The Frobenius problem for homomorphic embeddings of languages into the integers

Michel Dekking

DIAM, Delft University of Technology, December 12, 2017

1 Introduction

The Frobenius problem is also known as the ‘coin problem’. Since the value of a coin can only be positive, we will consider exclusively embeddings into the natural numbers \( \mathbb{N} = \{1, 2, 3, \ldots\} \). Let \( \mathcal{L} \) be a language, i.e., a sub-semigroup of the free semigroup generated by a finite alphabet under the concatenation operation.

A homomorphism of \( \mathcal{L} \) into the natural numbers is a map \( S : \mathcal{L} \to \mathbb{N} \) satisfying

\[
S(vw) = S(v) + S(w), \quad \text{for all } v, w \in \mathcal{L}.
\]

The two main questions to be asked about the image set \( S(\mathcal{L}) \) are

(Q1) Is the complement \( \mathbb{N} \setminus S(\mathcal{L}) \) finite or infinite?
(Q2) If the complement of \( S(\mathcal{L}) \) is finite, then what is the largest element in this set?

These two questions are known as the Frobenius problem in the special case that \( \mathcal{L} \) is the full language consisting of all words over a finite alphabet. In this case they have been posed as a problem (with solution) for an alphabet \( \{a, b\} \) of cardinality 2 by James Joseph Sylvester in 1884 [14]: \( \mathbb{N} \setminus S(\mathcal{L}) \) is finite, and its largest element is

\[
S(a)S(b) - S(a) - S(b).
\]

In this paper we will also restrict ourselves to the two symbol case: alphabet \( \{a, b\} \).

In Section 2 we prove that for the golden mean language (“no bb”) the set \( \mathbb{N} \setminus S(\mathcal{L}) \) is finite, with largest element

\[
S(a)^2 + S(a)S(b) - 3S(a) - S(b).
\]

Our main interest is however not in sofic languages\(^1\), but in languages with low complexity, where the complement of \( S(\mathcal{L}) \) can be infinite.

In Section 3 we analyse the case of Sturmian languages, and show that for the Fibonacci language a 0–∞ law holds: either the complement is empty or it has infinite cardinality.

In Section 4 we show that for any homomorphism \( S \) the image of the Thue-Morse language will consist of a union of 5 arithmetic sequences.

In Section 5 we consider two-dimensional embeddings, which behave quite differently.

---

\(^1\)Languages defined by the labelling of infinite paths of an automaton.
We usually suppose that $\gcd(S(a), S(b)) = 1$. First of all this is not a big loss since automatically the complement will have infinite cardinality in this case. Secondly, if $r$ divides both $S_1(a)$ and $S_1(b)$ for some homomorphism $S_1$, then

$$S_1(L^n) = r^n S_2(L^n), \text{ for } n = 1, 2, \ldots,$$

where $S_2(a) = \frac{S_1(a)}{r}, S_2(b) = \frac{S_1(b)}{r}$.

Our work is related to the work on abelian complexity, see, e.g., [3], [12], [8]. See Lemma 3.1 for such a connection.

Our work is also related to the notion of additive complexity, see [13] and [2]. The additive complexity of an infinite word $w$ over a finite set of integers (see [2]) is the function $n \rightarrow \phi^+(w, n)$ that counts the number of distinct sums obtained by summing $n$ consecutive symbols of $w$. In general we write $L^n_w$ for the set of words of length $n$ in a language $L$. Let $L_w$ be the language of all words occurring in the infinite word $w$. Then the additive complexity is $\phi^+(w, n) = \text{Card}\{S(u) : u \in L^n_w\}$, where $S$ is the identity map on the alphabet of $w$.

We finally mention that homomorphisms $S$ from a language to the natural numbers already occur in the 1972 paper [4, Section 6] in the context of the Fibonacci language, where they are called weights.

2 Homomorphic images of the golden-mean language

The golden mean language is the language $L_{GM}$ consisting of all words over $\{a, b\}$ in which $bb$ does not occur as a subword. Now if $S$ satisfies $S(a) = 1$ or $S(b) = 1$, then it is easily seen that $S(L_{GM}) = \mathbb{N}$, so for these homomorphisms the golden mean and the full language both map to $\mathbb{N}$. One could say they both have Frobenius number 0. In general however, the Frobenius number will increase substantially. If we take $S$ defined by

$$S(a) = 100, \ S(b) = 3,$$

then the Frobenius number of the full language under $S$ is $300 - 100 - 3 = 197$, and the Frobenius number of $S(L_{GM})$ is equal to $9997$. For arbitrary homomorphisms the solution of the Frobenius problem for the golden mean language is given by the following, where we write $S_a := S(a), S_b := S(b)$.

**Theorem 2.1** Let $S : L_{GM} \to \mathbb{N}$ be a homomorphism. Suppose $\gcd(S_a, S_b) = 1$, and both $S_a > 1$ and $S_b > 1$. Then the Frobenius number of $S(L_{GM})$ is equal to

$$\max \mathbb{N} \setminus S(L_{GM}) = S_a(S_a - 3) + S_b(S_a - 1).$$

**Proof:** Let an $S_a$-point be defined as a multiple $nS_a$, $n = 0, 1, \ldots$, and an $S_a$-interval as the set of numbers between two consecutive $S_a$-points. We also consider $S_b$-chains, defined for $n \geq 0$ by

$$C(n) = \{nS_a + S_b, \ nS_a + 2S_b, \ldots, \ nS_a + (n + 1)S_b\}.$$
Note that the union of the \( S_a \)-points and the \( S_b \)-chains will give \( \mathcal{L}_{GM} \).

The key observation is that the \( S_b \)-chain \( C(S_a - 2) \) has \( S_a - 1 \) elements, which are all different modulo \( S_a \). This is a consequence of \( \gcd(S_a, S_b) = 1 \). It follows that the \( S_b \)-chains fill in more and more points of the \( S_a \)-intervals. The last point to be filled in is modulo \( S_a \) equal to \( S_a - S_b \), produced by the last element of the chain \( C(S_a - 2) \).

This is the number
\[
P := (S_a - 2)S_a + (S_a - 1)S_b.
\]

But then the largest number in the complement of \( \mathcal{L}_{GM} \) is \( P - S_a \), which is the number as claimed in the theorem. In this argument we used that if a point in an \( S_a \)-interval is filled in, then the corresponding points modulo \( S_a \) in all later intervals will also be filled in, simply because the later chains will be extensions of the earlier ones. \( \Box \)

![Figure 1: Example with \( S(a) = 7, S(b) = 3 \): row \( n \) shows the \( S_a \)-points in blue, and the \( S_b \)-chain \( C(n - 1) \) in yellow and green, for \( n = 1, \ldots, 8 \) (truncated at 56).](image)

### 3 Sturmian languages

Sturmian words are infinite words over a two letter alphabet that have exactly \( n + 1 \) subwords for each \( n = 1, 2, \ldots \). We call the collection of these subwords a Sturmian language. There is a surprising characterization of Sturmian words: \( s \) is Sturmian if and only if \( s \) is irrational mechanical, which means that there exists an irrational number \( \alpha \in (0, 1) \) and a number \( \rho \) such that
\[
s(\alpha, \rho) = \left\lfloor (n + 1) \alpha + \rho \right\rfloor - \left\lfloor n \alpha + \rho \right\rfloor, \quad s'(\alpha, \rho) = \left\lceil (n + 1) \alpha + \rho \right\rceil - \left\lceil n \alpha + \rho \right\rceil
\]

See, e.g., [10, Prop. 2.1.13]. Because of this representation, we will use the alphabet \( \{0, 1\} \) instead of \( \{a, b\} \) in this section.

Of special interest are the Sturmian words \( s_\alpha := s(\alpha, 0) \) and \( s'_\alpha := s'(\alpha, 0) \) of intercept 0. These have the property that they only differ in the first element:
\[
s_\alpha = 0 c_\alpha, \quad s'_\alpha = 1 c_\alpha.
\]

Here \( c_\alpha := s_{\alpha, \alpha} \) is called the characteristic word of \( \alpha \). For \( n \geq 0 \) we have
\[
c_\alpha(n) = s_{\alpha, \alpha}(n) = [(n + 1) \alpha + \alpha] - [n \alpha + \alpha] = [(n + 1) \alpha] - [(n + 1) \alpha].
\]

The words \( s_\alpha, s'_\alpha \) and \( c_\alpha \) generate the same language ([10, Prop.2.1.18]), which we denote \( \mathcal{L}_\alpha \). Recall that \( \mathcal{L}^n_\alpha \) is the set of words of length \( n \) in \( \mathcal{L}_\alpha \).
Lemma 3.1 Let $L_\alpha$ be a Sturmian language, and let $S$ be a homomorphism with $S(0) \neq S(1)$. Then $\text{Card } S(L_\alpha^n) = 2$ for all $n \geq 1$.

Proof: This follows directly from the fact ([10, Th.2.1.5]) that Sturmian words are balanced, i.e., any two words of the same length can at most differ 1 in their number of ones. \[\square\]

A sequence $([n\alpha])$, where $[.]$ denotes integer part, is called a Beatty sequence if $\alpha > 1$, and a slow Beatty sequence if $0 < \alpha < 1$ (terminology from [9]).

Theorem 3.1 Let $\alpha$ be an irrational number from $(0, 1)$. Let $L_\alpha$ be the Sturmian language generated by $\alpha$, and let $(q_n)_{n \geq 0}$ be the slow Beatty sequence defined by $q_n = \lfloor (n+1)\alpha \rfloor$.

Let $S : L_\alpha \to \mathbb{N}$ be a homomorphism. Define $S_0 = S(0), S_1 = S(1)$. Then $S(L_\alpha) = \{(S_1 - S_0)q_n + nS_0 + S_0 : \ n = 0, \ldots \} \cup \{(S_1 - S_0)q_n + nS_0 + S_1 : \ n = 0, \ldots \}$.

Proof: If $S_0 = S_1$ then this is certainly true, so suppose $S_0 \neq S_1$ in the sequel. We denote $c_{\alpha}[i, j] := c_{\alpha}(i) \ldots c_{\alpha}(j)$ for integers $0 \leq i < j$. Let $N_\ell(w)$ denote the number of occurrences of the letter $\ell$ in a word $w$ for $\ell = 0, 1$. Then $N_1(c_\alpha[0, n-1]) = \sum_{k=0}^{n-1} c_\alpha(k) = [(n+1)\alpha] - [\alpha] = q_n, \quad N_0(c_\alpha[0, n-1]) = n - q_n.$

Of course all words $c_\alpha[0, n-1]$ are in the Sturmian language $L_\alpha$, but $L_\alpha$ also contains the words $0c_\alpha[0, n-1]$ and $1c_\alpha[0, n-1]$. It thus follows from Lemma 3.1 that $S(L_\alpha)$ is given by the union of all images $S(0c_\alpha[0, n-1])$ and $S(1c_\alpha[0, n-1])$. Since $S(0c_\alpha[0, n-1]) = S_0 + (n - q_n)S_0 + q_nS_1 = (S_1 - S_0)q_n + nS_0 + S_0,$ the result follows. \[\square\]

3.1 The Fibonacci language

Let $\Phi = (\sqrt{5} + 1)/2 = 1.61803 \ldots$ be the golden mean, and let $\alpha := 2 - \Phi$. We have $c_\alpha = ([(n+1)\alpha] - [n\alpha])_{n \geq 1} = 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, \ldots,$
the infinite Fibonacci word. We write $L_F := L_\alpha$.

Theorem 3.2 Let $S : L_F \to \mathbb{N}$ be a homomorphism. Then $S(L_F) = ((S_0 - S_1)|n\Phi| + (2S_1 - S_0)n + S_0 - S_1)_{n \geq 1} \cup ((S_0 - S_1)|n\Phi| + (2S_1 - S_0)n)_{n \geq 1}$. 

4
According to Lemma 3.2 the complement of $S(0)$ is empty for non-integer $x$:

\[(S_1 - S_0)q_{n-1} + nS_0 = (S_1 - S_0)[n\alpha] + nS_0 = (S_1 - S_0)[n(2 - \Phi)] + nS_0 = 2(S_1 - S_0)n + (S_1 - S_0)[-n\Phi] + nS_0 = (2S_1 - S_0)n + (S_1 - S_0)(-n\Phi - 1) = (S_0 - S_1)[n\Phi] + (2S_1 - S_0)n + S_0 - S_1.\]

\[\square\]

**Lemma 3.2** For $S(0) = 1, S(1) \leq 3$ or $S(0) = 2, S(1) = 1$ one has $S(\mathcal{L}_F) = \mathbb{N}$.

**Proof:** Take $(S_0, S_1) = (1, 1)$. Then obviously $S(\mathcal{L}_F) = \mathbb{N}$.

Take $(S_0, S_1) = (2, 1)$. Then $S(\mathcal{L}_F) = \mathbb{N}$, since by Theorem 3.2 $S(\mathcal{L}_F)$ is the union of $([n\Phi])$ and $([n\Phi] + 1)$, where the difference of two consecutive terms in $([n\Phi])$ is never more than $2$.

Take $(S_0, S_1) = (1, 2)$. Then $S(\mathcal{L}_F) = \mathbb{N}$, since $S(\mathcal{L}_F)$ is the union of $([n(3 - \Phi)])$ and $([n(3 - \Phi)] + 1)$, where the difference of two consecutive terms in $([n(3 - \Phi)])$ is never more than $2$.

Take $(S_0, S_1) = (1, 3)$. This case is more complicated. Let $u := (-2[n\Phi] + 5n - 2)_{n \geq 1}$, and $v := u + 2$. Then according to Theorem 3.2 the union of the sets determined by $u$ and $v$ is $S(\mathcal{L}_F)$. Let $\Delta u$ be the difference sequence defined by $\Delta u_n = u_{n+1} - u_n$ for $n \geq 0$. It is easy to see that the difference sequences $\Delta v$ and $\Delta u$ are both equal to the Fibonacci sequence $1, 3, 1, 1, 3, 1, \ldots$ on the alphabet $\{1, 3\}$ (cf. [1]). We claim that if two consecutive numbers $m, m + 1$ are missing in $u$, then these two do appear in $v$, implying that $S(\mathcal{L}_F) = \mathbb{N}$. Indeed the two missing numbers are characterized by $u_{n+1} - u_n = 3$ for some $n$, and the missing numbers are $m = u_n + 1$ and $u_n + 2$. The second number appears in $v$, simply because $v = u + 2$. The first number appears because $u_{n+1} - u_n = 3$ implies $u_n - u_{n-1} = 1$ (no $33$ in the $1$-$3$-Fibonacci sequence), and so $v_{n-1} = v_n - 1 = u_n + 1$. \[\square\]

We define $\mathcal{E} := \{(1, 1), (1, 2), (1, 3), (2, 1)\}$.

**Theorem 3.3** Let $S : \mathcal{L}_F \to \mathbb{N}$ be a homomorphism. Then $\mathbb{N} \setminus S(\mathcal{L}_F)$ has infinite cardinality, unless $(S(0), S(1)) \in \mathcal{E}$, in which case the complement is empty.

**Proof:** According to Lemma 3.2 the complement of $S(\mathcal{L}_F)$ is empty for $(S_0, S_1) \in \mathcal{E}$.

The density of the set $S(\mathcal{L}_F)$ in the natural numbers exists, and equals

$$\delta := \frac{2}{(S_0 - S_1)\Phi + 2S_1 - S_0}.$$  

The theorem will be proved if we show that $\delta < 1$ for $(S_0, S_1)$ not in $\mathcal{E}$. First we note that the denominator of $\delta$ is positive:

$$(S_0 - S_1)(\Phi - 1) + S_1 > -S_1(\Phi - 1) + S_1 = S_1(2 - \Phi) > 0,$$

where we used that $1 < \Phi < 2$. We now have

$$\delta < 1 \iff (S_0 - S_1)\Phi + 2S_1 - S_0 > 2 \iff (S_0 - S_1)\Phi > S_0 - S_1 + 2 - S_1.$$
If $S_0 > S_1$, this is satisfied, since under this condition $(2 - S_1)/(S_0 - S_1) \leq 0$, unless $(S_0, S_1) = (2, 1) \in \mathcal{E}$. If $S_0 < S_1$, we have to see that $\Phi < 1 + (2 - S_1)/(S_0 - S_1)$. This holds for $S_0 \geq 2$, since then $(2 - S_1)/(S_0 - S_1) \geq 1$. If $S_0 = 1$, then this does not hold for $S_1 = 1, 2, 3$, i.e., for pairs from $\mathcal{E}$, but it will hold for all $S_1 \geq 4$. 

For particular values of $S(0)$ and $S(1)$ the complement of the embedding of the language has a nice structure, as it can be expressed in the classical Beatty sequences $A(n) = [n\Phi]$ for $n \geq 1$, and $B(n) = [n\Phi^2]$ for $n \geq 1$. The sequences $A$ and $B$ are called the lower Wythoff sequence and upper Wythoff sequence; they are extremely well-studied.

**Example 1.** Let $S$ be given by $S(0) = 3$ and $S(1) = 2$. In the following we use the notation $pX + qY + r = (pX(n) + qY(n) + r)_{n \geq 1}$ for real numbers $p, q, r$ and functions $X, Y: \mathbb{N} \to \mathbb{N}$. Then

$$S(\mathcal{L}_F) = B(\mathbb{N}) \cup B + 1(\mathbb{N}), \quad \mathbb{N} \setminus S(\mathcal{L}_F) = \{1, 4, 9, 12, \ldots\} = 2A + \text{Id} + 1 (\mathbb{N} \cup \{0\}).$$

The first statement follows directly from Theorem 3.2. The second statement follows in a number of steps from the fact that $A$ and $B$ form a Beatty pair: $A(\mathbb{N}) \cap B(\mathbb{N}) = \emptyset$, and $A(\mathbb{N}) \cup B(\mathbb{N}) = \mathbb{N}$. This implies that $A(A(\mathbb{N})) \cup A(B(\mathbb{N})) \cup B(\mathbb{N}) = \mathbb{N}$, where the three sets are disjoint. But $AA = B - 1$ (see, e.g., Formula (3.2) in [4]). Adding 1 to all three sequences it follows that

$$B(\mathbb{N}) \cup B + 1(\mathbb{N}) \cup AB + 1(\mathbb{N}) = \mathbb{N} \setminus \{1\}.$$ 

Moreover, according to [4, Formula (3.5)] one has $AB = A + B = 2A + \text{Id}$. But then the three sequences $(\lfloor n\Phi \rfloor + n)_{n \geq 1}$, $(\lfloor n\Phi^2 \rfloor + n + 1)_{n \geq 1}$, $(\lfloor 2n\Phi \rfloor + n + 1)_{n \geq 0}$, form a complementary triple, i.e., as sets they are disjoint, and their union is $\mathbb{N}$.

A similar result holds for $S(0) = 4$, $S(1) = 3$.

**Example 2.** Let $S$ be given by $S(0) = 3$ and $S(1) = 1$, then by Theorem 3.2

$$S(\mathcal{L}_F) = 2A - \text{Id} (\mathbb{N}) \cup 2A - \text{Id} + 2(\mathbb{N}).$$

It is proved in [4] that

$$\mathbb{N} \setminus S(\mathcal{L}_F) = \{2, 9, 20, 27, 38, 49, \ldots\} = 4A + 3\text{Id} + 2 (\mathbb{N} \cup \{0\}),$$

and that the three sequences $(2\lfloor n\Phi \rfloor - n)_{n \geq 1}$, $(2\lfloor n\Phi^2 \rfloor - n + 2)_{n \geq 1}$, $(4\lfloor n\Phi \rfloor + 3n + 2)_{n \geq 0}$, form a complementary triple.

### 4 The Thue-Morse language

Let $\theta$ given by $\theta(a) = ab$, $\theta(b) = ba$ be the Thue-Morse morphism. Let $\mathcal{L}_\text{TM}$ be the language generated by this morphism. Let $R_{r,s} = \{s, r+s, 2r+s, \ldots\}$ be the set determined by the arithmetic sequence with terms $rn + s$ for $n = 0, 1, \ldots$.

**Theorem 4.1** Let $S : \mathcal{L}_\text{TM} \to \mathbb{N}$ be a homomorphism. Define $p = S(0), q = S(1)$. Then

$$S(\mathcal{L}_\text{TM}) = R_{p+q,0} \cup R_{p+q,p} \cup R_{p+q,q} \cup R_{p+q,2p} \cup R_{p+q,2q}. $$

---

2In these two cases $\mathbb{N} \setminus S(\mathcal{L}_F)$ is given by sequences A276885, respectively A276886 in OEIS ([3]). It is easily seen that the definitions of these sequences in OEIS are equivalent to the way in which we obtain them.
Proof: Let $L_{TM}^n$ be the set of words of length $n$ in the Thue-Morse language. Put $r = S(ab) = p + q$. It is clear (and for $p = 0, q = 1$ observed also in [12]) that since the Thue-Morse word is a non-periodic concatenation of $ab$ and $ba$ that for $n = 1, 2, \ldots$

$$S(L_{TM}^{2n}) = \{rn, rn + q - p, rn + p - q\}, \quad S(L_{TM}^{2n-1}) = \{rn + p, rn + q\}.$$ 

This implies the statement of the theorem.

\[ \square \]

**Theorem 4.2** Let $S : L_{TM} \rightarrow \mathbb{N}$ be a homomorphism. Then $\mathbb{N} \setminus S(L_F)$ has infinite cardinality if and only if $S(a) + S(b) \geq 6$. For $S(a) + S(b) < 6$, the complement is either empty or a singleton.

Proof: This follows directly from Theorem 4.1. If $S(a) + S(b) \geq 6$, then the density of $\mathbb{N} \setminus S(L_{TM})$, is at least $1/6$, so the set has infinite cardinality. The results for $S(a) + S(b) < 6$ follow also directly from the previous theorem.

\[ \square \]

**Remark** Let $\sigma$ given by $\sigma(a) = ab, \sigma(b) = aa$ be the period-doubling or Toeplitz morphism. The difficulty—see [8, Lemma 6]—of determining the abelian complexity of the period-doubling morphism already indicates that solving the Frobenius problem for the period-doubling language will be much more involved than for the Thue-Morse language.

5 Two dimensional embeddings

We learn from this that the alphabet is ‘too small’, and that we should rather consider embeddings in $\mathbb{Z} \times \mathbb{Z}$ instead of $\mathbb{N} \times \mathbb{N}$. We focus again on low complexity languages, in particular on those generated by a primitive morphism $\phi$ on an alphabet $A$. Such a morphism has a language $L_\phi$ associated to it, where each word $w \in L_\phi$ has a measure $\mu_\phi(w)$. For a given homomorphism $S : L_\phi \rightarrow \mathbb{Z} \times \mathbb{Z}$ we call the average

$$\Delta_\phi(S) := \sum_{a \in A} \mu_\phi(a)S(a)$$

the drift of $S$.

**Proposition 5.2** Let $L_\phi$ be a language generated by primitive morphism on an alphabet $A$, and let $S : L_\phi \rightarrow \mathbb{Z} \times \mathbb{Z}$ be a homomorphism. Then $\mathbb{Z} \times \mathbb{Z} \setminus S(L)$ has infinite cardinality if $\Delta_\phi(S) \neq (0, 0)$.
Proof: It is well-known that the measure \( \mu_\phi \) is strictly ergodic. Because of this, we have for words \( w \) from \( L_\phi \), where \( |w| \) denotes the length of \( w \),

\[
\frac{1}{|w|} S(w) = \frac{1}{|w|} \sum_{a \in A} N_a(w) S(a) \rightarrow \sum_{a \in A} \mu_\phi(a) S(a) = \Delta_\phi(S) \text{ as } |w| \rightarrow \infty.
\]

Thus for long words \( w \) the images \( S(w) \) will be concentrated around the line in the direction of the drift of \( S \), and so the complement of \( S(L_\phi) \) will have infinite cardinality if the drift is not \((0, 0)\).

Can we say something about the Frobenius problem for homomorphic images of morphic languages of an embedding with drift \((0, 0)\)? We shall give an infinite family of morphic languages \( L_\theta \) on an alphabet \( A = \{a,b,c,d\} \) of four letters where for the homomorphism \( S^\oplus \) given by

\[
S^\oplus(a) = (1,0), \quad S^\oplus(b) = (0,1), \quad S^\oplus(c) = (-1,0), \quad S^\oplus(d) = (0,-1)
\]

the homomorphic embedding is the whole \( \mathbb{Z} \times \mathbb{Z} \)—and thus the complement is empty. We shall make use of the paperfolding morphisms introduced in [6]. Let \( \sigma \) be the rotation morphism on the alphabet \( \{a,b,c,d\} \) given by \( \sigma(a) = b \), \( \sigma(b) = c \), \( \sigma(c) = d \), \( \sigma(d) = a \), and let \( \tau \) be the anti-morphism given by \( \tau(w_1 \ldots w_n) = w_n \ldots w_1 \).

A morphism \( \theta \) on \( \{a,b,c,d\} \) is called a paperfolding morphism if

1) \( \sigma \tau \theta = \theta \),

2) Letters from \( \{a,c\} \) alternate with letters from \( \{b,d\} \) in \( \theta(a) \).

A paperfolding morphism is called symmetric if \( \sigma \theta = \theta \). It is clear that this happens if and only if the word \( \theta(a) \) is a palindrome.

Let \( G \) be a (semi-) group with operation \( + \) and unit \( e \). In general an infinite word \( x = (x_n) \) over an alphabet \( A \) and a homomorphism \( S : A^* \rightarrow G \) generate a walk \( Z = (Z_n)_{n \geq 0} \) by (cf. [3])

\[
Z_0 = e, \quad Z_{n+1} = Z_n + S(x_n) = S(x_0 \ldots x_n), \text{ for } n \geq 0.
\]

A paperfolding morphism \( \theta \) with \( \theta(a) = a \ldots \) is called perfect if the four walks generated by the fixed point \( x = \theta^\infty(a) \), and its three rotations over \( \pi/2 \), \( \pi \) and \( 3\pi/2 \) visit every integer point in the plane exactly twice (except the origin, which is visited 4 times).

In [6] it is—not explicitly—proved that for any odd integer \( N \) that is the sum of two squares there exists a perfect symmetric paperfolding morphism of length \( N \). To make the proof explicit, one uses that according to the paragraph at the end of Section 7 in [6] there exists a symmetric planefilling and self-avoiding string for each such \( N \), and then one observes that the construction of such a string in the proof of [6, Theorem 4] always satisfies the perfectness criterion given in [6, Theorem 5].

The smallest length is \( N = 5 \), with morphism \( \theta \) given by

\[
\theta(a) = abcba, \quad \theta(b) = bcdcb, \quad \theta(c) = cdadc, \quad \theta(d) = dabad.
\]

\[\text{This corrects an omission in [6, Definition 1].}\]
Figure 2: The four images of the words $\theta^4(a), \ldots, \theta^4(d)$ under $S^\oplus$, where $\theta$ is the perfect symmetric 5-folding morphism. The origin is not covered, but it is the image of the word $abcd \in \mathcal{L}_\theta$.

**Proposition 5.3** Let $\mathcal{L}_\theta$ be the language generated by a perfect symmetric paperfolding morphism $\theta$. Then $S^\oplus(\mathcal{L}_\theta) = \mathbb{Z} \times \mathbb{Z}$.

**Proof:** This follows directly from Theorem 5 in [6], using the observation above.

**References**

[1] J.-P. Allouche and F. M. Dekking, Beatty sequences and complementary triples, In preparation.

[2] H. Ardal, T. Brown, V. Jungic, J. Sahasrabudhe, On additive and Abelian complexity in infinite Words, Integers 12 (2012).

[3] F. Blanchet-Sadri, Daniel Seita, and David Wise, Computing abelian complexity of binary uniform morphic words, Theoretical Computer Science 640 (2016) 41-51.

[4] L. Carlitz, Richard Scoville, and V.E.Hogatt jr, Fibonacci representations, Fib. Quart. 10 (1972),1–28.

[5] F.M.Dekking, Marches automatiques, J. Théor. Nombres Bordeaux 5 (1993), 93–100.

[6] F. Michel Dekking, Paperfolding morphisms, planefilling curves, and fractal tiles, Theoretical Computer Science 414 (2012), 20–37.

[7] F. Michel Dekking, Morphisms, symbolic sequences, and their standard forms, Journal of Integer Sequences 19 (2016), Article 16.1.1, 1-8.
[8] J. Karhumäki, A. Saarela, and L. Q. Zamboni. Variations of the Morse-Hedlund Theorem for k-abelian equivalence. Developments in Language Theory, Proceedings, Ekaterinburg, Russia, 26–29 August 2014, Volume 8633 of the series Lecture Notes in Computer Science, 203–214. (http://arxiv.org/abs/1302.3783, 2013.)

[9] C. Kimberling and K. B. Stolarsky, Slow Beatty sequences, devious convergence, and partitional divergence, Amer. Math. Monthly, 123 (No. 2, 2016), 267-273.

[10] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications 90, Cambridge University Press, 2002.

[11] The On-Line Encyclopedia of Integer Sequences, founded by N.J.A Sloane, sequences A276885 and A276886.

[12] Gwénaël Richomme, Kalle Saari, and Luca Q. Zamboni, Abelian complexity of minimal subshifts, J. Lond. Math. Soc. (2) 83 (2011), no. 1, 7995. MR 2763945, https://doi.org/10.1112/jlms/jdq063

[13] Julian Sahasrabudhe, Sturmian words and constant additive complexity, Integers 15 (2015).

[14] Sylvester, James Joseph, “Question 7382”. Mathematical Questions from the Educational Times 41 (1884), 21.