The Frobenius problem for homomorphic embeddings of languages into the integers

Dekking, Michel

DOI
10.1016/j.tcs.2018.04.023

Publication date
2018

Document Version
Accepted author manuscript

Published in
Theoretical Computer Science

Citation (APA)
Dekking, M. (2018). The Frobenius problem for homomorphic embeddings of languages into the integers. *Theoretical Computer Science*, *732*, 73-79. https://doi.org/10.1016/j.tcs.2018.04.023

Elsevier Editorial System(tm) for Theoretical Computer Science
Manuscript Draft

Manuscript Number: TCS-D-17-00876R1

Title: The Frobenius problem for homomorphic embeddings of languages into the integers

Article Type: VSI: WORDS 2017

Section/Category: A - Algorithms, automata, complexity and games

Keywords: Frobenius problem, golden mean language, Sturmian language, Thue-Morse language, paperfolding morphism

Corresponding Author: Professor F.Michel Dekking, Ph.D.

Corresponding Author's Institution: Delft University of Technology

First Author: F.Michel Dekking, Ph.D.

Order of Authors: F.Michel Dekking, Ph.D.

The Frobenius problem for homomorphic embeddings of languages into the integers

Michel Dekking

DIAM, Delft University of Technology, Faculty EEMCS, P.O. Box 5031, 2600 GA Delft, The Netherlands. Email: f.m.dekking@tudelft.nl

Abstract Let S be a map from a language $\mathcal{L}$ to the integers satisfying $\mathrm{S}(vw) = \mathrm{S}(v) + \mathrm{S}(w)$ for all $v, w \in \mathcal{L}$. The classical Frobenius problem asks whether the complement of $\mathrm{S}(\mathcal{L})$ in the natural numbers will be infinite or finite, and in the latter case the value of the largest element in this complement. This is also known as the 'coin-problem', and $\mathcal{L}$ is the full language consisting of all words over a finite alphabet. We solve the Frobenius problem for the golden mean language, any Sturmian language and the Thue-Morse language. We also consider two-dimensional embeddings.

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 $\mathbb{N} = \{1, 2, 3, \dots\}$. Let $\mathcal{L}$ be a language, i.e., a factor-closed subset 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 $\mathrm{S} : \mathcal{L} \to \mathbb{N}$ satisfying

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

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

(Q1) Is the complement $\mathbb{N} \setminus \mathrm{S}(\mathcal{L})$ finite or infinite?

(Q2) If the complement of $\mathrm{S}(\mathcal{L})$ 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 [15]: $\mathbb{N} \setminus \mathrm{S}(\mathcal{L})$ is finite if and only if $\mathrm{S}(a)$ and $\mathrm{S}(b)$ are relatively prime, in which case its largest element is

$$\mathrm{S}(a)\mathrm{S}(b) - \mathrm{S}(a) - \mathrm{S}(b).$$

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 $\mathbb{N} \setminus \mathrm{S}(\mathcal{L})$ is finite when $\mathrm{S}(a)$ and $\mathrm{S}(b)$ are relatively prime, with largest element

$$\mathrm{S}(a)^2 + \mathrm{S}(a)\mathrm{S}(b) - 3\mathrm{S}(a) - \mathrm{S}(b).$$

Our main interest is however not in sofic languages[1], but in languages with low complexity (slow growth of the number of subwords of length n), where the complement of $\mathrm{S}(\mathcal{L})$ 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.

We usually suppose that $\gcd(\mathrm{S}(a), \mathrm{S}(b)) = 1$. First of all this is not a big loss since automatically the complement will have infinite cardinality if this is not the case. Secondly, if r divides both $\mathrm{S}_1(a)$ and $\mathrm{S}_1(b)$ for some homomorphism S_1, then

$$\mathrm{S}_1(\mathcal{L}^n) = r^n \mathrm{S}_2(\mathcal{L}^n), \quad \text{for } n = 1, 2, \ldots, \text{ where } \mathrm{S}_2(a) = \frac{\mathrm{S}_1(a)}{r}, \mathrm{S}_2(b) = \frac{\mathrm{S}_1(b)}{r}.$$

Here, and throughout the paper, we write $\mathcal{L}^n$ for the set of words of length n in a language $\mathcal{L}$—not to be confused with the n-fold concatenation of $\mathcal{L}$.

Our work is related to the work on *abelian complexity*, see, e.g., [3], [13], [8]. See Lemma 3.1 for such a connection.

Our work is also related to the notion of *additive complexity*, see [14] and [2]. The *additive complexity* of an infinite word w over a finite set of integers (see [2]) is the function $n \to \phi^+(w, n)$ that counts the number of distinct sums obtained by summing n consecutive symbols of w. Let $\mathcal{L}_w$ be the language of all words occurring in the infinite word w. Then the additive complexity is $\phi^+(w, n) = \mathrm{Card}\{\mathrm{S}(u) : u \in \mathcal{L}_w^n\}$, where S is the identity map on the alphabet of w. Here we do, and may, assume that the alphabet of w is a subset of $\mathbb{N}$.

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 $\mathcal{L}_{\mathrm{GM}}$ consisting of all words over $\{a, b\}$ in which bb does not occur as a subword. Now if S satisfies $\mathrm{S}(a) = 1$ or $\mathrm{S}(b) = 1$, then it is easily seen that $\mathrm{S}(\mathcal{L}_{\mathrm{GM}}) = \mathbb{N}$, 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

$$\mathrm{S}(a) = 100, \; \mathrm{S}(b) = 3,$$

[1]Regular languages, or equivalently, languages defined by the labelling of paths of an automaton, see [10, Section 1.5]

then the Frobenius number of the full language under S is $300 - 100 - 3 = 197$, and the Frobenius number of $\mathrm{S}(\mathcal{L}_{\mathrm{GM}})$ 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 $\mathrm{S}_a := \mathrm{S}(a)$, $\mathrm{S}_b := \mathrm{S}(b)$.

Theorem 2.1 *Let* $\mathrm{S} : \mathcal{L}_{\mathrm{GM}} \to \mathbb{N}$ *be a homomorphism. Suppose* $\gcd(\mathrm{S}_a, \mathrm{S}_b) = 1$*, and both* $\mathrm{S}_a > 1$ *and* $\mathrm{S}_b > 1$*. Then the Frobenius number of* $\mathrm{S}(\mathcal{L}_{\mathrm{GM}})$ *is equal to*

$$\max \mathbb{N} \setminus \mathrm{S}(\mathcal{L}_{\mathrm{GM}}) = \mathrm{S}_a(\mathrm{S}_a - 3) + \mathrm{S}_b(\mathrm{S}_a - 1).$$

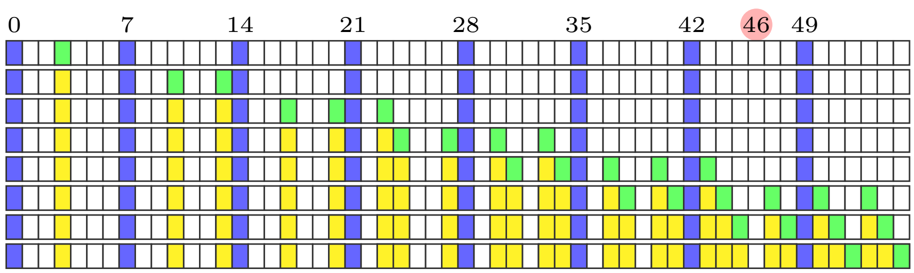
Proof: Let an S_a-*point* be defined as a multiple $n\mathrm{S}_a$, $n = 0, 1, \ldots$, and an S_a-*interval* as the set of numbers between two consecutive S_a-points. We also consider S_b-*chains*, defined for $n \geq 0$ by

$$C(n) = \{n\mathrm{S}_a + \mathrm{S}_b,\, n\mathrm{S}_a + 2\mathrm{S}_b, \ldots, n\mathrm{S}_a + (n+1)\mathrm{S}_b\}.$$

The union of the S_a-points and the S_b-chains will give $\mathcal{L}_{\mathrm{GM}}$: the S_a-points are the images of the words a^n, and the S_b-chains are the images of the words in which n letters a occur, and k letters b for $k = 1, 2, \ldots, n+1$. Here $n+1$ is the maximal number of letters b that can occur because of the "no bb" constraint.
The key observation is that the S_b-chain $C(\mathrm{S}_a - 2)$ has $\mathrm{S}_a - 1$ elements, which are all distinct modulo S_a. This is a consequence of $\gcd(\mathrm{S}_a, \mathrm{S}_b) = 1$. 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 equal to $\mathrm{S}_a - \mathrm{S}_b$ modulo S_a, produced by the last element of the chain $C(\mathrm{S}_a - 2)$. This is the number

$$P := (\mathrm{S}_a - 2)\mathrm{S}_a + (\mathrm{S}_a - 1)\mathrm{S}_b.$$

But then the largest number in the complement of $\mathcal{L}_{\mathrm{GM}}$ is $P - \mathrm{S}_a$, 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. $\square$



$S(a) = 7$, $S(b) = 3$: row n shows $C(n-1)$ in green, $n = 1, \ldots, 8$ (truncated at 56).

Figure 1: Example with $\mathrm{S}(a) = 7, \mathrm{S}(b) = 3$: 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).

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 exist an irrational number $\alpha \in (0,1)$ and a number ρ such that $s = s_{\alpha,\rho}$, or $s = s'_{\alpha,\rho}$, where

$$s_{\alpha,\rho} = \big([(n+1)\,\alpha+\rho] - [n\,\alpha+\rho]\big)_{n\geq 0},\ s'_{\alpha,\rho} = \big(\lceil(n+1)\,\alpha+\rho\rceil - \lceil n\,\alpha+\rho\rceil\big)_{n\geq 0}.$$

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:

$$s_\alpha = 0\,c_\alpha, \qquad s'_\alpha = 1\,c_\alpha.$$

Here $c_\alpha := s_{\alpha,\alpha}$ is called the *characteristic word* of α. For $n \geq 0$ we have

$$c_\alpha(n) = s_{\alpha,\alpha}(n) = [(n+1)\,\alpha+\alpha] - [n\,\alpha+\alpha] = [(n+2)\,\alpha] - [(n+1)\,\alpha].$$

The words s_α, s'_α and c_α 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 $\mathcal{L}_\alpha$ be a Sturmian language, and let* S *be a homomorphism with* $\mathrm{S}(0) \neq \mathrm{S}(1)$. *Then* $\mathrm{Card}\,\mathrm{S}(\mathcal{L}^n_\alpha) = 2$ *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 differ by at most 1 in their number of ones. □

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 α be an irrational number from $(0,1)$. Let $\mathcal{L}_\alpha$ be the Sturmian language generated by α, and let $(q_n)_{n\geq 0}$ be the slow Beatty sequence defined by*

$$q_n = [(n+1)\alpha].$$

Let $\mathrm{S} : \mathcal{L}_\alpha \to \mathbb{N}$ *be a homomorphism. Define* $\mathrm{S}_0 = \mathrm{S}(0), \mathrm{S}_1 = \mathrm{S}(1)$. *Then*

$$\mathrm{S}(\mathcal{L}_\alpha) = \{(\mathrm{S}_1-\mathrm{S}_0)q_n + n\mathrm{S}_0 + \mathrm{S}_0 :\ n = 0,\dots\} \cup \{(\mathrm{S}_1-\mathrm{S}_0)q_n + n\mathrm{S}_0 + \mathrm{S}_1 :\ n = 0,\dots\}.$$

Proof: If $\mathrm{S}_0 = \mathrm{S}_1$ then this is certainly true, so suppose $\mathrm{S}_0 \neq \mathrm{S}_1$ in the sequel. We denote $c_\alpha[i,j] := c_\alpha(i)\dots c_\alpha(j)$ for integers i,j with $0 \leq i < j$. Let $N_\ell(w)$ denote the number of occurrences of the letter ℓ 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 $\mathcal{L}_\alpha$, but $\mathcal{L}_\alpha$ also contains the words $0c_\alpha[0,n-1]$ and $1c_\alpha[0,n-1]$. It thus follows from Lemma 3.1 that $\mathrm{S}(\mathcal{L}_\alpha)$ is given by the union of all images $\mathrm{S}(0c_\alpha[0,n-1])$ and $\mathrm{S}(1c_\alpha[0,n-1])$. Since

$$\mathrm{S}(0c_\alpha[0,n-1]) = \mathrm{S}_0 + (n-q_n)\mathrm{S}_0 + q_n\mathrm{S}_1 = (\mathrm{S}_1-\mathrm{S}_0)q_n + n\mathrm{S}_0 + \mathrm{S}_0,$$

the result follows. □

3.1 The Fibonacci language

Let $\Phi = (\sqrt{5}+1)/2 = 1.61803\ldots$ be the golden mean, and let $\alpha := 2 - \Phi = \Phi^{-2}$. We have (see, e.g., [10, Example 2.1.24])

$$c_\alpha = ([(n+1)\alpha] - [n\alpha])_{n\geq 1} = 0,1,0,0,1,0,1,0,0,1,0,0,1,0,1,0,0,\ldots,$$

the infinite Fibonacci word. We write $\mathcal{L}_{\rm F} := \mathcal{L}_\alpha$.

Theorem 3.2 *Let* $\mathrm{S} : \mathcal{L}_{\rm F} \to \mathbb{N}$ *be a homomorphism. Then*

$$\mathrm{S}(\mathcal{L}_{\rm F}) = \big((\mathrm{S}_0-\mathrm{S}_1)[n\Phi]+(2\mathrm{S}_1-\mathrm{S}_0)n+\mathrm{S}_0-\mathrm{S}_1\big)_{n\geq 1} \cup \big((\mathrm{S}_0-\mathrm{S}_1)[n\Phi]+(2\mathrm{S}_1-\mathrm{S}_0)n\big)_{n\geq 1}.$$

Proof: This is a corollary to Theorem 3.1, using $[-x] = -[x]-1$ for non-integer x:

$$\begin{aligned}(\mathrm{S}_1-\mathrm{S}_0)q_{n-1}+n\mathrm{S}_0 &= (\mathrm{S}_1-\mathrm{S}_0)[n\alpha]+n\mathrm{S}_0 = (\mathrm{S}_1-\mathrm{S}_0)[n(2-\Phi)]+n\mathrm{S}_0\\ &= 2(\mathrm{S}_1-\mathrm{S}_0)n+(\mathrm{S}_1-\mathrm{S}_0)[-n\Phi]+n\mathrm{S}_0\\ &= (2\mathrm{S}_1-\mathrm{S}_0)n+(\mathrm{S}_1-\mathrm{S}_0)(-[n\Phi]-1)\\ &= (\mathrm{S}_0-\mathrm{S}_1)[n\Phi]+(2\mathrm{S}_1-\mathrm{S}_0)n+\mathrm{S}_0-\mathrm{S}_1.\end{aligned}$$

□

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

Proof: Take $(\mathrm{S}_0,\mathrm{S}_1) = (1{,}1)$. Then obviously $\mathrm{S}(\mathcal{L}_{\rm F})=\mathbb{N}$.
Take $(\mathrm{S}_0,\mathrm{S}_1) = (2{,}1)$. Then $\mathrm{S}(\mathcal{L}_{\rm F})=\mathbb{N}$, since by Theorem 3.2 $\mathrm{S}(\mathcal{L}_{\rm F})$ 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 $(\mathrm{S}_0,\mathrm{S}_1) = (1{,}2)$. Then $\mathrm{S}(\mathcal{L}_{\rm F})=\mathbb{N}$, since $\mathrm{S}(\mathcal{L}_{\rm F})$ 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 $(\mathrm{S}_0,\mathrm{S}_1) = (1{,}3)$. 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 $\mathrm{S}(\mathcal{L}_{\rm F})$. Let Δu 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 Δv and Δu 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 $\mathrm{S}(\mathcal{L}_{\rm F})=\mathbb{N}$. 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 $v_{n-1} = v_n - 1 = u_n + 1$. □

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

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

Proof: According to Lemma 3.2 the complement of $\mathrm{S}(\mathcal{L}_\mathrm{F})$ is empty for $(\mathrm{S}_0, \mathrm{S}_1) \in \mathcal{E}$. The density of the set $\mathrm{S}(\mathcal{L}_\mathrm{F})$ in the natural numbers exists, and equals

$$\delta := \frac{2}{(\mathrm{S}_0 - \mathrm{S}_1)\Phi + 2\mathrm{S}_1 - \mathrm{S}_0}.$$

The theorem will be proved if we show that $\delta < 1$ for $(\mathrm{S}_0, \mathrm{S}_1)$ not in $\mathcal{E}$. First we note that the denominator of δ is positive:

$$(\mathrm{S}_0 - \mathrm{S}_1)(\Phi - 1) + \mathrm{S}_1 > -\mathrm{S}_1(\Phi - 1) + \mathrm{S}_1 = \mathrm{S}_1(2 - \Phi) > 0,$$

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

$$\delta < 1 \;\Leftrightarrow\; (\mathrm{S}_0 - \mathrm{S}_1)\Phi + 2\mathrm{S}_1 - \mathrm{S}_0 > 2 \;\Leftrightarrow\; (\mathrm{S}_0 - \mathrm{S}_1)\Phi > \mathrm{S}_0 - \mathrm{S}_1 + 2 - \mathrm{S}_1.$$

If $\mathrm{S}_0 > \mathrm{S}_1$, this is satisfied, since under this condition $(2 - \mathrm{S}_1)/(\mathrm{S}_0 - \mathrm{S}_1) \le 0$, unless $(\mathrm{S}_0, \mathrm{S}_1) = (2, 1) \in \mathcal{E}$. If $\mathrm{S}_0 < \mathrm{S}_1$, we have to see that $\Phi < 1 + (2 - \mathrm{S}_1)/(\mathrm{S}_0 - \mathrm{S}_1)$. This holds for $\mathrm{S}_0 \ge 2$, since then $(2 - \mathrm{S}_1)/(\mathrm{S}_0 - \mathrm{S}_1) \ge 1$. If $\mathrm{S}_0 = 1$, then this does not hold for $\mathrm{S}_1 = 1, 2, 3$, i.e., for pairs from $\mathcal{E}$, but it will hold for all $\mathrm{S}_1 \ge 4$. □

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 $n \ge 1$, and $B(n) = [n\Phi^2]$ for $n \ge 1$. 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 consider A and B as functions from $\mathbb{N}$ to $\mathbb{N}$, and define functions $pX + qY + r$ by

$(pX + qY + r)(n) = pX(n) + qY(n) + r$

for real numbers p, q, r and functions $X, Y : \mathbb{N} \to \mathbb{N}$. Then

$\mathrm{S}(\mathcal{L}_\mathrm{F}) = B(\mathbb{N}) \cup (B{+}1)(\mathbb{N}), \quad \mathbb{N} \backslash \mathrm{S}(\mathcal{L}_\mathrm{F}) = \{1, 4, 9, 12, \dots\} = (2A{+}\mathrm{Id}{+}1)(\mathbb{N} \cup \{0\})$.

The first statement follows from Theorem 3.2, and the relationship $B = A + \mathrm{Id}$. 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 $AB = A + B = 2A + \mathrm{Id}$.
But then the three sequences $([n\Phi] + n)_{n\ge1}$, $([n\Phi] + n + 1)_{n\ge1}$, $(2[n\Phi] + n + 1)_{n\ge0}$, form a complementary triple, i.e., as sets they are disjoint, and their union is $\mathbb{N}$.
A similar result holds for[2] 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

$\mathrm{S}(\mathcal{L}_\mathrm{F}) = (2A - \mathrm{Id})(\mathbb{N}) \,\cup\, (2A - \mathrm{Id} + 2)(\mathbb{N})$.

It is proved in [1] that

$\mathbb{N} \setminus \mathrm{S}(\mathcal{L}_\mathrm{F}) = \{2, 9, 20, 27, 38, 49, \dots\} = (4A + 3\mathrm{Id} + 2)(\mathbb{N} \cup \{0\})$,

and that the three sequences $(2[n\Phi] - n)_{n\ge1}$, $(2[n\Phi] - n + 2)_{n\ge1}$, $(4[n\Phi] + 3n + 2)_{n\ge0}$, form a complementary triple.

[2]In these two cases $\mathbb{N} \setminus \mathrm{S}(\mathcal{L}_\mathrm{F})$ is given by sequences A276885, respectively A276886 in OEIS ([11]). It is easily seen that the definitions of these sequences in OEIS are equivalent to the way in which we obtain them.

4 The Thue-Morse language

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

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

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

Proof: Let $\mathcal{L}^n_{\mathrm{TM}}$ be the set of words of length n in the Thue-Morse language. Put $r = \mathrm{S}(ab) = p+q$. It is clear (and for $p = 0, q = 1$ also observed in [13]) that since the Thue-Morse word is a non-periodic concatenation of ab and ba that for $n = 1, 2, \dots$

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

This implies the statement of the theorem. □

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

Proof: This follows directly from Theorem 4.1. If $\mathrm{S}(a) + \mathrm{S}(b) \geq 6$, then the density of $\mathbb{N} \setminus \mathrm{S}(\mathcal{L}_{\mathrm{TM}})$, is at least 1/6, so the set has infinite cardinality.
When $\mathrm{S}(a) + \mathrm{S}(b) < 6$, then, because of symmetry, we only have to consider the four cases $(\mathrm{S}(a), \mathrm{S}(b)) = (1, 2)$, $(\mathrm{S}(a), \mathrm{S}(b)) = (1, 3)$, $(\mathrm{S}(a), \mathrm{S}(b)) = (1, 4)$, and $(\mathrm{S}(a), \mathrm{S}(b)) = (2, 3)$. In these cases one will find with Theorem 4.1 that the complement of $\mathrm{S}(\mathcal{L}_{\mathrm{F}})$ is empty in the first two cases, and equal to $\{3\}$, respectively $\{1\}$ in the last two. □

Remark Let σ 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

Here we consider homomorphisms $\mathrm{S} : \mathcal{L} \to \mathbb{N} \times \mathbb{N}$ and $\mathrm{S} : \mathcal{L} \to \mathbb{Z} \times \mathbb{Z}$. The situation changes drastically for this 'double-coin' problem.

Proposition 5.1 *Let* $\mathcal{L}$ *be a language on the alphabet* $\{a, b\}$, *and let* $\mathrm{S} : \mathcal{L} \to \mathbb{N} \times \mathbb{N}$ *be a homomorphism. Then* $\mathbb{N} \times \mathbb{N} \setminus \mathrm{S}(\mathcal{L})$ *has infinite cardinality for all pairs* $\{\mathrm{S}(a), \mathrm{S}(b)\}$ *which are not equal to the pair* $\{(0, 1), (1, 0)\}$.

Proof: It suffices to prove this for the full language $\mathcal{L}_{\text{full}}$. The image under S is an integer lattice, with a complement of infinite cardinality, unless $\mathrm{S}(a)$ and $\mathrm{S}(b)$ are the unit vectors. □

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 φ on an alphabet A. Such a morphism has a language $\mathcal{L}_\varphi$ associated to it, where each word $w \in \mathcal{L}_\phi$ has a measure $\mu_\varphi(w)$. For a given homomorphism $\mathrm{S} : \mathcal{L}_\varphi \to \mathbb{Z} \times \mathbb{Z}$ we call the average

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

the *drift* of S.

Proposition 5.2 *Let $\mathcal{L}_\varphi$ be a language generated by a primitive morphism on an alphabet A, and let $\mathrm{S} : \mathcal{L}_\varphi \to \mathbb{Z} \times \mathbb{Z}$ be a homomorphism. Then $\mathbb{Z} \times \mathbb{Z} \setminus \mathrm{S}(\mathcal{L})$ has infinite cardinality if $\Delta_\varphi(\mathrm{S}) \neq (0, 0)$.*

Proof: It is well-known that the measure μ_φ is strictly ergodic (see, e.g., [12]). Because of this, we have for words w from $\mathcal{L}_\varphi$, where $|w|$ denotes the length of w,

$$\frac{1}{|w|}\mathrm{S}(w) = \frac{1}{|w|}\sum_{a \in A} N_a(w)\mathrm{S}(a) \to \sum_{a \in A} \mu_\varphi(a)\mathrm{S}(a) = \Delta_\varphi(\mathrm{S}) \text{ as } |w| \to \infty.$$

Thus for long words w the images $\mathrm{S}(w)$ will be concentrated around the line in the direction of the drift of S, and so the complement of $\mathrm{S}(\mathcal{L}_\varphi)$ 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 $\mathcal{L}_\theta$ on an alphabet $A = \{a, b, c, d\}$ of four letters where for the homomorphism $\mathrm{S}^\oplus$ given by

$$\mathrm{S}^\oplus(a) = (1, 0),\ \mathrm{S}^\oplus(b) = (0, 1),\ \mathrm{S}^\oplus(c) = (-1, 0),\ \mathrm{S}^\oplus(d) = (0, -1)$$

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 σ 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 τ be the anti-morphism given by $\tau(w_1 \dots w_n) = w_n \dots w_1$.
A morphism θ on $\{a, b, c, d\}$ is called a *paperfolding* morphism if

1) $\sigma\tau\theta = \theta\sigma\tau$,
2) Letters from $\{a, c\}$ alternate[3] with letters from $\{b, d\}$ in $\theta(a)$.

A paperfolding morphism is called *symmetric* if $\sigma\theta = \theta\sigma$. 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 $\mathrm{S} : A^* \to G$ generate a *walk* $Z = (Z_n)_{n \geq 0}$ by (cf. [5])

$$Z_0 = e, \qquad Z_{n+1} = Z_n + \mathrm{S}(x_n) = \mathrm{S}(x_0 \dots x_n), \text{ for } n \geq 0.$$

[3] This corrects an omission in [6, Definition 1].

A paperfolding morphism θ with $\theta(a) = a...$ 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 plane-filling 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 θ given by

$$\theta(a) = abcba,\ \theta(b) = bcdcb,\ \theta(c) = cdadc,\ \theta(d) = dabad.$$

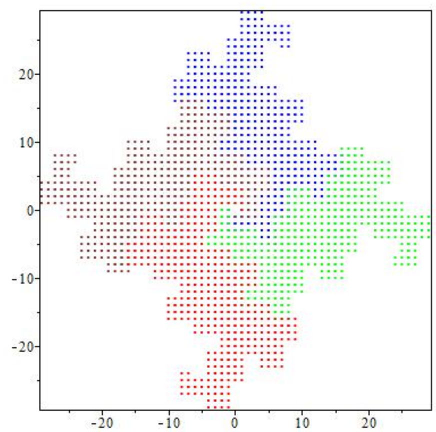


Figure 2: The four images of the words $\theta^4(a), \ldots, \theta^4(d)$ under $\mathrm{S}^\oplus$, where θ is the perfect symmetric 5-folding morphism. The origin is not covered, but it is the image of the word $abcd \in \mathcal{L}_\theta$.

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

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

Acknowledgement

I would like to express my gratitude to two reviewers for their excellent comments.

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), #A21, 1–8.

[3] F. Blanchet-Sadri, D. Seita, and D. Wise, Computing abelian complexity of binary uniform morphic words, Theor. Comput. Sci. 640 (2016) 41-51.

[4] L. Carlitz, R. 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. M. Dekking, Paperfolding morphisms, planefilling curves, and fractal tiles, Theor. Comput. Sci. 414 (2012), 20–37.

[7] F. M. 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] Martine Queffélec, *Substitution Dynamical Systems – Spectral Analysis*. Lecture Notes in Mathematics 1294, 2nd ed., Springer, Berlin 2010.

[13] G. Richomme, K. Saari, and L. 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

[14] J. Sahasrabudhe, Sturmian words and constant additive complexity, Integers 15 (2015), #A30, 1–8.

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