



REDUCED COMPLEXITIES FOR SEQUENCES OVER FINITE ALPHABETS

John M. Campbell

Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada
 jh241966@dal.ca

James D. Currie

Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, Manitoba, Canada
 j.currie@uwinnipeg.ca

Narad Rampersad

Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, Manitoba, Canada
 n.rampersad@uwinnipeg.ca

Received: 9/19/25, Accepted: 2/6/26, Published: 2/20/26

Abstract

Letting w denote a finite, nonempty word, let $\text{red}(w)$ denote the word obtained from w by replacing every subword s of w of the form $cc \cdots c$ for a given character c (such that there is no character immediately to the left or right of s equal to c) with c . Complexity functions for infinite words play important roles within combinatorics on words and the study of automatic sequences, and this leads us to introduce and investigate variants of the factor and abelian complexity functions using the given reduction operation. By enumerating words v and w of a given length $n \geq 0$ and associated with an infinite sequence over a finite alphabet such that $\text{red}(v)$ and $\text{red}(w)$ are equal or otherwise equivalent in some specified way, by analogy with the factor and abelian complexity functions, this may be seen as producing simplified versions of existing complexity functions. We prove a recursion for the reduced factor complexity function $\rho_{\mathbf{t}}^{\text{red}}$ for the Thue–Morse sequence \mathbf{t} , giving us that $(\rho_{\mathbf{t}}^{\text{red}}(n) : n \in \mathbb{N})$ is a 2-regular sequence, and we prove an explicit evaluation for the reduced factor complexity function $\rho_{\mathbf{f}}^{\text{red}}$ for the (regular) paperfolding sequence \mathbf{f} , together with an evaluation for the reduced abelian complexity function $\rho_{\mathbf{f}}^{\text{ab,red}}$ for \mathbf{f} . We conclude with open problems concerning $\rho_{\mathbf{t}}^{\text{ab,red}}$.

1. Introduction

For an infinite sequence \mathbf{x} over a finite alphabet, the *factor complexity function*

$$\rho_{\mathbf{x}} : \mathbb{N}_0 \rightarrow \mathbb{N} \tag{1}$$

maps $n \geq 0$ to the number of distinct factors of \mathbf{x} of length n . Similarly, the *abelian complexity function*

$$\rho_{\mathbf{x}}^{\text{ab}} : \mathbb{N}_0 \rightarrow \mathbb{N} \tag{2}$$

maps $n \geq 0$ to the number of distinct factors of \mathbf{x} of length n , up to equivalence by rearrangements of characters (possibly by an identity permutation) in words of the same length. To construct simplified versions of the complexity functions in (1) and (2) and of other existing complexity functions, we introduce counting functions based on factors that are reduced according to a function on nonempty, finite words defined below.

We henceforth adopt the convention whereby infinite sequences are denoted in boldface and whereby finite sequences may be referred to as (finite) *words* and are denoted without boldface. For a nonempty and finite word w , let w be written as

$$w = c_1 \cdots c_1 c_2 \cdots c_2 \cdots c_n \cdots c_n$$

for some positive integer n , where c_1, c_2, \dots, c_n are characters such that $c_i \neq c_{i+1}$ for all possible indices i . A maximal block of identical characters $c_i \cdots c_i$ is called a *run*. We then define the *reduction* $\text{red}(w)$ of w so that

$$\text{red}(w) = c_1 c_2 \cdots c_n.$$

For finite words w and v , we say that w and v are *reduced-equivalent*, writing $w \sim_{\text{red}} v$, if $\text{red}(w) = \text{red}(v)$. In the particular case when w is over the binary alphabet $\{0, 1\}$, we see that $\text{red}(w)$ must consist of alternating 0's and 1's and hence that the \sim_{red} equivalence class of w is uniquely determined by the first letter of w and the number of runs in w . For example, if $w = 0010110$ and $v = 0111010$ then $\text{red}(w) = \text{red}(v) = 01010$ and $w \sim_{\text{red}} v$.

For an infinite sequence \mathbf{x} over a finite alphabet, we then define the *reduced factor complexity function*

$$\rho_{\mathbf{x}}^{\text{red}} : \mathbb{N}_0 \rightarrow \mathbb{N} \tag{3}$$

by analogy with (1) so that (3) maps $n \geq 0$ to the number of distinct factors of \mathbf{x} of length n , up to equivalence by \sim_{red} . Similarly, we define the *reduced abelian complexity function*

$$\rho_{\mathbf{x}}^{\text{ab,red}} : \mathbb{N}_0 \rightarrow \mathbb{N} \tag{4}$$

by analogy with (2) so that (4) maps $n \geq 0$ to the number of distinct factors of \mathbf{x} of length n , where two factors v and w are considered to be equivalent if $\text{red}(v)$

is of the same length as $\text{red}(w)$, and $\text{red}(v)$ can be obtained by rearranging the characters of $\text{red}(w)$ (possibly by an identity permutation).

Informally, given a complexity function p defined on \mathbf{x} , a *reduced complexity function* associated with p may be defined by counting length- n factors of \mathbf{x} , up to an equivalence relation such that factors v and w of equal length are equivalent if $\text{red}(v)$ and $\text{red}(w)$ are equivalent according to how finite words are enumerated in the definition of p . The reduced abelian complexity function defined above provides a prototypical instance of this.

For an infinite sequence denoted as an infinite word, let the index of the initial term be 1. For $n \in \mathbb{N}$, set \mathbf{t}_n to be equal to the number of 1's, modulo 2, in the base-2 expansion of $n - 1$. This produces the famous *Thue–Morse sequence*

$$\mathbf{t} = 011010011001011010010110011010011001011001101001011010 \cdots \quad (5)$$

with reference to the work of Allouche and Shallit [4] on the ubiquitous nature of the sequence in (5). For a positive integer n , we write $n = n'2^k$ for an odd integer n' , and we then set $\mathbf{f}_n = 0$ if $n' \equiv 1 \pmod{4}$ and $\mathbf{f}_n = 1$ otherwise. This allows us to define the *ordinary* (or *regular*) *paperfolding sequence*

$$\mathbf{f} = 0010011000110110001001110011011000100110001101110010011 \cdots \quad (6)$$

The sequences in (5) and (6) may be seen as fundamentally important and prototypical instances of automatic sequences, with reference to the standard monograph on automatic sequences [3]. Past research on the evaluation of the factor complexity function $\rho_{\mathbf{t}}$ [5, 7, 10] together with the work of Allouche [2] on the evaluation of $\rho_{\mathbf{f}}$ and together with the work of Madill and Rampersad [13] on the evaluation of $\rho_{\mathbf{f}}^{\text{ab}}$ motivate problems concerning the evaluation of reduced complexity functions for both \mathbf{t} and \mathbf{f} .

As noted earlier, the reduced complexity function of an infinite sequence is closely related to the structure of the runs in the sequence. The sequence of run lengths of both \mathbf{t} and \mathbf{f} have previously been studied. For instance, Allouche, Allouche, and Shallit [1] showed that the run length sequence of \mathbf{t} , given by the fixed point of the map $1 \rightarrow 121, 2 \rightarrow 12221$, is not an automatic sequence. The run length sequence of \mathbf{f} has also been studied by Bunder, Bates, and Arnold [8] as well as by Shallit [16].

2. Main Results

We begin by considering the reduced factor complexity functions for \mathbf{t} and \mathbf{f} , and then consider the reduced abelian complexity function for \mathbf{f} . We conclude in Section 3 with open problems concerning the reduced abelian complexity function for \mathbf{t} .

2.1. Reduced Factor Complexity Functions

For convenience, we typically disregard the trivial case whereby a complexity function may or may not count the unique word of length 0, i.e., the empty or null word. The integer sequence

$$(\rho_{\mathbf{t}}(n) : n \in \mathbb{N}) = (2, 4, 6, 10, 12, 16, 20, 22, 24, 28, 32, 36, 40, 42, 44, \dots) \quad (7)$$

is indexed in the On-Line Encyclopedia of Integer Sequences (OEIS) [17] as A005942 and satisfies the recursion

$$\rho_{\mathbf{t}}(n) = \begin{cases} \rho_{\mathbf{t}}\left(\frac{n}{2}\right) + \rho_{\mathbf{t}}\left(\frac{n}{2} + 1\right) & \text{if } n \text{ is even,} \\ 2\rho_{\mathbf{t}}\left(\frac{n+1}{2}\right) & \text{otherwise,} \end{cases}$$

for $n \geq 1$. In contrast to (7), we find that the integer sequence

$$(\rho_{\mathbf{t}}^{\text{red}}(n) : n \in \mathbb{N}) = (2, 4, 4, 6, 4, 6, 6, 6, 4, 6, 6, 8, 6, 8, 6, 6, 4, 6, 6, 8, 6, 8, \dots)$$

is not currently included in the OEIS, and this suggests that the concept of “reduced complexity” functions has not been considered previously, and this motivates the recurrence introduced below.

Theorem 1. *For every positive integer n , we have that*

$$\rho_{\mathbf{t}}^{\text{red}}(n) = \begin{cases} \rho_{\mathbf{t}}^{\text{red}}\left(\frac{n+1}{2}\right) & \text{if } n \text{ is odd,} \\ \rho_{\mathbf{t}}^{\text{red}}(m + 1) + 2 & \text{if } n = 4m \text{ or } n = 4m + 2. \end{cases}$$

In order to prove this theorem, we recall that the Thue–Morse word \mathbf{t} can also be defined as a fixed point of the morphism μ that maps $0 \rightarrow 01$ and $1 \rightarrow 10$; i.e., we have $\mathbf{t} = \mu^\omega(0)$. Let \mathcal{F} be the set of factors of \mathbf{t} . For a word v , we denote by ^-v (resp., v^- , $^-v^-$) the result of deleting the first letter (resp., last letter, first and last letters) of v . If $w \in \mathcal{F}$ and $|w| \geq 4$ then the index $\iota(w)$ of w in \mathbf{t} is fixed modulo 2, and w can be parsed uniquely as one of

$$w = \begin{cases} \mu(u), & \text{some } u \in \mathcal{F}, \\ ^-\mu(u)^-, & \text{some } u \in \mathcal{F}, \\ \mu(u)^-, & \text{some } u \in \mathcal{F}, \text{ or} \\ ^-\mu(u), & \text{some } u \in \mathcal{F} \end{cases}$$

with these cases corresponding, in order, to the situations where $\langle |w|, \iota(w) \rangle$ modulo 2 is $\langle 0, 0 \rangle$, $\langle 0, 1 \rangle$, $\langle 1, 0 \rangle$, or $\langle 1, 1 \rangle$.

Call 01 and 10 *alternations*. By the number of alternations in a word w we mean $w_{\{01,10\}}$, the total number of occurrences of 01 and 10 in w .

Lemma 1. *If $w \in \{0, 1\}^*$ contains α alternations, then $\mu(w)$ contains $2|w| - 1 - \alpha$ alternations.*

Proof. From the definition of μ , each letter of w maps to an alternation. For each of the $|w| - 1$ length 2 factors ab of w there is an alternation in $\mu(w)$ formed from the last letter of $\mu(a)$ and the first letter of $\mu(b)$ exactly when ab is not itself an alternation. Thus $\mu(w)$ has $|w| + (|w| - 1 - \alpha) = 2|w| - 1 - \alpha$ alternations. \square

For a nonnegative integer n , let m_n (resp., M_n) be the minimum (resp., maximum) number of alternations in any $w \in \mathcal{F}$ with $|w| = n$.

Lemma 2. *For $n \geq 2$ we have*

$$\begin{aligned} m_{2n} &= 2n - 1 - M_{n+1}, \\ M_{2n} &= 2n - 1 - m_n, \\ m_{2n+1} &= 2n - M_{n+1}, \text{ and} \\ M_{2n+1} &= 2n - m_{n+1}. \end{aligned}$$

Proof. We begin by showing that $m_{2n} = 2n - 1 - M_{n+1}$. Suppose that $n \geq 2$ and w is a word of length $2n$ with the fewest alternations. Suppose w has even index ι . Sliding one place to the right, consider the word w' of length $2n$ and odd index $\iota + 1$. Comparing the count of alternations in w' with that in w , we see that w' omits the alternation formed by the first two letters of w , and may contain at most one new alternation, as a suffix. Thus w' can contain no more alternations than w . Thus w' must be a word of length $2n$ with the fewest alternations.

Write $w' = {}^-\mu(u)^-$ for some $u \in \mathcal{F}$ with $|u| = n + 1$. Let u contain α alternations. Thus w' contains $2|u| - 1 - \alpha$ alternations. We claim that u has the maximum number of alternations for a word of length $n + 1$, so that $M_{n+1} = \alpha$.

To get a contradiction, suppose that $u' \in \mathcal{F}$ with $|u'| = n + 1$ contains β alternations where $\beta > \alpha$. Then $\mu(u')$ contains $2|u'| - 1 - \beta = 2n + 1 - \beta$ alternations. The word ${}^-\mu(u')^-$ omits the first and last alternations in $\mu(u')$, and thus contains $2n - 1 - \beta < 2n - 1 - \alpha$ alternations, i.e., fewer than w' . Since $|{}^-\mu(u')^-| = 2n$, this contradicts the choice of w' . We conclude that $M_{n+1} = \alpha$, which means that $m_{2n} = 2n - 1 - M_{n+1}$, as claimed.

Next we show that $M_{2n} = 2n - 1 - m_n$. Suppose that $n \geq 2$ and w is a word of length $2n$ with the most alternations. Suppose w has odd index ι . Sliding one place to the right, consider the word w' of length $2n$ and even index $\iota + 1$. Comparing the count of alternations in w' with that in w , we see that w' may possibly have lost an alternation formed by the first two letters of w , but definitely contains a new alternation, as a suffix. Thus w' contains at least as many alternations as w . Thus w' must also be a word of length $2n$ with the most alternations.

Write $w' = \mu(u)$ for some $u \in \mathcal{F}$ with $|u| = n$. Let u contain α alternations. Thus w' contains $2|u| - 1 - \alpha = 2n - 1 - \alpha$ alternations. We claim that u has the minimum number of alternations for a word of length n , so that $m_n = \alpha$.

To get a contradiction, suppose that $u' \in \mathcal{F}$ with $|u'| = n$ contains β alternations where $\beta < \alpha$. Then $\mu(u')$ has length $2n$ and contains $2|u'| - 1 - \beta = 2n - 1 - \beta >$

$2n - 1 - \alpha$ alternations, i.e., more than w' . This contradicts the choice of w' . We conclude that $m_n = \alpha$, which means that $M_{2n} = 2n - 1 - m_n$, as claimed.

The third equality is proved analogously. Suppose that $n \geq 2$ and w is a word of length $2n + 1$ with the fewest alternations. Write $w = \mu(u)^-$ or $w = ^-\mu(u)$ for some $u \in \mathcal{F}$ with $|u| = n + 1$. Let u contain α alternations. Since w is obtained by deleting a letter from one end of $\mu(u)$, it has one fewer alternation than $\mu(u)$. Thus w contains $2|u| - 1 - \alpha - 1 = 2n - \alpha$ alternations. We claim that u has the maximum number of alternations for a word of length $n + 1$, so that $M_{n+1} = \alpha$.

To get a contradiction, suppose that $u' \in \mathcal{F}$ with $|u'| = n$ contains β alternations where $\beta > \alpha$. Then $\mu(u')^-$ is a word of length $2n + 1$ that contains $2|u'| - 1 - \beta - 1 = 2n - \beta < 2n - \alpha$ alternations. This contradicts the choice of w' . We thus conclude that $M_{n+1} = \alpha$, which means that $m_{2n+1} = 2n - M_{n+1}$, as claimed.

Finally, suppose that $n \geq 2$ and w is a word of length $2n + 1$ with the most alternations. Write $w = \mu(u)^-$ or $w = ^-\mu(u)$ for some $u \in \mathcal{F}$ with $|u| = n + 1$. Let u contain α alternations. Since w is obtained by deleting a letter from one end of $\mu(u)$, it has one fewer alternation than $\mu(u)$. Thus w contains $2|u| - 1 - \alpha - 1 = 2n - \alpha$ alternations. We claim that u has the minimum number of alternations for a word of length $n + 1$, so that $m_{n+1} = \alpha$.

To get a contradiction, suppose that $u' \in \mathcal{F}$ with $|u'| = n$ contains β alternations where $\beta < \alpha$. Then $\mu(u')^-$ is a word of length $2n + 1$ that contains $2|u'| - 1 - \beta - 1 = 2n - \beta > 2n - \alpha$ alternations. This contradicts the choice of w' . Consequently, we may conclude that $m_{n+1} = \alpha$, which means that $M_{2n+1} = 2n - m_{n+1}$, as claimed. \square

Lemma 3. For $n \geq 1$ we have

$$\begin{aligned} m_{4n} &= 2n - 1 + m_{n+1}, \\ M_{4n} &= 2n + M_{n+1}, \\ m_{4n+2} &= 2n + m_{n+1}, \text{ and} \\ M_{4n+2} &= 2n + 1 + M_{n+1}. \end{aligned}$$

Proof. To get each identity, we apply Lemma 2 twice. We have

$$\begin{aligned} m_{4n} &= 4n - 1 - M_{2n+1} \\ &= 4n - 1 - (2n - m_{n+1}) \\ &= 2n - 1 + m_{n+1}, \end{aligned}$$

$$\begin{aligned} M_{4n} &= 4n - 1 - m_{2n} \\ &= 4n - 1 - (2n - 1 - M_{n+1}) \\ &= 2n + M_{n+1}, \end{aligned}$$

$$\begin{aligned} m_{4n+2} &= 4n + 2 - 1 - M_{2n+2} \\ &= 4n + 1 - (2n + 2 - 1 - m_{n+1}) \\ &= 2n + m_{n+1}, \end{aligned}$$

and

$$\begin{aligned} M_{4n+2} &= 4n + 2 - 1 - m_{2n+1} \\ &= 4n + 1 - (2n - M_{n+1}) \\ &= 2n + 1 + M_{n+1}. \end{aligned}$$

□

We can now give the proof of Theorem 1.

Proof of Theorem 1. Recall that the \sim_{red} equivalence class of a binary word w is determined by the first letter of w and the number of runs in w . Moreover, the number of runs in w is one greater than the number of alternations in w . Now let w be a factor of length n of \mathbf{t} with i runs. Then $i \in [m_n + 1, M_n + 1]$. Furthermore, for every i in this interval, there is such a w in \mathbf{t} (sliding w to the left or right in \mathbf{t} can only increase or decrease the number of runs by at most 1). Since \mathbf{t} is closed under complementation, we see that

$$\rho_{\mathbf{t}}^{\text{red}}(n) = 2(M_n + 1 - m_n - 1 + 1) = 2(M_n - m_n + 1).$$

We now consider different cases for n and apply Lemmas 2 and 3.

If n is odd, say $n = 2m + 1$, we have

$$\begin{aligned} \rho_{\mathbf{t}}^{\text{red}}(n) &= 2(M_{2m+1} - m_{2m+1} + 1) \\ &= 2(2m - m_{m+1} - 2m + M_{m+1} + 1) \\ &= 2(M_{m+1} - m_{m+1} + 1) \\ &= \rho_{\mathbf{t}}^{\text{red}}(m + 1) \\ &= \rho_{\mathbf{t}}^{\text{red}}\left(\frac{n + 1}{2}\right). \end{aligned}$$

If $n = 4m$, we have

$$\begin{aligned} \rho_{\mathbf{t}}^{\text{red}}(n) &= 2(M_{4m} - m_{4m} + 1) \\ &= 2(2m + M_{m+1} - 2m + 1 - m_{m+1} + 1) \\ &= 2(M_{m+1} - m_{m+1} + 1) + 2 \\ &= \rho_{\mathbf{t}}^{\text{red}}(m + 1) + 2. \end{aligned}$$

- Repeat.

$0\ 0\ 1\ ?\ 0\ 1\ 1\ ?\ 0\ 0\ 1\ ?\ 0\ 1\ 1\ \dots$
 $0\ 0\ 1\ 0\ 0\ 1\ 1\ ?\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ \dots$
 $0\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ \dots$

In the limit, one obtains the paperfolding word \mathbf{f} .

Lemma 4. *Let $a, b \in \{0, 1\}$, $a \neq b$, and let w be a length- $(2n + 1)$ factor of \mathbf{f} of the form*

$$w = a?b?a?b? \dots a?b \text{ or } w = a?b?a?b? \dots b?a.$$

Then w has $n + 1$ runs.

Proof. The number of runs in w is one more than the number of occurrences of ab and ba in w . It is clear that there are n such occurrences. □

Lemma 5. *For every positive integer n , we have that $\rho_{\mathbf{f}}^{\text{red}}(2n) = 4$.*

Proof. Let w be a factor of \mathbf{f} of length $2n$. Then w has one of the forms

$$\begin{aligned} w &= a?b?a?b? \dots a?bx, & w &= a?b?a?b? \dots b?ax, \\ w &= xa?b?a?b? \dots a?b, & w &= xa?b?a?b? \dots b?a, \end{aligned}$$

for some $a, b, x \in \{0, 1\}$ with $a \neq b$.

Suppose $w = a?b?a?b? \dots a?bx$. By Lemma 4, the word w has n runs if $x = b$ and $n + 1$ runs if $x = a$. We need to show that all four choices for $a, x \in \{0, 1\}$ occur. Let i be the position of the x in some occurrence of w in \mathbf{f} .

If $a = 0$ then $i \equiv 0 \pmod{4}$. Choose $i \equiv 4 \pmod{16}$ to get $x = 0$ and choose $i \equiv 12 \pmod{16}$ to get $x = 1$.

If $a = 1$ then $i \equiv 2 \pmod{4}$. Choose $i \equiv 2 \pmod{8}$ to get $x = 0$ and choose $i \equiv 6 \pmod{8}$ to get $x = 1$.

Hence w has four possible \sim_{red} equivalence classes and all occur in \mathbf{f} . A similar analysis applies to the other cases for w and yields four equivalence classes in every case. □

Lemma 6. *For every positive integer n , we have that $\rho_{\mathbf{f}}^{\text{red}}(8n + 1) = 4$.*

Proof. Let w be a factor of \mathbf{f} of length $8n + 1$. Then w has one of the forms

$$w = a?b?a?b? \dots b?a, \quad w = xa?b?a?b? \dots a?by,$$

for some $a, b, x, y \in \{0, 1\}$ with $a \neq b$.

Suppose that $w = a?b?a?b? \dots b?a$. By Lemma 4, the word w has $4n + 1$ runs; the two possibilities for $a \in \{0, 1\}$ give two equivalence classes.

Now suppose that $w = xa?b?a?b? \cdots a?by$ and let i and j be the positions of the x and y , respectively, in some occurrence of w in \mathbf{f} . Note that $j - i = 8n$. If we write $w = xw'y$, we can apply Lemma 4 to w' to conclude that w' has $4n$ runs. Note that if $x = y$ then w has $4n + 1$ runs, but we have already accounted for these equivalence classes in the previous case. We therefore consider the case where $x \neq y$.

If $a = 0$ then $i \equiv 0 \pmod{4}$. Write $8n = (2n' + 1)2^k$. To get $xy = 01$, choose $i = 2^{k-1}$, which gives $j = (4n' + 3)2^{k-1}$. In this case w has $4n$ runs. To get $xy = 10$, choose $i = 3 \cdot 2^{k-1}$, which gives $j = (4n' + 5)2^{k-1}$. In this case w has $4n + 2$ runs.

If $a = 1$ then $i \equiv 2 \pmod{4}$. We therefore have either $i \equiv j \equiv 2 \pmod{8}$ or $i \equiv j \equiv 6 \pmod{8}$. Both cases give $x = y$, a contradiction.

In total then, w has four possible \sim_{red} equivalence classes and all occur in \mathbf{f} . \square

For the remaining equivalence classes modulo 8, we make use of the software Walnut [15] in the proofs of the following lemmas. Note that, while we have been indexing the terms of the paperfolding word starting with 1, Walnut expects all automatic sequences to be indexed starting with 0. The indices in the Walnut formulas used in the proofs below therefore are 1 less than you would expect from the surrounding analysis.

Lemma 7. *For every nonnegative integer n , we have that $\rho_{\mathbf{f}}^{\text{red}}(8n + 3) = 6$.*

Proof. Let w be a factor of \mathbf{f} of length $8n + 3$. Then w has one of the forms

$$w = a?b?a?b? \cdots a?b, \quad w = xa?b?a?b? \cdots b?ay,$$

for some $a, b, x, y \in \{0, 1\}$ with $a \neq b$.

Suppose that $w = a?b?a?b? \cdots a?b$. By Lemma 4, the word w has $4n + 2$ runs; the two possibilities for $a \in \{0, 1\}$ give two equivalence classes.

Now, suppose that $w = xa?b?a?b? \cdots b?ay$ and let i be the position of the x in some occurrence of w in \mathbf{f} . Note that if $a = 0$, then $i \equiv 0 \pmod{4}$, and if $a = 1$, then $i \equiv 2 \pmod{4}$. Also, if $x \neq y$, then w has $4n + 2$ runs, and we have already accounted for these equivalence classes above. We therefore consider $x = y$. There are two possibilities for a and two for x ; if these all occur in \mathbf{f} we get four additional equivalence classes. We verify this with the following Walnut commands, which all return TRUE:

```
eval pf_red_3mod8a "?msd_2 An Er P[4*r+3]=@0 & P[4*r+4]=@0 &
P[4*r+8*n+5]=@0":
eval pf_red_3mod8b "?msd_2 An Er P[4*r+3]=@1 & P[4*r+4]=@0 &
P[4*r+8*n+5]=@1":
eval pf_red_3mod8c "?msd_2 An Er P[4*r+1]=@0 & P[4*r+2]=@1 &
P[4*r+8*n+3]=@0":
```

```
eval pf_red_3mod8d "?msd_2 An Er P[4*r+1]=@1 & P[4*r+2]=@1 &
P[4*r+8*n+3]=@1":
```

□

Lemma 8. *For every nonnegative integer n , we have that $\rho_{\mathbf{f}}^{\text{red}}(8n + 5) = 6$.*

Proof. Let w be a factor of \mathbf{f} of length $8n + 5$. Then w has one of the forms

$$w = a?b?a?b? \cdots b?a, \quad w = xa?b?a?b? \cdots a?by,$$

for some $a, b, x, y \in \{0, 1\}$ with $a \neq b$.

Suppose that $w = a?b?a?b? \cdots b?a$. By Lemma 4, the word w has $4n + 3$ runs; the two possibilities for $a \in \{0, 1\}$ give two equivalence classes.

Now suppose that $w = xa?b?a?b? \cdots a?by$. In this case we get our four additional equivalence classes when $x \neq y$. We verify their existence for all n with the following Walnut commands, which all return TRUE:

```
eval pf_red_5mod8a "?msd_2 An Er P[4*r+3]=@0 & P[4*r+4]=@0 &
P[4*r+8*n+7]=@1":
eval pf_red_5mod8b "?msd_2 An Er P[4*r+3]=@1 & P[4*r+4]=@0 &
P[4*r+8*n+7]=@0":
eval pf_red_5mod8c "?msd_2 An Er P[4*r+1]=@0 & P[4*r+2]=@1 &
P[4*r+8*n+5]=@1":
eval pf_red_5mod8d "?msd_2 An Er P[4*r+1]=@1 & P[4*r+2]=@1 &
P[4*r+8*n+5]=@0":
```

□

Lemma 9. *For every nonnegative integer n , we have that $\rho_{\mathbf{f}}^{\text{red}}(8n + 7) = 6$.*

Proof. Let w be a factor of \mathbf{f} of length $8n + 7$. Then w has one of the forms

$$w = a?b?a?b? \cdots a?b, \quad w = xa?b?a?b? \cdots b?ay,$$

for some $a, b, x, y \in \{0, 1\}$ with $a \neq b$.

Suppose that $w = a?b?a?b? \cdots a?b$. By Lemma 4, the word w has $4n + 4$ runs; the two possibilities for $a \in \{0, 1\}$ give two equivalence classes.

Now suppose that $w = xa?b?a?b? \cdots b?ay$. In this case we get our four additional equivalence classes when $x = y$. We verify their existence for all n with the following Walnut commands, which all return TRUE:

```
eval pf_red_7mod8a "?msd_2 An Er P[4*r+3]=@0 & P[4*r+4]=@0 &
P[4*r+8*n+9]=@0":
eval pf_red_7mod8b "?msd_2 An Er P[4*r+3]=@1 & P[4*r+4]=@0 &
P[4*r+8*n+9]=@1":
```

```
eval pf_red_7mod8c "?msd_2 An Er P[4*r+1]=@0 & P[4*r+2]=@1 &
P[4*r+8*n+7]=@0":
eval pf_red_7mod8d "?msd_2 An Er P[4*r+1]=@1 & P[4*r+2]=@1 &
P[4*r+8*n+7]=@1":
```

□

Proof of Theorem 2. The statement immediately follows from Lemmas 5–9. □

2.2. Reduced Abelian Complexity Functions

Relative to our proof of Theorem 2, a similar approach can be used to evaluate the reduced abelian complexity function for \mathbf{f} . We later consider the problem of determining a recurrence for the reduced abelian complexity function for \mathbf{t} .

The integer sequence

$$(\rho_{\mathbf{f}}^{\text{ab}}(n) : n \in \mathbb{N}) = (2, 3, 4, 3, 4, 5, 4, 3, 4, 5, 6, 5, 4, 5, 4, 3, 4, 5, 6, 5, 6, \dots) \quad (9)$$

agrees with the OEIS entry [A214613](#) and was first shown to be a 2-regular sequence by Madill and Rampersad [13]. In contrast to (9), we find that

$$(\rho_{\mathbf{f}}^{\text{ab,red}}(n) : n \in \mathbb{N}) = (2, 3, 5, 3, 4, 3, 5, 3, 4, 3, 5, 3, 4, 3, 5, 3, 4, 3, 5, 3, 4, 3, \dots)$$

is eventually periodic.

Theorem 3. *For every positive integer n , we have that*

$$\rho_{\mathbf{f}}^{\text{ab,red}}(n) = \begin{cases} 3 & \text{if } n \text{ is even,} \\ 4 & \text{if } n > 1 \text{ and } n \equiv 1 \pmod{4}, \\ 5 & n \equiv 3 \pmod{4}. \end{cases}$$

Proof. Suppose two binary words w and w' begin with different letters but have the same number of runs. If w and w' have an even number of runs then $\text{red}(w)$ and $\text{red}(w')$ are abelian equivalent; whereas, if w and w' have an odd number of runs then $\text{red}(w)$ and $\text{red}(w')$ are not abelian equivalent.

To prove the result we go through the various lemmas covering the different cases of the proof of Theorem 2 and examine the parity of the number of runs in each \sim_{red} equivalence class.

If $n = 2m$ then the proof of Lemma 5 gives:

- $2 \sim_{\text{red}}$ equivalence classes with m runs, and
- $2 \sim_{\text{red}}$ equivalence classes with $m + 1$ runs.

One of m and $m + 1$ is even and the two corresponding \sim_{red} equivalence classes merge under the abelian reduced equivalence. Hence, we have $\rho_{\mathbf{f}}^{\text{ab,red}}(n) = 3$.

If $n = 8m + 1$, $m > 0$, then the proof of Lemma 6 gives:

- $1 \sim_{\text{red}}$ equivalence class with $4m$ runs,
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 1$ runs, and
- $1 \sim_{\text{red}}$ equivalence class with $4m + 2$ runs.

Hence, we have $\rho_{\mathbf{f}}^{\text{ab,red}}(n) = 4$.

If $n = 8m + 3$ then the proof of Lemma 7 gives:

- $2 \sim_{\text{red}}$ equivalence classes with $4m + 1$ runs,
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 2$ runs, and
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 3$ runs.

Hence, we have $\rho_{\mathbf{f}}^{\text{ab,red}}(n) = 5$.

If $n = 8m + 5$ then the proof of Lemma 8 gives:

- $2 \sim_{\text{red}}$ equivalence classes with $4m + 2$ runs,
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 3$ runs, and
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 4$ runs.

Hence, we have $\rho_{\mathbf{f}}^{\text{ab,red}}(n) = 4$.

If $n = 8m + 7$ then the proof of Lemma 9 gives:

- $2 \sim_{\text{red}}$ equivalence classes with $4m + 3$ runs, and
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 4$ runs, and
- $2 \sim_{\text{red}}$ equivalence classes with $4m + 5$ runs.

Hence, we have $\rho_{\mathbf{f}}^{\text{ab,red}}(n) = 5$. □

The problem of determining a recursion for

$$(\rho_{\mathbf{t}}^{\text{ab,red}}(n) : n \in \mathbb{N}) = (2, 3, 3, 4, 3, 5, 4, 5, 3, 4, 5, 6, 4, 6, 5, 4, 3, 5, 4, \dots) \quad (10)$$

appears to be much more challenging, relative to the above theorems. This problem is motivated by past research on the abelian complexity functions for Thue–Morse-like sequences [6, 9, 11, 12, 14], and leads us to provide, in the next section, open problems concerning $\rho_{\mathbf{t}}^{\text{ab,red}}$.

3. Conclusion

Although it appears that

$$\rho_{\mathbf{t}}^{\text{ab,red}}(2n+1) = \rho_{\mathbf{t}}^{\text{ab,red}}(n+1)$$

for nonnegative integers n , the problem of determining a full recursion for $\rho_{\mathbf{t}}^{\text{ab,red}}(n)$ seems to be challenging. It appears that

$$\left| \rho_{\mathbf{t}}^{\text{ab,red}}(4n+2) - \rho_{\mathbf{t}}^{\text{ab,red}}(4n) \right| = \begin{cases} 0 & \text{if } \mathbf{t}_{n+1} = \mathbf{t}_{3n+1}, \\ 1 & \text{otherwise,} \end{cases} \quad (11)$$

but it is unclear how the sign of $\rho_{\mathbf{t}}^{\text{ab,red}}(4n+2) - \rho_{\mathbf{t}}^{\text{ab,red}}(4n)$ could be evaluated in an explicit way for the nonzero case, and we leave this as an open problem. Also, it is unclear as to how a suitable recursion could be determined for $\rho_{\mathbf{t}}^{\text{ab,red}}(4n)$, and we leave this as an open problem. We also leave it as an open problem to prove (11). It also appears that the integer sequence in (10) is not k -automatic, and we leave it as an open problem to prove this.

Acknowledgements. The first author was supported by an NSERC Discovery Grant and thanks Jean-Paul Allouche for useful feedback. The second and third authors also acknowledge the support of the NSERC Discovery Grant program.

References

- [1] G. Allouche, J.-P. Allouche, and J. Shallit, Kolams indiens, dessins sur le sable aux îles Vanuatu, courbe de Sierpinski, et morphismes de monoïde, *Ann. Inst. Fourier (Grenoble)* **56** (2006), 2115-2130.
- [2] J.-P. Allouche, The number of factors in a paperfolding sequence, *Bull. Aust. Math. Soc.* **46** (1) (1992), 23-32.
- [3] J.-P. Allouche and J. Shallit, *Automatic Sequences*, Cambridge University Press, Cambridge, 2003.
- [4] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in *Sequences and their applications (Singapore, 1998)*, Springer, London, 1999, 1-16.
- [5] S. V. Avgustinovich, The number of different subwords of given length in the Morse-Hedlund sequence, *Sibirsk. Zh. Issled. Oper.* **1** (2) (1994), 3-7, 103.
- [6] F. Blanchet-Sadri, J. D. Currie, N. Rampersad, and N. Fox, Abelian complexity of fixed point of morphism $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$, *Integers* **14** (2014), Paper No. A11, 17.
- [7] S. Brlek, Enumeration of factors in the Thue-Morse word, *Discrete Appl. Math.* **24** (1-3) (1989), 83-96.

- [8] M. Bunder, B. Bates, and S. Arnold, The summed paperfolding sequence, *Bull. Aust. Math. Soc.* **110** (2024), 189–198.
- [9] J. Chen and Z.-X. Wen, On the abelian complexity of generalized Thue-Morse sequences, *Theoret. Comput. Sci.* **780** (2019), 66–73.
- [10] A. de Luca and S. Varricchio, Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups, *Theoret. Comput. Sci.* **63** (3) (1989), 333–348.
- [11] F. Greinecker, On the 2-abelian complexity of the Thue-Morse word, *Theoret. Comput. Sci.* **593** (2015), 88–105.
- [12] I. Kaboré and B. Kientéga, Abelian complexity of Thue-Morse word over a ternary alphabet, in *Combinatorics on Words*, Springer, Cham, 2017, 132–143.
- [13] B. Madill and N. Rampersad, The abelian complexity of the paperfolding word, *Discrete Math.* **313** (7) (2013), 831–838.
- [14] A. Parreau, M. Rigo, E. Rowland, and É. Vandomme, A new approach to the 2-regularity of the ℓ -abelian complexity of 2-automatic sequences, *Electron. J. Combin.* **22** (1) (2015), Paper 1.27, 44 pp.
- [15] J. Shallit, *The Logical Approach to Automatic Sequences—Exploring Combinatorics on Words with Walnut*, Cambridge University Press, Cambridge, 2003.
- [16] J. Shallit, Runs in paperfolding sequences, *Bull. Aust. Math. Soc.*, in press. Available at <https://doi.org/10.1017/S0004972725100440>.
- [17] OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, <https://oeis.org>.