#A44 INTEGERS 23 (2023)

PROOF OF A CONJECTURE OF KRAWCHUK AND RAMPERSAD ON THE CYCLIC COMPLEXITY OF THE THUE-MORSE SEQUENCE

Jeffrey Shallit[1]
School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada
shallit@uwaterloo.ca

Received: 1/30/23, Revised: 4/19/23, Accepted: 6/1/23, Published: 6/30/23

Abstract

We prove a 2018 conjecture of Krawchuk and Rampersad on the extremal behavior of $c(n)$, where $c(n)$ counts the number of length-n factors of the Thue-Morse sequence $\mathbf{t}$, up to cyclic rotation.

1. Introduction

Let $\mathbf{x}$ be an infinite word (or sequence) over a finite alphabet. Many different notions of the "complexity" of $\mathbf{x}$ have been explored. To name just five:

- *subword* or *factor complexity*, the number of distinct blocks of length n appearing in $\mathbf{x}$ [5];
- *Abelian complexity*, the number of distinct blocks of length n in $\mathbf{x}$, up to permutation [13];
- *palindrome complexity*, the number of distinct length-n palindromes appearing in $\mathbf{x}$ [1];
- *linear complexity*, the length of the shortest linear recurrence satisfied by a prefix of length n [12];
- *maximum order complexity*, the degree of a smallest nonlinear feedback shift register generating a length-n prefix of $\mathbf{x}$ [7].

In a 2017 paper, Cassaigne et al. [4] introduced yet another interesting measure of complexity, called *cyclic complexity*. The cyclic complexity function $c_{\mathbf{x}}(n)$ is defined

DOI: 10.5281/zenodo.8099771
[1]Research supported in part by NSERC grant 2018-04118.

to be the number of length-n factors of $\mathbf{x}$, where factors that are the same, up to cyclic shift, are only counted once. It was further studied in [6].

Cyclic complexity is a more mysterious measure than ordinary subword complexity. For ordinary subword complexity $\rho_{\mathbf{x}}(n)$, which counts the number of distinct length-n factors appearing in $\mathbf{x}$, it is known that if $\mathbf{x}$ is an automatic sequence, then there is an automaton that takes as input the representations of n and y in parallel, and accepts iff $y = \rho_{\mathbf{x}}(n)$. In other words, $\rho_{\mathbf{x}}(n)$ is *synchronized* for automatic sequences; see [8] for more details. This means that checking whether $\rho_{\mathbf{x}}(n) \leq An+B$ or $\rho_{\mathbf{x}}(n) \geq An+B$ for all n is, in general, decidable for automatic sequences, since we can express these assertions as a first-order logical formula [15].

However, for cyclic complexity the function $c_{\mathbf{x}}(n)$ is not, in general, synchronized. We can see this as follows: let $\mathbf{p} = 11010001\cdots$ be the characteristic sequence of the powers of 2. Then it is not hard to see that $c_{\mathbf{p}}(n) = O(\log n)$ and $c_{\mathbf{p}}(2^n) = n+2$ for $n \geq 0$. However, by a theorem about synchronized sequences [14], this kind of growth rate is impossible. This fundamental difference may help explain why it is so much harder to prove inequalities for cyclic complexity.

In a recent paper, Krawchuk and Rampersad [9] studied the cyclic complexity function $c_{\mathbf{x}}(n)$ where $\mathbf{x}$ is one of the most celebrated aperiodic binary words, the Thue-Morse sequence $\mathbf{t} = 01101001\cdots$ [2]. They showed that the function $c_{\mathbf{t}}(n)$ is 2-regular and is specified by a linear representation of rank 50. This means there are vectors v, w and a matrix-valued morphism γ such that $c_{\mathbf{t}}(n) = v\gamma(z)w$ for all strings z that are binary representations of n (allowing leading zeros). See [3] for more details. The first few terms of $c_{\mathbf{t}}(n)$ are given in Table 1; it is sequence A360104 in the *On-Line Encyclopedia of Integer Sequences* (OEIS) [11].

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$c_{\mathbf{t}}(n)$	1	2	3	2	4	4	6	8	12	8	12	16	14	18	18	18	28

Table 1: First few values of $c_{\mathbf{t}}(n)$.

Krawchuk and Rampersad conjectured that

$$\limsup c_{\mathbf{t}}(n)/n = 2 \quad \text{and} \quad \liminf c_{\mathbf{t}}(n)/n = \frac{4}{3}.$$

In this paper we prove these two conjectures. The conjectures follow from two inequalities: $c_{\mathbf{t}}(n) \leq 2n-4$ for $n \geq 3$ and $c_{\mathbf{t}}(n) \geq \frac{4}{3}n-4$ for $n \geq 0$, which we prove in Section 3 and 4, respectively.

The method of linear representations (as discussed in, for example, [15]) is extremely powerful for proving statements about automatic sequences, but it has some limitations. While it can often be used to prove various equalities, proving inequalities is typically more problematic. In this paper we use traditional techniques based on induction, together with linear representations, to prove the desired inequalities.

2. Preliminary Results

Throughout we abbreviate $c_{\mathbf{t}}(n)$ by $c(n)$. We start with a list of identities for $c(n)$ that will prove useful later.

Proposition 1. *We have*

$$c(2^k) = 2^{k+1} - 4, \qquad (k \geq 2) \tag{1}$$
$$c(2^k + 3) = \frac{5}{3} \cdot 2^k - \frac{2}{3}(-1)^k + 2, \qquad (k \geq 2) \tag{2}$$
$$c(2^k + 1) = \frac{4}{3} \cdot 2^k + \frac{2}{3}(-1)^k - 2, \qquad (k \geq 2) \tag{3}$$
$$c(2^k - 3) = \frac{5}{3} \cdot 2^k + \frac{1}{3}(-1)^k - 5, \qquad (k \geq 5) \tag{4}$$
$$c(2^k - 1) = \frac{4}{3} \cdot 2^k - \frac{1}{3}(-1)^k - 3, \qquad (k \geq 2) \tag{5}$$
$$c(2^k - 5) = \frac{3}{2} \cdot 2^k + (-1)^k - 7, \qquad (k \geq 5) \tag{6}$$
$$c(2^k - 7) = \frac{3}{2} \cdot 2^k - (-1)^k - 9, \qquad (k \geq 3) \tag{7}$$
$$c(12 \cdot 2^k - 3) = \frac{56}{3} \cdot 2^k - \frac{2}{3}(-1)^k - 10, \qquad (k \geq 0). \tag{8}$$

Proof. Equation (1) can be found in [9, Proposition 1]. The remaining equalities can be proved exactly as in that paper, using the same technique. Also see [15, Section 9.11.15]. □

The following three identities are crucial to our approach. Their usefulness resides in the fact that they express subsequences of $c(n)$ as non-negative linear combinations of other subsequences, modulo a term that is bounded in absolute value. This facilitates proving upper and lower bounds by induction.

Lemma 1. *There are 2-automatic sequences a_0, a_1, a_3 such that*

$$c(2i) = 2c(i) + a_0(i) \tag{9}$$
$$c(4i + 1) = 2c(i + 1) + c(2i + 1) + a_1(i) \tag{10}$$
$$c(4i + 3) = \frac{1}{2}c(2i) + c(2i + 3) + \frac{1}{2}c(2i + 4) + a_3(i) \tag{11}$$

and furthermore

$$2 \leq a_0(i) \leq 6, \qquad (i \geq 3) \tag{12}$$
$$0 \leq a_1(i) \leq 10, \qquad (i \geq 2) \tag{13}$$
$$-1 \leq a_3(i) \leq 3, \qquad (i \geq 1). \tag{14}$$

Proof. For each relation, we compute a linear representation for each term except a_0 (resp., a_1, a_3) using `Walnut` [10]; then we compute a linear representation for the difference using block matrices. Then we minimize the linear representation and use the "semigroup trick" (see, e.g., [15, Section 4.11]) to verify that the sequence is automatic and find a deterministic finite automaton with output (DFAO) for a_0 (resp., a_1, a_3).

We provide more details about the computation of a_0 (in part because we will need them in what follows). We start with the following `Walnut` code:

```
def tmfactoreq "At t<n => T[i+t]=T[j+t]":
def tmconj "Et t<=n & $tmfactoreq(j,i+t,n-t) &
   $tmfactoreq(i,(j+n)-t,t)":
def tmc "Aj j<i => ~$tmconj(i,j,n)":
```

Here

- `tmfactoreq(i,j,n)` asserts that the two length-n factors $\mathbf{t}[i..i+n-1]$ and $\mathbf{t}[j..j+n-1]$ are identical.
- `tmconj(i,j,n)` asserts that $\mathbf{t}[i..i+n-1]$ is a cyclic shift of $\mathbf{t}[j..j+n-1]$.
- `tmc(i,n)` asserts that $\mathbf{t}[i..i+n-1]$ is the first occurrence of a factor that is cyclically equivalent to it.

Hence, by counting the number of i for which `tmc` evaluates to `TRUE`, we determine the number of length-n factors up to cyclic shift. We can then find the appropriate linear representations using `Walnut`:

```
eval tmcc n "$tmc(i,n)":
eval tmcc2 n "$tmc(i,2*n)":
```

The first two commands produce linear representations for $c(n)$ and $c(2n)$, in `Maple` format. They are of rank 50 and 60, respectively. From this we can easily compute a linear representation for $a_0(n) := c(2n) - 2c(n)$, of rank 110. Using Schützenberger's algorithm [3, Chap. 2], this representation can be minimized into a linear representation (v, γ, w) of rank 7, as follows:

$$v^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \gamma(0) = \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1 \end{bmatrix} \quad \gamma(1) = \begin{bmatrix} 0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&1/2&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&1/2&0&0 \end{bmatrix} \quad w = \begin{bmatrix} -1 \\ -1 \\ -2 \\ 2 \\ 4 \\ 2 \\ 6 \end{bmatrix}.$$

We can now use the "semigroup trick" to show that $a_0(n)$ is 2-automatic and find a DFAO for it. It has 8 states and is displayed in Figure 1. From examining the result, we see that $a_0(n) \in \{-2, -1, 2, 4, 6\}$ and furthermore $a_0(n) \in \{2, 4, 6\}$ for $n \geq 3$. This proves Equations (9) and (12).

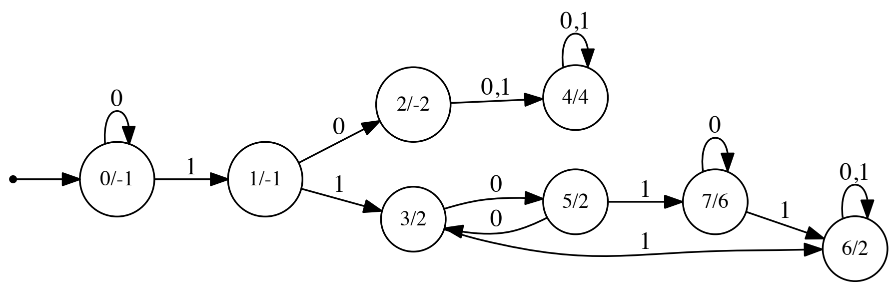

Figure 1: DFAO for $a_0(n)$.

Using a similar procedure, and the following Walnut commands, we can find the DFAO's for $a_1(n)$ and $a_3(n)$, which are given in Figures 2 and 3.

```
eval tmcc41 n "$tmc(i,4*n+1)":
eval tmcc1 n "$tmc(i,n+1)":
eval tmcc21 n "$tmc(i,2*n+1)":
eval tmcc43 n "$tmc(i,4*n+3)":
eval tmcc23 n "$tmc(i,2*n+3)":
eval tmcc24 n "$tmc(i,2*n+4)":
```

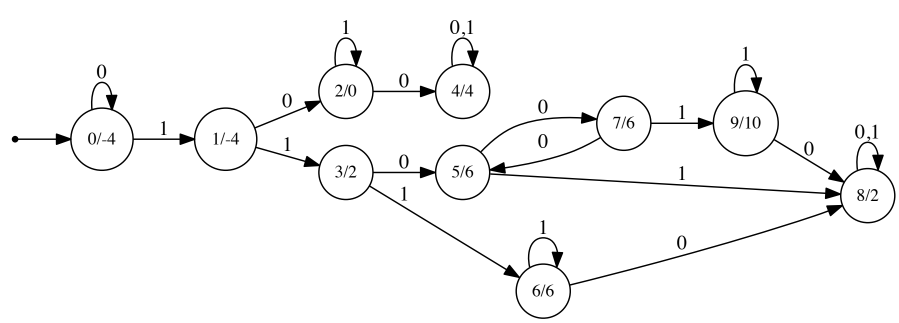

Figure 2: DFAO for $a_1(n)$.

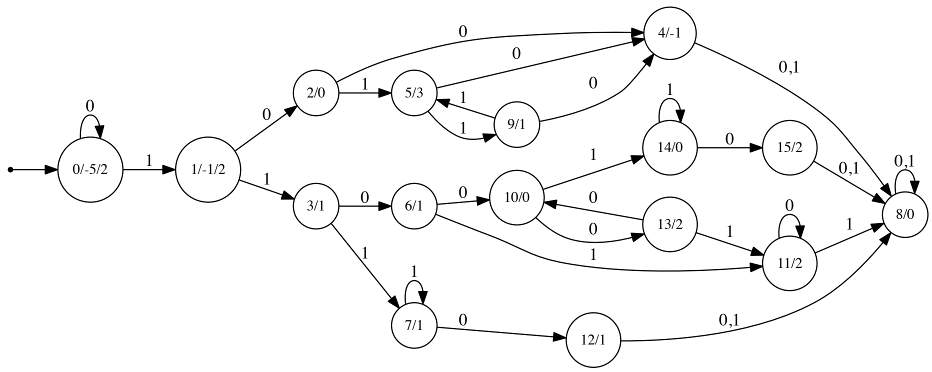

Figure 3: DFAO for $a_3(n)$.

Examining the results proves Equations (10), (11), (13), and (14). □

3. Upper Bound

Theorem 1. *We have $c(n) \le 2n - 4$ for all $n \ge 3$.*

Proof. The proof proceeds by induction on n, using Equations (9)–(11) and (13)–(14). However, the claim of the theorem does not seem to be strong enough to carry out the induction, so we actually prove the following stronger claim by induction:

$$c(n) \le 2n - 7 \text{ if } n \ge 12 \text{ and } n \notin P_2, \tag{15}$$

where $P_2 = \{2^i : i \ge 0\} = \{1, 2, 4, 8, 16, \ldots\}$. The base case is $n \le 44$; we can easily check that Assertion (15) holds for these n.

Now assume $n \ge 45$. There are three cases to consider.

Case 1: $n \equiv 0 \pmod 2$, $n = 2i$.

`Case 1a:` Assume $n/2 \notin P_2$. Then

$$\begin{aligned} c(n) &= c(2i) = 2c(i) + a_0(i) \\ &= 2c(n/2) + a_0(n/2) \\ &\le 2c(n/2) + 6 \\ &\le 2(n-7) + 6 \qquad \text{(by induction, since } n/2 \ge 22) \\ &= 2n - 8 \le 2n - 7. \end{aligned}$$

`Case 1b:` If $n/2 \in P_2$, then $n = 2^k$ for some $k \geq 1$. Since $n \geq 45$ we have $k \geq 6$. Hence the desired bound $c(n) \leq 2n - 4$ follows from Equation (1).

`Case 2:` $n \equiv 1 \pmod 4$, $n = 4i + 1$.

`Case 2a:` Assume $(n+3)/4 \notin P_2$ and $(n+1)/2 \notin P_2$. Then we have

$$\begin{aligned} c(n) &= c(4i+1) = 2c(i+1) + c(2i+1) + a_1(i) \\ &\leq 2c((n+3)/4) + c((n+1)/2) + a_1((n-1)/4) \\ &\leq 2((n+3)/2 - 7) + (n + 1 - 7) + 10 \\ &\qquad \text{(by induction, since } (n+1)/2 \geq (n+3)/4 \geq 12) \\ &= 2n - 7. \end{aligned}$$

`Case 2b:` If $(n+3)/4 \in P_2$, then $n = 2^k - 3$ for some $k \geq 2$. Since $n \geq 45$ we must have $k \geq 6$. Then Equation (4) implies the desired bound.

`Case 2c:` If $(n+1)/2 \in P_2$ then $n = 2^k - 1$ for some $k \geq 1$. Since $n \geq 45$ we have $k \geq 6$. Hence by Equation (5) we get the desired bound.

`Case 3:` $n \equiv 3 \pmod 4$, $n = 4i + 3$.

`Case 3a:` Assume $(n-3)/2 \notin P_2$ and $(n+3)/2 \notin P_2$ and $(n+5)/2 \notin P_2$. Then we have

$$\begin{aligned} c(n) = c(4i+3) &= \frac{1}{2}c(2i) + c(2i+3) + \frac{1}{2}c(2i+4) + a_3(i) \\ &= \frac{1}{2}c((n-3)/2) + c((n+3)/2) + \frac{1}{2}c((n+5)/2) + a_3((n-3)/4) \\ &\leq \frac{1}{2}(n - 10) + (n + 3 - 7) + \frac{1}{2}(n + 5 - 7) + 3 \\ &\qquad \text{(by induction, since } (n+5)/2 \geq (n+3)/2 \geq (n-3)/2 \geq 21) \\ &= 2n - 7. \end{aligned}$$

`Case 3b:` If $(n-3)/2 \in P_2$ then $n = 2^k + 3$ for some $k \geq 1$. Since $n \geq 45$ we have $k \geq 6$. So the desired bound follows from Equation (2).

`Case 3c:` If $(n+3)/2 \in P_2$ then $n = 2^k - 3$ for some $k \geq 1$. Since $n \geq 45$ we have $k \geq 6$. So the desired bound follows from Equation (4).

`Case 3d:` If $(n+5)/2 \in P_2$ then $n = 2^k - 5$ for some $k \geq 1$. Since $n \geq 45$ we have $k \geq 6$. So the desired bound follows from Equation (6).

We have now completed the proof of Assertion (15). To finish the proof of the theorem, we only need observe that if $n \geq 8$ is a power of 2, then $c(n) = 2n - 4$, and check that $c(n) \leq 2n - 4$ for $3 \leq n \leq 11$. □

Combining our upper bound with Equation (1), we now get the first conjecture of Krawchuk and Rampersad as an immediate corollary.

Corollary 1. *We have* $\limsup_{n\to\infty} c(n)/n = 2$.

4. Lower Bound

In this section we prove the the corresponding lower bound on $c(n)$.

Theorem 2. *We have* $c(n) \geq \frac{4}{3}n - 4$ *for* $n \geq 0$.

The ideas are similar to those in the proof of the upper bound, but a bit more complicated because the various exceptional sets are more intricate.

We need a lemma. Define the following exceptional sets.

$$A = \{2^k - 1 : k \geq 1\} = \{1, 3, 7, 15, 31, \ldots\}$$
$$B = \{2^k + 1 : k \geq 2\} = \{5, 9, 17, 33, \ldots\}$$
$$D = \{12 \cdot 2^k - 3 : k \geq 0\} = \{9, 21, 45, 93, \ldots\}$$
$$J = \{(2^{2i+1} + 1)2^j : i \geq 1, \ j \geq 0\} = \{9, 18, 33, 36, 66, 72, 129, 132, 144, 258, \ldots\}.$$

Lemma 2. *The following hold:*

(i) *If* $n \in A$, *then* $c(n) \geq \frac{4}{3}n - 2$.

(ii) *If* $n \in B$, *then* $c(n) \geq \frac{4}{3}n - 4$.

(iii) *If* $n \in D$, *then* $c(n) = 2c((n+3)/4) + c((n+1)/2)$.

(iv) *If* $n \in J$, *say* $n = (2^{2i+1} + 1)2^j$, *then* $c(n) = \frac{8}{3}2^{2i+j} + \frac{4}{3}2^j - 4 = \frac{4}{3}n - 4$ *for* $i \geq 1$, $j \geq 0$.

(v) *If* $n \in 4J + 3$, *say* $n = (2^{2i+1} + 1)2^{j+2} + 3$, *and* $(i, j) \neq (1, 0)$ *(i.e.,* $n \neq 39$*), then* $c(n) = (104 \cdot 2^{2i+j} + 64 \cdot 2^j + 4 \cdot 2^{2i}(-1)^j - 10 \cdot (-1)^j + 18)/9 \geq (4n + 16)/3$.

(vi) *If* $n \in 2J + 3$, *say* $n = (2^{2i+1} + 1)2^{j+1} + 3$, *and furthermore* $j \geq 1$, *then* $c(n) = (52 \cdot 2^{2i+j} + 32 \cdot 2^j - 4 \cdot 2^{2i}(-1)^j + 10 \cdot (-1)^j + 18)/9$.

(vii) *If* $n \in 2J - 5$, *say* $n = (2^{2i+1} + 1)2^{j+1} - 5$, *then* $c(n) = 6 \cdot 2^{2i+j} + 4 \cdot 2^j + \frac{8}{3} \cdot 2^{2i}(-1)^j - \frac{2}{3}(-1)^j - 14$ *for* $i \geq 1$ *and* $j \geq 2$.

Proof. Items (i) and (ii) follow immediately from Equations (3) and (5).

For item (iii), take $i = (n-1)/4$ in Equation (10). Then $c(n) = 2c((n+3)/4) + c((n+1)/2)$ iff $a_1((n-1)/4) = 0$. But from the DFAO in Figure 2 we see $a_i(m) = 0$ iff $m = 3 \cdot 2^k - 1$ for $k \geq 0$. Hence $a_1((n-1)/4) = 0$ iff $(n-1)/4 = 3 \cdot 2^k - 1$ for $k \geq 0$, iff $n = 12 \cdot 2^k - 3$ for $k \geq 0$.

For item (iv), we use a variant of the linear representation trick. Let $n = (2^{2i+1} + 1)2^j$. The base-2 representation of n is $10^{2i}10^j$, so $c(n) = v\gamma(10^{2i}10^j)w =$

$v\gamma(1)\gamma(0)^{2i}\gamma(1)\gamma(0)^j w$. The minimal polynomial of $\gamma(0)$ is $X^2(X-1)(X-2)(X+1)$, so each entry of $\gamma(0)^j$ for $j \geq 2$ is a linear combination of $2^j, (-1)^j$, and 1. The same is then true for $\gamma(1)\gamma(0)^j w$. Similarly, each entry of $\gamma(0)^{2i}$ for $i \geq 2$ is a linear combination of 2^{2i} and 1, and the same is true for $v\gamma(1)\gamma(0)^{2i}$. Hence the entries of the product $v\gamma(1)\gamma(0)^{2i}\gamma(1)\gamma(0)^j w$ are linear combinations of 2^{2i+j}, 2^j, 2^{2i}, 1, $2^{2i}(-1)^j$, and $(-1)^j$. We can deduce the particular constants by substituting small values of i and j and solving the resulting linear system. The result now follows.

Parts (v), (vi), and (vii) follow from the same technique. □

We are now ready to prove Theorem 2.

Proof of Theorem 2. Again, the statement of the theorem does not seem strong enough to make the induction go through.

We will prove the following three claims below by simultaneous induction on n.

(i) For all n we have $c(n) \geq \frac{4}{3}n - 4$.

(ii) If n is even and $n \notin J$ then $c(n) \geq \frac{4}{3}n - 2$.

(iii) If n is odd, $n \geq 47$, and $n \notin A \cup B$, then $c(n) \geq (4n + 16)/3$.

The base case is $n < 191$, which we can check by a short computation. Now assume $n \geq 191$.

`Case 1:` $n \equiv 0 \pmod 2$, $n = 2i$.

`Case 1a:` Suppose $n \in J$. Then from Lemma 2 (iv) we have $c(n) = \frac{4}{3}n - 4$.

`Case 1b:` Suppose $n \notin J$. Then

$$\begin{aligned} c(n) &= c(2i) = c(i) + a_0(i) \\ &= 2c(n/2) + a_0(n/2) \\ &\geq 2c(n/2) + 2 \\ &\geq 2 \cdot (\frac{4}{3}(n/2) - 2) + 2 \qquad \text{(by induction, since } n/2 \geq 47) \\ &= \frac{4}{3}n - 2. \end{aligned}$$

`Case 2:` $n \equiv 1 \pmod 4$, $n = 4i+1$. If $n \in A \cup B$ then the inequality $c(n) \geq \frac{4}{3}n - 4$ follows from Equations (3) and (5). So assume $n \notin A \cup B$.

`Case 2a:` Suppose $n \notin D$, $\frac{n+3}{4} \notin J \cup A \cup B$, $\frac{n+1}{2} \notin A \cup B$.

Then

$$\begin{aligned}
c(n) = c(4i+1) &= 2c(i+1) + c(2i+1) + a_1(i) \\
&= 2c((n+3)/4) + c((n+1)/2) + a_1((n-1)/4) \\
&\geq 2c((n+3)/4) + c((n+1)/2) + 2 \qquad \text{(because } n \notin D) \\
&\geq 2 \cdot (\frac{4}{3} \cdot ((n+3)/4) - 2) + (4\frac{n+1}{2} + 16)/3 + 2 \\
&\qquad \text{(because } (n+3)/4 \notin J \cup A \cup B \text{ and } (n+1)/2 \notin A \cup B \\
&\qquad \text{and } (n+1)/2 \geq (n+3)/4 \geq 47 \text{ by induction)} \\
&\geq \frac{4}{3}n + 6 \geq (4n+16)/3.
\end{aligned}$$

`Case 2b:` If $n \in D$ then Equation (8) implies that $c(n) \geq (4n/3) - 4$. If further $n \geq 47$ then it implies that $c(n) \geq (4n+16)/3$.

`Case 2c:` Suppose $\frac{n+3}{4} \in J$. Then from Lemma 2 (v) we have a closed form for $c(n)$. Using a routine calculation and the fact that $n \neq 39$, we get $c(n) \geq (4n+16)/3$.

`Case 2d:` Suppose $\frac{n+3}{4} \in A$. Then $n = 2^k - 7$ for $k \geq 3$, and then by Equation (7) we have $c(n) \geq (4n+16)/3$ for $k \geq 5$.

`Case 2e:` Suppose $\frac{n+3}{4} \in B$. Then $n \in B$, a contradiction.

`Case 2f:` Suppose $\frac{n+1}{2} \in A$. Then $n = 2^k - 3$ for $k \geq 2$, and by Equation (4) we have $c(n) \geq (4n+16)/3$ for $k \geq 5$.

`Case 2g:` Suppose $\frac{n+1}{2} \in B$. Then $n \in B$, a contradiction.

`Case 3:` $n \equiv 3 \pmod 4$, $n = 4i+3$. If $n \in A \cup B$ then $c(n) \geq \frac{4}{3}n - 4$ follows from Equations (3) and (5). So assume $n \notin A \cup B$.

`Case 3a:` Suppose $(n-3)/2 \notin J \cup A \cup B$ and $(n+3)/2 \notin J \cup A \cup B$ and $(n+5)/2 \notin J \cup A \cup B$. Then

$$\begin{aligned}
c(n) = c(4i+3) &= \frac{1}{2}c(2i) + c(2i+3) + \frac{1}{2}c(2i+4) + a_3(i) \\
&= \frac{1}{2}c((n-3)/2) + c((n+3)/2) + c((n+5)/2) + a_3((n-3)/4) \\
&\geq \frac{1}{2}(\frac{4}{3}(n-3)/2 - 2) + (4\frac{n+3}{2} + 16)/3 + \frac{1}{2}(\frac{4}{3}(n+5)/2 - 2) - 1 \\
&\qquad \text{(by conditions on } (n-3)/2, (n+3)/2, (n+5)/2 \\
&\qquad \text{and } (n+5)/2 \geq (n+3)/2 \geq (n-3)/2 \geq 47) \\
&= \frac{4}{3}n + 6 \geq (4n+16)/3.
\end{aligned}$$

`Case 3b:` Suppose $(n-3)/2 \in J$. Then $n \in 2J+3$, so $n = (2^{2i+1}+1)2^{j+1}+3$.

Since $n \equiv 3 \pmod 4$ we must have $j \geq 1$. Then by a routine calculation using Lemma 2 (vi), we have $c(n) \geq (4n + 16)/3$ since $n > 39$.

`Case 3c:` Suppose $(n-3)/2 \in A \cup B$. But $(n-3)/2 = 2n$ is even, a contradiction.

`Case 3d:` Suppose $(n + 3)/2 \in J$. Then $n \in 2J - 3$. But $n \equiv 3 \pmod 4$, so it is easy to see that this forces $n = 2^{2k} - 1 \in B$ for $k \geq 2$, a contradiction.

`Case 3e:` Suppose $(n + 3)/2 \in A$. Then $n = 2^k - 5$, and by Equation (6) we have $c(n) \geq (4n + 16)/3$ for $k \geq 6$.

`Case 3f:` Suppose $(n + 3)/2 \in B$. Then $n \in A$, a contradiction.

`Case 3g:` Suppose $(n + 5)/2 \in J$. Then $n \in 2J - 5$. Then from Lemma 2 (vii) it follows by a routine computation that $c(n) \geq (4n + 16)/3$.

`Case 3h:` Suppose $(n+5)/2 \in A \cup B$. But $(n+5)/2 = 2n+4$ is even, a contradiction.

This completes the proof by induction of (i), (ii), and (iii). □

Combining the lower bound with Equation (3) now gives us the following result.

Corollary 2. *We have* $\liminf_{n\to\infty} c(n)/n = 4/3$.

We also have enough to prove the following result.

Theorem 3. *We have* $c(n) = \frac{4}{3}n - 4$ *iff* $n \in J$.

Proof. The $\Longleftarrow$ direction follows from Lemma 2 (v), while the $\Longrightarrow$ direction is as follows: If n is even and $n \notin J$ then $c(n) \geq \frac{4}{3}n - 2 > \frac{4}{3}n - 4$ by above. If $n \geq 47$ is odd and $n \notin A \cup B$, then $c(n) \geq (4n + 16)/3 > \frac{4}{3}n - 4$ by above. It remains to check $n < 47$ and $n \in A \cup B$. For $n \in A$, we know from Lemma 2 (i) that $c(n) \geq \frac{4}{3}n - 2 > \frac{4}{3}n - 4$. For $n \in B$, it follows from Equation (3) that $c(n) = \frac{4}{3}n - 4$ iff $n = 2^{2k+1} + 1 \in J$. Finally, $n < 47$ can be checked with a computation. □

Acknowledgment. I thank Narad Rampersad and the referee for helpful remarks.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.* **292** (2003), 9–31.

[2] J.-P. Allouche and J. O. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in *Sequences and Their Applications, Proceedings of SETA '98*, Springer-Verlag, 1999, pp. 1–16.

[3] J. Berstel and C. Reutenauer, *Noncommutative Rational Series With Applications*, Vol. 137 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, 2011.

[4] J. Cassaigne, G. Fici, M. Sciortino, and L. Q. Zamboni, Cyclic complexity of words, *J. Combin. Theory Ser. A* **145** (2017), 36–56.

[5] J. Cassaigne and F. Nicolas, Factor complexity, in *Combinatorics, Automata and Number Theory*, Vol. 135 in *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, 2010, pp. 163–247.

[6] É. Charlier, S. Puzynina, and L. Q. Zamboni, On a group theoretic generalization of the Morse-Hedlund theorem, *Proc. Amer. Math. Soc.* **145** (2017), 3381–3394.

[7] C. J. A. Jansen, The maximum order complexity of sequence ensembles, in *EUROCRYPT '91*, Vol. 547, *Lecture Notes in Computer Science*, Springer-Verlag, 1991, pp. 153–159.

[8] D. Goč, L. Schaeffer, and J. Shallit, Subword complexity and k-synchronization, in *DLT 2013*, Vol. 7907 of *Lecture Notes in Computer Science*, Springer-Verlag, 2013, pp. 252–263.

[9] C. Krawchuk and N. Rampersad, Cyclic complexity of some infinite words and generalizations, *INTEGERS—Elect. J. Comb. Numb. Theory* **18A** (2018), #A12 (electronic), math.colgate.edu/~integers/vol18a.html.

[10] H. Mousavi, Automatic theorem proving in Walnut, Arxiv preprint arXiv:1603.06017 [cs.FL], available at http://arxiv.org/abs/1603.06017, 2016.

[11] OEIS Foundation, Inc., The on-line encyclopedia of integer sequences, 2023, available at https://oeis.org.

[12] H. Niederreiter, Linear complexity and related complexity measures for sequences, in *INDOCRYPT 2003*, Vol. 2904 of *Lecture Notes in Computer Science*, Springer-Verlag, 2003, pp. 1–17.

[13] G. Richomme, K. Saari, and L. Q. Zamboni, Abelian complexity in minimal subshifts, *J. London Math. Soc.* **83** (2011), 79–95.

[14] J. Shallit, Synchronized sequences, in *WORDS 2021*, Vol. 12847 of *Lecture Notes in Computer Science*, Springer-Verlag, 2021, pp. 1–19.

[15] J. Shallit, *The Logical Approach To Automatic Sequences: Exploring Combinatorics on Words with* Walnut, Vol. 482 of *London Math. Society Lecture Note Series*, Cambridge University Press, 2022.