$

In either case, we may take $C = \lceil \max \{1 + \frac{a+2b}{3}, b, \frac{b-a}{3}, 4\} \rceil$ and since

$$n(\lfloor t_0 \rfloor) \geq n(t_0) - |a - b| = \frac{3M}{a + 2b} - |a - b| \geq 132 \cdot 5^{C-4} = N_C,$$

by Corollary 1 we have that there exists a subword w of Φ such that $|w| = n(\lfloor t_0 \rfloor)$ and $\psi(w) = (x - \lfloor t_0 \rfloor b, y + \lfloor t_0 \rfloor b)$. \square

As noted, Theorem 3 follows directly from Lemma 4. However, the bound on M described in Lemma 4 is certainly not optimal; the maximum non-representable integer may be much smaller than $M_{a,b}$. We therefore now compute exactly the largest value of $\mathbb{N} \setminus S(\mathcal{L}_\Phi)$ for several small values of a, b .

We compute the complement of $S(\mathcal{L}_\Phi)$ based on the Parikh vectors of factors of length up to

$$r_{a,b} = \frac{M_{a,b}}{\min\{a, b\}}$$

and thus for any integer $M < M_{a,b}$, if it is representable then its representation should appear among the Parikh vectors of factors up to length $r_{a,b}$. For convenience, we collected the Parikh vectors of factors up to length $r_0 = \max\{r_{a,b} : 1 \leq a, b \leq 6\} = 16500 < 5^7$ and then computed $S(\mathcal{L}_\Phi)$ and its complement only using the Parikh vectors of factors of the appropriate lengths. The results are reported in Table 1.

(a, b)	$\lceil M_{a,b} \rceil$	$\mathbb{N} \setminus S(\mathcal{L}_\Phi)$
(1,1)	132	$\{\}$
(1,2)	222	$\{\}$
(1,3)	313	$\{\}$
(1,4)	405	$\{3\}$
(1,5)	2435	$\{3,4,9\}$
(1,6)	14322	$\{3,4,5,10,11\}$
(2,1)	178	$\{\}$
(2,3)	355	$\{1\}$
(2,5)	2652	$\{1,3,6,8,13\}$
(3,1)	244	$\{\}$
(3,2)	311	$\{1\}$
(3,4)	2424	$\{1,2,5,9\}$
(3,5)	14309	$\{1,2,4,7,9,12,17\}$
(4,1)	270	$\{\}$
(4,3)	2204	$\{1,2,5\}$
(4,5)	15405	$\{1, 2, 3, 6, 7, 11, 12, 16, 21, 25\}$
(5,1)	318	$\{\}$
(5,2)	405	$\{1,3\}$
(5,3)	2428	$\{1,2,4,7,15\}$
(5,4)	14305	$\{1, 2, 3, 6, 7, 11, 15, 20, 24\}$
(5,6)	93506	$\{1, 2, 3, 4, 7, 8, 9, 13, 14, 15, 19, 20, 25, 26, 30, 31, 36, 42, 59\}$
(6,1)	366	$\{5\}$
(6,5)	88006	$\{1, 2, 3, 4, 7, 8, 9, 13, 14, 18, 19, 24, 25, 29, 30, 35\}$

Table 1: $\mathbb{N} \setminus S(\mathcal{L}_\Phi)$ for small values of a, b

4. A Ternary Word With Constant Abelian Complexity

Dekking [6] proved that Sturmian words do not have the Frobenius property. If \mathbf{s} is a Sturmian word, then \mathbf{s} is *balanced*, i.e., for all letters $a \in \{0, 1\}$, we have $||u|_a - |v|_a| \leq 1$ whenever u and v are factors of \mathbf{s} of the same length. Furthermore, as noted in the introduction, we have $\rho_{\mathbf{s}}(n) = 2$ for all $n \geq 1$, and indeed, the aperiodic words with this abelian complexity function are exactly the Sturmian words. Dekking also performed a detailed analysis of $S(\mathcal{L}_{\mathbf{f}})$ for the Fibonacci word \mathbf{f} defined as follows.

Definition 1 (Fibonacci Word). Let $\phi = \frac{1}{2}(1 + \sqrt{5}) = 1.618\dots$ and let $\alpha = 2 - \phi = 0.38196\dots$. We define

$$\mathbf{f} = (|(n + 1)\alpha| - |n\alpha|)_{n \geq 1} = 01001010010010100\dots$$

We also note that

$$(|(n + 1)\phi| - |n\phi|)_{n \geq 1} = 21221212212212122\dots$$

is the sequence obtained from \mathbf{f} by applying the map $0 \rightarrow 2$.

Dekking showed that $S(\mathcal{L}_{\mathbf{f}})$ is co-finite except when

$$(S(0), S(1)) \in \{(1, 1), (1, 2), (1, 3), (2, 1)\}.$$

If one wished to extend Dekking’s analysis to ternary words, then in this setting, the natural ternary analogue of Sturmian words are aperiodic ternary words \mathbf{x} with abelian complexity $\rho_{\mathbf{x}}(n) = 3$ for $n \geq 1$. Currently there is no complete characterization of such words; however, Richomme, Saari, and Zamboni [14] proved the following:

Theorem 4. ([14, Theorem 4.2]). *If \mathbf{x} is aperiodic, ternary, and balanced, then $\rho_{\mathbf{x}}(n) = 3$ for $n \geq 1$.*

Hubert [8] gave a useful characterization of aperiodic balanced words. The reader may consult Hubert’s paper for more details. Here, we will use his characterization to construct a word \mathbf{t} from the Fibonacci word \mathbf{f} with abelian complexity 3 for all lengths. For ease of notation, let T be the operation that sends $1 \rightarrow 1$ and every *second* $0 \rightarrow 2$, starting with the *second* 0. Similarly, let \bar{T} be the operation that sends $1 \rightarrow 1$ and every *second* $0 \rightarrow 2$, starting with the *first* 0.

Example 3. Let $\chi = 01010101\dots$. Then $T(\chi) = 01210121\dots$ and $\bar{T}(\chi) = 21012101\dots$.

We define

$$\mathbf{t} = T(\mathbf{f}) = 01201210210210120\dots$$

and we immediately have the following.

Lemma 5. $\rho_{\mathbf{t}}^{ab}(n) = 3$ for all $n \geq 1$.

Proof. By [8] (and its English explanation in [13, Section 4]), the word \mathbf{t} is an aperiodic, uniformly recurrent, balanced word on $\{0, 1, 2\}$, so the result follows from Theorem 4. \square

We will also make use of the following property.

Definition 2 (WELLD OC Property [3]). We say that an infinite aperiodic word λ on $A = \{0, 1, \dots, d - 1\}$ has *well distributed occurrences (WELLD OC)* if for every $m \in \mathbb{N}$ and every subword w of λ we have

$$\{(|u|_0, |u|_1, \dots, |u|_{d-1}) \bmod m : \lambda = uwv\} = \mathbb{Z}_m^d.$$

Sturmian words have the WELLD OC property [3, Theorem 3.3].

Definition 3. For a subset $A \subseteq \mathbb{R}$ and a constant $c \in \mathbb{R}$ we define $c + A := \{c + a : a \in A\}$.

Lemma 6. $\mathcal{L}_{\mathbf{t}} = T(\mathcal{L}_{\mathbf{f}}) \cup \overline{T}(\mathcal{L}_{\mathbf{f}})$.

Proof. Certainly $\mathcal{L}_{\mathbf{t}} \subseteq T(\mathcal{L}_{\mathbf{f}}) \cup \overline{T}(\mathcal{L}_{\mathbf{f}})$, since any factor of \mathbf{t} is obtained by taking a factor of \mathbf{f} and replacing every other 0 with a 2. Let $t_0 \in T(\mathcal{L}_{\mathbf{f}}) \cup \overline{T}(\mathcal{L}_{\mathbf{f}})$. Without loss of generality, say $t_0 = T(w)$ for some $w \in \mathcal{L}_{\mathbf{f}}$. Then by the WELLD OC property (with $m = 2$), there is an occurrence of w in \mathbf{f} where it is preceded by an even number of 0's and an occurrence where it is preceded by an odd number of 0's. Then $T(w)$ and $\overline{T}(w)$ both occur as subwords of t . \square

It is well-known that $0\mathbf{f}[1, n] \in \mathcal{L}_{\mathbf{f}}$ and $1\mathbf{f}[1, n] \in \mathcal{L}_{\mathbf{f}}$. Thus we have $T(0\mathbf{f}[1, n])$, $T(1\mathbf{f}[1, n])$, $\overline{T}(0\mathbf{f}[1, n])$, and $\overline{T}(1\mathbf{f}[1, n])$ in $\mathcal{L}_{\mathbf{t}}$. We will refer to these as the *generating prefixes* later on. Since we only have 3 possible Parikh vectors for each n , exactly two of these must be equal. This equality depends on the parity of $|\mathbf{f}[1, n]|_0$.

Theorem 5. For $n \geq 1$ define $h(n) = \lfloor (n + 1)\alpha \rfloor$. If $|\mathbf{f}[1, n]|_0$ is odd then

$$\begin{aligned} \psi(T(0\mathbf{f}[1, n])) &= \psi(\overline{T}(0\mathbf{f}[1, n])) = \left(\frac{n - h(n)}{2} + \frac{1}{2}, h(n), \frac{n - h(n)}{2} + \frac{1}{2} \right) \\ \psi(T(1\mathbf{f}[1, n])) &= \left(\frac{n - h(n)}{2} + \frac{1}{2}, h(n) + 1, \frac{n - h(n)}{2} - \frac{1}{2} \right) \\ \psi(\overline{T}(1\mathbf{f}[1, n])) &= \left(\frac{n - h(n)}{2} - \frac{1}{2}, h(n) + 1, \frac{n - h(n)}{2} + \frac{1}{2} \right). \end{aligned}$$

If $|\mathbf{f}[1, n]_0|$ is even then

$$\begin{aligned} \psi(T(\mathbf{0f}[1, n])) &= \left(\frac{n-h(n)}{2} + 1, h(n), \frac{n-h(n)}{2}\right) \\ \psi(\overline{T}(\mathbf{0f}[1, n])) &= \left(\frac{n-h(n)}{2}, h(n), \frac{n-h(n)}{2} + 1\right) \\ \psi(T(\mathbf{1f}[1, n])) &= \psi(\overline{T}(\mathbf{1f}[1, n])) = \left(\frac{n-h(n)}{2}, h(n) + 1, \frac{n-h(n)}{2}\right). \end{aligned}$$

Proof. First note that

$$|\mathbf{f}[1, n]_1| = \sum_{i=1}^n ([(i+1)\alpha] - [i\alpha]) = [(n+1)\alpha] - [\alpha] = [(n+1)\alpha] - 0 = h(n).$$

If $|\mathbf{f}[1, n]_0|$ is odd, it is clear that

$$\psi(T(\mathbf{0f}[1, n])) = \left(\frac{n-h(n)+1}{2}, h(n), \frac{n-h(n)+1}{2}\right) = \psi(\overline{T}(\mathbf{0f}[1, n]))$$

since exactly half of the 0's in $\mathbf{0f}[1, n]$ will become 2's after we apply T . For $\mathbf{1f}[1, n]$, we have

$$\psi(T(\mathbf{1f}[1, n])) = \left(\frac{n-h(n)-1}{2} + 1, h(n), \frac{n-h(n)-1}{2}\right).$$

By Lemma 6, we get the third Parikh vector by swapping the first and last components.

If $|\mathbf{f}[1, n]_0|$ is even, we apply a similar line of reasoning to $\psi(T(\mathbf{1f}[1, n])) = \psi(\overline{T}(\mathbf{1f}[1, n]))$, $\psi(T(\mathbf{0f}[1, n]))$, and $\psi(\overline{T}(\mathbf{0f}[1, n]))$, which gives the above. \square

Let $S : \mathcal{L}_t \rightarrow \mathbb{N}$ be a morphism with $S(0) = S_0$, $S(1) = S_1$, and $S(2) = S_2$. As always, we assume that $\gcd(S_0, S_1, S_2) = 1$. Define

$$m(n) = \frac{1}{2} \lfloor n\phi \rfloor (S_0 - 2S_1 + S_2) - \frac{1}{2} n(S_0 - 4S_1 + S_2). \tag{12}$$

(Note that $2m(n)$ is a *generalized Beatty sequence*, in the sense of Allouche and Dekking [1].) Using the fact that $\lfloor -x \rfloor = -\lfloor x \rfloor - 1$ for $x \notin \mathbb{Z}$, we see that $\lfloor n\alpha \rfloor = 2n - \lfloor n\phi \rfloor - 1$. Using this identity and the fact that $S(w) = S_0|w|_0 + S_1|w|_1 + S_2|w|_2$, we obtain (after some algebra) the following corollary of Theorem 5.

Corollary 2. *If $|\mathbf{f}[1, n-1]_0|$ is odd then*

$$\begin{aligned} S(T(\mathbf{0f}[1, n-1])) &= S(\overline{T}(\mathbf{0f}[1, n-1])) = m(n) + \frac{1}{2}S_0 - S_1 + \frac{1}{2}S_2 \\ S(T(\mathbf{1f}[1, n-1])) &= S(T(\mathbf{0f}[1, n-1])) - S_2 + S_1 = m(n) + \frac{1}{2}S_0 - \frac{1}{2}S_2 \\ S(\overline{T}(\mathbf{1f}[1, n-1])) &= S(\overline{T}(\mathbf{0f}[1, n-1])) - S_0 + S_1 = m(n) - \frac{1}{2}S_0 + \frac{1}{2}S_2. \end{aligned}$$

If $|\mathbf{f}[1, n - 1]|_0$ is even then

$$\begin{aligned} S(T(\mathbf{Of}[1, n - 1])) &= m(n) + S_0 - S_1 \\ S(\overline{T}(\mathbf{Of}[1, n - 1])) &= S(T(\mathbf{Of}[1, n - 1])) - S_0 + S_2 = m(n) - S_1 + S_2 \\ S(\overline{T}(\mathbf{1f}[1, n - 1])) &= S(T(\mathbf{1f}[1, n - 1])) = S(T(\mathbf{Of}[1, n - 1])) - S_0 + S_1 = m(n). \end{aligned}$$

Define

$$\begin{aligned} k_1 = o_1 &= \frac{1}{2}S_0 - S_1 + \frac{1}{2}S_2 & k_2 = o_2 &= \frac{1}{2}S_0 - \frac{1}{2}S_2 & k_3 = o_3 &= -\frac{1}{2}S_0 + \frac{1}{2}S_2 \\ k_4 = e_1 &= S_0 - S_1 & k_5 = e_2 &= -S_1 + S_2 & k_6 = e_3 &= 0. \end{aligned}$$

We will refer to the $m(n)$'s as *main terms* and the k_i 's as *offsets*.

Theorem 6. Define $\mu(n) = [(n - 1 - \lfloor(n - 1)\alpha\rfloor) \bmod 2]$. Then $S(\mathcal{L}_{n,\mathbf{t}}) = \{g_1(n), g_2(n), g_3(n)\}$, where

$$\begin{aligned} g_1(n) &= m(n) + e_1 + o_3\mu(n) \\ g_2(n) &= m(n) + e_2 + (o_2 - e_2)\mu(n) \\ g_3(n) &= m(n) + o_3\mu(n). \end{aligned}$$

Proof. Note that $e_i + (o_i - e_i)\mu(n)$ is o_i when $|\mathbf{f}[1, n - 1]|_0$ is odd and e_i when $|\mathbf{f}[1, n - 1]|_0$ is even. We therefore obtain the equations

$$\begin{aligned} g_1(n) &= m(n) + e_1 + (o_1 - e_1)[(n - 1 - \lfloor(n - 1)\alpha\rfloor) \bmod 2] \\ g_2(n) &= m(n) + e_2 + (o_2 - e_2)[(n - 1 - \lfloor(n - 1)\alpha\rfloor) \bmod 2] \\ g_3(n) &= m(n) + e_3 + (o_3 - e_3)[(n - 1 - \lfloor(n - 1)\alpha\rfloor) \bmod 2] \end{aligned}$$

from Corollary 2. □

Theorem 7. The word \mathbf{t} does not have the Frobenius property.

Proof. From Theorem 6 we see that among the first $\max\{g_1(n), g_2(n), g_3(n)\}$ natural numbers, at most $3n$ are in $S(\mathcal{L}_{\mathbf{t}})$. From (12) and Theorem 6 we find that there is a constant C such that for $n \geq 1$, we have

$$\max\{g_1(n), g_2(n), g_3(n)\} \geq \frac{1}{2}n\phi(S_0 - 2S_1 + S_2) - \frac{1}{2}n(S_0 - 4S_1 + S_2) + C.$$

Let

$$\delta := \lim_{n \rightarrow \infty} \frac{|S(\mathcal{L}_{\mathbf{t}}) \cap \{1, \dots, n\}|}{n}$$

denote the natural density of $S(\mathcal{L}_{\mathbf{t}})$. Then

$$\begin{aligned} \delta &\leq \lim_{n \rightarrow \infty} \frac{3n}{\frac{1}{2}n\phi(S_0 - 2S_1 + S_2) - \frac{1}{2}n(S_0 - 4S_1 + S_2) + C} \\ &= \frac{6}{(\phi - 1)(S_0 + S_2) + 2(2 - \phi)S_1}. \end{aligned}$$

The denominator of this last expression is approximately $0.618(S_0 + S_2) + 0.764S_1$. Since each S_i is at least 1, we see that if any S_i is at least 8, this denominator is larger than 6 and hence $\delta < 1$. It follows that if $S_i \geq 8$ for some i , then $S(\mathcal{L}_t)$ has an infinite complement. Thus t does not have the Frobenius property. \square

Next, we determine the maps S for which $S(\mathcal{L}_t)$ is co-finite. We only have to consider those S for which $S_i \leq 7$ for $i = 1, 2, 3$. We will show that it is possible to determine if $S(\mathcal{L}_t)$ is co-finite by checking (by computer) a finite initial segment of the sequence $m(n)$. We begin with an analysis of the first difference sequence

$$\begin{aligned} \Delta m(n) &= m(n + 1) - m(n) \\ &= \frac{1}{2}([\!(n + 1)\phi] - \lfloor n\phi\rfloor)(S_0 - 2S_1 + S_2) - \frac{1}{2}(S_0 - 4S_1 + S_2) \\ &= ([\!(n + 1)\phi] - \lfloor n\phi\rfloor)k_1 - k_1 + S_1. \end{aligned}$$

Recalling that $([\!(n + 1)\phi] - \lfloor n\phi\rfloor)_{n \geq 1}$ is equal to the Fibonacci sequence over $\{2, 1\}$, we see that $\Delta m(n)$ is equal to the Fibonacci sequence over $\{k_1 + S_1, S_1\}$. Let $F = (\Delta m(n))_{n \geq 1}$; i.e, $F[n] = k_1 + S_1$ if $\mathbf{f}[n] = 0$ and $F[n] = S_1$ if $\mathbf{f}[n] = 1$. There is one degenerate case to consider here, namely, the case where $k_1 = 0$. In this case F is constant with each term equal to S_1 . However, the analysis below is not affected by this degenerate situation.

Let

$$k = \max\{|k_i| : i = 1, 2, \dots, 6\},$$

and for a given factor $F[i, j]$ of F , let

$$I(F[i, j]) = \left[k + 1, \sum_{q=i}^{j+1} F[q] - (k + 1) \right].$$

Definition 4 (Semi-image). We define the *even semi-image* of $F[i, j]$ as

$$\mathbb{S}^0(F[i, j]) = \left\{ \sum_{q=i}^s F[q] + e_r + (o_r - e_r) [|\mathbf{f}[i, s]|_0 \bmod 2] : \right. \\ \left. r = 1, 2, 3, \text{ and } s = i, \dots, j \right\}$$

and the *odd semi-image* of $F[i, j]$ as $(k_1 + S_1) + \mathbb{S}^1(F[i, j])$ where

$$\mathbb{S}^1(F[i, j]) = \left\{ \sum_{q=i}^s F[q] + e_r + (o_r - e_r) [1 - |\mathbf{f}[i, s]|_0 \bmod 2] : \right. \\ \left. r = 1, 2, 3, \text{ and } s = i, \dots, j \right\}.$$

These formulas are analogous to the ones from Theorem 6, but instead of using the generating prefixes we can use *any* factor of F . Since, by the WELLDOC property, each factor of F appears with either parity of $(k_1 + S_1)$ -steps prior to it, we must have two semi-images; the even (resp. odd) semi-image represents the image of the factor with an even (resp. odd) number of $(k_1 + S_1)$ -steps before it. The odd semi-image is shifted by $k_1 + S_1$ to account for non-integral k_1 but the same lines of reasoning will apply.

Definition 5 (Semi-complement). We define the *even semi-complement* as

$$\mathbb{K}^0(F[i, j]) = (I(F[i, j]) \setminus \mathbb{S}^0(F[i, j])) \cap \mathbb{N}$$

and the *odd semi-complement* as

$$\mathbb{K}^1(F[i, j]) = ((k_1 + S_1) + I(F[i, j]) \setminus [(k_1 + S_1) + \mathbb{S}^1(F[i, j])]) \cap \mathbb{N}$$

Example 4. Consider the triple $(1, 1, 2)$. The odd offsets are $\{0.5, -0.5, 0.5\}$, the even offsets are $\{0, 1, 0\}$,

$$(m(n))_{n \geq 1} = (1, 2.5, 3.5, 5, 6.5, 7.5, 9, 10, 11.5, 13, 14, \dots),$$

and

$$F = (1.5, 1, 1.5, 1.5, 1, 1.5, 1, 1.5, 1.5, 1, \dots).$$

Let $w = F[1, 4] = (1.5, 1, 1.5, 1.5)$. Then we have $k = 1$, $I(w) = [2, 4.5]$, $\mathbb{S}^0(w) = \{1, 2, 3, 4, 5, 6\}$, and $\mathbb{K}^0(w) = \{2, 3, 4\} \setminus \{1, 2, 3, 4, 5, 6\} = \emptyset$. We also have $\mathbb{S}^1(w) = \{1.5, 2.5, 3.5, 4.5, 5.5, 6.5\}$, and $\mathbb{K}^1(w) = \{4, 5, 6\} \setminus \{3, 4, 5, 6, 7, 8\} = \emptyset$.

Theorem 8. Fix (S_0, S_1, S_2) and let $l = \left\lceil \frac{2(k+1)}{\min\{S_1, k_1 + S_1\}} \right\rceil$. Then the complement of $S(\mathcal{L}_t)$ is finite if and only if $\mathbb{K}^0(F[i, i+l-1]) = \mathbb{K}^1(F[i, i+l-1]) = \emptyset$ for all $i \geq 1$.

We need two preliminary lemmas.

Lemma 7. Let $R(F[i, i+l-1]) = \sum_{q=1}^{i-1} F[q] + I(F[i, i+l-1])$. Then

$$\bigcup_{i \geq 1} R(F[i, i+l-1]) \supseteq \{n \in \mathbb{N} : n > k\}.$$

Proof. It suffices to show that

$$\sum_{q=1}^i F[q] + k + 1 \leq \sum_{q=1}^{i+l} F[q] - (k + 1),$$

which happens if and only if

$$2(k + 1) \leq \sum_{q=1}^{i+l} F[q] - \sum_{q=1}^i F[q] = \sum_{q=i+1}^{i+l} F[q].$$

Since we have $\sum_{q=i+1}^{i+l} F[q] \geq l \min\{S_1, k_1 + S_1\} \geq 2(k + 1)$, we are done. □

Lemma 8. *If $x \in R(F[i, i + l - 1])$ then*

$$x \neq \sum_{q=1}^{i-1-s} F[q] + k_j \quad \text{and} \quad x \neq \sum_{q=1}^{i+l+s} F[q] + k_j$$

for every $s \geq 0$ and $j = 1, 2, \dots, 6$.

Proof. For any $s \geq 0$ and $j = 1, 2, \dots, 6$ we have

$$\sum_{q=1}^{i+l+s} F[q] + k_j \geq \sum_{q=1}^{i+l} F[q] - k > x > \sum_{q=1}^{i-1} F[q] + k \geq \sum_{q=1}^{i-1-s} F[q] + k_j,$$

as required. □

Proof of Theorem 8. We begin with the converse. First note that if x is a factor of F and $|x| = l$ then $\sum x_i > 2(k + 1)$ so $I(x)$ is nonempty. If every semi-complement is empty, then there exists a sequence $(r(i))_{i \geq 1}$ on $\{0, 1\}$ such that

$$\begin{aligned} \mathbb{N} \cap R(F[i, i + l - 1]) &= \mathbb{N} \cap \left(\sum_{q=1}^{i-1} F[q] + I(F[i, i + l - 1]) \right) \\ &= \mathbb{N} \cap \left(\sum_{q=1}^{i-1} F[q] + \mathbb{S}^{r(i)}(F[i, i + l - 1]) \right). \end{aligned}$$

By Lemma 7, we get that $S(\mathcal{L}_t)$ is co-finite.

Now suppose that for some i the set $\mathbb{K}^0(F[i, i + l - 1])$ (resp. $\mathbb{K}^1(F[i, i + l - 1])$) is non-empty, and so one of the semi-images ‘misses’ an integer x_i . By the WELLDOC property, there exist infinitely many indices $\{i_r : r \in \mathbb{N}\}$ where $F[i_r, i_r + l - 1] = F[i, i + l - 1]$ and $|F[1, i_r - 1]|_0$ is even (resp. odd). Thus, for each r there exists an integer $x_{i_r} \in R(F[i_r, i_r + l - 1])$ such that $x_{i_r} \notin \sum_{q=1}^{i_r-1} F[q] + \mathbb{S}^0(F[i_r, i_r + l - 1])$ (resp. $x_{i_r} \notin \sum_{q=1}^{i_r-1} F[q] + \mathbb{S}^1(F[i_r, i_r + l - 1])$). By Lemma 8, $x_{i_r} \notin S(\mathcal{L}_t)$. Thus the complement of $S(\mathcal{L}_t)$ is infinite. □

Note that by Lemma 6, our results are symmetric with respect to S_0 and S_2 and if $S_0 = S_2$ then all of the results in [6] hold. As well, any triple with a greatest common divisor greater than one will have infinitely many elements in the complement of $S(\mathcal{L}_t)$. Thus, in all of the following calculations we skip any triple (x, y, z) where $\gcd(x, y, z) > 1$, $x = z$, or where (z, y, x) has already been evaluated.

For each triple, we first calculate $l = \left\lceil \frac{2(k+1)}{\min\{k_1+S_1, S_1\}} \right\rceil$ and then calculate all $l + 2$ factors³ of length $l + 1$ in⁴ F . We then calculate the semi-complements of each

³In the cases where $S_0 + S_2 = 2S_1$, i.e. F is constant, we merely check the semi-image for the single factor $F[1, l + 1]$.

⁴Different letters may follow different occurrences of each factor. The extra term at the end allows us to account for all possible values of $F[j + 1]$ when calculating $I(F[i, j])$.

factor of F , and by Theorem 8, if we find a non-empty semi-complement we know that the complement of $S(\mathcal{L}_t)$ is infinite; otherwise, the complement of $S(\mathcal{L}_t)$ is finite. We found 13 triples with finite complements. These are listed in Table 2.

(S_0, S_1, S_2)	$\mathbb{N} \setminus S(\mathcal{L}_t)$
(1, 1, 2)	$\{\}$
(1, 1, 3)	$\{\}$
(1, 1, 4)	$\{\}$
(1, 2, 2)	$\{\}$
(1, 2, 3)	$\{\}$
(1, 2, 4)	$\{\}$
(1, 3, 2)	$\{\}$
(1, 3, 5)	$\{2\}$
(1, 4, 2)	$\{\}$
(2, 1, 3)	$\{\}$
(2, 1, 4)	$\{\}$
(2, 1, 5)	$\{\}$
(2, 3, 4)	$\{1\}$

Table 2: Maps S for which $S(\mathcal{L}_t)$ has a finite complement

5. Futher Work

We have just given some examples of infinite words that either have or do not have the Frobenius property. In general, we would like to have a theorem that classifies an infinite word as either having or not having the Frobenius property based on its abelian complexity. For instance, is it true that if \mathbf{w} has abelian complexity $\rho_{\mathbf{w}}(n) = \Omega(n^r)$ for some $r > 0$, or perhaps even $\rho_{\mathbf{w}}(n) = \Omega(\log n)$, then \mathbf{w} has the Frobenius property? What happens when we move to ternary or larger alphabets?

References

- [1] J.-P. Allouche, M. Dekking, Generalized Beatty sequences and complementary triples. Preprint. <https://arxiv.org/abs/1809.03424>
- [2] H. Ardal, T. Brown, V. Jungić, and J. Sahasrabudhe, On abelian and additive complexity in infinite words, *Integers*, **12** (2012), #A21.
- [3] Ľ. Balková, M. Bucci, A. De Luca, J. Hladký, S. Puzynina, Aperiodic pseudorandom number generators based on infinite words, *Theoret. Comput. Sci.* **647** (2016), 85–100.
- [4] F. Blanchet-Sadri, N. Fox, N. Rampersad, On the asymptotic abelian complexity of morphic words, *Adv. Appl. Math.* **61** (2014), 46–84.
- [5] E. M. Coven and G. A. Hedlund, Sequences with minimal block growth, *Mathematical Systems Theory* **7** (1973), 138–153.

- [6] M. Dekking, The Frobenius problem for homomorphic embeddings of languages into the integers, *Theoret. Comput. Sci.* **732** (2018), 73–79.
- [7] M. Dekking, M. Mendès France, A. van der Poorten, Folds I–III, *Math. Intelligencer* **4** (1982), 130–138, 172–181, 190–195.
- [8] P. Hubert, Suites équilibrées, *Theoret. Comput. Sci.* **242** (2000), 91–108.
- [9] B. Madill, N. Rampersad, The abelian complexity of the paperfolding word, *Discrete Math.* **313** (2013), 831–838.
- [10] J. Ramírez Alfonsín, *The Diophantine Frobenius Problem*, Oxford Univ. Press, 2005.
- [11] J. Steuding, P. Stumpf, On the Frobenius problem for Beatty sequences, *Indag. Mathematicae* **28** (2017), 132–137.
- [12] J. Sylvester, Mathematical questions with their solutions, *Educational Times* **41** (1884), 21.
- [13] L. Vuillon, Balanced words, *Bull. Belg. Math. Soc. Simon Stevin* **10** (2003), 787–805.
- [14] G. Richomme, K. Saari, and L. Q. Zamboni, Abelian complexity of minimal subshifts, *J. London Math Soc.* **83** (2010), 79–95.