On the Asymptotic Abelian Complexity of Morphic Words*

Francine Blanchet-Sadri[1] and Nathan Fox[2]

[1] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402–6170, USA
blanchet@uncg.edu

[2] Department of Mathematics, Rutgers University, Hill Center for the Mathematical Sciences, 110 Frelinghuysen Rd., Piscataway, NJ 08854–8019, USA
fox@math.rutgers.edu

Abstract. The subword complexity of an infinite word counts the number of subwords of a given length, while the abelian complexity counts this number up to letter permutation. Although a lot of research has been done on the subword complexity of morphic words, i.e., words obtained as fixed points of iterated morphisms, little is known on their abelian complexity. In this paper, we undertake the classification of the asymptotic growths of the abelian complexities of fixed points of binary morphisms. Some general results we obtain stem from the concept of factorization of morphisms. We give an algorithm that yields all canonical factorizations of a given morphism, describe how to use it to check quickly whether a binary morphism is Sturmian, discuss how to fully factorize the Parry morphisms, and finally derive a complete classification of the abelian complexities of fixed points of uniform binary morphisms.

1 Introduction

The *subword complexity* of an infinite word w, denoted ρ_w, is the function mapping each positive integer n to the number of distinct subwords of w of length n. On the other hand, the *abelian complexity* of w, denoted ρ_w^{ab}, is the function mapping each positive integer n to the number of distinct Parikh vectors of subwords of w of length n. Here, we assume the standard alphabet $A_k = \{0, \ldots, k-1\}$, and the *Parikh vector* of a finite word over A_k is the vector whose ith entry is the number of occurrences of letter $i-1$ in the word.

An infinite word is a *morphic word* if it is the fixed point of some morphism at some letter. For compactness of notation, we frequently denote a morphism φ over A_k, $\varphi : A_k^* \to A_k^*$, as an ordered k-tuple $\varphi = (w_0, \ldots, w_{k-1})$, where

* This material is based upon work supported by the National Science Foundation under Grant No. DMS–1060775. A World Wide Web server interface has been established at www.uncg.edu/cmp/research/abeliancomplexity2 for automated use of the programs.

M.-P. Béal and O. Carton (Eds.): DLT 2013, LNCS 7907, pp. 94–105, 2013.

$\varphi(a) = w_a$ for each $a \in A_k$. The *fixed point of* φ *at* a, denoted $\varphi^\omega(a)$, is the limit as $n \to \infty$ of $\varphi^n(a)$. The fixed point exists precisely when the limit exists.

A lot of research has been done on the subword complexity of morphic words, e.g., Ehrenfeucht and Rozenberg [9] showed that the fixed points of *uniform morphisms*, i.e., morphisms φ over A_k satisfying $|\varphi(a)| = |\varphi(b)|$ for all $a, b \in A_k$, have at most linear subword complexity, Berstel and Séébold [3] gave a characterization of *Sturmian morphisms*, i.e., morphisms φ over the binary alphabet $A_2 = \{0, 1\}$ such that $\varphi^\omega(0)$ exists and is Sturmian, in other words, $\rho_{\varphi^\omega(0)}(n) = n+1$ for all n, and Frid [11] obtained a formula for the subword complexity of the fixed points of binary uniform morphisms. On the other hand, abelian complexity is a relatively new research topic. Balková, Břinda, and Turek [2] studied the abelian complexity of infinite words associated with quadratic Parry numbers, Currie and Rampersad [6] studied recurrent words with constant abelian complexity, and Richomme, Saari, and Zamboni [14] investigated abelian complexity of minimal subshifts.

In this paper, we are interested in classifying the asymptotic growths of the abelian complexities of words obtained as fixed points of iterated morphisms over A_2 at 0. This classification has already been done for the subword complexity of morphisms over A_k [7–10, 12] (see also [4, Section 9.2]). Pansiot's classification of the asymptotic growths of the subword complexity of morphic words not only depends on the type of morphisms but also on the distribution of so-called bounded letters [12]. We assume without loss of generality that the first letter in the image of 0 is 0 and we assume that all of our morphisms are nonerasing. Also for conciseness, we frequently use the term "abelian complexity of a morphism" (when referring to its fixed point at 0).

As mentioned above, we are mainly concerned with the asymptotic behaviors of the abelian complexities rather than their specific values. Some general results we obtain stem from the concept of *factorization of morphisms*. We mainly examine the monoid of binary morphisms under composition, but we also consider factorization in a more general setting.

The binary morphism types whose asymptotic abelian complexities we classify are the following: morphisms with ultimately periodic fixed points (Proposition 1(10)), Sturmian morphisms (Proposition 2(2)), morphisms with equal ratios of zeroes to ones in both images (Theorem 3), Parry morphisms, i.e., morphisms of the type $(0^p1, 0^q)$ with $p \geq q \geq 1$ or of the type $(0^p1, 0^q1)$ with $p > q \geq 1$ studied in [2] and cyclic shifts of their factorizations (Proposition 2(4) along with Corollaries 2 and 3), most morphisms where the image of 1 contains only ones (Theorem 4), and most uniform morphisms and cyclic shifts of their factorizations (Theorem 6).

All of the asymptotic abelian complexity classes we obtain, where $f(n) = 1$ if $\log_2 n \in \mathbb{Z}$, and $f(n) = \log n$ otherwise, are the following (listed in increasing order of growth): $\Theta(1)$, e.g., $(01, 10)$, $\tilde{\Theta}(f(n))$, e.g., $(01, 00)$, $\Theta(\log n)$, e.g., $(001, 110)$, $\Theta\left(n^{\log_a b}\right)$ for $a > b > 1$, e.g., $(0001, 0111)$, $\Theta\left(\frac{n}{\log n}\right)$, e.g., $(001, 11)$, and $\Theta(n)$, e.g., $(0001, 11)$.

The contents of our paper are as follows. In Section 2, we discuss some preliminaries and simple results. In Section 3, we study morphism factorizations. We give an algorithm that yields all canonical factorizations of a morphism φ over A_k into two morphisms each over an alphabet of at most k letters and we describe how to use it for checking quickly whether a binary morphism is Sturmian. In Section 4, we obtain our main results. Among other things, we show how to fully factorize the Parry morphisms and we also derive a complete classification of the abelian complexities of fixed points of uniform binary morphisms. Finally in Section 5, we conclude with suggestions for future work. Some proofs have been omitted and others have only been outlined due to the 12-page space constraint.

2 Preliminary Definitions and Results

Given $C \geq 0$, a (finite or infinite) word w over A_k is called *C-balanced* if for all letters a in w and for all integers $0 < n \leq |w|$ (or just $0 < n$ if w is infinite), the difference between the maximum and the minimum possible counts of letter a in a length-n subword of w is less than or equal to C.

Given an infinite word w over A_2, $z_M(n)$ (resp., $z_m(n)$) denotes the maximum (resp., minimum) number of zeroes in a length-n subword of w. For ease of notation, $z(v)$ denotes the number of zeroes in a binary word v (as opposed to the standard $|v|_0$). The number of ones in v is then $|v| - z(v)$.

Here are some facts about abelian complexity and zero counts.

Proposition 1. *If w is an infinite word over A_k, then the following hold:*

1. $\rho_w^{ab}(n) = \Theta(1)$ *if and only if w is C-balanced for some C [14, Lemma 2.2];*
2. *If w is Sturmian, then* $\rho_w^{ab}(n) = 2$ *for all n [5];*
3. $\rho_w^{ab}(n) = O\left(n^{k-1}\right)$ *[14, Theorem 2.4];*
4. $\rho_w^{ab}(n) \leq \rho_w(n)$*;*
5. *If $k = 2$, then* $\rho_w^{ab}(n) = z_M(n) - z_m(n) + 1$*;*
6. *If $k = 2$, then* $z_M(m+n) \leq z_M(m) + z_M(n)$*;*
7. *If $k = 2$, then* $z_m(m+n) \geq z_m(m) + z_m(n)$*;*
8. *If $k = 2$, then* $z_M(n+1) - z_M(n) \in \{0,1\}$ *and* $z_m(n+1) - z_m(n) \in \{0,1\}$*;*
9. *If $k = 2$, then* $\left|\rho_w^{ab}(n+1) - \rho_w^{ab}(n)\right| \leq 1$ *for all positive integers n;*
10. *If w is ultimately periodic, then* $\rho_w^{ab}(n) = \Theta(1)$.

Here are some morphisms that are classified based on prior results. For 3, any such word is ultimately periodic.

Proposition 2. *The fixed points at 0 of the following morphisms over A_2 have $\Theta(1)$ abelian complexity:*

1. *The Thue-Morse morphism* $(01, 10)$ *[14, Theorem 3.1];*
2. *Any Sturmian morphism (this includes* $(01, 0)$*) [14, Theorem 1.2];*
3. *Any morphism whose fixed point contains finitely many zeroes or ones (this includes* $(01, 11)$*);*

4. *Any morphism of the form* $(0^p1, 0^q)$ *with* $p \geq q \geq 1$ *or of the form* $(0^p1, 0^q1)$ *with* $p > q \geq 1$ *[2, Corollary 3.1].*

Let $f(x)$ be a real function, and define

$$f_m(x) := \inf_{a \geq x} f(a), f_M(x) := \sup_{a \leq x} f(a).$$

Now, let $g(x)$ be also a real function. We write $f(x) = \tilde{\Omega}(g(x))$ if $f_m(x) = \Omega(g_m(x))$ and $f_M(x) = \Omega(g_M(x))$, $f(x) = \tilde{O}(g(x))$ if $f_m(x) = O(g_m(x))$ and $f_M(x) = O(g_M(x))$, and $f(x) = \tilde{\Theta}(g(x))$ if $f_m(x) = \Theta(g_m(x))$ and $f_M(x) = \Theta(g_M(x))$.

3 Morphism Factorizations

Some more general results we obtain stem from the concept of factorization of morphisms. We mainly examine the monoid of binary morphisms under composition, but we also consider factorization in a more general setting.

Let φ be a morphism over A_k. If φ cannot be written as $\phi \circ \zeta$ for two morphisms $\phi : A_{k'}^* \to A_k^*$, $\zeta : A_k^* \to A_{k'}^*$ where neither is a permutation of $A_{k'}$, then φ is *irreducible* over $A_{k'}$. Otherwise, φ is *reducible*. (We use the convention that a permutation of the alphabet is not irreducible.) If $\varphi = \phi \circ \zeta$ for some morphisms $\phi : A_{k'}^* \to A_k^*$, $\zeta : A_k^* \to A_{k'}^*$, we say that φ *factors* as $\phi \circ \zeta$ over $A_{k'}$.

Here are two propositions that together lead to a factorization algorithm.

Proposition 3. *Let* $\varphi = (w_0, w_1, \ldots, w_{k-1})$ *be a morphism over* A_k*. If there exist* $v_0, v_1, \ldots, v_{k'-1} \in A_k^+$ *such that* $w_0, w_1, \ldots, w_k \in \{v_0, v_1, \ldots, v_{k'-1}\}^*$*, then there exists a morphism* $\zeta : A_k^* \to A_{k'}^*$ *such that* $\varphi = \phi \circ \zeta$*, where* $\phi = (v_0, v_1, \ldots, v_{k'-1})$*. Conversely every factorization* $\varphi = \phi \circ \zeta$*, where* $\phi : A_{k'}^* \to A_k^*$ *and* $\zeta : A_k^* \to A_{k'}^*$*, corresponds to* $v_i = \phi(i)$ *for* $i = 0, 1, \ldots, k' - 1$*, where* $w_0, w_1, \ldots, w_{k-1} \in \{v_0, v_1, \ldots, v_{k'-1}\}^*$*.*

Proposition 4. *Let* σ *be a function that permutes elements of a* k*-tuple and let* ψ_σ *be the morphism corresponding to the* k*-tuple obtained by applying* σ *to* $(0, 1, \ldots, k-1)$*. Let* φ *be a morphism over* A_k *that factors as* $\varphi = \phi \circ \zeta$ *for some morphisms* ϕ *and* ζ*. Then,* $\varphi = \sigma(\phi) \circ \psi_\sigma(\zeta)$*.*

Proposition 4 allows us to define the notion of a *canonical factorization.* Let φ be a morphism over A_k that factors as $\varphi = \phi \circ \zeta$ for some morphisms ϕ and ζ. Let $v = \zeta(0)\zeta(1)\cdots\zeta(k-1)$. We say that the factorization $\varphi = \phi \circ \zeta$ is *canonical* if $v[0] = 0$ and the first occurrence of each letter a, $a \neq 0$, in v is after the first occurrence of letter $a - 1$ in v. It is clear from Proposition 4 that given a factorization $\phi \circ \zeta$ of φ we can put it in canonical form by applying some permutation σ to the letters in the images in ζ and to the order of the images in ϕ. Hence, every factorization of φ corresponds to one in canonical form.

Before we give our factorization algorithm, here is an important note: the monoid of binary morphisms does not permit unique factorization into irreducible morphisms, even if the factorizations are canonical. Indeed, letting

$\varphi = (00, 11)$, we have $\varphi = (0, 11) \circ (00, 1) = (00, 1) \circ (0, 11)$. These are distinct canonical factorizations of φ into irreducibles.

We now give an algorithm, Algorithm 1, that yields all canonical factorizations over $A_{k'}$ of a given morphism φ into two morphisms. The basis of this algorithm is the subroutine it calls, Algorithm 2, which creates a factorization by recursively finding $v_0, v_1, \ldots, v_{k'-1}$, as specified in Proposition 3, and then backtracking to find more factorizations. It always attempts to match or create a v_i at the beginning of the shortest image in the morphism, as that has the fewest possibilities to consider. Algorithm 1 works as follows:

– Call Algorithm 2 with φ, k', and an empty list. Given morphism φ', integer k'', and a list of words $v_0, \ldots, v_m$, Algorithm 2 does the following:
  1. If φ' has no letters, return $\{(v_0, \ldots, v_m) \circ \varphi'\}$;
  2. If $k'' > 0$, try each prefix of a minimal-length image in φ' as v_{m+1}. Call this same subroutine each time with $k'' - 1$, pruning that prefix off that image. Consolidate the results of the recursive call and add to the set of factorization pairs with appropriate right factor;
  3. Try matching each v_i to a prefix of a minimal-length image in φ'. If there is a match, call this same subroutine with k'', pruning that prefix off that image. Consolidate the results of the recursive call and add to the set of factorization pairs with appropriate right factor;
  4. Return the set of factorization pairs.

– Put the resulting factorizations into canonical form.

Theorem 1. *Algorithm 1 can be applied recursively (and optionally along with a lookup table to avoid recomputing things) to obtain complete (canonical) factorizations of a given morphism into irreducible morphisms.*

Proof. Algorithm 1's correctness follows from Propositions 3 and 4. To obtain all factorizations (not just canonical ones), run Algorithm 1 and then apply all possible permutations to the resulting factorizations. □

Given as input $\varphi = (01, 00)$ and $k' = 2$, Algorithm 1 outputs the canonical factorizations $(0, 1) \circ (01, 00)$, $(01, 0) \circ (0, 11)$, and $(01, 00) \circ (0, 1)$:

φ'	k''	v_0	v_1	φ'	k''	v_0	v_1	φ'	k''	v_0	v_1
$(01, 00)$	2			$(01, 00)$	2			$(01, 00)$	2		
$(1, 00)$	1	0		$(\varepsilon, 00)$	1	01		$(\varepsilon, 00)$	1	01	
$(\varepsilon, 00)$	0	0	1	$(\varepsilon, 0)$	1	01	0	$(\varepsilon, \varepsilon)$	1	01	00
$(\varepsilon, 0)$	0	0	1	$(\varepsilon, \varepsilon)$	0	01	0				
$(\varepsilon, \varepsilon)$	0	0	1								

We conclude this section with a discussion on checking whether a binary morphism φ is Sturmian. Berstel and Séébold in [3] prove that φ is Sturmian if and only if $\varphi(10010010100101)$ is 1-balanced and primitive (not a power of a shorter word). This leads to an algorithm for deciding whether a given morphism is Sturmian. While the resulting algorithm is typically fast to give a negative answer, a positive answer requires computing (essentially) $|\varphi(10010010100101)|$ balance values and checking for primitivity. (Also, a check that our morphism's fixed point does not contain finitely many zeroes is needed.)

Richomme in [13] gives a note that leads to an alternative approach. The note says that a binary morphism is Sturmian if and only if it can be written as compositions of the morphisms $(1,0)$, $(0,01)$, $(10,1)$, $(0,10)$, and $(01,1)$. No canonical factorization of a Sturmian morphism ever contain $(1,0)$ or $(10,1)$, but these two can be combined to form $(01,0)$, which we must add to our list. Hence, we have the criterion that a binary morphism is Sturmian if and only if it has a canonical factorization that is a composite of $(0,01)$, $(01,0)$, $(0,10)$, and $(01,1)$. (We also disallow composites that are powers of the morphisms with ultimately periodic or finite fixed points so we can keep our fixed points aperiodic.) The task of factoring a Sturmian morphism is well suited to repeated application of Algorithm 1. In fact, we can speed up the algorithm specifically in this case by pre-seeding v_0 and v_1 with each of the possible factors for a Sturmian morphism (as in, directly calling Algorithm 2 with φ, 2, and $[v_0, v_1]$, where (v_0, v_1) is equal to each of the four possible morphisms). This algorithm is fast in both cases where the given morphism is or is not Sturmian.

4 Main Results

We have the following theorem which we can prove using the following lemma, commonly known as Fekete's Lemma.

Lemma 1. *Let* $\{a_n\}_{n\geq 1}$ *be a sequence such that* $a_{m+n} \geq a_m + a_n$ *(resp.,* $a_{m+n} \leq a_m + a_n$*). Then,* $\lim_{n\to\infty} \frac{a_n}{n}$ *exists and equals* $\sup \frac{a_n}{n}$ *(resp.,* $\inf \frac{a_n}{n}$*).*

Theorem 2. *Let* w *be an infinite binary word and* ψ *be a binary morphism. Then,* $\rho^{ab}_{\psi(w)}(n) = \tilde{O}\left(\rho^{ab}_{w}(n)\right)$.

This leads to the following corollary.

Corollary 1. *Let* ϕ *and* ψ *be binary morphisms. Then,*

$$\rho^{ab}_{(\psi\circ\phi)^{\omega}(0)}(n) = \tilde{\Theta}\left(\rho^{ab}_{(\phi\circ\psi)^{\omega}(0)}(n)\right).$$

The following result is a generalization of one direction of [14, Theorem 3.3]. Note that it holds for alphabets of any size.

Theorem 3. *Let* ψ *be a morphism over* A_k *such that there exist positive integers* $n_0, n_1, \ldots, n_{k-1}$ *such that for all* $a, b \in A_k$*,* $\psi(a)^{n_a}$ *and* $\psi(b)^{n_b}$ *are abelian equivalent (have the same Parikh vector). Then, for any infinite word* w *over* A_k*,* $\rho^{ab}_{\psi(w)}(n) = \Theta(1)$*. In particular, if the fixed point of* ψ *at* 0 *exists,* $\rho^{ab}_{\psi^{\omega}(0)}(n) = \Theta(1)$.

The following criterion allows the classification of more morphisms.

Theorem 4. *Let φ be a binary morphism such that $\varphi(1) = 1^m$ for some $m \geq 1$. Let c be the number of zeroes in $\varphi(0)$. Assume that $c + m > 2$ (so that the fixed point at 0 can exist), and, if $m = 1$, then assume $\varphi(0)$ ends in 1. Then, one of the following cases holds: $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta(n)$ if $c > m$, $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta\left(\frac{n}{\log n}\right)$ if $c = m$, $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta\left(n^{\log_m c}\right)$ if $1 < c < m$, and $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta(1)$ if $c = 1$.*

Proof. First, the case where $\Theta(1)$ if $c = 1$ follows from the fact that $\varphi^\omega(0) = 01^\omega$. Next, all other cases use the fact that $\rho^{ab}_{\varphi^\omega(0)}$ is monotone increasing ($\varphi^\omega(0)$ contains arbitrarily long blocks of ones). Finally, in each case, we consider limits of ratios of the maximal number of zeroes in a subword and the target complexity and can show that they exist and are between 0 and ∞. □

4.1 Factorization of Parry Morphisms

When we say Parry morphisms, we mean those studied in [2] that we stated have fixed points with bounded abelian complexity in Proposition 2(4). We describe all canonical factorizations of such morphisms, which allow us to construct additional morphisms with bounded abelian complexity, due to Corollary 1.

The following theorem states how to fully factor the two types of Parry morphisms.

Theorem 5.
- *If $\varphi = (0^p1, 0^q1)$ with $1 \leq q < p$, then all factorizations of φ are of the form $(\prod_{i=1}^m \phi_i) \circ (01, 1)$, where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$.*
- *If $\varphi = (0^p1, 0^q)$ with $1 \leq q \leq p$, then for all choices of a nonnegative odd integer N, of a sequence of nonnegative integers $a_0, a_1, \ldots, a_N$ with all but possibly the last positive, and of integers q_0, q' with $q_0 \geq 0$ and $q' > 0$ where*

  $$\prod_{i=0}^{\frac{N-1}{2}} a_{2i} + q_0 + \sum_{i=0}^{\frac{N-1}{2}} \left(a_{2i+1} \prod_{j=0}^{i} a_{2j} \right) = p, \qquad q' \prod_{i=0}^{\frac{N-1}{2}} a_{2i} = q,$$

  there exists a complete canonical factorization:

  $$\varphi = (0, 01)^{q_0} \circ \left(\prod_{j=0}^{\frac{N-1}{2}} \left(\left(\prod_{i=1}^{m_{2j}} (0^{p_{i,2j}}, 1) \right) \circ (0, 01)^{a_{2j+1}} \right) \right) \circ (01, 0) \circ \left(\prod_{i=1}^{m'} \left(0, 1^{q'_i}\right) \right),$$

  where each of the m_{2j}'s is a positive integer, all of the $p_{i,2j}$'s are prime, $\prod_{i=1}^{m_{2j}} p_{i,2j} = a_{2j}$, all of the q'_i's are prime, and $\prod_{i=1}^{m'} q'_i = q'$.

In both cases, any composites of the necessary forms yield a Parry morphism of the proper type (where for the complicated case, all we require is that the complicated p value exceed the complicated q value).

Proof. To prove this result, we need to show how to completely canonically factor various types of morphisms:

1. Every complete canonical factorization of the morphism $(0, 1^q)$ has the form $(0, 1^{p_1}) \circ (0, 1^{p_2}) \circ \cdots \circ (0, 1^{p_m})$ for $p_1, \ldots, p_m$ primes such that $\prod_{i=1}^m p_i = q$. Also, if $\prod_{i=1}^m p_i = q$, then $(0, 1^{p_1}) \circ (0, 1^{p_2}) \circ \cdots \circ (0, 1^{p_m}) = (0, 1^q)$.
2. Every complete canonical factorization of $(0^p, 1)$ has the form $(0^{p_1}, 1) \circ (0^{p_2}, 1) \circ \cdots \circ (0^{p_m}, 1)$ for $p_1, \ldots, p_m$ primes such that $\prod_{i=1}^m p_i = p$, and if $\prod_{i=1}^m p_i = p$, then $(0^{p_1}, 1) \circ (0^{p_2}, 1) \circ \cdots \circ (0^{p_m}, 1) = (0^p, 1)$.
3. Every complete canonical factorization of $(0^p, 0^q1)$ has the form $\prod_{i=1}^m \phi_i$, where $\phi_i = (0^{p_i}, 1)$ for some prime p_i, or $\phi_i = (0, 01)$. Also, any composite of the form $\prod_{i=1}^m \phi_i$, where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$, yields a morphism of the form $(0^p, 0^q1)$.
4. Every complete canonical factorization of $(0^p1, 1)$ has the form $(0^{p_1}, 1) \circ (0^{p_2}, 1) \circ \cdots \circ (0^{p_m}, 1) \circ (01, 1)$ for $p_1, \ldots, p_m$ primes such that $\prod_{i=1}^m p_i = p$, and if $\prod_{i=1}^m p_i = p$, then $(0^{p_1}, 1) \circ (0^{p_2}, 1) \circ \cdots \circ (0^{p_m}, 1) \circ (01, 1) = (0^p1, 1)$.
5. Every complete canonical factorization of $(0^p1, 0^q)$ has the form

   $$\left(\prod_{i=1}^m \phi_i\right) \circ (01, 0) \circ \left(\prod_{j=1}^{m'} (0, 1^{q_j})\right),$$

   where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$, and each of the q_j's is prime (we allow the second product to be empty, in which case $m' = 0$). Also, any composite of the form

   $$\left(\prod_{i=1}^m \phi_i\right) \circ (01, 0) \circ \left(\prod_{j=1}^{m'} (0, 1^{q_j})\right),$$

   where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$, and each of the q_j's is prime (and $m' = 0$ is allowed), yields a morphism of the form $(0^p1, 0^q)$.
6. Every complete canonical factorization of $(0^p1, 0^q1)$ with $p > q$ has the form $(\prod_{i=1}^m \phi_i) \circ (01, 1)$, where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$. Also, any composite of the form $(\prod_{i=1}^m \phi_i) \circ (01, 1)$, where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$ yields a morphism of the form $(0^p1, 0^q1)$ with $p > q$.

The result for the first type of Parry morphism follows directly from item 6. We now prove the result for the second type of Parry morphism. By item 5, all complete canonical factorizations of $\varphi = (0^p1, 0^q)$ are of the form

$$\varphi = \left(\prod_{i=1}^m \phi_i\right) \circ (01, 0) \circ \left(\prod_{j=1}^{m'} (0, 1^{q_j})\right),$$

where $\phi_i = (0^{p_i}, 1)$ for some prime p_i or $\phi_i = (0, 01)$, and each of the q_j's is prime. We can use item 2, item 1, and the fact that $(0, 01)^m = (0, 0^m 1)$ to assert

$$\varphi = (0, 0^{q_0} 1) \circ \left(\prod_{j=0}^{\frac{N-1}{2}} ((0^{a_{2j}}, 1) \circ (0, 0^{a_{2j+1}} 1)) \right) \circ (01, 0) \circ \left(0, 1^{q'}\right), \qquad (1)$$

for some odd $N \geq 0$, some $q_0 \geq 0$, some $q' > 0$, and some sequence $a_1, a_2, \ldots, a_N$ all positive except for possibly the last, which is nonegative (all of these are integers). This factorization is (probably) not complete. Item 2, item 1, and the fact that $(0, 01)^m = (0, 0^m 1)$ combine to give the complete form, which is precisely what the theorem requires (and is not restated here).

We begin by defining two sequences: $p_i = 1$ if $i = 0$, $a_{i-1} p_{i-1}$ if i is odd, and p_{i-1} otherwise, and $q_i = q_0$ if $i = 0$, $a_{i-1} p_{i-1} + q_{i-1}$ if i is even, and q_{i-1} otherwise. We can prove by induction on m that

$$(0, 0^{q_0} 1) \circ \left(\prod_{j=0}^{m} ((0^{a_{2j}}, 1) \circ (0, 0^{a_{2j+1}} 1)) \right) = (0^{p_{2m+2}}, 0^{q_{2m+2}} 1) .$$

As this is the beginning of the factorization in Eq. (1), this implies that $\varphi = (0^{p_{N+1}}, 0^{q_{N+1}} 1) \circ (01, 0) \circ \left(0, 1^{q'}\right) = (0^{p_{N+1}}, 0^{q_{N+1}} 1) \circ \left(01, 0^{q'}\right)$, which is equal to $\left(0^{p_{N+1} + q_{N+1}} 1, 0^{q' p_{N+1}}\right)$. So, we have $p = p_{N+1} + q_{N+1}$ and $q = q' p_{N+1}$. We can then prove by induction on m that

$$p_{2m} = \prod_{i=0}^{m-1} a_{2i}, \qquad q_{2m} = q_0 + \sum_{i=0}^{m-1} \left(a_{2i+1} \prod_{j=0}^{i} a_{2j} \right) .$$

Substituting $\frac{N+1}{2}$ for m proves the desired formulas:

$$\prod_{i=0}^{\frac{N-1}{2}} a_{2i} + q_0 + \sum_{i=0}^{\frac{N-1}{2}} \left(a_{2i+1} \prod_{j=0}^{i} a_{2j} \right) = p_{N+1} + q_{N+1} = p,$$

$$q' \prod_{i=0}^{\frac{N-1}{2}} a_{2i} = q' p_{N+1} = q,$$

thereby completing this direction of the proof.

The converses follow from the various preceding items. □

Theorem 5, when combined with Corollary 1, yields the following corollaries.

Corollary 2. *Let φ be a morphism with a complete canonical factorization of the form*

$$\left(\prod_{i=1}^{m_0} \phi_{0,i} \right) \circ (01, 1) \circ \left(\prod_{j=1}^{m_1} \phi_{1,j} \right) ,$$

for some integers $m_0, m_1 \geq 0$*, where* $\phi_{m,n} = (0^{p_n}, 1)$ *for some prime* p_n *or* $\phi_{m,n} = (0, 01)$*. Then* $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta(1)$*.*

Corollary 3. *Let* φ *be a morphism with a (probably not complete) canonical factorization that is a cyclic shift of the following composite:*

$$(0, 0^{q_0}1) \circ \left(\prod_{j=0}^{\frac{N-1}{2}} \left((0^{a_{2j}}, 1) \circ (0, 0^{a_{2j+1}}1) \right) \right) \circ (01, 0) \circ \left(0, 1^{q'}\right)$$

for some odd $N \geq 0$ *(in case* $N = -1$*, the product term is absent), some* $q_0 \geq 0$*, some* $q' > 0$*, and some sequence* $a_0, a_1, \ldots, a_N$ *all nonnegative. If*

$$\prod_{i=0}^{\frac{N-1}{2}} a_{2i} + q_0 + \sum_{i=0}^{\frac{N-3}{2}} \left(a_{2i+1} \prod_{j=0}^{i} a_{2j} \right) \geq q' \prod_{i=0}^{\frac{N-1}{2}} a_{2i},$$

then $\rho^{ab}_{\varphi^\omega(0)}(n) = \Theta(1)$*.*

An example of a morphism classifiable by Corollary 2 is $(001001, 00101)$, which has a complete canonical factorization $(0, 01) \circ (01, 1) \circ (0, 01) \circ (00, 1)$, which satisfies the conditions of Corollary 2. An example of a morphism classifiable by Corollary 3 is $(0011, 0)$, which has a complete canonical factorization $(0, 11) \circ (0, 01) \circ (01, 0)$. This is a cyclic shift of $(0, 01) \circ (01, 0) \circ (0, 11)$, so we have $q_0 = 1$, $q' = 2$, and $N = -1$.

4.2 Classification of Uniform Morphisms

We now derive a complete classification of the abelian complexities of fixed points of uniform binary morphisms.

Let φ be a uniform binary morphism with fixed point at 0. The *length* of φ (denoted $\ell(\varphi)$ or just ℓ if φ is unambiguous) is equal to $|\varphi(0)|$ (which equals $|\varphi(1)|$). The *difference* of φ (denoted $d(\varphi)$ or just d if φ is unambiguous) equals $|z(\varphi(0)) - z(\varphi(1))|$. The *delta* of φ (denoted $\Delta(\varphi)$ or just Δ if φ is unambiguous) equals $z_M(\ell(\varphi)) - \max\{z(\varphi(0)), z(\varphi(1))\}$, where $z_M(\ell(\varphi))$ denotes the maximum number of zeroes in a subword of length $\ell(\varphi)$ of $\varphi^\omega(0)$. Also, if φ is unambiguous, we denote $z(\varphi(0))$ by z_0, $z(\varphi(1))$ by z_1, and $\rho^{ab}_{\varphi^\omega(0)}(n)$ by $\rho^{ab}(n)$.

Theorem 6. *Let* φ *be a 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:*

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

Proof. We prove the fourth case. Assume that $d = 1$, $\Delta = 0$, and we are not in the second or third case. We can show that there exist integers i and j such that $\varphi^2(0)[i] = \varphi^2(1)[i] = 0$ and $\varphi^2(0)[j] = \varphi^2(1)[j] = 1$. We can also show that $d(\varphi^2) = 1$ and $\Delta(\varphi^2) = 0$, so $\rho^{ab}(\ell^2 n - 1) = \rho^{ab}(\ell^2 n + 1) = \rho^{ab}(\ell^2 n) + 1$. Consider the sequence defined by $a_0 = 1$ and $a_i = \ell^2 a_{i-1} + 1$. It can be easily shown by induction that $a_i = \sum_{j=0}^{i} \ell^{2j}$. Also, we know that $\rho^{ab}(a_i) = i$. Hence, we have $\rho^{ab}\left(\sum_{j=0}^{i} \ell^{2j}\right) = i$, so the a_i's give a subsequence of $\rho^{ab}(n)$ with logarithmic growth. We now show that $\rho^{ab}(n)$ is $O(\log n)$ completing our proof that it is $\tilde{\Theta}(f(n))$. Let c be the maximum value of $\rho^{ab}(n)$ for $n < \ell$. For $r \in \{0, \ldots, \ell - 1\}$, we can show that that

$$\rho^{ab}(\ell n + r) \leq d\rho^{ab}(n) - d + 2 + 2\Delta + \rho^{ab}(r) \leq \rho^{ab}(n) + 1 + c.$$

As we increase by (approximately) a factor of ℓ in the argument, we can increase by at most a constant in value. This is $\log n$ behavior, as required.

In the fifth case, $d = 1$ and $\Delta > 0$, and the inequality $\rho^{ab}(\ell n + r) \leq d\rho^{ab}(n) - d + 2 + 2\Delta + \rho^{ab}(r)$ leads similarly to $\rho^{ab}(n) = O(\log n)$. The same inequality leads to the $O(n^{\log_\ell d})$ bound in the sixth case, and a similar inequality yields the $\Omega(n^{\log_\ell d})$ bound. □

Note that some uniform morphisms have nontrivial factorizations. Hence, Theorem 6 gives a classification of the abelian complexities of some nonuniform morphisms as well via Corollary 1. For example, $(01, 00) = (01, 0) \circ (0, 11)$ and $(0, 11) \circ (01, 0) = (011, 0)$. Let $\varphi = (01, 00)$ and $\psi = (011, 0)$. Since $\rho^{ab}_{\varphi^\omega(0)}(n) = \tilde{\Theta}(f(n))$, $\rho^{ab}_{\psi^\omega(0)}(n) = \tilde{\Theta}(f(n))$ as well, though ψ is not uniform.

Referring to the fifth case of Theorem 6, we conjecture an $\Omega(\log n)$ bound abelian complexity for all uniform binary morphisms with $d = 1$ and $\Delta > 0$.

Conjecture 1. Let φ be a uniform binary morphism with $d = 1$ and $\Delta > 0$. For all $h \geq 1$ and $n \geq \ell^h$, $\rho^{ab}(n) \geq h + 2$.

5 Future Work

Problems to be considered in the future include: prove (or disprove) the conjectured $\Omega(\log n)$ bound for uniform binary morphisms with $d = 1$ and $\Delta > 0$, carry out worst and average case running time analyses on the factorization algorithm, examine additional classes of morphisms, and attempt to extend some results to $k > 2$. Most of our results about abelian complexity are about binary words. A notable exception is Theorem 3.

In general, if the alphabet size k is greater than 2, we lose the property that for an infinite word w, $\left|\rho^{ab}_w(n+1) - \rho^{ab}_w(n)\right| \leq 1$. We also can no longer reduce questions about abelian complexity to simply counting zeroes. In general, if w is an infinite word over a k-letter alphabet, Proposition 1(3) says that $\rho^{ab}_w(n) = O(n^{k-1})$. If w is required to be the fixed point of a morphism, we can give a better bound. Corollary 10.4.9 in [1] says that if w is the fixed point of a morphism, then $\rho_w(n) = O(n^2)$. Hence, by Proposition 1(4), if w is the fixed point of a morphism, then $\rho^{ab}_w(n) = O(n^2)$, no matter how large k is.

In many cases, we can give an even better upper bound. Allouche and Shallit [1] define a *primitive morphism* as a morphism φ for which there exists an integer $n \geq 1$ such that given any letters a and b in the alphabet, a occurs in $\varphi^n(b)$. Then [1, Theorem 10.4.12] states that if w is the fixed point of a primitive morphism, then $\rho_w(n) = O(n)$. Hence if w is the fixed point of a primitive morphism, then $\rho_w^{ab}(n) = O(n)$, no matter how large k is.

Finally, we note that the truth value of Corollary 1 has not been examined in depth for alphabets of size greater than 2. Our proof of Theorem 2 certainly depends on the alphabet size, but we have not yet seen a counterexample to it for a larger alphabet. Since binary morphisms can be factorized over larger alphabets, the truth of Corollary 1 would allow us to classify the abelian complexities of the fixed points of many morphisms with $k > 2$ simply based on the results we have here for binary morphisms.

References

1. Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press (2003)
2. Balková, L., Břinda, K., Turek, O.: Abelian complexity of infinite words associated with quadratic Parry numbers. Theoret. Comput. Sci. 412, 6252–6260 (2011)
3. Berstel, J., Séébold, P.: A characterization of Sturmian morphisms. In: Borzyszkowski, A.M., Sokolowski, S. (eds.) MFCS 1993. LNCS, vol. 711, pp. 281–290. Springer, Heidelberg (1993)
4. Choffrut, C., Karhumäki, J.: Combinatorics of Words. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, vol. 1, ch. 6, pp. 329–438. Springer, Berlin (1997)
5. Coven, E.M., Hedlund, G.A.: Sequences with minimal block growth. Math. Systems Theory 7, 138–153 (1973)
6. Currie, J., Rampersad, N.: Recurrent words with constant abelian complexity. Adv. Appl. Math. 47, 116–124 (2011)
7. Ehrenfeucht, A., Lee, K.P., Rozenberg, G.: Subword complexities of various classes of deterministic developmental languages without interactions. Theoret. Comput. Sci. 1, 59–75 (1975)
8. Ehrenfeucht, A., Rozenberg, G.: On the subword complexity of D0L languages with a constant distribution. Inform. Process. Lett. 13, 108–113 (1981)
9. Ehrenfeucht, A., Rozenberg, G.: On the subword complexity of square-free D0L languages. Theoret. Comput. Sci. 16, 25–32 (1981)
10. Ehrenfeucht, A., Rozenberg, G.: On the subword complexity of locally catenative D0L languages. Inform. Process. Lett. 16, 121–124 (1983)
11. Frid, A.E.: The subword complexity of fixed points of binary uniform morphisms. In: Chlebus, B.S., Czaja, L. (eds.) FCT 1997. LNCS, vol. 1279, pp. 179–187. Springer, Heidelberg (1997)
12. Pansiot, J.J.: Complexité des facteurs des mots infinis engendrés par morphismes itérés. In: Paredaens, J. (ed.) ICALP 1984. LNCS, vol. 172, pp. 380–389. Springer, Heidelberg (1984)
13. Richomme, G.: Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words. Theoret. Comput. Sci. 380, 393–400 (2007)
14. Richomme, G., Saari, K., Zamboni, L.Q.: Abelian complexity in minimal subshifts. J. London Math. Soc. 83, 79–95 (2011)